

GROUP ACTIONS, DIVISORS, AND PLANE CURVES

ARACELI BONIFANT AND JOHN MILNOR

ABSTRACT. After a general discussion of group actions, orbifolds, and “weak orbifolds” this note will provide elementary introductions to two basic moduli spaces over the real or complex numbers: First the moduli space of effective divisors with finite stabilizer on the projective space \mathbb{P}^1 , modulo the group of projective transformations of \mathbb{P}^1 ; and then the moduli space of curves (or more generally effective algebraic 1-cycles) with finite stabilizer in \mathbb{P}^2 , modulo the group of projective transformations of \mathbb{P}^2 . It also discusses automorphisms of curves, and the topological classification of smooth real curves in \mathbb{P}^2 .

CONTENTS

| | |
|--|----|
| 1. Introduction | 1 |
| 2. Group Actions and Orbifolds. | 4 |
| 3. The Moduli Space for Effective Divisors of Degree n . | 21 |
| 4. The Moduli Space for Real or Complex Plane Curves. | 39 |
| 5. Curves of Degree Three. | 42 |
| 6. Complex Curves of Degree at least Four | 51 |
| 7. Singularity Genus and Proper Action. | 59 |
| 8. Curves with Infinite Stabilizer | 69 |
| 9. Automorphisms of Smooth Curves and of Riemann Surfaces. | 77 |
| 10. Real Curves: The Harnack-Hilbert Problem. | 87 |
| Appendix A. Remarks on the Literature. | 91 |
| References | 92 |

1. INTRODUCTION

A basic objective of this paper is to provide an elementary introduction to the moduli space of curves of degree n in the real or complex projective plane, modulo the action of the group of projective transformations. However, in order to prepare for this, we first give an exposition of the general theory of smooth group actions and orbifolds. As a further introduction, we describe the theory of effective divisors on the real or complex projective

2010 *Mathematics Subject Classification*. Primary:14L30, 14H50, 57R18, 14H10; Secondary: 08A35, 14P25.

Key words and phrases. moduli space of curves, smooth complex curves, effective 1-cycles, stabilizers of curves, algebraic set, group actions, proper action, improper action, weakly proper action, orbifolds, weak orbifolds, rational homology manifold, tree-of-spheres, W-curves, moduli space of divisors, Deligne-Mumford compactification.

The first author wants to thank the Institute for Mathematical Sciences at Stony Brook University, where she spent her sabbatical year, for its support to this project.

line modulo projective transformations. We also discuss a number of related topics, including automorphism groups of curves, and the topology of smooth curves in the real projective plane. The different sections are largely independent of each other.

§2. Group Actions. We consider smooth actions of Lie groups on smooth Hausdorff manifolds. Since these actions are not always proper, we introduce the concepts of: *locally proper action* and *weakly locally proper action*. We introduce the usual concept of orbifold,¹ as well the concept of *weak orbifold* for quotient spaces which have only some of the standard orbifold properties; and prove the following:

For a *proper/locally proper/or weakly proper* action the quotient space is respectively a *Hausdorff orbifold/locally Hausdorff orbifold/ or locally Hausdorff weak orbifold*.

§3. Divisors on \mathbb{P}^1 and the Quotient Space \mathfrak{M}_n . Working either over the real or complex numbers, the space \mathfrak{M}_n of effective divisors of degree n with finite stabilizer on the projective line \mathbb{P}^1 , modulo projective transformations, is a T_1 space for every n ; but is a Hausdorff orbifold only for $n \leq 4$. In the complex case the space $\mathfrak{M}_4(\mathbb{C})$ is homeomorphic to $\mathbb{P}^1(\mathbb{C})$; but in the real case $\mathfrak{M}_4(\mathbb{R})$ is homeomorphic to a closed line segment in $\mathbb{P}^1(\mathbb{R})$. In both the real and complex cases, there is just one point of \mathfrak{M}_4 which is “improper” in the sense that the action of the group of projective transformation at corresponding divisors is not proper. On the other hand, for $n > 4$, there is a unique maximal open subset $\mathfrak{M}_n^{\text{Haus}} \subset \mathfrak{M}_n$ which is a Hausdorff orbifold; but \mathfrak{M}_n is not even locally Hausdorff at points outside of $\mathfrak{M}_n^{\text{Haus}}$. This subset $\mathfrak{M}_n^{\text{Haus}}$ is compact for n odd, but not for n even.

The proofs make use of a *Distortion Lemma* for automorphisms of \mathbb{P}^1 which describes the way in which an automorphism which lies outside a large compact subset of PGL_2 must distort the geometry. The proofs also make use of two familiar projective invariants associated with a 4-tuple of points in \mathbb{P}^1 . The cross-ratio depends on the ordering of the four points; while the shape invariant $\mathbf{J} = \mathbf{J}(x, y, z, w)$ (the classical j -invariant times a convenient constant) is invariant under permutations of the four points.

In the complex case, a close relative of our $\mathfrak{M}_n(\mathbb{C})$ is the classical moduli space $\mathcal{M}_{0,n}$ consisting of closed Riemann surfaces of genus zero provided with an ordered list of $n \geq 3$ distinct points, where two such marked Riemann surfaces are identified if there is a conformal isomorphism taking one to the other. We also discuss the compactification $\overline{\mathcal{M}}_{0,n}(\mathbb{C})$, a beautiful object introduced by Knudsen, based on ideas of Grothendieck, Deligne and Mumford. (See also Keel [Ke]; and see Etingof, Henriques, Kamnitzer and Rains [EHKR] for the analogous space $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$.) In both the real and complex cases, there are $\binom{n}{p}$ associated embeddings

$$\overline{\mathcal{M}}_{0,p+1} \times \overline{\mathcal{M}}_{0,q+1} \hookrightarrow \overline{\mathcal{M}}_{0,n},$$

where $p + q = n$. As an example, in the real case for $p = 2$ and $q = 3$, there are ten essentially different embeddings of the circle $\overline{\mathcal{M}}_{0,3} \times \overline{\mathcal{M}}_{0,4}$ into the surface $\overline{\mathcal{M}}_{0,5}$, cutting it up into twelve hyperbolic pentagons, and thus presenting it as a hyperbolic analogue of the dodecahedron. The space $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$ is a 3-manifold which has a Jaco-Shalen-Johanssen decomposition, showing that it has some combination of hyperbolic and flat geometry. More explicitly, if we remove the ten tori, corresponding to the embeddings of $\overline{\mathcal{M}}_{0,4} \times \overline{\mathcal{M}}_{0,4}$ into

¹Most authors require *orbifolds* to be Hausdorff spaces, however we will allow orbifolds which are only *locally Hausdorff*.

$\overline{\mathcal{M}}_{0,6}$, then the remainder can be given the structure of a complete hyperbolic 3-manifold with twenty infinite cusps.

§4. Curves (or Effective 1-Cycles) in \mathbb{P}^2 . The space of all curves of degree n in \mathbb{P}^2 can be conveniently considered as a dense open subset of a projective space of dimension $n(n+3)/2$ whose elements are formal linear combinations of curves in \mathbb{P}^2 . These are called *effective [algebraic] 1-cycles*. For $n \geq 3$ we study the moduli space \mathbb{M}_n consisting of all projective equivalence classes of effective 1-cycles \mathcal{C} of degree n with finite stabilizer. (For $n < 3$, there are no curves with finite stabilizer.) This moduli space has been studied by many authors. (See for example Mumford’s discussion of hypersurfaces [Mu, p. 79], which includes not only curves in \mathbb{P}^2 but also divisors in \mathbb{P}^1 .) However, our presentation is more elementary, and is addressed to non-specialists.

§5. Cubic Curves. The space $\mathbb{M}_3(\mathbb{C})$ is homeomorphic to the topological 2-sphere consisting of all ratios $(a^3 : b^2) \in \mathbb{P}^1(\mathbb{C})$, corresponding to curves $y^2 = x^3 + ax + b$. It is isomorphic as an orbifold to the moduli space $\mathfrak{M}_4(\mathbb{C})$ for divisors of degree four on $\mathbb{P}^1(\mathbb{C})$. In the real case, the space $\mathbb{M}_3(\mathbb{R})$ is diffeomorphic to the unit circle. We describe the resulting circle of real cubic curves in several different ways, not only in terms of the normal form $y^2 = x^3 + ax + b$ with $a^2 + b^2 = 1$, but also in terms of the Hesse normal form $x^3 + y^3 + z^3 = 3kxyz$, with $k \in \mathbb{R} \cup \{\infty\}$, and in terms of a convenient “flex-slope normal form”, parametrized by the slope s at a flex point.

§6, §7. Degree at least Four. When $n \geq 5$, we prove that the moduli space $\mathbb{M}_n(\mathbb{C})$ is not a Hausdorff space. On the other hand, we provide two different procedures for describing large open subsets of $\mathbb{M}_n(\mathbb{C})$ which are Hausdorff. One (in §6) is based on studying the distribution of “virtual flex points”, that is points which are either flex points or which yield flex points under perturbation. The other (in §7) is based on a study of a genus invariant for the various singular points. Both proofs make use of a Distortion Lemma for automorphisms of \mathbb{P}^2 which lie outside a large compact set of group elements. These arguments apply only to points of moduli space which correspond to (possibly singular) curves; but there is a brief discussion of extending the proof to more general 1-cycles.

The open subset $\mathbb{M}_n^{\text{sm}} \subset \mathbb{M}_n(\mathbb{C})$ corresponding to smooth curves in $\mathbb{P}^2(\mathbb{C})$ is a Hausdorff space, which maps naturally into the classical moduli space $\mathcal{M}_{g(n)}$ consisting of all isomorphism classes of compact Riemann surfaces of genus $g(n) = \binom{n-1}{2}$. This map is injective in all cases (compare Remark 9.10), but is an isomorphism only for $n = 3$.

§8. Infinite Automorphism Groups. Although we are primarily concerned with curves which have finite stabilizer, this section studies the opposite case of curves with infinitely many automorphisms. Following Klein and Lie, we call these W-curves. We provide an explicit catalog of all such curves. (Compare Aluffi and Faber [AF1].)

§9. Finite Automorphism Groups. The study of finite automorphism groups of curves is closely related to the study of the moduli space \mathbb{M}_n , since any curve with extra automorphisms gives rise to a singular point in moduli space, or at least to a ramified point. For each degree d and prime p , we provide an explicit criterion for deciding whether there exists a smooth curve of degree d with a projective automorphism of period p ; and compare this with a corresponding statement for automorphisms of arbitrary compact Riemann surfaces. We show that an arbitrary finite subgroup of $\text{PGL}_3(\mathbb{C})$ is the full projective automorphism

group for some smooth curve in $\mathbb{P}^2(\mathbb{C})$; but show (following Chang [Ch]) that a generic² curve of degree four has no non-trivial projective automorphism.

§10. Real Curves. In the real case, the most studied question about smooth curves in $\mathbb{P}^2(\mathbb{R})$ is the Hilbert-Harnack problem concerning the number and topological arrangement of the connected components, each of which is a topological circle. We provide a report about this problem, as well as a suggested reformulation. The paper concludes with a brief Appendix describing some of the literature on the moduli spaces that we consider.

The authors are grateful to the referees for important suggestions and corrections (particularly in Section 9). We also want to thank Robert Lazarsfeld, Dennis Sullivan and Oleg Viro for their suggestions.

2. GROUP ACTIONS AND ORBIFOLDS.

This section will provide a general introduction to quotient spaces under a smooth group action. In the best case, with a proper action, the quotient space is Hausdorff, with an orbifold structure. Since the group actions that we consider are not always proper, we also introduce a modified requirement of “weakly proper” action, which suffices to prove that the quotient is locally Hausdorff, with a “weak orbifold structure” which includes only some of the usual orbifold properties. The section will conclude by discussing the special case of the projective general linear group, and its action on projective space.

First consider the complex case. Let \mathbf{X} be a metrizable complex manifold and \mathbf{G} a complex Lie group³ which acts on the left by a holomorphic map $\mathbf{G} \times \mathbf{X} \rightarrow \mathbf{X}$,

$$(\mathbf{g}, \mathbf{x}) \mapsto \mathbf{g}(\mathbf{x}),$$

where $\mathbf{g}_1(\mathbf{g}_2(\mathbf{x})) = (\mathbf{g}_1\mathbf{g}_2)(\mathbf{x})$. (A manifold is metrizable if and only if it is paracompact and Hausdorff.) We will always assume that the action is *effective* in the sense that

$$\mathbf{g}(\mathbf{x}) = \mathbf{x} \text{ for all } \mathbf{x} \iff \mathbf{g} \text{ is the identity element } \mathbf{e} \in \mathbf{G}.$$

The *quotient space* (or orbit space) in which \mathbf{x} is identified with \mathbf{x}' if and only if $\mathbf{x}' = \mathbf{g}(\mathbf{x})$ for some \mathbf{g} will be denoted⁴ by \mathbf{X}/\mathbf{G} .

In the real case, the definitions are completely analogous, although we could equally well work either in the C^∞ category or in the real analytic category. To fix ideas, let us choose the real analytic category. Thus in the real case, we will assume that \mathbf{G} is a real Lie group, that \mathbf{X} is a metrizable real analytic manifold, and that $\mathbf{G} \times \mathbf{X} \rightarrow \mathbf{X}$ is a real analytic map. It will often be convenient to use the word “analytic”, by itself, to mean real analytic in the real case, or complex analytic in the complex case.

Definition 2.1. For each $\mathbf{x} \in \mathbf{X}$ the set $((\mathbf{x}))$ consisting of all images $\mathbf{g}(\mathbf{x})$ with $\mathbf{g} \in \mathbf{G}$ is called the *\mathbf{G} -orbit* of \mathbf{x} . We will also use the notation

$$((\mathbf{x})) = \mathbf{F} = \mathbf{F}_{\mathbf{x}}$$

if we are thinking of $((\mathbf{x}))$ as a “*fiber*” of the *projection map* $\pi : \mathbf{X} \rightarrow \mathbf{Y} = \mathbf{X}/\mathbf{G}$.

²We say that a statement is true for a *generic curve* if it is true for all curves in some set which is dense and open in the Zariski topology. (Some authors prefer the term “general curve”; since the word “generic” has a more technical meaning in the theory of schemes.)

³Although our Lie groups are always positive dimensional, the discussion would apply equally well to the case of a discrete group, which we might think of as a zero-dimensional Lie group.

⁴Since \mathbf{G} acts on the left, many authors would use the notation $\mathbf{G} \backslash \mathbf{X}$.

Remark 2.2. In other words, each fiber is an equivalence class, where two points of \mathbf{X} are equivalent if and only if they belong to the same orbit under the action of \mathbf{G} . More generally, given any equivalence relation \sim on \mathbf{X} , we can form the quotient space $\mathbf{Y} = \mathbf{X}/\sim$. Such a quotient space \mathbf{Y} always has a well defined *quotient topology*, defined by the condition that a set $\mathbf{U} \subset \mathbf{Y}$ is open if and only if the preimage $\pi^{-1}(\mathbf{U})$ is open as a subset of \mathbf{X} . (We will always assume that \mathbf{X} is a Hausdorff space; but it does not necessarily follow that \mathbf{Y} is Hausdorff.)

Remark 2.3 (Closed Orbits and the T_1 Condition). By definition, a topological space \mathbf{Y} is a T_1 -*space* if every point of \mathbf{Y} is closed as a subset of \mathbf{Y} . Evidently a quotient space \mathbf{X}/\mathbf{G} (or more generally \mathbf{X}/\sim) is a T_1 -space if and only if each orbit (or each equivalence class) is closed as a subset of \mathbf{X} .

Orbifolds and Weak Orbifolds.

The concept and basic properties of orbifolds are due to I. Satake, who called them V-manifolds (see [Sa1, Sa2]). They were later studied by W. Thurston (see [Th]), who introduced the term “orbifold”.

Let \mathbb{F} stand for either the real or the complex numbers.

Definition 2.4. By a d -dimensional \mathbb{F} -*orbifold chart* around a point \mathbf{y} of a topological space \mathbf{Y} will be meant the following:

- (1) a finite group $\mathfrak{R} \subset \mathrm{GL}_d(\mathbb{F})$ acting linearly on \mathbb{F}^d ;
- (2) an \mathfrak{R} -invariant open neighborhood W of the point $\mathbf{0} \in \mathbb{F}^d$; and
- (3) a homeomorphism h from the quotient space W/\mathfrak{R} onto an open neighborhood U of \mathbf{y} in \mathbf{Y} such that the zero vector in W maps to \mathbf{y} .

The group \mathfrak{R} will be called the *ramification group* at \mathbf{y} , and its order $\tau_{\mathbf{y}} \geq 1$ will be called the *ramification index*. A point \mathbf{y}' is *ramified* if $\tau_{\mathbf{y}'} > 1$ and *unramified* if $\tau_{\mathbf{y}'} = 1$.

The space \mathbf{Y} together with an integer valued function $\mathbf{y} \mapsto \tau_{\mathbf{y}} \geq 1$ will be called a d -dimensional *weak orbifold* over \mathbb{F} if there exists such an orbifold chart $W/\mathfrak{R} \cong U$ around every point $\mathbf{y} \in \mathbf{Y}$, such that the associated ramification function from U to the positive integers coincides with the specified function $\mathbf{y}' \mapsto \tau_{\mathbf{y}'}$ restricted to U .

Lemma 2.5. *Every weak orbifold is a locally Hausdorff space; and the function $\mathbf{y} \mapsto \tau_{\mathbf{y}} \geq 1$ from \mathbf{Y} to the set of positive integers is always upper semicontinuous, taking the value $\tau_{\mathbf{y}} = 1$ on a dense open set. More precisely, for any orbifold chart $W/\mathfrak{R} \xrightarrow{\cong} U = U_{\mathbf{y}}$, we have $\tau_{\mathbf{y}'} \leq \tau_{\mathbf{y}}$ for every $\mathbf{y}' \in U$, with $\tau_{\mathbf{y}'} = 1$ on a dense open subset of U .*

Of course in good cases \mathbf{Y} will be a Hausdorff space; but even a locally Hausdorff space can be quite useful.⁵

⁵Perhaps the most startling application of locally Hausdorff spaces in science would be to the “Many Worlds” interpretation of quantum mechanics, in which the space-time universe continually splits into two or more alternate universes. (See for example [Bec].) The resulting object is possibly best described as a space which is locally Hausdorff, but wildly non-Hausdorff. It can be constructed mathematically out of infinitely many copies of the Minkowski space $\mathbb{R}^{3,1}$ by gluing together corresponding open subsets. (Of course it does not make any objective sense to ask whether these alternate universes “really exist”. The only legitimate question is whether a mathematical model including such alternate universes can provide a convenient and testable model for the observable universe.)

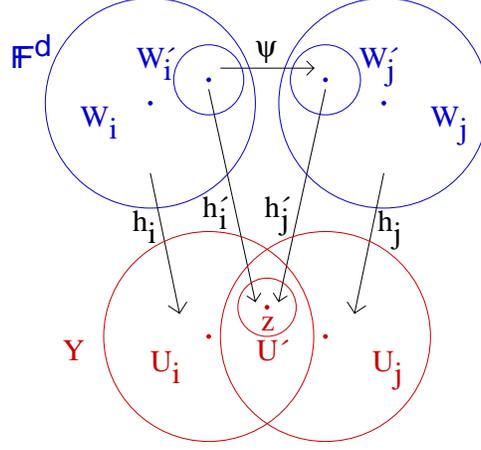


FIG. 1. Compatibility for two orbifold charts.

Proof of Lemma 2.5. For any chart $W/\mathfrak{R} \xrightarrow{\cong} U$ around \mathbf{y} , and any non-identity element $\mathbf{g} \in \mathfrak{R}$, the set of elements of \mathbb{F}^d fixed by \mathbf{g} must be a linear subspace of \mathbb{F}^d of dimension at most $d-1$. The complement of this finite union of linear subspaces within W is a dense open subset $W_0 \subset W$. If $U_0 \cong W_0/\mathfrak{R}$ is its image, then the associated mapping $W_0 \rightarrow U_0$ is locally bijective, and is precisely $\mathbf{r}_{\mathbf{y}}$ -to-one. Since this map is one-to-one in a small neighborhood of any point of W_0 , it follows that all points of U_0 are unramified. Each \mathfrak{R} -orbit in U is compact and non-empty, so we can use the Hausdorff metric for compact subsets of \mathbb{F}^d to make $U \cong W/\mathfrak{R}$ into a metric space. In particular, it follows that \mathbf{Y} is locally Hausdorff. \square

Definition 2.6. An orbifold chart $W/\mathfrak{R} \cong U$ around \mathbf{y} gives rise to a smaller orbifold chart around any point $\mathbf{y}' \in U$ which is called the *restriction* of this chart to a neighborhood of \mathbf{y}' . In fact, choosing a representative point $\mathbf{w}' \in W$ over \mathbf{y}' , let $\mathfrak{R}_{\mathbf{w}'}$ be the stabilizer, consisting of all $\mathbf{g} \in \mathfrak{R}$ for which $\mathbf{g}(\mathbf{w}') = \mathbf{w}'$. Evidently $\mathfrak{R}_{\mathbf{w}'}$ acts linearly on W , fixing the point \mathbf{w}' . Choose a $\mathfrak{R}_{\mathbf{w}'}$ -invariant neighborhood W' of \mathbf{w}' which is small enough so that the various images $\mathbf{g}(W')$ with $\mathbf{g} \in \mathfrak{R} \setminus \mathfrak{R}_{\mathbf{w}'}$ are all disjoint from W' . Then the projection from W' to $W'/\mathfrak{R}_{\mathbf{w}'} \subset U$ is the required restriction to an orbifold chart around \mathbf{y}' .

Definition 2.7. An *orbifold atlas* on Y is a collection of orbifold chart homeomorphisms

$$h_j : W_j/\mathfrak{R}_j \xrightarrow{\cong} U_j \subset Y,$$

where the U_j are open sets covering Y , which satisfy the following.

Compatibility Condition. For each point \mathbf{z} in an overlap $U_i \cap U_j$ and each sufficiently small neighborhood U' of \mathbf{z} , the restriction of the i -th and j -th orbifold charts to U' are *isomorphic* in the following sense. Let

$$h'_i : W'_i/\mathfrak{R}_{i,w_i} \xrightarrow{\cong} U' \quad \text{and} \quad h'_j : W'_j/\mathfrak{R}_{j,w_j} \xrightarrow{\cong} U'$$

be the two restrictions. Then we require that there should be an analytic isomorphism $\psi : W'_i \xrightarrow{\cong} W'_j$ so that $h'_i = h'_j \circ \psi$. (Compare Figure 1.)

Furthermore there should be a group isomorphism $\phi : \mathfrak{A}_{i,w_i} \xrightarrow{\cong} \mathfrak{A}_{j,w_j}$ so that

$$\psi(\mathbf{g}(w)) = \phi(\mathbf{g})(\psi(w))$$

for every $\mathbf{g} \in \mathfrak{A}_{i,w_i}$ and every $w \in W'_i$.

Two such atlases are *equivalent* if their union also satisfies this compatibility condition. The space \mathbf{Y} together with an equivalence class of such atlases is called an *orbifold*.⁶ Compare [Th], [BMP].

There may well be examples of weak orbifolds which are not compatible with any orbifold structure; but we do not know any such examples.

Example 2.8. On any Riemann surface, we can choose any function $\mathbf{y} \mapsto \mathbf{r}_{\mathbf{y}} \geq 1$ which takes the value $\mathbf{r}_{\mathbf{y}} = 1$ except at finitely many points. A corresponding collection of complex orbifold charts is easily constructed; and the compatibility condition is easily verified.

Here is a pair of more interesting examples.

Example 2.9. Let $\mathbf{X} \subset \mathbb{R}^3$ be the plane consisting of all (x, y, z) with $x + y + z = 0$, and let \mathfrak{G}_3 be the group of permutations of the three coordinates. Then under the action of \mathfrak{G}_3 each point of \mathbf{X} can be put into a unique normal form with $x \leq y \leq z$. If we set $u = y - x \geq 0$ and $v = z - y \geq 0$, then we can solve easily for x, y, z as linear functions of u, v . Thus the quotient space $\mathbf{Y} = \mathbf{X}/\mathfrak{G}_3$ can be identified with the positive quadrant in the (u, v) -plane. The ramification index is $\mathbf{r}(u, v) = 1$ throughout the interior of the quadrant, with $\mathbf{r}(u, 0) = \mathbf{r}(0, v) = 2$ along the two edges, and with $\mathbf{r}(0, 0) = 6$ at the origin. Note that the compatibility is automatically satisfied, since we have specified only one coordinate chart.

The analogous example with $(x, y, z) \in \mathbb{C}^3$ looks quite different. In this case, the quotient space is best described as the complex (σ_2, σ_3) -plane, where

$$\sigma_2 = xy + yz + zx \quad \text{and} \quad \sigma_3 = xyz$$

are elementary symmetric functions, with $\sigma_1 = x + y + z = 0$. Again the ramification index of a point in the quotient space is $\mathbf{r} = 1$ if x, y, z are all distinct, $\mathbf{r} = 2$ if only two are equal to each other, and $\mathbf{r} = 6$ if $x = y = z = 0$. Any triple with $\mathbf{r} = 2$ takes the form $(x, x, -2x)$, up to permutation of the coordinates, with $\sigma_2 = -3x^2$ and $\sigma_3 = -2x^3$. Thus $\mathbf{r} = 1$ if the expression⁷ $4\sigma_2^3 + 27\sigma_3^2$ is non-zero; with $\mathbf{r} = 2$ for most points where this expression is zero; but with $\mathbf{r} = 6$ if $\sigma_2 = \sigma_3 = 0$.

The main object of this section will be to describe conditions on the group action which guarantee that the quotient will be an orbifold or weak orbifold.

Proper and Weakly Proper Actions.

⁶Caution. Most authors require orbifolds to be Hausdorff spaces; but as noted in the introduction we will allow orbifolds which are only locally Hausdorff.

⁷Up to sign, this expression is just the discriminant of the polynomial $t^3 + \sigma_2 t - \sigma_3 = (t-x)(t-y)(t-z)$.

Definition 2.10. A continuous action of \mathbf{G} on a locally compact space \mathbf{X} is called *proper* if the associated map

$$(\mathbf{g}, \mathbf{x}) \mapsto (\mathbf{g}(\mathbf{x}), \mathbf{x})$$

from $(\mathbf{G} \times \mathbf{X})$ to $(\mathbf{X} \times \mathbf{X})$ is a proper map (in the usual sense that the preimage of any compact set is compact). A completely equivalent requirement is the following.

For any pair of compact subsets $K, K' \subset \mathbf{X}$ the set of all \mathbf{g} with $\mathbf{g}(K) \cap K' \neq \emptyset$ is compact.

Here is another completely equivalent condition.

For every pair of points \mathbf{x} and \mathbf{x}' in \mathbf{X} , there exist neighborhoods U of \mathbf{x} and U' of \mathbf{x}' which are small enough so that the set of *all* $\mathbf{g} \in \mathbf{G}$ with $\mathbf{g}(U) \cap U' \neq \emptyset$ has compact closure.

The proof that these three forms of the definition are equivalent is straightforward, and will be left to the reader.⁸

The action is *locally proper* at \mathbf{x} if this condition is satisfied for the special case where $\mathbf{x} = \mathbf{x}'$. It then follows that the action is proper throughout some \mathbf{G} -invariant open neighborhood of \mathbf{x} . In fact, if each \mathbf{g} which maps some point of U into U is contained in a compact set $K \subset \mathbf{G}$, then it follows that each \mathbf{g} which maps a point of $\mathbf{g}_1(U)$ to a point of $\mathbf{g}_2(U)$ is contained in the compact set $\mathbf{g}_2 K \mathbf{g}_1^{-1} \subset \mathbf{G}$.

The action will be called *weakly (locally) proper* at \mathbf{x} if the following still weaker condition is satisfied:

There should be a neighborhood U of \mathbf{x} and a compact set $K \subset \mathbf{G}$ such that, whenever two points \mathbf{x}_1 and \mathbf{x}_2 of U belong to the same \mathbf{G} -orbit, there exists at least one element $\mathbf{g} \in K$ with $\mathbf{g}(\mathbf{x}_1) = \mathbf{x}_2$.

It will be convenient to call a point in \mathbf{X}/\mathbf{G} either *proper* or *improper* according as the action of \mathbf{G} on corresponding points of \mathbf{X} is or is not locally proper. Similarly an improper point in \mathbf{X}/\mathbf{G} will be called *weakly proper* if the action of \mathbf{G} on corresponding points of \mathbf{X} is weakly proper. (*Caution:* Even when a point is improper, the quotient space may have a perfectly good orbifold structure.)

Definition 2.11. Given an action of \mathbf{G} on \mathbf{X} , the *stabilizer* $\mathbf{G}_{\mathbf{x}}$ of a point $\mathbf{x} \in \mathbf{X}$ is the closed subgroup of \mathbf{G} consisting of all $\mathbf{g} \in \mathbf{G}$ for which $\mathbf{g}(\mathbf{x}) = \mathbf{x}$. Note that points on the same fiber have isomorphic stabilizers, since

$$\mathbf{G}_{\mathbf{g}(\mathbf{x})} = \mathbf{g} \mathbf{G}_{\mathbf{x}} \mathbf{g}^{-1}.$$

If the stabilizer $\mathbf{G}_{\mathbf{x}}$ is finite, then it follows easily that the fiber \mathbf{F} through \mathbf{x} (consisting of all images $\mathbf{g}(\mathbf{x})$ with $\mathbf{g} \in \mathbf{G}$) is a smoothly embedded submanifold which is locally diffeomorphic to \mathbf{G} .

Under the hypothesis that all stabilizers are finite, we will prove the following in Theorem 2.18 together with Lemma 2.13 and Corollary 2.28:

For a proper action the quotient space is a Hausdorff orbifold.

For a locally proper action the quotient is a locally Hausdorff orbifold.

For a weakly proper action the quotient is a locally Hausdorff weak orbifold.

⁸In the special case of a discrete group, such an action is called *properly discontinuous*. (For more about proper actions, see the discussion beginning with Lemma 2.25).

(See Figure 2 for an example of a smooth locally proper action with trivial stabilizers where the quotient is not a Hausdorff space.)

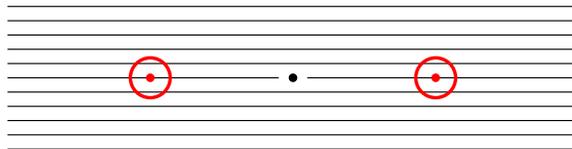


FIG. 2. Example: The additive group of real numbers acts on the punctured (x, y) -plane $\mathbb{R}^2 \setminus \{(0, 0)\}$ by an action $(x, y) \mapsto \mathbf{g}_t(x, y)$ for $t \in \mathbb{R}$ which satisfies the differential equation

$$d\mathbf{g}_t(x, y)/dt = (\sqrt{x^2 + y^2}, 0),$$

Since $\sqrt{x^2 + y^2}$ is strictly positive throughout the punctured plane, it follows that \mathbf{g}_t moves every point to the right for $t > 0$; although no orbit can reach the origin. (Note that \mathbf{g}_t acts on the real axis by $\mathbf{g}_t(x, 0) = (e^t x, 0)$ when $x > 0$; but $\mathbf{g}_t(x, 0) = (e^{-t} x, 0)$ when $x < 0$.) The action is locally proper but not proper; and the quotient space is locally Hausdorff but not Hausdorff. In fact, within every neighborhood of a point on the negative real axis and every neighborhood of a point on the positive real axis (as illustrated by circles in the figure), we can choose points which belong to the same orbit under the action.

Remark 2.12. Although there are examples which are weakly proper but not locally proper, they seem to be hard to find. Lemma 3.11 will show that divisors of degree four with only three distinct points give rise to such examples; and the proof of Lemma 5.4 will show that curves of degree three with a simple double point provide closely related examples. However, these are the only examples we know.

The following is well known.

Lemma 2.13. *If the action is proper, then the quotient \mathbf{X}/\mathbf{G} is a Hausdorff space.*

It follows as an immediate Corollary that a locally proper action yields a quotient space which is locally Hausdorff. However the quotient under a locally proper action need not be Hausdorff. (Compare Figure 2.) If stabilizers are finite, then we will see in Theorem 2.18 that even a weakly proper action yields a quotient space which is locally Hausdorff.

Remark 2.14. Note that every locally Hausdorff space is T_1 . In fact if one point \mathbf{p} belongs to the closure of a different point \mathbf{q} , then no neighborhood of \mathbf{p} is Hausdorff.

Proof of Lemma 2.13. It will be convenient to choose a metric on \mathbf{X} . Given \mathbf{x} and \mathbf{x}' there are two possibilities. If we can choose neighborhoods U and U' so that no translate $\mathbf{g}(U)$ intersects U' , then the images $\pi(U)$ and $\pi(U')$ in the quotient space are disjoint open sets.

On the other hand, taking U_j and U'_j to be a sequence of neighborhoods of \mathbf{x} and \mathbf{x}' respectively of radius $1/j$, if we can choose a group element \mathbf{g}_j for each j with $\mathbf{g}_j(U_j) \cap U'_j \neq \emptyset$, then by compactness we can pass to an infinite subsequence so that the

\mathbf{g}_j converge to a limit \mathbf{g} . It follows easily that $\mathbf{g}(\mathbf{x}) = \mathbf{x}'$, so that \mathbf{x} and \mathbf{x}' map to the same point in the quotient space. \square

Remark 2.15. The converse to Lemma 2.13 is false: A quotient space may be Hausdorff even when the action is not proper. Compare the discussions of $\mathfrak{M}_4(\mathbb{C})$ and $\mathfrak{M}_4(\mathbb{R})$ in Lemmas 3.4, 3.5 and Lemma 3.11, and of the quotient spaces $\mathbb{M}_3(\mathbb{C})$ and $\mathbb{M}_3(\mathbb{R})$ in Section 5. These quotients are Hausdorff orbifolds, even though the associated group action fails to be proper everywhere. (See Example 2.8, as well as Lemma 5.4 below.) However, for larger n we will have to deal with moduli spaces which are definitely not Hausdorff. (See Theorem 3.2 for \mathfrak{M}_n and Theorem 6.1 for \mathbb{M}_n .)

Weak Orbifold Structures. The passage from weakly proper actions to weak orbifold structures will be based on the following. Given any fiber \mathbf{F} , and given any point $\mathbf{x} \in \mathbf{F}$, we will refer to the quotient of tangent vector spaces

$$V_{\mathbf{x}} = T_{\mathbf{x}}\mathbf{X}/T_{\mathbf{x}}\mathbf{F}$$

as the *transverse vector space* to \mathbf{F} at \mathbf{x} . (If \mathbf{X} is provided with a Riemannian metric, then $V_{\mathbf{x}}$ can be identified with the normal vector space at \mathbf{x} .)

In order to describe a weak orbifold structure on the quotient, we must first construct the associated ramification groups. Note that the stabilizer $\mathbf{G}_{\mathbf{x}}$ acts linearly on both $T_{\mathbf{x}}\mathbf{X}$ and $T_{\mathbf{x}}\mathbf{F}$, and hence acts linearly on the d -dimensional quotient space $V_{\mathbf{x}}$, where d is the codimension of \mathbf{F} in \mathbf{X} . In practice, we will always assume that the stabilizer is finite, so that d is equal to the difference $\dim(\mathbf{X}) - \dim(\mathbf{G})$. However, this action is not always effective: The group $\mathbf{G}_{\mathbf{x}}$ may act non-trivially on $T_{\mathbf{x}}\mathbf{F}$, while leaving the transverse vector space pointwise fixed.

Definition 2.16. Let $\mathbf{H}_{\mathbf{x}}$ be the normal subgroup of $\mathbf{G}_{\mathbf{x}}$ consisting of all group elements which act as the identity map on $V_{\mathbf{x}}$ (that is, all $\mathbf{h} \in \mathbf{G}_{\mathbf{x}}$ such that $\mathbf{h}(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V_{\mathbf{x}}$). The quotient group

$$\mathfrak{R}_{\mathbf{x}} = \mathbf{G}_{\mathbf{x}}/\mathbf{H}_{\mathbf{x}}$$

will be called the *ramification group* at \mathbf{x} . Note that by its very definition, $\mathfrak{R}_{\mathbf{x}}$ comes with a linear action on the vector space $V_{\mathbf{x}}$, which is isomorphic to \mathbb{R}^d or \mathbb{C}^d . It is not hard to check that different points on the same fiber \mathbf{F} have isomorphic ramification groups. Let \mathbf{y} be the image of the fiber \mathbf{F} in \mathbf{Y} . As in Definition 2.4, the order of this finite group $\mathfrak{R}_{\mathbf{x}}$ will be called the *ramification index* $\tau = \tau(\mathbf{y}) \geq 1$; and \mathbf{y} will be called *unramified* if $\tau(\mathbf{y}) = 1$.

We will need the following.

Lemma 2.17 (Invariant Metrics). *In the real case, given any finite subgroup $\Gamma \subset \mathbf{G}$ there exists a smooth Γ -invariant Riemannian metric on the space \mathbf{X} . Similarly, in the complex case \mathbf{X} has a smooth Γ -invariant Hermitian metric.*

Proof. Starting with an arbitrary smooth Riemannian or Hermitian metric, average over its transforms⁹ by elements of Γ . Then each element of Γ will represent an isometry for the averaged metric. \square

⁹A Riemannian metric can be described as a smooth function μ which assigns to each $\mathbf{x} \in \mathbf{X}$ a symmetric positive definite inner product $\mu_{\mathbf{x}}(v, w)$ on the vector space $T_{\mathbf{x}}\mathbf{X}$ of tangent vectors at \mathbf{x} . Given any diffeomorphism $f : \mathbf{X} \rightarrow \mathbf{X}'$, and given a Riemannian metric μ on \mathbf{X}' , we can use the first

Now let \mathbf{F} be any fiber with finite stabilizers, and let $\mathbf{x}_0 \in \mathbf{F}$ be an arbitrary base point. Since the stabilizer $\mathbf{G}_{\mathbf{x}_0}$ is a finite group acting on \mathbf{X} , we can choose a $\mathbf{G}_{\mathbf{x}_0}$ -invariant metric on X . Using this metric, the transverse vector space $V_{\mathbf{x}_0} = T_{\mathbf{x}_0}\mathbf{X}/T_{\mathbf{x}_0}\mathbf{F}$ can be identified with the normal vector space consisting of all tangent vectors to \mathbf{X} at \mathbf{x}_0 which are orthogonal to the fiber at \mathbf{x}_0 . Given $\varepsilon > 0$, we can consider geodesics of length ε starting at \mathbf{x}_0 which are orthogonal to \mathbf{F} at \mathbf{x}_0 . If ε is small enough, these geodesics will sweep out a smooth d -dimensional disk D_ε which meets \mathbf{F} transversally, where d is the codimension of \mathbf{F} in \mathbf{X} . Since the transverse disk D_ε is canonically diffeomorphic to the ε -disk in $V_{\mathbf{x}_0}$ by this construction, it follows that the group $\mathfrak{R}_{\mathbf{x}_0}$ acts effectively on D_ε .

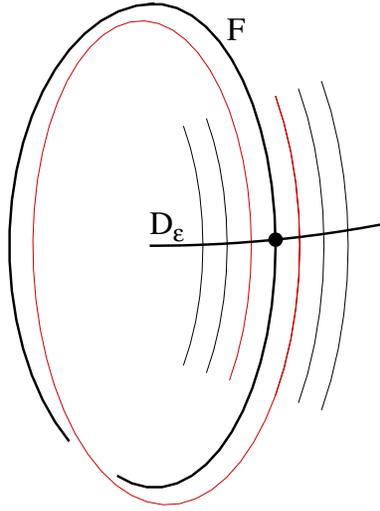


FIG. 3. A transverse D_ε to the fiber \mathbf{F} and several nearby fibers. In this example, a neighborhood of \mathbf{F} within \mathbf{X} is a Möbius band.

We will prove the following.

Theorem 2.18 (Weak Orbifold Theorem). *If the action of \mathbf{G} on \mathbf{X} is weakly proper, with finite stabilizers, then the quotient space \mathbf{Y} is a weak orbifold. More explicitly, given $\mathbf{x} \in \mathbf{X}$, let $\mathbf{y} = \pi(\mathbf{x})$ be its image in the quotient space $\mathbf{Y} = \mathbf{X}/\mathbf{G}$. Then \mathbf{Y} is locally homeomorphic at \mathbf{y} to the quotient of the d -dimensional transverse vector space $V_{\mathbf{x}}$ by the action of the finite group $\mathfrak{R}_{\mathbf{x}}$, which acts linearly on $V_{\mathbf{x}}$. In particular, \mathbf{Y} is locally Hausdorff and metrizable near \mathbf{y} , and also locally compact. Furthermore, the projection map from a small transverse disk D_ε to \mathbf{Y} is τ -to-one throughout a dense open subset of D_ε (or over a dense open subset of \mathbf{Y}), where τ is the order of $\mathfrak{R}_{\mathbf{x}}$.*

derivative map $f_* : T_{\mathbf{x}}\mathbf{X} \xrightarrow{\cong} T_{f(\mathbf{x})}\mathbf{X}'$ to pull back the metric, setting

$$(f^*\mu)_{\mathbf{x}}(v, w) = \mu_{f(\mathbf{x})}(f_*(v), f_*(w)) \in \mathbb{R}.$$

In particular, given any finite group Γ consisting of r diffeomorphisms from \mathbf{X} to itself, we can form the average $\hat{\mu} = \frac{1}{r} \sum_{\mathbf{g} \in \Gamma} \mathbf{g}^*\mu$. The construction in the complex case is similar, using Hermitian inner products.

Corollary 2.19. *Given an action of a Lie group \mathbf{G} on a manifold \mathbf{X} , with finite stabilizers, there are three well defined open subsets*

$$U_{\text{LP}} \subset U_{\text{WP}} \subset U_{\text{LHaus}} \subset \mathbf{X}/\mathbf{G} .$$

Here U_{LP} is the set of locally proper points, U_{WP} is the set of weakly proper points, and U_{LHaus} is the set of all locally Hausdorff points.

This Corollary follows easily from the Theorem and the discussion above. The proof of Theorem 2.18 will depend on the following.

Lemma 2.20 (Local Product Structure). *Using the \mathbf{G}_x -invariant metric on \mathbf{x} , let D_ε be the open disk swept out by normal geodesics of length less than ε at \mathbf{x} . Then any translated disk $\mathbf{g}(D_\varepsilon)$ is determined uniquely by its center point $\mathbf{x}' = \mathbf{g}(\mathbf{x})$. For \mathbf{x}' near \mathbf{x} in \mathbf{F} , two such translated disks are disjoint, unless they have the same center point. It follows that some neighborhood of \mathbf{x} in \mathbf{X} is diffeomorphic to the product of D_ε with a neighborhood of the identity element in \mathbf{G} .*

Proof. (Compare Figure 3.) If $\mathbf{g}_1(\mathbf{x}) = \mathbf{g}_2(\mathbf{x})$, then evidently $\mathbf{g}_1^{-1}\mathbf{g}_2 \in \mathbf{G}_x$. Since elements of \mathbf{G}_x map D_ε to itself, it follows that $\mathbf{g}_1(D_\varepsilon) = \mathbf{g}_2(D_\varepsilon)$. Finally, if W is a small neighborhood of the identity in \mathbf{G} , and if ε is small enough, then since the tangent space to \mathbf{X} at \mathbf{x} is the direct sum of the tangent space to \mathbf{F} and the space of normal vectors at \mathbf{x} , it follows from the Inverse Function Theorem that the map $(\mathbf{g}, \delta) \mapsto \mathbf{g}(\delta)$ sends $W \times D_\varepsilon$ diffeomorphically onto an open subset of \mathbf{X} . \square

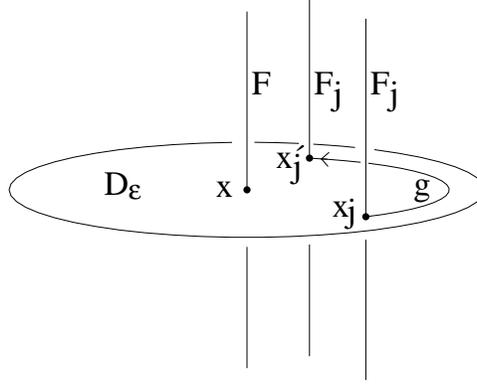


FIG. 4. In the weakly proper case, two points of D_ε belong to the same fiber only if there is an element $\mathbf{g} \in \mathbf{G}_x$ carrying one to the other.

Proof of Theorem 2.18. Note first that elements of \mathbf{G}_x carry geodesics to geodesics, mapping a small disk D_ε onto itself, and mapping each fiber onto itself. We must first show that two points of a sufficiently small disk D_ε will belong to the same fiber only if some element of \mathbf{G}_x maps one to the other. (Compare Figure 4.) Suppose, for arbitrarily large $j > 0$, that there exist points \mathbf{x}_j and \mathbf{x}_j' in $D_{1/j}$ which belong to the same fiber, so that $\mathbf{x}_j' = \mathbf{g}_j(\mathbf{x}_j)$ for some $\mathbf{g}_j \in \mathbf{G}$; but so that \mathbf{x}_j and \mathbf{x}_j' are not in the same orbit under \mathbf{G}_x . Since the action is weakly proper, we can choose these group elements \mathbf{g}_j within a

compact subset $K \subset \mathbf{G}$. After passing to an infinite subsequence, we can assume that these elements \mathbf{g}_j tend to a limit $\mathbf{g} \in K$. Since $\mathbf{g}_j(\mathbf{x}_j) = \mathbf{x}_j'$ with both \mathbf{x}_j and \mathbf{x}_j' tending to \mathbf{x} , it follows by continuity that $\mathbf{g}(\mathbf{x}) = \mathbf{x}$. Therefore $\mathbf{g}_j(\mathbf{x})$ tends to \mathbf{x} as $j \rightarrow \infty$ within the subsequence. Thus, by Lemma 2.20, each image $\mathbf{g}_j(D_\varepsilon)$ is either equal to or disjoint from D_ε . Since $\mathbf{g}_j(\mathbf{x}_j) = \mathbf{x}_j' \in D_\varepsilon$, it follows that $\mathbf{g}_j \in \mathbf{G}_\mathbf{x}$ whenever j is sufficiently large, as required.

This shows that the quotient D_ε/\mathbf{G} maps bijectively to its image in \mathbf{X}/\mathbf{G} . If ε is small enough, the same will be true for the compact disk \overline{D}_ε . It is easy to see that the quotient of any compact metric space by a finite group action is compact, with an induced metric. In fact, we can use the Hausdorff metric on the space of all non-empty compact subsets; and each \mathbf{G} -orbit is such a compact subset. Therefore, under the hypothesis of Theorem 2.18 it follows that the quotient space is locally compact, metric, and hence Hausdorff, near \mathbf{y} . Since the action of $\mathfrak{R}_\mathbf{x}$ on the transverse vector space $V_\mathbf{x}$ is linear and effective, it follows, for each non-trivial cyclic subgroup of $\mathfrak{R}_\mathbf{x}$, that the action is free except on some proper linear subspace of $V_\mathbf{x}$. Therefore the projection map from D_ε to \mathbf{Y} is \mathfrak{r} -to-one except on a finite union of linear subspaces. \square

After a few pages of discussion, we will return to the construction of a full orbifold structure, beginning with Lemma 2.25 below.

Miscellaneous Remarks and Examples.

Remark 2.21. In the unramified case, of course the quotient $\mathbf{Y} = \mathbf{X}/\mathbf{G}$ inherits the structure of a real or complex analytic manifold from \mathbf{X} . However, in general \mathbf{Y} need not be even a topological manifold. Perhaps the simplest non-manifold example is the quotient of the Euclidean space \mathbb{R}^3 by the two element group $\{\pm 1\}$ acting by $\mathbf{g}(\mathbf{x}) = \pm \mathbf{x}$. In this case, the quotient is not locally orientable near the origin. Similarly the quotient of \mathbb{C}^2 by the corresponding group is not a manifold, since a small neighborhood of the origin in the quotient space, with the origin removed, is not simply-connected.

Remark 2.22 (Rational Homology Manifolds). We will show that:

*Any complex weak orbifold or any locally orientable real weak orbifold is a **rational homology manifold**, in the sense that any point of such an orbifold has a neighborhood homeomorphic to the cone over a space with the rational homology of a sphere.*

Here the local orientability condition in the real case means that the space \mathbf{Y} must be locally of the form D_ε/Γ , where the action of the finite group Γ preserves orientation. In other words Γ must be contained in the rotation group $\text{SO}(d)$, rather than the full orthogonal group $\text{O}(d)$.

The statement then follows from the following more general principle:

Lemma 2.23. *If a finite group Γ acts on a finite cell complex K , then the rational homology $H_*(K/\Gamma; \mathbb{Q})$ is isomorphic to the subgroup*

$$H_*(K; \mathbb{Q})_\Gamma \subset H_*(K; \mathbb{Q})$$

consisting of all elements which are fixed under the induced action of Γ .

Proof. After passing to a suitable subdivision of the cell complex K we may assume that each group element which maps a cell onto itself acts as the identity map on this cell. Choosing some orientation for each cell, the associated chain complex $C_*(K) = C_*(K; \mathbb{Q})$

is the graded rational vector space with one basis element for each cell. The projection map $\pi : K \rightarrow K/\Gamma$ induces a chain mapping $\pi_* : C_*(K) \rightarrow C_*(K/\Gamma)$ between these rational chain complexes. That is, π_n maps $C_n(K)$ to $C_n(K/\Gamma)$, and commutes with the boundary operator $\partial : C_n(K) \rightarrow C_{n-1}(K)$ i.e., we obtain the following commutative diagram

$$\begin{array}{ccc} C_n(K) & \xrightarrow{\pi_n} & C_n(K/\Gamma) \\ \partial \downarrow & & \downarrow \partial \\ C_{n-1}(K) & \xrightarrow{\pi_{n-1}} & C_{n-1}(K/\Gamma) . \end{array}$$

But there is also a less familiar chain map

$$\pi^* : C_*(K/\Gamma) \rightarrow C_*(K)$$

in the other direction, which sends each cell of K/Γ to the weighted sum of the cells of K which lie over it. Here each such cell σ of K is to be weighted by the number of elements in the stabilizer $\Gamma_\sigma \subset \Gamma$. Then the composition

$$C_*(K/\Gamma) \xrightarrow{\pi^*} C_*(K) \xrightarrow{\pi_*} C_*(K/\Gamma)$$

is just multiplication by the order of Γ ; so the same is true of the induced composition

$$H_*(K/\Gamma) \xrightarrow{\pi^*} H_*(K) \xrightarrow{\pi_*} H_*(K/\Gamma)$$

of rational homology groups. Since this composition is bijective, it follows easily that $H_*(K/\Gamma)$ maps isomorphically onto its image in $H_*(K)$; and also that $H_*(K)$ splits as the direct sum of the image of π^* and the kernel of π_* . On the other hand the other composition

$$H_*(K) \xrightarrow{\pi_*} H_*(K/\Gamma) \xrightarrow{\pi^*} H_*(K)$$

maps each element of $H_*(K)$ to the sum of its images under the various elements of Γ . It follows that the kernel of π_* is the subspace consisting of all elements $\eta \in H_*(K)$ such that $\sum_{\gamma \in \Gamma} \gamma_*(\eta) = 0$. Since every element of the image of π^* is Γ -invariant, and no non-zero element of the kernel of π_* is Γ -invariant, the conclusion follows. (We thank Dennis Sullivan for supplying this argument.) \square

In particular, if K is a rational homology sphere and the action of Γ preserves orientation, then it follows that K/Γ is also a rational homology sphere.

Now suppose as in Theorem 2.18 that the quotient space \mathbf{Y} is locally homeomorphic to \mathbb{R}^d/Γ , where Γ is now the ramification group. Then we can choose a Γ -invariant simplicial structure on \mathbb{R}^d . Taking K to be the star boundary of the origin; that is the boundary of the union of all closed simplexes which contain the origin, it follows that K/Γ is a homology $(d-1)$ -sphere, and hence that \mathbf{Y}/Γ is a rational homology d -manifold. The corresponding statement in the complex case follows easily.

Remark 2.24 (Quotient Analytic Structures). Whether or not the quotient \mathbf{Y} is a (possibly non-Hausdorff) topological manifold, we can put some kind of ‘‘analytic structure’’ on it as follows. (Recall that we use the word analytic as an abbreviation for real analytic in the real case, and complex analytic in the complex case.) To every open subset $\mathbf{Y}' \subset \mathbf{Y}$ assign the algebra $\mathcal{A}(\mathbf{Y}')$ consisting of all functions $f : \mathbf{Y}' \rightarrow \mathbb{R}$ in the real case (or $f : \mathbf{Y}' \rightarrow \mathbb{C}$ in the complex case), such that the composition $f \circ \pi$ mapping $\pi^{-1}(\mathbf{Y}')$ to \mathbb{R} (or \mathbb{C}) is analytic, where $\pi : \mathbf{X} \rightarrow \mathbf{Y}$ is the projection map.

Definition. We say that the orbifold \mathbf{Y} is a d -dimensional smooth *analytic manifold* if for every point of \mathbf{Y} there is a neighborhood \mathbf{Y}' and functions $f_1, \dots, f_d \in \mathcal{A}(\mathbf{Y}')$ such that:

- (1) The correspondence $\mathbf{y} \mapsto (f_1(\mathbf{y}), \dots, f_d(\mathbf{y}))$ maps \mathbf{Y}' homeomorphically onto an open subset of¹⁰ \mathbb{R}^d or \mathbb{C}^d ; and
- (2) Every element $f \in \mathcal{A}(\mathbf{Y}')$ can be expressed as an analytic function of f_1, \dots, f_d .

In other words, for every such f there must be an analytic function F , defined on some open subset of \mathbb{R}^d or \mathbb{C}^d , such that

$$f(\mathbf{y}) = F(f_1(\mathbf{y}), \dots, f_d(\mathbf{y})) \quad \text{for all } \mathbf{y} \in \mathbf{Y}' .$$

Smooth Examples. If a cyclic group acts on a one-dimensional complex manifold, then it is easy to see that the quotient space is always an analytic manifold. Another classical example is the following: Let \mathfrak{S}_n be the symmetric group on n elements acting on \mathbb{C}^n by permuting the n coordinates. Then the quotient $\mathbb{C}^n/\mathfrak{S}_n$ is a smooth manifold which is isomorphic to \mathbb{C}^n itself.¹¹ We can simply choose f_1, \dots, f_n to be the elementary symmetric functions of the n coordinates.

Such smooth examples seem to be rather rare when $n \geq 2$. Here is more typical example of a ramified action with a topological manifold as quotient.

A Non-Smooth Manifold Example. Let the two element group $\{\pm 1\}$ act on \mathbb{R}^2 by

$$(x, y) \mapsto \pm(x, y) .$$

Then the quotient space \mathbf{Y} is clearly homeomorphic to \mathbb{R}^2 . In fact if we introduce the complex variable $z = x + iy$, then $z^2 = x^2 - y^2 + 2ixy$ provides a good complex parametrization. However, as a real orbifold \mathbf{Y} does not have a smooth analytic structure. To see this, note that the set $\mathcal{A}(\mathbf{Y})$ consists of all maps $\mathbf{Y} \rightarrow \mathbb{R}$ which can be expressed as real analytic functions of

$$x^2, y^2, \quad \text{and} \quad xy .$$

There is no way of choosing just two functions f_1 and f_2 in $\mathcal{A}(\mathbf{Y})$ so that every element of $\mathcal{A}(\mathbf{Y})$ can be expressed as a smooth function of f_1 and f_2 . In fact, the correspondence

$$(x, y) \mapsto (\xi, \eta, \zeta) = (x^2, y^2, xy)$$

sends the real plane \mathbb{R}^2 homeomorphically onto a topological submanifold of \mathbb{R}^3 which is clearly not smooth, since it projects onto the positive quadrant in the (ξ, η) -plane. It follows from this that Condition (1) of the Definition cannot be satisfied at the origin.

A Wild Topological Manifold Example. Let \mathbb{H} be the space of quaternions. We will give an example of a finite group G_{120} acting smoothly on \mathbb{H} with the following rather startling property. The quotient space $\mathbb{R} \times (\mathbb{H}/G_{120}) = (\mathbb{R} \times \mathbb{H})/G_{120}$ is homeomorphic to \mathbb{R}^5 ; but the set of ramified points $\mathbb{R} \times \mathbf{0}$ corresponds to a line in this quotient space which is so wildly embedded that its complement is not simply connected.

¹⁰In the real case, it would be natural to also include manifolds-with-boundary by allowing a closed half-space as model space in Item (1) above. In fact, one could also include “manifolds with corners” by allowing a convex polyhedron as model space. (Compare Example 2.9.)

¹¹However the real case is quite different, since $\mathbb{R}^n/\mathfrak{S}_n$ is not isomorphic to \mathbb{R}^n . In fact the quotient $\mathbb{R}^n/\mathfrak{S}_n$ can best be identified with the convex polyhedron $x_1 \leq x_2 \leq \dots \leq x_n$ in \mathbb{R}^n .

To begin the construction, note that the unit sphere $S^3 \subset \mathbb{H}$ is isomorphic to the universal covering group of the rotation group $\mathrm{SO}(3)$. The 60 element icosahedral subgroup of $\mathrm{SO}(3)$ is covered by the 120 element **double icosahedral group** $G_{120} \subset S^3$. The quotient space S^3/G_{120} is the ‘‘Poincaré fake sphere’’, with the homology of the standard 3-sphere, but with fundamental group G_{120} . If we let the group G_{120} act on \mathbb{H} by left multiplication, then the quotient \mathbb{H}/G_{120} is not a manifold, since a punctured neighborhood of the origin is not simply-connected. However, the ‘‘double suspension theorem’’ of Cannon and Edwards implies that the product

$$\mathbb{R} \times (\mathbb{H}/G_{120}) \cong (\mathbb{R} \times \mathbb{H})/G_{120}$$

is a simply-connected manifold, homeomorphic to \mathbb{R}^5 . (Compare [Ca] or [Ed].) This product cannot be given any differentiable structure such that the subset $\mathbb{R} \times \mathbf{0}$ of ramified points is a differentiable submanifold. This follows since the complement of this one-dimensional topological submanifold has fundamental group G_{120} .

Locally Proper Actions and Orbifolds.

Recall that the action is **locally proper** at \mathbf{x} if it is proper throughout some \mathbf{G} -invariant neighborhood of \mathbf{x} . (Definition 2.10.) One important property of locally proper actions is the following.

Lemma 2.25. *If the action is locally proper at \mathbf{x} with finite stabilizers, then for all \mathbf{x}' sufficiently close to \mathbf{x} the stabilizer $\mathbf{G}_{\mathbf{x}'}$ is isomorphic to a subgroup of $\mathbf{G}_{\mathbf{x}}$. In particular, the order $|\mathbf{G}_{\mathbf{x}}|$ of the stabilizer is upper semi-continuous as a function of \mathbf{x} , so that $|\mathbf{G}_{\mathbf{x}'}| \leq |\mathbf{G}_{\mathbf{x}}|$ for all \mathbf{x}' sufficiently close to \mathbf{x} .*

Proof. Suppose that there were points \mathbf{x}'_j arbitrarily close to \mathbf{x} with $\mathbf{G}_{\mathbf{x}'_j}$ not isomorphic to a subgroup of $\mathbf{G}_{\mathbf{x}}$. Since the action is locally proper, there is a compact set $K \subset \mathbf{G}$ such that the stabilizer $\mathbf{G}_{\mathbf{x}'}$ is contained in K for all \mathbf{x}' in some neighborhood of \mathbf{x} . The collection of all compact subsets of K forms a compact metric space, using the Hausdorff metric. Therefore, given any sequence of such points \mathbf{x}'_j converging to \mathbf{x} , after passing to an infinite subsequence we can assume that the sequence of finite groups $\mathbf{G}_{\mathbf{x}'_j} \subset K$ converges to a Hausdorff limit set $\mathbf{G}' \subset \mathbf{G}_{\mathbf{x}}$ as j tends to infinity. This limit \mathbf{G}' must be a subgroup of $\mathbf{G}_{\mathbf{x}}$: Otherwise, for any $\mathbf{g}' \in \mathbf{G}' \setminus \mathbf{G}_{\mathbf{x}}$, we would have points \mathbf{x}'_j converging to \mathbf{x} and group elements $\mathbf{g}_j \in \mathbf{G}$ converging to \mathbf{g}' , with $\mathbf{g}_j(\mathbf{x}'_j) = \mathbf{x}'_j$ but $\mathbf{g}'(\mathbf{x}) \neq \mathbf{x}$.

This group $\mathbf{G}_{\mathbf{x}'_j}$ must be isomorphic to \mathbf{G}' for large j . In fact, since $\mathbf{G}_{\mathbf{x}'_j}$ converges to the finite group \mathbf{G}' as $j \rightarrow \infty$, the correspondence which maps each $\mathbf{g} \in \mathbf{G}_{\mathbf{x}'_j}$ to the closest point of \mathbf{G}' must be a surjective homomorphism for large j . Since we are assuming that $\mathbf{G}_{\mathbf{x}'_j}$ is not isomorphic to any subgroup of $\mathbf{G}_{\mathbf{x}}$, it follows that the kernel of this surjection $\mathbf{G}_{\mathbf{x}'_j} \rightarrow \mathbf{G}'$ must contain some non-identity element \mathbf{g}'_j of \mathbf{G} . But the sequence $\{\mathbf{g}'_j\}$ must converge to the identity element. Now consider the exponential map $\exp : \mathcal{L} \rightarrow \mathbf{G}$, which maps a neighborhood of the zero element in the Lie algebra to a neighborhood of the identity. (Recall that the Lie algebra \mathcal{L} can be identified with the tangent space to \mathbf{G} at the identity element.) We can set $\mathbf{g}'_j = \exp(v_j)$ where v_j tends to zero. Thus the group generated by \mathbf{g}'_j corresponds to the set of all images $\exp(kv_j)$ with $k \in \mathbb{Z}$. Clearly these images fill out the corresponding one-parameter subgroup more and more densely as $v_j \rightarrow 0$, so that the Hausdorff limit could not be a finite group. This contradicts our hypothesis that $\mathbf{G}_{\mathbf{x}}$ is finite; and hence completes the proof. \square

Here is a more precise statement.

Lemma 2.26. *If the action is locally proper, then each stabilizer $\mathbf{G}_{\mathbf{x}}$ contains a subgroup $\mathbf{G}_{\mathbf{x}}^0$ which varies continuously with \mathbf{x} , and such that $\mathbf{G}_{\mathbf{x}'}^0 = \mathbf{G}_{\mathbf{x}'}$ for all \mathbf{x}' in a dense open set. If \mathbf{X} is connected, then it follows that the various subgroups $\mathbf{G}_{\mathbf{x}}^0$ are all mutually isomorphic.*

As a typical example, in the space of smooth cubic curves in $\mathbb{P}^2(\mathbb{C})$, every stabilizer $\mathbf{G}_{\mathbf{x}}$ contains a subgroup $\mathbf{G}_{\mathbf{x}}^0$ of order eighteen, with $\mathbf{G}_{\mathbf{x}}$ equal to $\mathbf{G}_{\mathbf{x}}^0$ except along two fibers where there are extra symmetries. (See [BM], and compare Section 5.)

Proof of Lemma 2.26. Recall that the ramification index $r_{\mathbf{y}}$ at the image of \mathbf{x} in the quotient space is the order of the quotient group $\mathfrak{R}_{\mathbf{y}} = \mathbf{G}_{\mathbf{x}}/\mathbf{H}_{\mathbf{x}}$. Since $r_{\mathbf{y}'} = 1$ throughout a dense open subset of a neighborhood of \mathbf{y} by Lemma 2.5, it follows that $\mathbf{G}_{\mathbf{x}'} = \mathbf{H}_{\mathbf{x}'}$ for \mathbf{x}' in a dense open subset of a neighborhood of \mathbf{x} . On the other hand, arguing as above, we see that the groups $\mathbf{H}_{\mathbf{x}'}$ for this dense open set converge to a group $\mathbf{G}_{\mathbf{x}}^0 \subset \mathbf{H}_{\mathbf{x}}$ as \mathbf{x}' converges to \mathbf{x} . In fact it follows easily that $\mathbf{H}_{\mathbf{x}'}$ contains a subgroup isomorphic to $\mathbf{G}_{\mathbf{x}}^0$ for every \mathbf{x}' in the neighborhood. \square

G-Invariant Tubular Neighborhoods. Given any smooth \mathbf{G} -action with finite stabilizers, and given any fiber $\mathbf{F} \subset \mathbf{X}$, it is not difficult to construct arbitrarily small \mathbf{G} -invariant neighborhoods \mathbf{E}_{ε} of \mathbf{F} in \mathbf{X} . Simply choose a transverse disk $D_{\varepsilon} = D_{\varepsilon}(\mathbf{x}_0)$ to the fiber \mathbf{F} at \mathbf{x}_0 , and let \mathbf{E}_{ε} be the union of its images $\mathbf{g}(D_{\varepsilon})$ as \mathbf{g} varies over \mathbf{G} . Recall from Lemma 2.20 that the image disk $\mathbf{g}(D_{\varepsilon})$ depends only on its center point $\mathbf{x} = \mathbf{g}(\mathbf{x}_0)$. We will use the alternate notation

$$D_{\varepsilon}(\mathbf{x}) = \mathbf{g}(D_{\varepsilon}) \quad \text{whenever} \quad \mathbf{x} = \mathbf{g}(\mathbf{x}_0).$$

If the action is locally proper, we can give a much more precise description.

Theorem 2.27. *If the action is locally proper with finite stabilizers, and if ε is small enough, then the neighborhood \mathbf{E}_{ε} is diffeomorphic to the quotient space of $\mathbf{G} \times D_{\varepsilon}(\mathbf{x}_0)$ by the equivalence relation*

$$(\mathbf{g}\mathbf{g}_0, \delta) \sim (\mathbf{g}, \mathbf{g}_0(\delta))$$

for every $\mathbf{g} \in \mathbf{G}$, $\delta \in D_{\varepsilon}$, and every \mathbf{g}_0 in the stabilizer $\mathbf{G}_{\mathbf{x}_0}$. It follows that \mathbf{E}_{ε} can be described as the total space of a locally trivial fiber bundle, with projection map $\mathbf{E}_{\varepsilon} \rightarrow \mathbf{F}$ which carries each $D_{\varepsilon}(\mathbf{x}) \subset \mathbf{E}_{\varepsilon}$ to its center point $\mathbf{x} \in \mathbf{F}$. Furthermore, the \mathbf{G} -orbits provide a foliation of \mathbf{E}_{ε} which is transverse to every $D_{\varepsilon}(\mathbf{x})$, therefore providing a local product structure around any $D_{\varepsilon}(\mathbf{x})$.

Proof. (Compare [Mei], [DK].) We must first show that the various disks $D_{\varepsilon}(\mathbf{x})$ with $\mathbf{x} \in \mathbf{F}$ are pairwise disjoint.

Step 1. Since the stabilizer is finite, it follows that \mathbf{F} is locally diffeomorphic to \mathbf{G} . Therefore it follows as in Lemma 2.20 that the various image disks $\mathbf{g}(D_{\varepsilon}(\mathbf{x}))$, with \mathbf{x} close to \mathbf{x}_0 in \mathbf{F} , are all disjoint, provided that ε is small enough.

Step 2. Taking $\varepsilon > 0$ as in Step 1. Suppose that there is a sequence of numbers $\varepsilon > \varepsilon_1 > \varepsilon_2 > \dots$ tending to zero such that for each j there are points $\mathbf{x}_j \neq \mathbf{x}'_j$ on \mathbf{F} and group elements \mathbf{g}_j and \mathbf{g}'_j with $\mathbf{g}_j(\mathbf{x}_0) = \mathbf{x}_j$ and $\mathbf{g}'_j(\mathbf{x}_0) = \mathbf{x}'_j$ such that the two disks

$$D_{\varepsilon_j}(\mathbf{x}_j) = \mathbf{g}_j(D_{\varepsilon_j}(\mathbf{x}_0)) \quad \text{and} \quad D_{\varepsilon_j}(\mathbf{x}'_j) = \mathbf{g}'_j(D_{\varepsilon_j}(\mathbf{x}_0))$$

intersect each other at some point \mathbf{x}^* . (See Figure 5.) Then we can write

$$\mathbf{x}^* = \mathbf{g}_j(\delta_j) = \mathbf{g}'_j(\delta'_j)$$

for appropriate points $\delta_j, \delta'_j \in D_{\varepsilon_j} = D_{\varepsilon_j}(\mathbf{x}_0)$. Now setting $\mathbf{g}_j^* = \mathbf{g}_j^{-1}\mathbf{g}'_j$, it follows that the disk $\mathbf{g}_j^*(D_{\varepsilon})$ intersects D_{ε} at the point $\mathbf{g}_j^*(\delta'_j) = \delta_j$.

Since the action is locally proper, it follows that all group elements \mathbf{g} which satisfy $\mathbf{g}(\overline{D_{\varepsilon}}) \cap \overline{D_{\varepsilon}} \neq \emptyset$ must be contained in some compact set $K \subset \mathbf{G}$. After passing to an infinite subsequence, we may assume that the group elements \mathbf{g}_j^* tend to a limit in $\mathbf{g}^* \in K$. Taking the limit of the equation $\mathbf{g}_j^*(\delta'_j) = \delta_j$ as $j \rightarrow \infty$, since both δ_j and δ'_j must tend to \mathbf{x}_0 , it follows that $\mathbf{g}^*(\mathbf{x}_0) = \mathbf{x}_0$. Therefore the sequence $\mathbf{g}_j^*(\mathbf{x}_0)$ must tend to \mathbf{x}_0 . Thus for some fixed $\varepsilon > 0$ we have constructed disks $\mathbf{g}_j^*(D_{\varepsilon})$ with center point arbitrarily close to \mathbf{x}_0 which intersect D_{ε} but are not equal to D_{ε} . This contradicts Step 1, and proves that *all* of the disks $D_{\varepsilon}(\mathbf{x})$ must be pairwise disjoint.

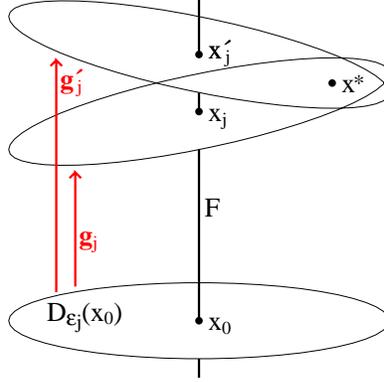


FIG. 5. Illustrating the proof of Theorem 2.27.

It follows from Lemma 2.20 that the mapping $\mathbf{E}_{\varepsilon} \rightarrow \mathbf{F}$ has a local product structure near the disk D_{ε} . We can use translation by any group element \mathbf{g} to translate this product structure to a neighborhood of any disk $\mathbf{g}(D_{\varepsilon})$. Further details of the proof of Theorem 2.27 are straightforward. \square

Corollary 2.28. *If the action is locally proper with finite stabilizers, then the quotient is an orbifold (Definition 2.7).*

Proof. We know from Theorem 2.18 that the quotient is a weak orbifold. Choose a disk D_{ε} as in Theorem 2.27, and let \mathbf{E}_{ε} be the associated tubular neighborhood: the union of all translates of D_{ε} by elements of \mathbf{G} . For any fiber \mathbf{F}' which intersects D_{ε} , we must study how a sufficiently small tubular neighborhood of \mathbf{F}' is related to \mathbf{E}_{ε} . Let \mathbf{x}' be an arbitrary base point on \mathbf{F}' , and let $D'_{\varepsilon'}$ be a small transverse disk to \mathbf{F}' at \mathbf{x}' , using a $\mathbf{G}_{\mathbf{x}'}$ -invariant metric. Choose some group element \mathbf{g}_0 which maps \mathbf{x}' to a point $\mathbf{x}'' \in D_{\varepsilon} \cap \mathbf{F}'$. Then $D'_{\varepsilon'}$ maps to a disk which is transverse to \mathbf{F}' at \mathbf{x}'' . Using the local product structure, if ε' is small enough, we can project $\mathbf{g}_0(D'_{\varepsilon'})$ diffeomorphically onto a subdisk $D'' \subset D_{\varepsilon}$ as illustrated in Figure 6.

Now consider the orbifold chart $D_{\varepsilon} \rightarrow D_{\varepsilon}/\mathfrak{R}_{\mathbf{x}} \subset \mathbf{Y} = \mathbf{X}/\mathbf{G}$, as well as the corresponding chart $D'_{\varepsilon'} \rightarrow D'_{\varepsilon'}/\mathfrak{R}_{\mathbf{x}'}$. Using Lemmas 2.25 and 2.26, we see that for any $\mathbf{g} \in \mathfrak{R}_{\mathbf{x}'}$ the

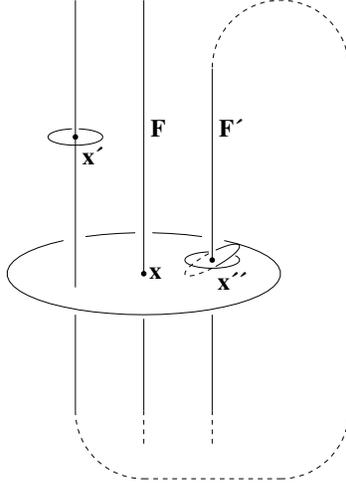


FIG. 6. Illustrating the proof of Corollary 2.28.

action of \mathbf{g} on $D'_{\epsilon'}$ will correspond to the action of some uniquely defined $\phi(\mathbf{g}) \in \mathfrak{A}_{\mathbf{x}}$ on the image disk in D_{ϵ} . Thus the compatibility condition of Definition 2.7 is satisfied. \square

Remark 2.29 (The Projective Linear Group and Projective Space). For the rest of this paper, the group \mathbf{G} will be the real or complex projective linear group PGL_m , which acts on the corresponding projective space \mathbb{P}^{m-1} , with $m = 2$ in Section 3, or $m = 3$ in later sections.

Over any field, the corresponding group PGL_m can be defined as the quotient GL_m/N , where GL_m is the group of linear automorphisms of an m -dimensional vector space, and N is the normal subgroup consisting of scalar transformations $\mathbf{x} \mapsto t\mathbf{x}$. Here t can be any fixed non-zero field element. Writing the linear transformation as

$$(x_1, x_2, \dots, x_m) \mapsto (x'_1, x'_2, \dots, x'_m),$$

there is an associated automorphism $(x_1 : x_2 : \dots : x_m) \mapsto (x'_1 : x'_2 : \dots : x'_m)$ of the $(m - 1)$ -dimensional projective space over the field. Automorphisms obtained in this way are called **projective automorphisms**. Thus:

Over any field, PGL_m can be identified with the group of all projective automorphisms of the projective space \mathbb{P}^{m-1} .

Equivalently, PGL_m can be described as the group of all equivalence classes of non-singular $m \times m$ matrices over the field, where two matrices are equivalent if one can be obtained from the other by multiplication by a non-zero constant. Let \mathbb{P}^{m^2-1} be the projective space consisting of all lines through the origin in the m^2 -dimensional vector space consisting of $m \times m$ matrices. Then it follows easily that:

Over any field, the group PGL_m can be considered as a Zariski open subset of the projective space \mathbb{P}^{m^2-1} .

Specializing to the real or complex case, it follows that PGL_m is a smooth real or complex manifold of dimension $m^2 - 1$, with a smooth product operation. Hence it is a Lie group.

We will need the following statement in Sections 3 and 6.

Lemma 2.30. *Every element of $\mathrm{PGL}_m(\mathbb{R})$ or $\mathrm{PGL}_m(\mathbb{C})$ can be written as a composition*

$$\mathbf{g} = \mathbf{r} \circ \mathbf{d} \circ \mathbf{r}' ,$$

where \mathbf{r} and \mathbf{r}' are isometries, that is elements of the projective orthogonal group PO_m in the real case or the projective unitary group PU_m in the complex case, and where \mathbf{d} is a diagonal transformation of the form

$$\mathbf{d}(x_1 : \cdots : x_m) = (a_1 x_1 : \cdots : a_m x_m)$$

where the a_j are real numbers with $a_1 \geq a_2 \geq \cdots \geq a_m > 0$. Furthermore, the ratios $(a_1 : a_2 : \cdots : a_m)$ are uniquely determined by \mathbf{g} (although \mathbf{r} and \mathbf{r}' may not be uniquely determined).

In particular, the invariant $a_1/a_m \geq 1$ provides a rough measure of how far the automorphism \mathbf{g} is from being an isometry with respect to the standard metric for \mathbb{P}^{m-1} . (See Definition 2.31.)

Proof of Lemma 2.30. To fix ideas we will discuss only the complex case; but the real case is completely analogous. This is proved by applying the Gram-Schmidt process to a corresponding linear transformation $\ell : V \rightarrow V'$, where V and V' are m -dimensional complex vector spaces with Hermitian inner product and with associated norm $\|v\| = \sqrt{v \cdot v}$. Given a linear bijection $\ell : V \rightarrow V'$, choose a unit vector $u_1 \in V$ which maximizes the norm $\|\ell(u_1)\|$. Then $\ell(u_1)$ can be written as a product $a_1 u'_1$ where $a_1 > 0$ is this maximal norm, and where u'_1 is a unit vector in V' . Note that ℓ maps any unit vector v orthogonal to u_1 in V to a vector v' orthogonal to u'_1 in V' . In fact, each linear combination $u_1 \cos(\theta) + v \sin(\theta)$ is another unit vector in V , which maps to $a_1 u'_1 \cos(\theta) + v' \sin(\theta)$ in V' . A brief computation shows that the derivative of the squared norm of this image vector with respect to θ at $\theta = 0$ is $2a_1 u'_1 \cdot v'$. Since the derivative at a maximum point must be zero, this proves that $u'_1 \cdot v' = 0$, as asserted.

Thus ℓ maps the orthogonal complement of u_1 to the orthogonal complement of u'_1 . Repeating the same argument for this map of orthogonal complements we find unit vectors u_2 orthogonal to u_1 and u'_2 orthogonal to u'_1 so that $\ell(u_2) = a_2 u'_2$ with $a_1 \geq a_2 > 0$. Continuing inductively, we find an orthonormal basis $\{u_j\}$ for V and an orthonormal basis $\{u'_j\}$ for V' so that

$$\ell(u_j) = a_j u'_j \quad \text{with} \quad a_1 \geq a_2 \geq \cdots \geq a_m > 0 .$$

Now taking V and V' to be copies of the standard \mathbb{C}^m , it follows that ℓ is the composition of:

- (1) a unitary transformation which takes the standard basis for \mathbb{C}^m to the basis $\{u_j\}$,
- (2) a diagonal transformation of the required form, and
- (3) a unitary transformation taking $\{u'_j\}$ to the standard basis.

This statement about the general linear group GL_m clearly implies the required statement about the projective linear group PGL_m . This proves Lemma 2.30. \square

Definition 2.31 (The Fubini-Study Metric). We will sometimes refer to distances between points of projective space. Except in the one-dimensional case, this will mean distance in the *Fubini-Study metric*, which can be described as follows. It clearly suffices to consider the complex case. Let \mathbf{u} be a unit vector in \mathbb{C}^{n+1} , so that $\mathbf{u} \cdot \mathbf{u} = 1$, using the standard Hermitian inner product. A point in \mathbb{P}^n can be described as an equivalence class (\mathbf{u}) of such unit vectors, where \mathbf{u} is equivalent to \mathbf{u}' if and only if $\mathbf{u}' = e^{it}\mathbf{u}$ for some real t . Now let \mathbf{u} and \mathbf{v} be fixed orthogonal unit vectors, $\mathbf{u} \cdot \mathbf{v} = 0$. If we think of \mathbf{u} and \mathbf{v} as points on a real sphere of dimension $2n + 1$, then there is a unique geodesic $\theta \mapsto \mathbf{u} \cos(\theta) + \mathbf{v} \sin(\theta)$ joining these two points, parametrized by arclength, with $0 \leq \theta \leq \pi/2$. By definition, the corresponding locus

$$\ell(\theta) = (\mathbf{u} \cos(\theta) + \mathbf{v} \sin(\theta))$$

in \mathbb{P}^n is a geodesic parameterized by arclength for the Fubini-Study metric.¹² It follows that the Fubini-Study distance between two points on such a geodesic is given by

$$D(\ell(\theta), \ell(\theta')) = |\theta' - \theta|.$$

More generally, since

$$\ell(\theta) \cdot \ell(\theta') = \cos(\theta) \cos(\theta') + \sin(\theta) \sin(\theta') = \cos(\theta' - \theta),$$

it follows easily that we can define the distance between any two points (\mathbf{u}_1) and (\mathbf{u}_2) by the closely related formula

$$D((\mathbf{u}_1), (\mathbf{u}_2)) = \arccos(|\mathbf{u}_1 \cdot \mathbf{u}_2|) \in [0, \pi/2].$$

The triangle inequality for distance defined in this way follows from the triangle inequality for the standard sphere.

3. THE MODULI SPACE FOR EFFECTIVE DIVISORS OF DEGREE n .

We first look at a basic family of moduli spaces which are relatively easy to understand. Since the discussions in the real and complex cases are very similar, it will be convenient to use the symbol \mathbb{F} to denote either \mathbb{R} or \mathbb{C} . The projective line $\mathbb{P}^1 = \mathbb{P}^1(\mathbb{F})$ is a circle in the real case, or a Riemann sphere in the complex case. It is often convenient to identify \mathbb{P}^1 with the union $\widehat{\mathbb{F}} = \mathbb{F} \cup \{\infty\}$. More precisely, each point $(x : y) \in \mathbb{P}^1$ can be identified with the quotient $x/y \in \widehat{\mathbb{F}} = \mathbb{F} \cup \{\infty\}$. Note that the group $\mathbf{G} = \mathrm{PGL}_2(\mathbb{F})$ acting on $\mathbb{P}^1(\mathbb{F})$ corresponds to the group of fractional linear transformations,

$$z \mapsto \frac{az + b}{cz + d} \quad \text{with} \quad a, b, c, d \in \mathbb{F}, \quad ad - bc \neq 0,$$

acting on $\widehat{\mathbb{F}}$.

By definition, an *effective divisor* of degree n on \mathbb{P}^1 is a formal sum of the form

$$\mathcal{D} = m_1 \langle \mathbf{p}_1 \rangle + \cdots + m_k \langle \mathbf{p}_k \rangle,$$

where the \mathbf{p}_j are distinct points of \mathbb{P}^1 , and where the *multiplicities* $m_j \geq 1$ are integers, with $\sum m_j = n$. The set $|\mathcal{D}| = \{\mathbf{p}_1, \dots, \mathbf{p}_k\} \subset \mathbb{P}^1$ will be called the *support* of \mathcal{D} .

Let $\widehat{\mathcal{D}}_n = \widehat{\mathcal{D}}_n(\mathbb{F})$ be the space of all effective divisors of degree n on $\mathbb{P}^1 = \mathbb{P}^1(\mathbb{F})$. (The notation \mathcal{D}_n will be reserved for the open subset consisting of n -tuples of distinct points.) In the complex case, if we think of a divisor as the set of roots of a homogeneous polynomial

¹²There is actually a one-parameter family of minimal geodesics between (\mathbf{u}) and (\mathbf{v}) , since we can replace \mathbf{v} by $e^{it}\mathbf{v}$.

of degree n in two variables, then it follows easily that the space $\widehat{\mathcal{D}}_n(\mathbb{C})$ can be given the structure of a complex projective space $\mathbb{P}^n(\mathbb{C})$. In the real case, $\widehat{\mathcal{D}}_n(\mathbb{R})$ can be identified with the closed subset of $\mathbb{P}^n(\mathbb{R})$ corresponding to those real homogeneous polynomials which have only real roots.

The group $\mathbf{G} = \mathrm{PGL}_2(\mathbb{F})$ acts on \mathbb{P}^1 , and hence on the space $\widehat{\mathcal{D}}_n$ of formal sums. Note that the action on \mathbb{P}^1 is *three point simply transitive*. That is, there is one and only one group element which take any ordered set of three distinct points of \mathbb{P}^1 to any other ordered set of three distinct points.

It follows that the stabilizer $\mathbf{G}_{\mathcal{D}}$ for the action at a point $\mathcal{D} \in \widehat{\mathcal{D}}_n$ is finite if and only if the number k of points in $|\mathcal{D}|$ satisfies $k \geq 3$. In fact $\mathbf{G}_{\mathcal{D}}$ maps injectively into the group of all permutations of the set $|\mathcal{D}|$ whenever $k \geq 3$.

Definition 3.1. Let $\widehat{\mathcal{D}}_n^{\mathrm{fs}}$ be the open subset of $\widehat{\mathcal{D}}_n$ consisting of effective divisors with finite stabilizer, or in other words with at least three distinct points, and define the *moduli space* for divisors to be the quotient $\mathfrak{M}_n = \widehat{\mathcal{D}}_n^{\mathrm{fs}}/\mathbf{G}$.

One basic invariant for the \mathbf{G} -orbit of a divisor is the *maximum multiplicity*

$$1 \leq \max_j \{m_j\} \leq n$$

of the points in $|\mathcal{D}|$.

Theorem 3.2. \mathfrak{M}_n is a T_1 -space for every n ; but it is a Hausdorff space only for $n \leq 4$. For any n , the open subset of \mathfrak{M}_n consisting of \mathbf{G} -equivalence classes of divisors with maximum multiplicity satisfying

$$\max_j \{m_j\} < n/2$$

is a Hausdorff space and an orbifold. However if $n \geq 5$, then any point for which $\max_j \{m_j\} \geq n/2$ is not even locally Hausdorff.

For $n \geq 5$, we will use the notation $\mathfrak{M}_n^{\mathrm{Haus}}$ for this maximal open Hausdorff subset of \mathfrak{M}_n .

Theorem 3.3. For $n \geq 5$, the space $\mathfrak{M}_n^{\mathrm{Haus}}$ is compact for n odd; but not for n even.

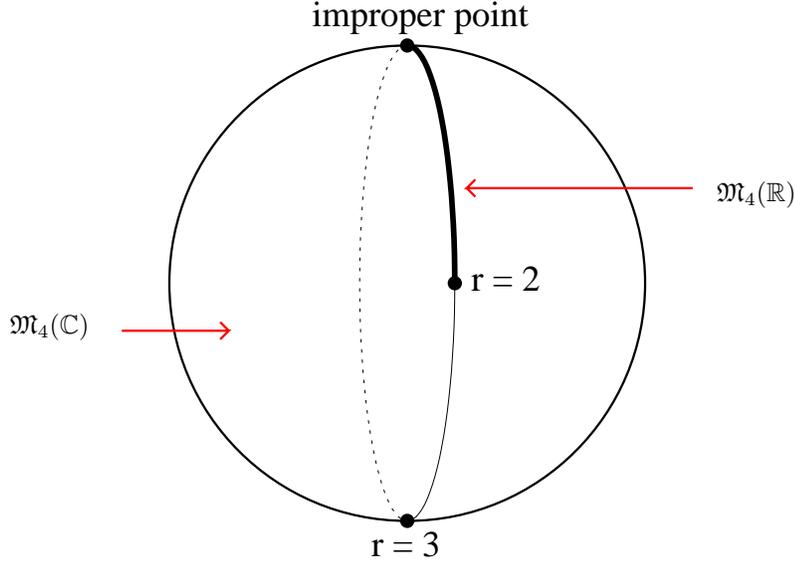
The case $n = 5$ is particularly striking, since the non-Hausdorff space \mathfrak{M}_5 consists of a compact Hausdorff space $\mathfrak{M}_5^{\mathrm{Haus}}$ together with just one “bad” point of the form

$$((3\langle \mathbf{p} \rangle + \langle \mathbf{q} \rangle + \langle \mathbf{r} \rangle)),$$

with $\max_j \{m_j\} = 3 > 5/2$.

To begin the proof of Theorem 3.2, we will study the cases $n \leq 4$. It is easy to check that \mathfrak{M}_n is empty for $n < 3$. For $n = 3$, since the action of \mathbf{G} on \mathbb{P}^1 is three-point simply transitive, it follows easily that $\mathfrak{M}_3(\mathbb{R}) = \mathfrak{M}_3(\mathbb{C})$ consists of a single point; with stabilizer the symmetric group \mathfrak{S}_3 . For $n = 4$ we have the following. (Recall that each point of \mathfrak{M}_n corresponds to an entire \mathbf{G} -orbit of divisors $\mathcal{D} \in \widehat{\mathcal{D}}_n^{\mathrm{fs}}$.)

Lemma 3.4. *The moduli space $\mathfrak{M}_4(\mathbb{C})$ is an orbifold diffeomorphic to the Riemann sphere $\mathbb{P}^1(\mathbb{C}) \cong \widehat{\mathbb{C}}$ with three ramified points. One of these is the unique improper point corresponding to divisors of degree four with only three distinct points. The stabilizer $\mathbf{G}_{\mathcal{D}}$ is isomorphic to:*

FIG. 7. The moduli spaces $\mathfrak{M}_4(\mathbb{R}) \subset \mathfrak{M}_4(\mathbb{C})$.

- $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ if \mathcal{D} represents an unramified point of this orbifold.
- $\mathbb{Z}/2$ if \mathcal{D} represents the improper point, which has index $\tau = 2$,
- the dihedral group of order eight for the other point of index $\tau = 2$, and
- the tetrahedral group of order twelve for the ramified point of index $\tau = 3$.

Lemma 3.5. *In the real case, the moduli space $\mathfrak{M}_4(\mathbb{R})$ is an orbifold isomorphic to a closed interval bounded by the two ramified points of index $\tau = 2$ in the circle $\mathbb{P}^1(\mathbb{R}) \subset \mathbb{P}^1(\mathbb{C}) \cong \mathfrak{M}_4(\mathbb{C})$. (Compare Figure 7. There is no point of index $\tau = 3$.) The stabilizers in the real case are isomorphic to the corresponding stabilizers in the complex case.*

The proof of these two lemmas will make use of two familiar projective invariants associated with a 4-tuple of points in \mathbb{P}^1 . The first is the cross-ratio, which depends on the ordering of the four arguments, and the second is the “shape invariant” \mathbf{J} which is independent of order.

We will use cross-ratios of the form¹³

$$\rho(x, y, z, w) = \rho \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \frac{(x-y)(z-w)}{(x-z)(y-w)}, \quad (1)$$

where x, y, z, w are distinct real or complex numbers. This expression is well defined and continuous on the space of ordered 4-tuples of distinct points of \mathbb{R} or \mathbb{C} , taking values in $\widehat{\mathbb{R}}$ or $\widehat{\mathbb{C}}$. In either case it extends uniquely to the case where any one of the four points is allowed to take the value ∞ . For example as $w \rightarrow \infty$ Equation (1) tends to the limit

$$\rho \begin{bmatrix} x & y \\ z & \infty \end{bmatrix} = \frac{x-y}{x-z}. \quad (2)$$

It is not hard to check that ρ is unchanged if we apply an affine automorphism $x \mapsto az+b$, or apply the inversion map $x \mapsto 1/x$, to each of the four variables. Since the group PGL_2

¹³This notation is non-standard, but is convenient and easy to remember.

of fractional linear transformations is generated by affine transformations and inversion, it follows that ρ is invariant under the action of this group.

Lemma 3.6. *There is a necessarily unique projective automorphism carrying one ordered set of four distinct points of \mathbb{P}^1 to another if and only if they have the same cross-ratio, which can take any value other than 0, 1, or ∞ .*

Proof. It suffices to consider the special case where the second 4-tuple has the form $(0, y, 1, \infty)$, so that the cross-ratio is y by equation (2). Using three point transitivity, there is a unique projective automorphism taking the appropriate points to 0, 1 and ∞ ; and it follows that the remaining point must map to the cross-ratio y . \square

However, as two of the four points come together (so that only three are distinct), the cross-ratio will tend to a limit belonging to the complementary set $\{0, 1, \infty\}$. (If only two of the four points are distinct, then the cross-ratio cannot be defined in any useful way.)

Note that the cross-ratio is always unchanged as we interchange the two rows, or the two columns, of the matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$. Thus we obtain the following. (Compare Lemma 2.26.)

Lemma 3.7. *For any 4-tuple (x, y, z, w) of four distinct points, there is a transitive 4 element group of permutations of the four points, isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, which preserves their cross-ratio. Hence the stabilizer $\mathbf{G}_{\mathcal{D}}$ for the associated divisor $\mathcal{D} = \langle x \rangle + \langle y \rangle + \langle z \rangle + \langle w \rangle$ always contains $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ as a subgroup.*

Proof. In both the real and complex cases, this follows immediately from the discussion above. \square

The Shape Invariant \mathbf{J} . We next describe a number $\mathbf{J} = \mathbf{J}(x, y, z, w)$ which is invariant not only under projective automorphisms of \mathbb{P}^1 , but also under permutations of the four variables.

First consider the generic case where all four points are distinct. After a projective transformation, we may assume that $w = \infty$ and that x, y, z are finite. Then x, y, z are uniquely determined up to a simultaneous affine transformation. Therefore the differences

$$\alpha = x - y, \quad \beta = y - z, \quad \gamma = z - x \quad (3)$$

are uniquely determined up to multiplication by a common non-zero constant. Next consider the elementary symmetric functions

$$\sigma_1 = \alpha + \beta + \gamma = 0, \quad \sigma_2 = \alpha\beta + \alpha\gamma + \beta\gamma, \quad \sigma_3 = \alpha\beta\gamma.$$

If we multiply α, β, γ by a common constant $t \neq 0$, then each σ_j will be multiplied by t^j . Therefore the ratio¹⁴

$$\mathbf{J} := -\frac{4}{27} \frac{\sigma_2^3}{\sigma_3^2}. \quad (4)$$

will remain unchanged. It might seem that w plays a special role in this construction; but remember from Lemmas 3.6 and 3.7 that there is a transitive group of projective automorphisms permuting the four variables. Therefore it doesn't matter which of the four variables we put at infinity.

¹⁴Here the factor of $-4/27$ has been inserted so that \mathbf{J} will take the value $+1$ in the case of dihedral symmetry, where two of the three numbers α, β, γ are equal.

If only three of the four variables x, y, z, w are distinct, then we set $\mathbf{J} = \infty$. For example if $x = y$ so that $\alpha = 0$, then $\sigma_3 = 0$, hence $\mathbf{J} = \infty$. It is not difficult to check that the resulting function

$$(x, y, z, w) \mapsto \mathbf{J}(x, y, z, w) \in \widehat{\mathbb{F}}$$

is continuous as a function of four variables in $\widehat{\mathbb{F}}$, provided that we require that at least three of the four variables remain distinct.

In general, four points of \mathbb{P}^1 determine six different cross-ratios, according to the order in which they are listed (compare Remark 3.10), but only one shape invariant.

Lemma 3.8. *Any one of these six cross-ratios determines the shape invariant according to the formula*

$$\mathbf{J} = \frac{4(\rho^2 - \rho + 1)^3}{27\rho^2(1 - \rho)^2}. \quad (5)$$

Proof. Since both sides of equation (5) are invariant under affine transformations of the plane, it suffices to consider the special case where the 4-tuple (x, y, z, w) is equal to $(0, t, 1, \infty)$, with cross-ratio t . (In fact one can choose an affine transformation which maps x to zero and z to one, while keeping ∞ fixed. The point y will then necessarily map to the cross-ratio.) We then have

$$\alpha = -t, \quad \beta = t - 1, \quad \gamma = 1,$$

hence $\sigma_2 = -(t^2 - t + 1)$ and $\sigma_3 = t(1 - t)$; and the required identity (5) follows immediately. \square

Remark 3.9. The shape invariant \mathbf{J} is just the classical j -invariant of an associated cubic curve, divided by a constant factor of $12^3 = 1728$. (Compare Section 5.) To see the relationship, first subtract the average $(x + y + z)/3$ from x, y , and z , in order to obtain a triple (X, Y, Z) with $X + Y + Z = 0$. These corrected variables will then be the roots of a uniquely defined cubic equation $X^3 + AX + B = 0$. If we express the σ_2 and σ_3 of equation (4) as functions of these three variables, then computation shows that

$$\mathbf{J} = \frac{4A^3}{4A^3 + 27B^2}.$$

Here the denominator is the classical expression for the discriminant of a cubic polynomial, up to sign. Details of the computation will be omitted.

Proof of Lemma 3.4. The discussion above shows that every divisor

$$\mathcal{D} = \langle x \rangle + \langle y \rangle + \langle z \rangle + \langle w \rangle$$

with at least three distinct elements determines a point $\mathbf{J}(x, y, z, w)$ in the Riemann sphere $\widehat{\mathbb{C}}$, and that this image point is invariant under the action of the group $\mathbf{G} = \mathrm{PGL}_2(\mathbb{C})$ on the divisor. It is easy to check that the resulting correspondence

$$\mathfrak{M}_4(\mathbb{C}) \longrightarrow \widehat{\mathbb{C}}$$

is continuous and bijective, and hence is a homeomorphism.

To describe the precise stabilizers for the various points of $\mathfrak{M}_4(\mathbb{C})$ we will need the following.

Remark 3.10 (Symmetries of Cross-Ratios). For an arbitrary permutation σ of a set $\{x, y, z, w\}$ of four distinct points the cross-ratio $\rho(x, y, z, w)$ will be transformed by some corresponding rational map $\mathbf{g}(\sigma)$. By Lemma 3.7, the 4 element subgroup consisting of permutations which preserve cross-ratios acts transitively on the four points. Thus any given permutation can be composed with an element of this subgroup so as to map w to itself. Therefore, in studying cross-ratios, it suffices to consider only the subgroup \mathfrak{S}_3 consisting of permutations of $\{x, y, z\}$ with w fixed. This group \mathfrak{S}_3 consists of a cyclic subgroup of order three, together with three elements of order two. It is not hard to check that the elements of order two correspond to the involutions which takes ρ to either

$$1/\rho \quad \text{or} \quad 1 - \rho \quad \text{or} \quad \rho/(\rho - 1); \quad (6)$$

while the two elements of order three correspond to the rational maps

$$\rho \mapsto 1/(1 - \rho) \quad \text{and} \quad \rho \mapsto 1 - 1/\rho. \quad (7)$$

Thus a generic element $\mathcal{D} \in \widehat{\mathcal{D}}_4$ has six different associated cross-ratios, and any one of the six determines the other five.

If x, y, z, w are distinct then by Lemma 3.7, the stabilizer $\mathbf{G}_{\mathcal{D}}$ of the associated divisor

$$\mathcal{D} = \langle x \rangle + \langle y \rangle + \langle z \rangle + \langle w \rangle$$

can be identified with the group of all permutations of $\{x, y, z, w\}$ which preserve the cross-ratio. In particular, it always contains a subgroup isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ by Lemma 3.7. If the six cross-ratios are all distinct, then the stabilizer is equal to this commutative subgroup of order 4; but there are three exceptional cases (including the degenerate case), corresponding to equalities between various of the numbers (6) and (7) and ρ .

Dihedral Symmetry. If the shape invariant is $\mathbf{J} = 1$, then there are only three associated cross-ratios, namely $-1, 1/2$ and 2 . (Each of these is fixed under one of the involutions of equation (6).) As an example, a corresponding divisor can be chosen as

$$\mathcal{D} = \langle -1 \rangle + \langle 0 \rangle + \langle 1 \rangle + \langle \infty \rangle.$$

The associated stabilizer is the dihedral group of order eight, generated by the rotation

$$x \mapsto \frac{1+x}{1-x} \quad \text{with} \quad 0 \mapsto 1 \mapsto \infty \mapsto -1 \mapsto 0,$$

together with the reflection $x \mapsto -x$. The ramification index is $\tau = 2$.

Tetrahedral Symmetry. If $\mathbf{J} = 0$ then there are only two associated cross-ratios, namely $\rho = \frac{1 \pm \sqrt{-3}}{2}$. A corresponding divisor can be obtained by placing the four points at the vertices of a tetrahedron on the Riemann sphere (identified with the unit sphere in Euclidean 3-space). Then evidently the corresponding stabilizer is the tetrahedral group of order 12 (the group of orientation preserving isometries of the tetrahedron). Since the cross-ratios are not real, this possibility can occur only in the complex case.

It is not hard to check from the equations (6) and (7) that these are the only non-degenerate examples for which there are not six distinct cross-ratios.

The Degenerate Case. If two of the four points come together, then the possible cross-ratios are $0, 1, \infty$. (Compare Lemma 3.6.) Much of the discussion above breaks down in this case. In particular, it is easy to check that the stabilizer has only two elements instead of four. According to Lemma 2.25, this implies that the action of \mathbf{G} is not locally proper at such points.

Lemma 3.11. *Although the action of \mathbf{G} is not locally proper at such degenerate points, it is still weakly proper. (Compare Definition 2.10).*

Proof. It will be convenient to use the notation $x \approx y$ to mean that x and y are very close or are equal to each other. Then any two divisors in a neighborhood of an improper divisor $2\langle x \rangle + \langle z \rangle + \langle w \rangle$ will have the form

$$\mathcal{D}_j = \langle x_j \rangle + \langle y_j \rangle + \langle z_j \rangle + \langle w_j \rangle \quad \text{for } j = 1, 2;$$

where $x_j \approx y_j$, but where $z_1 \approx z_2$ and $w_1 \approx w_2$ are well separated. If there is a group element taking \mathcal{D}_1 to \mathcal{D}_2 , then since $\rho(x_j, y_j, z_j, w_j) \approx 0$, it must take $\{x_1, y_1\}$ to either $\{x_2, y_2\}$ or $\{z_2, w_2\}$. After composing with a element of the central subgroup $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, we may assume that

$$\{x_1, y_1\} \mapsto \{x_2, y_2\} \quad \text{and that } z_1 \mapsto z_2, w_1 \mapsto w_2 .$$

This shows that we can choose the group element to belong to a compact subgroup of PGL_2 , which proves that the action is weakly proper. \square

This, together with Theorem 2.18 and Example 2.8, completes the proof of Lemma 3.4. \square

Proof of Lemma 3.5. In the real case, a completely analogous argument shows that the correspondence $\mathcal{D} \mapsto \mathbf{J}(\mathcal{D})$ gives rise to an injective map from $\mathfrak{M}_4(\mathbb{R})$ into $\widehat{\mathbb{R}}$. However the image of \mathfrak{M}_4 is no longer the entire $\widehat{\mathbb{R}}$. In fact the image is precisely equal to the closed interval $1 \leq \mathbf{J} \leq \infty$. To see this, consider the correspondence $\rho \mapsto \mathbf{J}(\rho)$ of equation (5) on the interval $0 < \rho < 1$. Evidently $\mathbf{J}(\rho)$ tends to infinity as ρ tends to zero or one. It follows from Equation (6) that this correspondence must satisfy the identity

$$\mathbf{J}(\rho) = \mathbf{J}(1 - \rho) ,$$

so that the point $\rho = 1/2$ with $\mathbf{J}(1/2) = 1$ must be a local extremum. Using the other symmetry relation we see that the behavior in the intervals $-\infty < \rho < 0$ and $1 < \rho < \infty$ is completely analogous. Since this rational function has degree six, it is at most six-to one. It follows that it must maps each of the six intervals

$$[-\infty, -1], \quad [-1, 0], \quad [0, 1/2], \quad [1/2, 1], \quad [1, 2], \quad \text{and } [2, +\infty]$$

homeomorphically onto $[1, +\infty]$.

The rest of the proof of Lemma 3.5 can easily be completed, since the arguments are almost the same as those in the complex case. \square

Next we must study the case $n > 4$.

Lemma 3.12. *If $\mathcal{D} = \sum_j m_j \langle \mathbf{p}_j \rangle$ is a divisor of degree $n > 4$ with*

$$\max_j \{m_j\} \geq n/2,$$

then the quotient space \mathfrak{M}_n is not locally Hausdorff at the image point $\pi(\mathcal{D})$.

For the rest of this section, the real and complex cases are completely analogous, so it will suffice to concentrate on the complex case.

Proof of Lemma 3.12. First consider the special case of divisors of even degree $n = 2h \geq 6$, with $\max_j \{m_j\} = h \geq 3$. Identifying \mathbb{P}^1 with $\mathbb{C} \cup \{\infty\}$, let \mathcal{D}_1 and \mathcal{D}_2 be of the form

$$\mathcal{D}_j = \mathcal{D}'_j + h\langle\infty\rangle ,$$

where both \mathcal{D}'_1 and \mathcal{D}'_2 are divisors of degree $h = n/2$ with support consisting of h distinct points in the finite plane, and where $h\langle\infty\rangle$ is the divisor consisting only of the point ∞ with multiplicity h . Thus the support $|\mathcal{D}_j|$ has $h + 1 \geq 4$ elements. Since $4 > \dim(\mathbf{G}) = 3$, we can always choose two such divisors \mathcal{D}_1 and \mathcal{D}_2 which do not belong to the same \mathbf{G} -orbit.

Now consider the projective involution $\mathbf{g}_r(z) = r^2/z$, where r is a large real number. Note that \mathbf{g}_r maps the neighborhood $|z| < r$ of zero onto the neighborhood $|z| > r$ of infinity. Then the two divisors

$$\mathcal{D}'_1 + \mathbf{g}_r(\mathcal{D}'_2) \quad \text{and} \quad \mathbf{g}_r(\mathcal{D}'_1) + \mathcal{D}'_2$$

belong to the same \mathbf{G} -orbit. Yet by choosing r sufficiently large we can place the first arbitrarily close to \mathcal{D}_1 and the second arbitrarily close to \mathcal{D}_2 . This proves that the quotient $\mathfrak{M}_{2h} = \widehat{\mathfrak{D}}_{2h}^{\text{fs}}/\mathbf{G}$ is not a Hausdorff space. In fact, since \mathcal{D}_2 can be arbitrarily close to \mathcal{D}_1 , it follows that \mathfrak{M}_{2h} is not even locally Hausdorff at $((\mathcal{D}_1))$. Furthermore, since any divisor with $\max_j \{m_j\} > h$ can be approximated by one with $\max_j \{m_j\} = h$, it follows that \mathfrak{M}_{2h} is not locally Hausdorff at any point with $\max_j \{m_j\} \geq h$.

The proof for $n = 2h + 1 \geq 5$ is similar. For this case we take

$$\mathcal{D}_1 = \mathcal{D}'_1 + (h + 1)\langle\infty\rangle \quad \text{and} \quad \mathcal{D}_2 = \mathcal{D}'_2 + h\langle\infty\rangle ,$$

where \mathcal{D}'_1 has degree $h \geq 2$, but \mathcal{D}'_2 has degree $h + 1 \geq 3$. It then follows as above that \mathfrak{M}_{2h+1} is not Hausdorff. Again \mathcal{D}_1 and \mathcal{D}_2 can be arbitrarily close to each other: Starting with any \mathcal{D}_1 , it is only necessary to replace the point of multiplicity $h + 1$ for \mathcal{D}_1 by a point of multiplicity h , together with a nearby point of multiplicity one, in order to obtain an appropriate \mathcal{D}_2 . It follows easily that \mathfrak{M}_{2h+1} is not locally Hausdorff at any point with $\max_j \{m_j\} \geq h + 1$. This completes the proof of Lemma 3.12. \square

The proof of Theorem 3.2 will also require a study of group elements which are “close to infinity” in \mathbf{G} (or in other words, outside of a large compact subset of \mathbf{G}). Choose some metric on \mathbb{P}^1 , for example the standard spherical metric, and let $N_\varepsilon(\mathbf{p})$ be the open ε -neighborhood of \mathbf{p} .

Lemma 3.13 (Distortion Lemma for Automorphisms of \mathbb{P}^1). *Given any $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon \subset \mathbf{G}$ with the following property. For any $\mathbf{g} \in \mathbf{G} \setminus K$ there exist two (not necessarily distinct) open ε -disks $N^+ = N_\varepsilon(\mathbf{p})$ and $N^- = N_\varepsilon(\mathbf{q})$ such that*

$$\mathbf{g}(N^+) \cup N^- = \mathbb{P}^1 .$$

It follows that \mathbf{g} maps every point outside of N^+ into N^- . (Roughly speaking, we can think of N^+ as a repelling disk and N^- as an attracting disk.)

Proof of Lemma 3.13. First consider the corresponding statement for the group of diagonal automorphisms

$$\mathbf{d}_\kappa(x : y) = (\kappa x : y) , \quad \text{with } \kappa \in \mathbb{C} \setminus \{0\} ,$$

or in affine coordinates with $z = x/y$, $z \mapsto \kappa z$. Interchanging the coordinates x and y if necessary, we may assume that $|\kappa| \geq 1$. The condition that \mathbf{d}_κ lies outside of a large

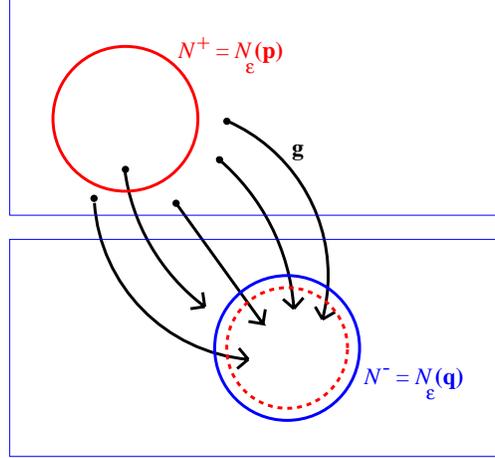


FIG. 8. Illustrating the Distortion Lemma. The two rectangles represent copies of \mathbb{P}^1 . The image $g(N^+)$ covers everything outside of the dotted circle. Hence everything outside of N^+ must map into N^- .

compact set then means that $|\kappa|$ is large (as illustrated in Figure 8). The proof can then easily be completed, choosing $\mathbf{p} = 0$ and $\mathbf{q} = \infty$. (For example, if $\kappa = 1/\varepsilon^2$ then \mathbf{d}_κ maps the small disk $|z| < \varepsilon$ onto the large disk $|z| < 1/\varepsilon$.)

The proof for the group of projective transformations \mathbf{G} then follows immediately, using Lemma 2.30. \square

Proof of Theorem 3.2. We will first show that \mathfrak{M}_n is a T_1 -space. This means that every point of \mathfrak{M}_n is a closed set; or equivalently that every \mathbf{G} -orbit in $\widehat{\mathfrak{D}}_n^{\text{fs}}$ is a closed set. In other words, we must show that every limit point of such an orbit within the larger space $\widehat{\mathfrak{D}}_n$ either belongs to the orbit or else has infinite stabilizer, so that it is outside of $\widehat{\mathfrak{D}}_n^{\text{fs}}$. (Compare the proof of Lemma 6.2 for a different argument.)

Given any $\mathcal{D} \in \widehat{\mathfrak{D}}_n^{\text{fs}}$, let \mathcal{D}' be a divisor which can be expressed as the limit

$$\mathcal{D}' = \lim_{j \rightarrow \infty} \mathbf{g}_j(\mathcal{D})$$

of points $\mathbf{g}_j(\mathcal{D})$ in the \mathbf{G} -orbit $((\mathcal{D}))$. If \mathcal{D}' itself does not belong to this \mathbf{G} -orbit, then we will show that the support $|\mathcal{D}'|$ can have only one or two elements, so that $\mathcal{D}' \notin \widehat{\mathfrak{D}}_n^{\text{fs}}$.

Choose $\varepsilon > 0$ small enough so that any two points of $|\mathcal{D}|$ have distance $\geq 2\varepsilon$, or in other words so that any $N_\varepsilon(\mathbf{p})$ can contain at most one point of $|\mathcal{D}|$. The group elements \mathbf{g}_j must tend to infinity in \mathbf{G} (that is, they must eventually lie outside of any compact subset), since otherwise the limit point would be in the \mathbf{G} -orbit $((\mathcal{D}))$. Hence we can choose corresponding ε_j tending to zero. For each j with $\varepsilon_j < \varepsilon$, we can choose ε_j -disks N_j^+ and N_j^- as in the Distortion Lemma 3.13, it follows that all but at most one point of $|\mathbf{g}_j(\mathcal{D})|$ must lie in the disk N_j^- . Passing to the limit, it follows that $|\mathcal{D}'|$ can have at most two points, as asserted.

Thus it follows that $((\mathcal{D}))$ is closed as a subset of $\widehat{\mathfrak{D}}_n^{\text{fs}}$, and hence that $\pi(\mathcal{D})$ is closed as a subset of \mathfrak{M}_n . Therefore the quotient $\mathfrak{M}_n = \widehat{\mathfrak{D}}_n^{\text{fs}}/\mathbf{G}$ is a T_1 -space.

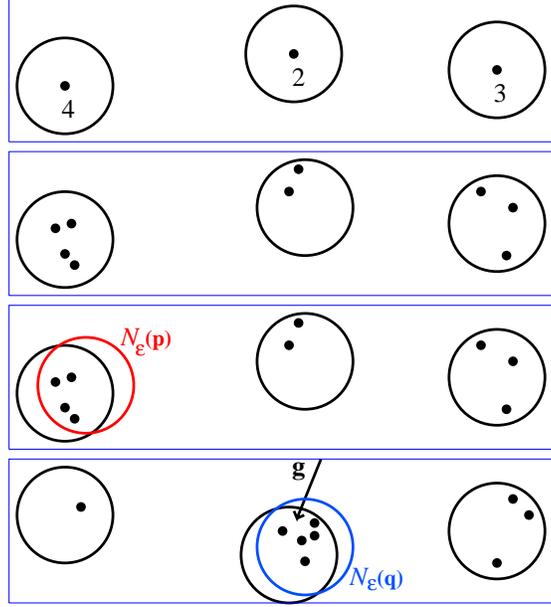


FIG. 9. Illustrating the proof of 3.15. The top frame illustrates a typical divisor $\mathcal{D}_1 \in \mathfrak{V}_9$, showing the ε -balls around points of multiplicity 4, 2, and 3. The next two frames illustrate a divisor $\mathcal{D}'_1 \in \mathfrak{N}_\varepsilon(\mathcal{D}_1)$, and the last frame illustrates a divisor \mathcal{D}'_2 whose points are contained in ε -balls around the points of \mathcal{D}_2 . Assuming that $\mathbf{g} \notin K_\varepsilon$ maps \mathcal{D}'_1 to \mathcal{D}'_2 , the last two frames show associated balls $N^+ = N_\varepsilon(\mathbf{p})$ and $N^- = N_\varepsilon(\mathbf{q})$ such that the five points outside of N^+ must all map into N^- ; yielding a contradiction.

Finishing the proof of Theorem 3.2 will require one further preliminary step.

Definition 3.14. For any n , let \mathfrak{V}_n be the open set consisting of all divisors in $\widehat{\mathfrak{D}}_n$ such that

$$\max_j \{m_j\} < n/2.$$

(For $n \geq 5$, \mathfrak{V}_n is just the set we have called $\mathfrak{M}_n^{\text{Haus}}$.)

Proposition 3.15. For any n , the action of $\mathbf{G} = \text{PGL}_2$ on this open set $\mathfrak{V}_n \subset \widehat{\mathfrak{D}}_n$ is proper.

Proof. Consider two divisors $\mathcal{D}_1, \mathcal{D}_2 \in \mathfrak{V}_n$. We must construct neighborhoods U_1 of \mathcal{D}_1 and U_2 of \mathcal{D}_2 and a compact set $K \subset \mathbf{G}$ so that any $\mathbf{g} \in \mathbf{G}$ which maps a point $\mathcal{D}'_1 \in U_1$ to a point $\mathcal{D}'_2 \in U_2$ must belong to K .

Let ε be small enough so that any two distinct points of $|\mathcal{D}_\ell|$ have distance

$$\text{dist}(\mathbf{p}, \mathbf{p}') > 4\varepsilon \tag{8}$$

from each other, both for $\ell = 1$ and for $\ell = 2$. Thus no ball of radius ε can intersect more than one of the ε balls around the points of $|\mathcal{D}_\ell|$.

Let K_ε be a corresponding compact subset of \mathbf{G} , as described in the Distortion Lemma 3.13. Let $\mathfrak{N}_\varepsilon(\mathcal{D}_\ell) \subset \widehat{\mathfrak{D}}_n$ be the neighborhood of \mathcal{D}_ℓ consisting of all $\mathcal{D}'_\ell \in \widehat{\mathfrak{D}}_n$ such that, for each $\mathbf{p} \in |\mathcal{D}_\ell|$, the number of points of \mathcal{D}'_ℓ in $N_\varepsilon(\mathbf{p})$ counted with multiplicity, is precisely equal to the multiplicity of \mathbf{p} as a point of \mathcal{D}_ℓ . (In other words, a point \mathbf{p} of multiplicity m_j for \mathcal{D}_ℓ is allowed to split into as many as m_j distinct points in \mathcal{D}'_ℓ ; but these points are not allowed to move out of the ε -neighborhood of \mathbf{p} . Compare Figure 9.)

We must prove that any $\mathbf{g} \in \mathbf{G}$ which maps some $\mathcal{D}'_1 \in \mathfrak{N}_\varepsilon(\mathcal{D}_1)$ to a $\mathcal{D}'_2 \in \mathfrak{N}_\varepsilon(\mathcal{D}_2)$ must belong to the compact set K_ε . Suppose to the contrary that $\mathbf{g} \notin K_\varepsilon$. Then we could construct corresponding ε -balls $N^+ = N_\varepsilon(\mathbf{p}_0)$ and $N^- = N_\varepsilon(\mathbf{q}_0)$, so that $\mathbf{g}(N^+) \cup N^- = \mathbb{P}^1$. Since any ε -ball intersects at most one of the $N_\varepsilon(\mathbf{p})$ with $\mathbf{p} \in |\mathcal{D}|$, the ball N^+ must contain fewer than $n/2$ points of \mathcal{D}'_1 , counted with multiplicity. Hence its complement must contain more than $n/2$ such points. Since \mathbf{g} maps the complement of N^+ into N^- , this means that N^- contains more than $n/2$ points of \mathcal{D}'_2 , counted with multiplicity. But this is impossible since N^- can intersect at most one of the ε -balls around points of $|\mathcal{D}_2|$. This contradiction completes the proof of Proposition 3.15. \square

Since proper action implies Hausdorff quotient, it also completes the proof of Theorem 3.2. \square

Proof of Theorem 3.3 (Compactness). Consider the set $\mathfrak{Y}_n \subset \widehat{\mathfrak{D}}_n$ consisting of divisors with maximum multiplicity strictly less than $n/2$. (Definition 3.14.) We will prove that the set \mathfrak{Y}_n is compact if and only if n is odd.

First consider the case of even degree $n = 2k \geq 4$. Let $\mathcal{D} \in \mathfrak{D}_k$ be some divisor consisting of k distinct non-zero complex numbers. (Here we are identifying \mathbb{P}^1 with $\mathbb{C} \cup \{\infty\}$.) Setting $\mathbf{g}_\lambda(z) = \lambda z$, consider the sequence of divisors $\mathcal{D}_j = \mathbf{g}_j(\mathcal{D}) + \mathbf{g}_{1/j}(\mathcal{D})$, which converges to $k\langle\infty\rangle + k\langle 0\rangle$ as $j \rightarrow \infty$. We can spread out the last k points by passing to the sequence $\mathbf{g}_j(\mathcal{D}_j)$, which converges to $k\langle\infty\rangle + \mathcal{D}$; or we can bring the first k points away from infinity by passing to the sequence $\mathbf{g}_{1/j}(\mathcal{D}_j)$ which converges to $\mathcal{D} + k\langle 0\rangle$. However no matter what automorphism we apply to this sequence, there will always be at least one point with multiplicity k or larger. Therefore, the corresponding sequence in \mathfrak{Y}_n has no limit in \mathfrak{Y}_n ; which proves non-compactness.

For the proof in the odd case we will need the following.

Lemma 3.16. *If $n = 2k - 1$ is odd, then there exists a compact set $\mathfrak{K}_n \subset \widehat{\mathfrak{D}}_n$ with the following property. A divisor $\mathcal{D} \in \widehat{\mathfrak{D}}_n$ belongs to \mathfrak{Y}_n if and only if its orbit $((\mathcal{D})) \subset \widehat{\mathfrak{D}}_n$ intersects the set \mathfrak{K}_n .*

Proof. It will be convenient to use the standard spherical metric $\frac{2|dz|}{1+|z|^2}$, so that the distance between any two antipodal points is π . For any set $S \subset \mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$, let $0 \leq \mathbf{diam}(S) \leq \pi$ denote the diameter. For any divisor \mathcal{D} , define the *weight* of a subset $S \subset |\mathcal{D}|$ to be the sum of the multiplicities of the points of S . Let \mathfrak{K}_n be the set of all divisors \mathcal{D} of degree n such that any subset $S \subset |\mathcal{D}|$ of weight $\geq k$, has diameter $\mathbf{Diam}(S) \geq \pi/4$.

Clearly \mathfrak{K}_n is a closed subset of $\widehat{\mathfrak{D}}_n$, and hence is compact. We must show that the orbit $((\mathcal{D}))$ of a divisor intersects \mathfrak{K}_n if and only if \mathcal{D} contains no point of multiplicity $\geq k$. Certainly if \mathcal{D} does contain a point of multiplicity $\geq k$, then the same is true for every divisor in the orbit $((\mathcal{D}))$, so $((\mathcal{D}))$, must be disjoint from \mathfrak{K}_n .

Let $\Theta(\mathcal{D})$ be the minimum of $\mathbf{Diam}(S)$ over all subsets $S \subset |\mathcal{D}|$ which have weight $\geq k$. Then Θ is a continuous function from $\widehat{\mathcal{D}}_n$ to \mathbb{R} , so it must take a maximum value on the compact set (\mathcal{D}) . We claim that this maximum must be either zero (when $|\mathcal{D}|$ contains a point of multiplicity k or more), or else $\geq \pi/4$ (when every point has multiplicity less than k). Suppose to the contrary that Θ takes a maximum of Θ_0 which is strictly between zero and $\pi/4$ at some divisor $\mathcal{D}_0 \in (\mathcal{D})$. This means that $|\mathcal{D}_0|$ has one or more subsets of weight $\geq k$ with diameter Θ_0 , and all others with larger diameter. Since the intersection of two subsets of weight $\geq k$ must be non-empty, it follows that the union of these smallest sets of weight $\geq k$ must have diameter $\leq 2\Theta_0 < \pi/2$. Hence this union is strictly contained in the hemisphere centered at any one of its points. After rotating the Riemann sphere, we may assume that this hemisphere is centered at zero. Using affine coordinates, this corresponds to the unit disk $|z| < 1$.

Given any compact subset of the open unit disk, we can choose λ so that the map $\mathbf{g}_\lambda(z) = \lambda z$ will strictly increase the spherical distance between any two points of this compact set. In fact we will show that the map \mathbf{g}_λ from the disk $|z| < \lambda$ onto the unit disk strictly increases spherical distances. To prove this, note that the inverse map $\mathbf{g}_{1/\lambda}$ multiplies infinitesimal distances near w by a factor of

$$\frac{\lambda(1 + |w|^2)}{1 + |\lambda w|^2}.$$

This expression is strictly less than one for $|w| < 1$, as one can check by multiplying the inequality $\lambda|w|^2 < 1$ by $1 - \lambda$ and then rearranging terms. It follows that every geodesic segment maps to a strictly shorter curve under $\mathbf{g}_{1/\lambda}$; and the conclusion follows.

Thus every one of our purportedly smallest sets $|\mathcal{D}_0|$ will map to a strictly larger set. But this contradicts the assumption that Θ_0 is the maximum. (Note that any other set of weight $\geq k$, with diameter strictly greater than Θ_0 , will still have diameter greater than Θ_0 after a small perturbation.) This contradiction completes the proof of Lemma 3.16. \square

We can now complete the proof of Theorem 3.3. Since the set $\mathfrak{K}_n \subset \widehat{\mathcal{D}}_n$ is compact, its image in the moduli space \mathfrak{M}_n is also compact. Evidently this image is precisely the required set \mathfrak{A}_n . \square

Remark 3.17 (Trivial Stabilizers). For $n \geq 5$, a generic divisor $\mathcal{D} \in \mathcal{D}_n$ has trivial stabilizer. (See Theorem 9.3 for a similar statement for curves in \mathbb{P}^2 .) To see this, let us temporarily work with the space $\mathfrak{D}_n^{\text{ord}}$ consisting of ordered n -tuples $\vec{\mathbf{p}} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$ of distinct points in \mathbb{P}^1 . Every such $\vec{\mathbf{p}}$ determines a corresponding divisor

$$\mathcal{D}(\vec{\mathbf{p}}) = \langle \mathbf{p}_1 \rangle + \dots + \langle \mathbf{p}_n \rangle;$$

and every 4-element subset $\Sigma = \{j_1, \dots, j_4\} \subset \{1, \dots, n\}$ determines a corresponding degree four divisor

$$\mathcal{D}(\vec{\mathbf{p}}, \Sigma) = \langle \mathbf{p}_{j_1} \rangle + \dots + \langle \mathbf{p}_{j_4} \rangle.$$

Let $\mathbf{J}(\vec{\mathbf{p}}, \Sigma) \in \mathbb{C}$ be the shape invariant associated with this divisor. Then for each $\Sigma \neq \Sigma'$ the equation

$$\mathbf{J}(\vec{\mathbf{p}}, \Sigma) = \mathbf{J}(\vec{\mathbf{p}}, \Sigma')$$

determines a proper algebraic subvariety of $\mathfrak{D}_n^{\text{ord}}$. The complement of the union of these finitely many subvarieties is a dense open subset of $\mathfrak{D}_n^{\text{ord}}$; and any element in this dense open set corresponds to a divisor $\langle \mathbf{p}_1 \rangle + \dots + \langle \mathbf{p}_n \rangle$ which has trivial stabilizer. This follows

since the stabilizer of $\mathcal{D}(\vec{\mathbf{p}})$ consists of all permutations of $\{1, \dots, n\}$ which map $\mathcal{D}(\vec{\mathbf{p}})$ to itself. But any non-trivial permutation must clearly map some four point subset to a different four point subset.

Remark 3.18 (Comparison with the Moduli Space $\mathcal{M}_{0,n}$). Let $\mathcal{M}_{0,n}$ be the classical moduli space consisting of closed¹⁵ Riemann surfaces of genus zero which are provided with an ordered list of $n \geq 3$ distinct points; where two such marked Riemann surfaces are identified if there is a conformal isomorphism taking one to the other.

We can unorder these n points by taking the quotient $\mathcal{M}_{0,n}^{\text{un}} = \mathcal{M}_{0,n}/\mathfrak{S}_n$ under the action of \mathfrak{S}_n , the symmetric group of permutations of the ordered list. This quotient can be identified with an open subset

$$\mathfrak{D}_n/\mathbf{G} \subset \widehat{\mathfrak{D}}_n^{\text{fs}}/\mathbf{G} = \mathfrak{M}_n(\mathbb{C})$$

of our moduli space for divisors. Here $\mathfrak{D}_n \subset \widehat{\mathfrak{D}}_n$ is the set of divisors with $\max_j \{m_j\} = 1$, so that its n points are all distinct.

Following Deligne, Mumford, Knudsen and many others, the space $\mathcal{M}_{0,n}$ has a very interesting compactification (Remark 3.20). We can easily apply their construction to the unordered case. Let \mathbb{F} stand for either \mathbb{R} or \mathbb{C} , so that we can describe the real and complex cases in parallel.

Definition 3.19. By an (unordered) *tree-of-marked-spheres* (or *circles* in the real case), we will mean a connected space \mathbf{T} which is the union $\mathbf{S}_1 \cup \dots \cup \mathbf{S}_k$ of one or more copies \mathbf{S}_j of the projective line $\mathbb{P}^1(\mathbb{F})$, together with a finite subset of \mathbf{T} which we will call the set of *marked points*. We require that:

- (1) each marked point belongs to only one of the \mathbf{S}_j ;
- (2) each non-empty intersection $\mathbf{S}_i \cap \mathbf{S}_j$ with $i \neq j$ must consist of a single point (called a *nodal point*), which does not belong to any other \mathbf{S}_k ;
- (3) each \mathbf{S}_j must contain at least three points¹⁶ which are either marked or nodal, and that;
- (4) the abstract graph with one vertex for each \mathbf{S}_j and one edge for each nodal point should be a tree; that is, it must be connected and acyclic.

(Compare Figure 10.) Two such trees-of-marked-spheres are *isomorphic* if there is a homeomorphism between the underlying spaces \mathbf{T} and \mathbf{T}' which preserves the marked points and which is fractional linear on each \mathbf{S}_j . By definition, the *degree* $n \geq 3$ of a tree-of-marked-spheres is the total number of marked points.

By definition, each point of the *compactification* $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ corresponds to a unique isomorphism class of trees-of-marked-spheres of degree n . (Compare [Ar1, Ar2].) In the complex case, we can think of this construction intuitively as follows. Starting with a tree-of-marked-spheres \mathbf{T} , if we remove a small round neighborhood of each intersection point and glue the resulting boundary circles together, then we obtain a Riemann surface of genus zero with n marked points, corresponding to a nearby element of $\mathcal{M}_{0,n}^{\text{un}}$. Conversely, suppose that we start from a surface of genus zero with $n \geq 3$ punctures, provided with its natural hyperbolic metric. If this surface has one or more very short closed geodesics, then we can

¹⁵By definition, a Riemann surface is closed if it is compact without boundary.

¹⁶This is Mumford's stability condition. In the complex case, it means that \mathbf{S}_j with these points removed must be a hyperbolic Riemann surface.

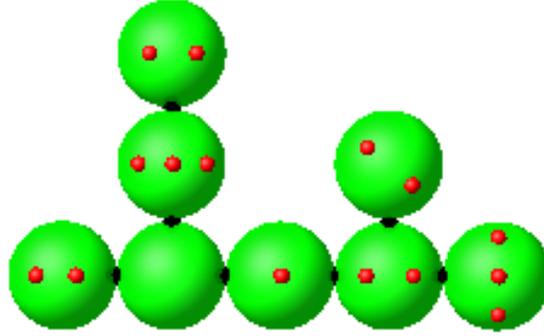


FIG. 10. Showing a tree-of-marked-spheres of degree 15. Each red dot represents a marked point, and each black dot represents a nodal point. (For the count of 15, it is assumed that all red dots are visible in the figure.)

obtain a “nearby” tree-of-marked-surfaces of genus zero, by replacing each such geodesic by a single point.

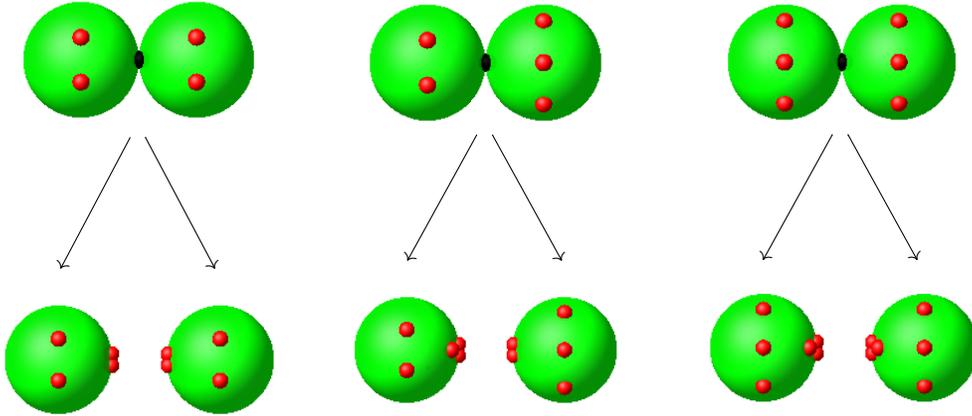


FIG. 11. The top row represents three examples of trees-of-marked-spheres, corresponding to points in the compactification $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ for n equal to 4, 5 and 6 respectively. Corresponding to each sphere S_j in \mathbf{T} , there is a canonical retraction map from \mathbf{T} to S_j , indicated by an arrow in the figure, which maps the n marked points of \mathbf{T} to a divisor $\mathcal{D}_j \in \widehat{\mathcal{D}}_n$. Here each former intersection point is to be weighted by the number of marked points which map to it. These image divisors are shown in the bottom row. (Here a point of multiplicity two or three is indicated schematically by a cluster of two or three overlapping dots.)

There is a “many-valued map” from $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ to \mathfrak{M}_n . More precisely, for each point of $\overline{\mathcal{M}}_{0,n}^{\text{un}}$, represented as a tree of k spheres, there are k associated points of \mathfrak{M}_n , defined

as follows. (Compare Figure 11.) For each sphere $\mathbf{S}_j \subset \mathbf{T}$, there is a unique continuous retraction $\mathbf{r}_j : \mathbf{T} \rightarrow \mathbf{S}_j$ which maps every point of \mathbf{S}_j to itself, and which maps each \mathbf{S}_i with $i \neq j$ to the closest nodal point of \mathbf{S}_j . The collection of all marked points of \mathbf{T} corresponds under \mathbf{r}_j to a divisor \mathcal{D}_j of degree n on the copy \mathbf{S}_j of \mathbb{P}^1 , and hence on the standard \mathbb{P}^1 .

Note that the non-Hausdorff nature of $\widehat{\mathfrak{D}}_n$ is closely related to this construction. In fact, whenever the spheres \mathbf{S}_j and \mathbf{S}_k have an intersection point, it is not hard to see that the corresponding divisors \mathcal{D}_j and \mathcal{D}_k in $\widehat{\mathfrak{D}}_n$ represent points of \mathfrak{M}_n which do not have disjoint neighborhoods. For example, our proof in Lemma 3.12 that \mathfrak{M}_5 and \mathfrak{M}_6 are not Hausdorff makes use precisely of the divisors illustrated in Figure 11. (However, in the degree four case, the corresponding divisors \mathcal{D}_1 and \mathcal{D}_2 actually represent the same point of \mathfrak{M}_4 , which is a Hausdorff space.)

It is not too difficult to prove that $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ can be identified with \mathfrak{M}_n for the cases $n = 3$ and $n = 4$, and with $\mathfrak{M}_n^{\text{Haus}}$ for $n = 5$. Similarly $\mathfrak{M}_6^{\text{Haus}}$ can be identified with an open subset of $\overline{\mathcal{M}}_{0,6}^{\text{un}}$. However there seems to be no such direct relationship when $n \geq 7$.

Remark 3.20 (The Compactification of $\mathcal{M}_{0,n}$). We can learn more about $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ by noting that it is equal to the quotient of the compactification $\overline{\mathcal{M}}_{0,n}$ of the moduli space for ordered n -tuples by the action of the symmetric group \mathfrak{S}_n . This compactification $\overline{\mathcal{M}}_{0,n}$ is a beautiful object which was introduced by Knudsen [Knu], based on ideas of Grothendieck, Deligne, and Mumford. (See Fulton and MacPherson, [FMcP] for a generalization.) It has been studied by many authors, and is well understood.

We can work over either the real numbers or the complex numbers. The construction of this compactification in terms of trees of marked spheres (or marked circles in the real case) is just like the description of $\overline{\mathcal{M}}_{0,n}^{\text{un}}$ as given above, except that the n marked points must now be given n distinct labels, using for example the integers between 1 and n .

First consider the complex case. (In Mumford's terminology, such a tree of labeled spheres is called a "stable curve of genus zero". Compare [Mu, p.228].) Knudsen showed that the compactification $\overline{\mathcal{M}}_{0,n}(\mathbb{C})$ can be constructed out of a smooth variety by iterated blow-ups, and hence that it is a smooth complex variety. (For an alternative proof, using cross-ratios to embed $\overline{\mathcal{M}}_{0,n}(\mathbb{C})$ smoothly in a product of many spheres, see [MS1].) It follows immediately that $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$ is also a smooth manifold, since it is just the fixed point set for complex conjugation on the complex manifold.

The topology of $\overline{\mathcal{M}}_{0,n}(\mathbb{C})$ has been studied by Keel [Ke]. He showed for example that these manifolds are simply-connected, with homology only in even dimensions, and with no torsion. For the simplest cases: $\overline{\mathcal{M}}_{0,3}$ is a point; $\overline{\mathcal{M}}_{0,4}$ is a copy of the sphere \mathbb{P}^1 ; and $\overline{\mathcal{M}}_{0,5}$ is the connected sum of one copy of \mathbb{P}^2 with its standard orientation, together with four copies with reversed orientation.

The topology of $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$ has been studied by Etingof, Henriques, Kamnitzer and Rains [EHKR], who showed for example that there is an isomorphism of mod two cohomology rings of the form

$$H^k(\overline{\mathcal{M}}_{0,n}(\mathbb{R}); \mathbb{Z}/2) \cong H^{2k}(\overline{\mathcal{M}}_{0,n}(\mathbb{C}); \mathbb{Z}/2).$$

The manifold $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$ is non-orientable with a non-abelian fundamental group for all $n > 4$.

Consider a partition of the index set $\{1, 2, \dots, n\}$ into two disjoint subsets I and J , where I has $p \geq 2$ elements and J has $q \geq 2$ elements. In either the real or the complex

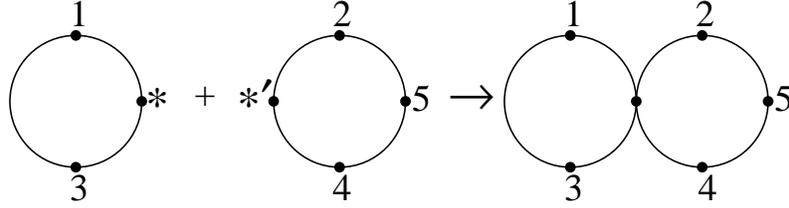


FIG. 12. Illustrating the map $\mathcal{M}_{0,2+1} \times \mathcal{M}_{0,3+1} \hookrightarrow \mathcal{M}_{0,5}$ of Equation (9).

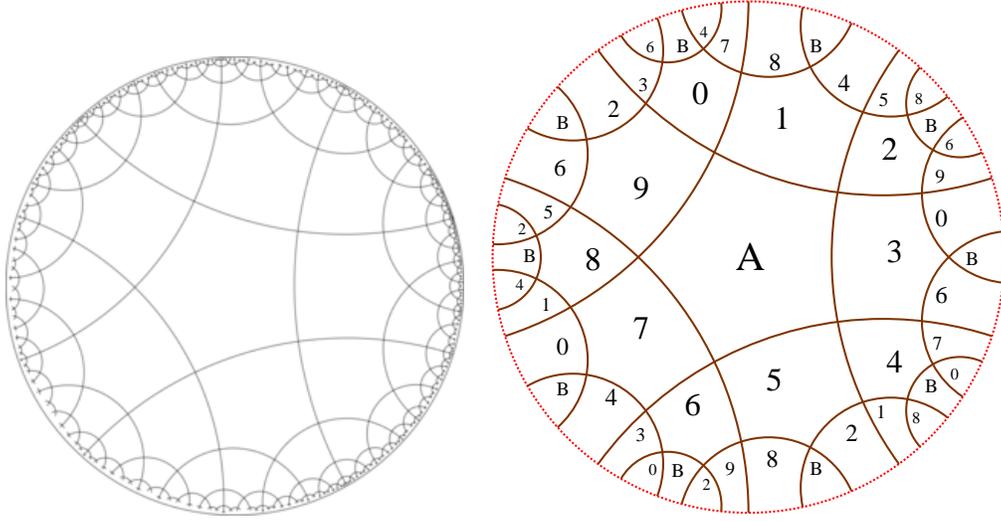


FIG. 13. On the left, the universal covering space of $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$ yields a tessellation of the hyperbolic plane by regular pentagons. On the right, regions with the same label map to a common pentagon in $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$. (It is very hard to visualize the quotient space $\mathcal{M}_{0,5}(\mathbb{R})$ in which pentagons with the same label have been identified.)

case, there is an associated embedding

$$\overline{\mathcal{M}}_{0,p+1} \times \overline{\mathcal{M}}_{0,q+1} \hookrightarrow \overline{\mathcal{M}}_{0,n}, \quad (9)$$

where $n = p + q$. If we restrict this map to the dense open subset $\mathcal{M}_{0,p+1} \times \mathcal{M}_{0,q+1}$ then the associated trees of labeled spheres can be described quite explicitly as illustrated in Figure 12. Label the first sphere by the elements of I together with one additional element $*$. Similarly, label the second sphere by the elements of J together with one additional element $*'$. Now construct the third tree by gluing $*$ onto $*'$.

Every element of the ideal boundary $\overline{\mathcal{M}}_{0,n} \setminus \mathcal{M}_{0,n}$ is contained in the image of at least one such embedding (9). Furthermore, using $\text{mod } 2$ coefficients in the real case, every homology group of $\overline{\mathcal{M}}_{0,n}$, except in the top dimension, is generated by the classes which correspond to these embedded submanifolds.

As a simplest example, there are three distinct ways of partitioning $\{1, 2, 3, 4\}$ into subsets with two elements. Correspondingly, in both the real and complex cases, there are

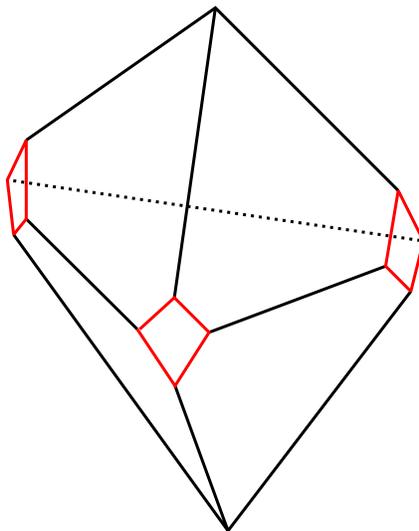


FIG. 14. Each of the sixty 3-cells of $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$ is bounded by six pentagons and 3 squares. (The ninety squares fit together in ten groups of nine to form ten tori.) If we place this cell in hyperbolic 3-space and then stretch the three squares out to points on the circle at infinity, then we can obtain a convex hyperbolic polyhedron with three ideal vertices at infinity together with two finite vertices, and with all dihedral angles equal to 90° . The resulting sixty polyhedra can then be glued together to form a hyperbolic 3-manifold M with twenty infinite cusps.

three distinct ways of embedding the point $\overline{\mathcal{M}}_{0,3} \times \overline{\mathcal{M}}_{0,3}$ into the circle or 2-sphere $\overline{\mathcal{M}}_{0,4}$. The complement of this set of three points is the dense open subset $\mathcal{M}_{0,4}$.

Example 3.21 (The “hyperbolic dodecahedron” $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$). There are ten ways of partitioning $\{1, 2, 3, 4, 5\}$ into subsets of order two and three, and correspondingly ten ways of embedding $\overline{\mathcal{M}}_{0,3} \times \overline{\mathcal{M}}_{0,4} \cong \overline{\mathcal{M}}_{0,4}$ into $\overline{\mathcal{M}}_{0,5}$. In the real case, the ten embedded circles divide this surface into twelve pentagons, which represent the twelve connected components of $\mathcal{M}_{0,5}(\mathbb{R})$. (Compare Figure 13.) This surface can be given a hyperbolic metric, so that these circles are geodesics. Like the standard dodecahedron in Euclidean 3-space, the surface admits a group of 120 isometries such that any isometry from one pentagon to another extends uniquely to a global isometry. However, the two isometry groups are not isomorphic: one has a direct summand of order two generated by the antipodal map, while the other is the symmetric group \mathfrak{S}_5 . Like the standard dodecahedron, $\overline{\mathcal{M}}_{0,5}$ has 12 faces and 30 edges; but it has only 15 vertices, so that the Euler characteristic is $12 - 30 + 15 = -3$. Just as in the standard dodecahedron, every face has a unique opposite face. However, unlike the standard dodecahedron, it cannot admit any fixed point free involution, since its Euler characteristic is odd.

Example 3.22 (The space $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$). This is an interesting example for 3-manifold theory. This manifold is highly symmetric, with a group of 720 automorphisms. (In fact any $\overline{\mathcal{M}}_{0,n}$ with $n \geq 5$ has the full symmetric group on n elements as a group of automorphisms.)

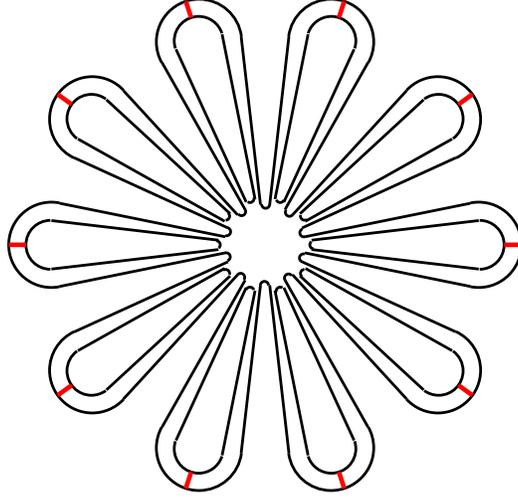


FIG. 15. A cartoon of the 3-manifold $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$, which has an automorphism group of order 720. If we cut along the ten tori (represented by short transverse line segments) then the remainder can be given the structure of a complete hyperbolic manifold with 20 infinite cusps.

The ten partitions of $\{1, 2, 3, 4, 5, 6\}$ into two subsets with three elements each yield ten embeddings of the torus $\overline{\mathcal{M}}_{0,4} \times \overline{\mathcal{M}}_{0,4}$ into $\overline{\mathcal{M}}_{0,6}$. If we remove these ten tori, then the remainder M can be given the structure of a complete hyperbolic manifold of finite volume, as described in Figure 14. (See <http://math.stonybrook.edu/~jack/scgp18.pdf> for further details.) This is an example of a JSJ-decomposition, as first introduced by Jaco and Shalen [JS] and by Johanssen [J]. See Figure 15 for a cartoon of the resulting 3-manifold.

Using this decomposition, it is not hard to check that the fundamental group $\pi_1(\overline{\mathcal{M}}_{0,6}(\mathbb{R}))$ maps onto a free group on ten generators. This fundamental group also contains copies of the abelian group $\mathbb{Z} \oplus \mathbb{Z}$ corresponding to the various tori. To see this, first note that a topological manifold homeomorphic to $\overline{\mathcal{M}}_{0,6}$ can be constructed from the hyperbolic manifold M by cutting off an open neighborhood N_j of each cusp, to obtain a submanifold M' with the same homotopy type, and then pasting the resulting boundary tori together in pairs.

For each torus $T_j = \partial N_j$, the fundamental group $\pi_1(T_j) \cong \pi_1(N_j)$ injects isomorphically into $\pi_1(M)$. To see this, note that the universal covering space \widetilde{M} (isomorphic to hyperbolic 3-space) has the property that any subset of diameter δ can be shrunk to a point without moving any point further than δ . Any closed loop in N_j can be deformed within N_j towards the corresponding cusp point so as to have arbitrarily small length L , and hence so as to have diameter less than its distance from T_j . If such a loop is contractible within M , then it can be lifted to a closed curve of length L in the universal covering space, which can be contracted within a small neighborhood. It follows that the original curve must be contractible within N_j , as required.

Now if we form a closed manifold M'' from M' by gluing pairs of boundary tori together, then it follows by iterated use of the HNN theorem¹⁷ that $\pi_1(M')$ injects into $\pi_1(M'')$. Therefore each $\pi_1(T_j)$ injects into $\pi_1(M'') \cong \pi_1(\overline{\mathcal{M}}_{0,6})$.

4. THE MODULI SPACE FOR REAL OR COMPLEX PLANE CURVES.

This section will be an outline of basic definitions and notations for curves or 1-cycles of arbitrary degree n . (See §5 for a more detailed discussion of the degree three case, and see §6 and §7 for higher degrees.)

Let \mathbb{F} be either \mathbb{R} or \mathbb{C} . It will be convenient to define an *irreducible curve of degree $n \geq 1$ over \mathbb{F}* as an equivalence class of irreducible homogeneous polynomials $\Phi(x, y, z)$ of degree n with coefficients in \mathbb{F} , where two such polynomials are equivalent if one is a non-zero constant multiple of the other. In the complex case it follows from the Hilbert Nullstellensatz that two irreducible curves are equal if and only if they have the same zero locus

$$\{(x : y : z) \in \mathbb{P}^2(\mathbb{C}) ; \Phi(x, y, z) = 0\} .$$

However, in the real case, the analogous zero locus in $\mathbb{P}^2(\mathbb{R})$ is no longer a complete invariant. As an example, we will consider the equivalence classes of $x^2 + y^2 + az^2$ and $x^2 + y^2 + bz^2$ to be distinct curves if $0 < a < b$, even though the corresponding zero loci in $\mathbb{P}^2(\mathbb{R})$ are both empty. In such cases, it is necessary to look at the corresponding complex zero locus in order to distinguish between two different irreducible real curves.

More generally, the term *curve* will be used for any union of finitely many distinct irreducible curves. The notation $\mathcal{C}_n = \mathcal{C}_n(\mathbb{F})$ will be used for the space of all curves of total degree n in $\mathbb{P}^2(\mathbb{F})$. However, this space of curves is not compact. For example, the ellipse $x^2 + ay^2 = a$ converges, as $a \rightarrow 0$, to the locus $x^2 = 0$, which we can think of as the line $x = 0$ counted twice. To allow for such degenerate limits, we can proceed as follows.

Definition 4.1. An *effective (algebraic) 1-cycle* of degree n over \mathbb{F} is a formal sum

$$\mathcal{C} = m_1 \cdot \mathcal{C}_1 + \cdots + m_k \cdot \mathcal{C}_k ,$$

where the \mathcal{C}_j are distinct irreducible curves defined over \mathbb{F} , and where the multiplicities m_j are strictly positive integer coefficients with

$$n = \sum_{j=1}^k m_j \cdot \text{degree}(\mathcal{C}_j) .$$

The notation $\widehat{\mathcal{C}}_n$ will be used for the space of all effective 1-cycles.

The vector space consisting of all polynomials $\Phi : \mathbb{F}^3 \rightarrow \mathbb{F}$ which are homogeneous of degree n has a basis consisting of the $\binom{n+2}{2}$ monomials $x^i y^j z^k$ with $i + j + k = n$. Each such Φ factors as a product of powers $\Phi_j^{m_j}$ of irreducible polynomials, which are uniquely defined up to a non-zero constant factor, and hence corresponds to a unique effective 1-cycle over \mathbb{F} . This proves the following.

¹⁷See [HNN], [Ser2]. This is a purely group theoretic result with the following topological consequence: Let K be a cell complex with two disjoint subcomplexes and an isomorphism f between them, let $K' = K/f$ be the quotient complex in which these subcomplexes are glued together. If the fundamental groups of both subcomplexes inject into $\pi_1(K)$, then the fundamental group of the image subcomplex injects into $\pi_1(K')$.

Lemma 4.2. *The space $\widehat{\mathcal{C}}_n(\mathbb{F})$ consisting of all effective 1-cycles of degree n can be given the structure of a projective space of dimension*

$$\binom{n+2}{2} - 1 = \frac{n(n+3)}{2}$$

over \mathbb{F} .

(This space $\widehat{\mathcal{C}}_n$ is known as the *Chow variety* for 1-cycles of degree n . Compare [Harr, p. 272].) The disjoint union

$$\{0\} \sqcup \widehat{\mathcal{C}}_1 \sqcup \widehat{\mathcal{C}}_2 \sqcup \widehat{\mathcal{C}}_3 \sqcup \cdots$$

of the $\widehat{\mathcal{C}}_n$ can be described as the free additive semigroup with one generator for each irreducible curve over \mathbb{F} .

For any 1-cycle $\mathcal{C} = \sum m_j \cdot \mathcal{C}_j$ over \mathbb{F} , the zero set

$$|\mathcal{C}| = |\mathcal{C}|_{\mathbb{F}} = \{(x : y : z) \in \mathbb{P}^2(\mathbb{F}) ; \Phi(x, y, z) = 0\} = |\mathcal{C}_1| \cup \cdots \cup |\mathcal{C}_k| \subset \mathbb{P}^2(\mathbb{F})$$

will be called the *support* of \mathcal{C} . Note that this definition ignores multiplicities. In the real case, it is also useful to consider the complex support $|\mathcal{C}|_{\mathbb{C}} \subset \mathbb{P}^2(\mathbb{C})$, which provides more information.

Thus a curve is just a 1-cycle

$$\mathcal{C} = \mathcal{C}_1 + \cdots + \mathcal{C}_k ,$$

where the \mathcal{C}_j are distinct irreducible curves, and where all of the multiplicities are $+1$. The space \mathcal{C}_n of curves is a dense open subset of the projective space $\widehat{\mathcal{C}}_n$. In the complex case, a curve is uniquely determined by its support, so that there is no need to distinguish between \mathcal{C} and $|\mathcal{C}|$. But for 1-cycles or for real curves, it is important to make the distinction.

Let $\mathbf{G} = \mathrm{PGL}_3(\mathbb{F})$ be the 8-dimensional Lie group consisting of all projective automorphisms of the projective plane $\mathbb{P}^2 = \mathbb{P}^2(\mathbb{F})$, as discussed in Remark 2.29. Each such projective automorphism acts¹⁸ on the space of curves in \mathbb{P}^2 , and hence acts on the space $\widehat{\mathcal{C}}_n$ of effective 1-cycles of degree n . The notation $((\mathcal{C})) \subset \widehat{\mathcal{C}}_n$ will be used for the orbit of \mathcal{C} under the action of \mathbf{G} .

Definition 4.3. The subgroup $\mathbf{G}_{\mathcal{C}} \subset \mathbf{G}$ consisting of all automorphisms $\mathbf{g} \in \mathbf{G}$ which map \mathcal{C} to itself, is called the *stabilizer* of \mathcal{C} . In the case of a smooth complex curve of degree $n > 1$, it can be identified with the group of all conformal automorphisms of \mathcal{C} which extend to projective automorphisms of \mathbb{P}^2 . (Here the condition $n > 1$ is needed to guarantee that there is at most one such extension.)

Definition 4.4. An *algebraic set* (or Zariski closed set) in a projective space (such as in $\widehat{\mathcal{C}}_n$) over \mathbb{F} will mean any subset¹⁹ defined by finitely many homogeneous polynomial equations with coefficients in \mathbb{F} . Such a set is *irreducible* if it is not the union of two proper algebraic subsets. (The term *variety* will also be used for an irreducible algebraic set.) Any algebraic set can be expressed uniquely as a union of finitely many maximal irreducible algebraic subsets, which are called its *irreducible components*.

¹⁸If \mathcal{C} is defined by the equation $\Phi(x : y : z) = 0$, then $\mathbf{g}(\mathcal{C})$ is defined by

$$\Phi \circ \mathbf{g}^{-1}(x : y : z) = 0 .$$

¹⁹Caution: Here in the real case, we consider only real points of projective space. This contrasts with our terminology for curves, where we implicitly use complex points to distinguish between different real curves.

Remark 4.5. If the stabilizer $\mathbf{G}_{\mathcal{C}}$ is finite, then it is clearly an algebraic set. However, if it is infinite, then it is non-compact, and hence cannot be described as an algebraic subset of a projective space. However, it is still locally closed in the Zariski topology. More explicitly, we can think of PGL_3 as a Zariski open set in the projective space \mathbb{P}^8 consisting of lines through the origin in the space of 3×3 matrices. Its complement in \mathbb{P}^8 is the Zariski closed subset corresponding to singular matrices. Now the stabilizer of \mathcal{C} is locally closed, since it can be described as the intersection of the open set PGL_3 with a closed set described by polynomial equations in the entries a_{ij} of the matrices.

It follows that $\mathbf{G}_{\mathcal{C}}$ is either a finite group, or a Lie group with finitely many connected components. The distinction between these two possibilities is of fundamental importance. Note that a 1-cycle \mathcal{C} has infinite stabilizer if and only if its \mathbf{G} -orbit $((\mathcal{C})) \subset \widehat{\mathcal{C}}_n$ has dimension strictly less than $\dim(\mathbf{G}) = 8$. In fact there is a natural fibration with projection map $\mathbf{g} \mapsto \mathbf{g}(\mathcal{C})$ from \mathbf{G} to the subset $((\mathcal{C})) \subset \widehat{\mathcal{C}}_n$ with fiber $\mathbf{G}_{\mathcal{C}}$. It follows that

$$\dim(\mathbf{G}_{\mathcal{C}}) + \dim((\mathcal{C})) = \dim(\mathbf{G}), \quad (10)$$

where $\dim(\mathbf{G}) = 8$ in our case.

Definition 4.6. Let \mathfrak{W}_n be the subset of $\widehat{\mathcal{C}}_n$ consisting of all 1-cycles with infinite stabilizer. (Compare §8.) We will be particularly concerned with the complementary set

$$\widehat{\mathcal{C}}_n^{\mathrm{fs}} = \widehat{\mathcal{C}}_n \setminus \mathfrak{W}_n,$$

consisting of 1-cycles with *finite stabilizer*. The quotient space

$$\mathbb{M}_n = \mathbb{M}_n(\mathbb{F}) = \widehat{\mathcal{C}}_n^{\mathrm{fs}}/\mathbf{G}$$

consisting of all projective equivalence classes of such 1-cycles will be called the *moduli space* for 1-cycles of degree n over \mathbb{C} . Thus each element $((\mathcal{C})) \in \mathbb{M}_n$ is an equivalence class of effective 1-cycles in $\widehat{\mathcal{C}}_n^{\mathrm{fs}}$, where two 1-cycles \mathcal{C} and \mathcal{C}' are equivalent if and only if $\mathbf{g}(\mathcal{C}) = \mathcal{C}'$ for some $\mathbf{g} \in \mathbf{G}$, or in other words if and only if they belong to the same \mathbf{G} -orbit in the space of 1-cycles.

To begin the discussion, note that \mathfrak{W}_n is a closed subset of $\widehat{\mathcal{C}}_n$, or equivalently that its complement $\widehat{\mathcal{C}}_n^{\mathrm{fs}}$ is an open set. In fact, for any \mathcal{C} in \mathfrak{W}_n we can choose a one parameter subgroup of \mathbf{G} which maps \mathcal{C} to itself. Since the set of all one parameter subgroups of \mathbf{G} is compact, it follows easily that \mathfrak{W}_n is also closed and compact.

See §8 for a detailed study of $\mathfrak{W}_n(\mathbb{C})$, including a proof that it is a (Zariski closed) algebraic set.

Another important algebraic set, at least in the complex case, is the locus \mathcal{R}_n consisting of all reducible 1-cycles in $\widehat{\mathcal{C}}_n$. By definition, a 1-cycle is *reducible* if and only if it is in the image of the smooth map

$$(\mathcal{C}, \mathcal{C}') \mapsto \mathcal{C} + \mathcal{C}' \quad \text{from} \quad \widehat{\mathcal{C}}_k \times \widehat{\mathcal{C}}_{n-k} \quad \text{to} \quad \widehat{\mathcal{C}}_n$$

for some $0 < k < n$. In the complex case, each such image is an irreducible variety. In fact the product of two irreducible varieties is irreducible (see for example [Harr, Exercise 5.9]), while the image of an irreducible variety under projection is algebraic ([Harr, Thm. 3.5]), and is clearly irreducible. It follows that the set $\mathcal{R}_n(\mathbb{C})$ of all reducible 1-cycles is an algebraic set. However, in the real case the best one can say is that $\mathcal{R}_n(\mathbb{R})$ is a closed

semi-algebraic set, defined by polynomial equalities and inequalities.²⁰ (As an example the 1-cycle

$$x^2 + 2bxy + y^2 = 0$$

in $\mathbb{P}^2(\mathbb{R})$ is reducible if and only if $|b| \geq 1$, in which case it is equal to a union of two (not necessarily distinct) lines $y/x = -b \pm \sqrt{b^2 - 1}$ through the origin $(0 : 0 : 1)$ of the affine plane. In the case $|b| < 1$ this 1-cycle is irreducible, with real zero locus $|\mathcal{C}|_{\mathbb{R}}$ equal to the empty set.

We will be particularly interested in the complementary open set

$$\mathfrak{C}_n^{\text{irr}} = \widehat{\mathfrak{C}}_n \setminus \mathcal{R}_n \subset \mathfrak{C}_n$$

consisting of irreducible 1-cycles of degree n . Evidently every irreducible 1-cycle has all multiplicities equal to one, and hence is actually an irreducible curve.

Note that the dimension function $n \mapsto \dim(\widehat{\mathfrak{C}}_n) = n(n+3)/2$ is convex, in the sense that $\dim(\widehat{\mathfrak{C}}_n) > \dim(\widehat{\mathfrak{C}}_k) + \dim(\widehat{\mathfrak{C}}_{n-k})$ for $0 < k < n$. Here are a few values:

$$\begin{array}{cccccccc} n & 1 & 2 & 3 & 4 & 5 & \cdots & \\ \dim(\widehat{\mathfrak{C}}_n) & 2 & 5 & 9 & 14 & 20 & \cdots & . \end{array} \quad (11)$$

It follows by an easy computation that $\dim(\widehat{\mathfrak{C}}_n) = \dim(\widehat{\mathfrak{C}}_{n-1}) + n + 1$. Similarly it follows, for $n \geq 2$, that the dimension of the space of reducible cycles is

$$\dim(\mathcal{R}_n) = \dim(\widehat{\mathfrak{C}}_{n-1}) + 2 = \dim(\widehat{\mathfrak{C}}_n) - n + 1 .$$

In other words, the closed subset consisting of reducible cycles in $\widehat{\mathfrak{C}}_n$ has codimension $n-1$; provided that $n \geq 2$.

5. CURVES OF DEGREE THREE.

This section will study the moduli spaces $\mathbb{M}_3(\mathbb{R})$ and $\mathbb{M}_3(\mathbb{C})$ for curves of degree three. At first we will not distinguish between the real and complex cases, since the arguments are exactly the same for both.

Note that $n = 3$ is the first case where $\mathbb{M}_n \neq \emptyset$. In fact, for $n < 3$ it follows easily from Equation (10) and the Table in (11) that there are no 1-cycles with finite stabilizer; so that the corresponding moduli space \mathbb{M}_n is empty. Similarly, since $\dim(\mathcal{R}_3) = 7 < 8$, it follows that:

Every reducible curve or 1-cycle of degree 3 has infinite stabilizer. Thus, when studying \mathbb{M}_3 , it suffices to work with the open subset $\mathfrak{C}_3^{\text{irr}} \subset \mathfrak{C}_3$ consisting of irreducible curves.

Note that any irreducible curve of degree three (or more generally of odd degree) is uniquely determined by its real support $|\mathcal{C}|_{\mathbb{R}}$. In fact any curve of odd degree has many real points, since every real line intersects it in at least one point. (However, a reducible curve such as $x(y^2 + z^2) = 0$ is not necessarily determined by its real support.)

The following statement is surely well known, although we don't know any explicit reference which considers completely arbitrary cubic curves.

²⁰ See [BCR] for semi-algebraic sets.

Proposition 5.1. *If the field \mathbb{F} is either \mathbb{C} or \mathbb{R} , then a cubic curve $\mathcal{C} \subset \mathbb{P}^2(\mathbb{F})$ is irreducible if and only if it can be transformed, by an \mathbb{F} -projective change of coordinates, to the standard normal form, which can be written in affine coordinates $(x : y : 1)$ as*

$$y^2 = x^3 + ax + b. \quad (12)$$

Furthermore:

(a) *This associated normal form is unique up to the transformation*

$$(a, b) \mapsto (t^4a, t^6b) \quad (13)$$

where t can be any non-zero element of \mathbb{F} .

(b) *The curve defined by (12) has finite stabilizer if and only if $(a, b) \neq (0, 0)$.*

(c) *This reduction to normal form can be carried out uniformly over some neighborhood \mathcal{U} of any given \mathcal{C}_0 . That is, there is a smooth map $\mathcal{C} \mapsto \mathbf{g}_{\mathcal{C}}$ from \mathcal{U} to $\mathrm{PGL}_3(\mathbb{F})$ so that for any $\mathcal{C} \in \mathcal{U}$ the automorphism $\mathbf{g}_{\mathcal{C}} : \mathbb{P}^2 \rightarrow \mathbb{P}^2$ maps \mathcal{C} to a curve in standard normal form.*

The proof will depend on the following.

Lemma 5.2. *Every irreducible real or complex cubic curve contains at least one flex point.*

Proof of Lemma 5.2. In the complex case, we will make use of the classical Plücker formulas, which compare an irreducible curve $\mathcal{C} \subset \mathbb{P}^2(\mathbb{C})$ with its dual curve $\mathcal{C}^* \subset \mathbb{P}^{2*}$. (See for example [Nam] or [GH, p.278].) Here \mathbb{P}^{2*} is the dual complex projective plane consisting of all lines in \mathbb{P}^2 ; and $\mathcal{C}^* \subset \mathbb{P}^{2*}$ is the subset consisting of lines which are tangent to \mathcal{C} .

Consider a curve \mathcal{C} of degree n in $\mathbb{P}^2 = \mathbb{P}^2(\mathbb{C})$ with no singularities other than simple double points and cusps, and with only simple flex points and bitangent lines (that is lines which are tangent at two different points). These conditions are certainly satisfied in the cubic case. (In particular, a cubic curve can have no bitangents.) Let:

- f be the number of flex points,
- δ the number of double points,
- κ the number of cusp points, and
- b the number of bitangents;

and let f^* , δ^* , κ^* , and b^* be the corresponding numbers for the dual curve. Then

$$f = \kappa^* \iff f^* = \kappa \quad \text{and} \quad \delta = b^* \iff \delta^* = b.$$

Furthermore the degrees n^* of \mathcal{C}^* and n of \mathcal{C} are given by the formulas:

$$\begin{aligned} n^* &= n(n-1) - 2\delta - 3\kappa, \\ n &= n^*(n^*-1) - 2b - 3f. \end{aligned}$$

Now let us specialize to the case $n = 3$. Note that $\delta + \kappa \leq 1$, since otherwise it would follow from the last equation and its dual that $n^* \leq 2$ hence $n \leq 2$. Thus there are only three possible cases to consider.

In the smooth case $\delta = \kappa = 0$ these equations yield:

$$n^* = 6 \quad \text{and} \quad n = 3 = 30 - 3f; \quad \text{hence there are } f = 9 \text{ flex points.}$$

If there is a simple double point so that $\delta = 1$ and $\kappa = 0$, they yield:

$$n^* = 6 - 2 = 4 \quad \text{and} \quad 3 = 12 - 3f \quad \text{with } f = 3 \text{ flex points.}$$

In the case of a cusp point, with $\delta = 0$ and $\kappa = 1$, they yield:

$$n^* = 6 - 3 = 3 \quad \text{and} \quad 3 = 6 - 3f \quad \text{with } f = 1 \text{ flex point.}$$

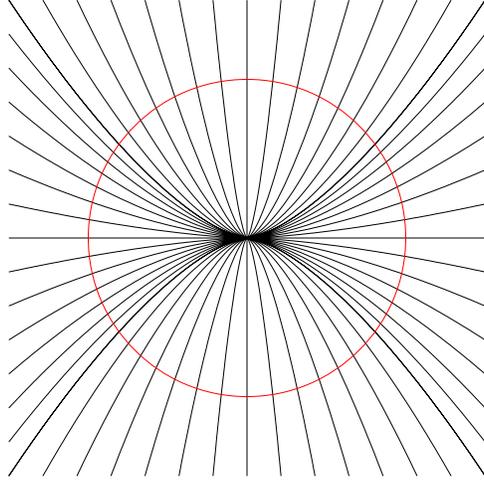


FIG. 16. Foliation of the real (a, b) -plane, with the origin removed, by the connected components of the curves $(a^3 : b^2) = \text{constant}$. More explicitly, each curve can be parametrized as $a = a_0 t^2$ and $b = b_0 t^3$ with $t > 0$, where (a_0, b_0) can be any point on one of these curves. The unit circle has a single transverse intersection with each curve.

Thus the number of flex points of an irreducible complex cubic is always odd.

It follows that every irreducible real cubic must have at least one flex point. In fact, the associated complex curve must also be irreducible. (Otherwise it would have a real factor.) Since the non-real flex points occur in complex conjugate pairs, there must be at least one real flex point; which proves Lemma 5.2. \square

Outline Proof of Proposition 5.1. (Compare [BM, §3].) Given a single flex point, choose a projective transformation which moves this point to $(0 : 1 : 0)$, and moves the tangent line to \mathcal{C} at this point to the line $z = 0$. The tangent line has a triple intersection point with the curve \mathcal{C} at $(0 : 1 : 0)$; hence it can have no other intersection point. This means that the defining equation for \mathcal{C} must now consist of an x^3 term, plus other terms which are all divisible by z . Furthermore, it must include an $y^2 z$ term, since otherwise all of its partial derivatives at $(0 : 1 : 0)$ would be zero. After switching to affine coordinates $(x : y : 1)$, and after multiplying x and y by suitable constants, the equation will have the form

$$y^2 = x^3 + ax + b \quad \text{plus terms in } x^2, xy, \text{ and } y.$$

However, the last two terms can be eliminated by adding a suitable $px + q$ to the y coordinate, and the x^2 term can then be eliminated by adding a suitable constant to the x coordinate. \square

Using Proposition 5.1, it follows easily that the quotient space $\mathfrak{C}_n^{\text{irr}}/\mathbf{G}$ of irreducible curves modulo the action of \mathbf{G} can be identified with the quotient of the plane consisting of pairs $(a, b) \in \mathbb{F}^2$, by the equivalence relation

$$(a, b) \sim (t^4 a, t^6 b) \quad \text{for any } t \neq 0. \quad (14)$$

(Compare Figure 16.) This entire quotient space has a rather nasty topology, since the center point $a = b = 0$, corresponding to the cusp curve

$$y^2 = x^3,$$

belongs to the closure of every other point. However, we will eliminate this problem by considering only curves with finite stabilizer, and hence removing the center point.

Theorem 5.3. *If the field \mathbb{F} is either \mathbb{R} or \mathbb{C} , then the moduli space*

$$\mathbb{M}_3(\mathbb{F}) = \mathfrak{C}_3^{\text{fs}}(\mathbb{F})/\mathbf{G}(\mathbb{F})$$

can be identified with the quotient of the punctured (a, b) -plane $\mathbb{F}^2 \setminus \{(0, 0)\}$ by the equivalence relation (14). In the real case this moduli space is real analytically diffeomorphic to the unit circle, with no ramification. In the complex case it is an orbifold, biholomorphic to the 2-sphere consisting of all ratios $(a^3 : b^2) \in \mathbb{P}^1(\mathbb{C})$, but with three ramified points. In both cases there is just one improper point, corresponding to curves with a simple self-crossing singularity, where the action of \mathbf{G} is only weakly proper.

In fact it is easy to check that every orbit in $\mathbb{R}^2 \setminus \{(0, 0)\}$ intersects the unit circle $a^2 + b^2 = 1$ exactly once. Therefore, the moduli space $\mathbb{M}_3(\mathbb{R})$ is real analytically diffeomorphic to the unit circle. In the complex case, each orbit corresponds to a fixed ratio $(a^3 : b^2) \in \mathbb{P}^1(\mathbb{C})$. (Equivalently, this ratio is captured by the shape invariant

$$\mathbf{J} = \frac{4a^3}{4a^3 + 27b^2} \in \widehat{\mathbb{C}}$$

of Remark 3.9.) Thus $\mathbb{M}_3(\mathbb{C})$ is biholomorphic to the Riemann sphere $\mathbb{P}^1(\mathbb{C}) \cong \widehat{\mathbb{C}}$. (Compare Remark 2.24.) Here is an alternative description.

Lemma 5.4. *The moduli space $\mathbb{M}_3(\mathbb{C})$ for complex cubic curves is a Hausdorff orbifold which is canonically isomorphic to the moduli space $\mathfrak{M}_4(\mathbb{C})$ for divisors of degree four. In particular both spaces have three ramified points, one of which is improper with index $\mathfrak{r} = 2$, the others having index $\mathfrak{r} = 2$ and $\mathfrak{r} = 3$ respectively.*

Proof. Given four distinct points on the Riemann sphere, there is a unique 2-fold covering curve, branched at these four points. In fact, if we place one of the four points at infinity, and let $f(x)$ be the monic cubic polynomial with roots at the three finite points, then the locus $y^2 = f(x)$ in the affine plane, or $y^2z = f(x, z)$ in the projective plane, maps to \mathbb{P}^1 by the required branched covering map

$$(x : y : z) \mapsto (x : z),$$

branched at the four specified points. It follows from Proposition 5.1 that every smooth cubic curve can be put in this form. Furthermore, as two of the four points come together, the corresponding cubic curve will tend to a curve with a simple double point. Any symmetry fixing the point at infinity will give rise to a corresponding symmetry of the cubic curve, so that there is a precise correspondence between ramified points for divisors of degree four and for curves of degree three.

However, this does not mean that the corresponding stabilizers are isomorphic. We have seen that a generic divisor of degree four has stabilizer $\mathbb{Z}/2 \oplus \mathbb{Z}/2$. On the other hand, the stabilizer of a generic complex cubic curve is non-abelian of order eighteen. (See the paragraph following Theorem 9.3.) For the two proper ramified points, with ramification index two [or three], the order of the stabilizer is 8 [or 12] for divisors, but 36 [or 54] for cubic curves.

Now consider the degenerate (improper) case where two points of the divisor come together, or where the cubic curve acquires a self-crossing point. Then the stabilizer has order two in the divisor case, and order 6 in the cubic curve case. (Compare [BM, Figure 10], where the six symmetries of one real form of this curve are generated by 120° rotations, and reflections on the vertical axis.) By Lemma 2.25, it follows that the class of the complex cubic curve with a self-crossing point is not even locally proper.

However, the action is weakly locally proper. We can write the singular curve in standard normal form as

$$\Phi(x, y) = x^3 - 3x + 2 - y^2 = 0 ,$$

with $\partial\Phi/\partial x = \partial\Phi/\partial y = 0$ at the singular point $(1, 0)$, and with invariant

$$(a^3 : b^2) = (-27 : 4) .$$

It follows from Proposition 5.1(c) that any curve which is close to this singular curve can be reduced to standard normal form by an automorphism close to the identity. Furthermore, if two such curves belong to the same \mathbf{G} -orbit, then it follows easily from Proposition 5.1(a) that we can transform one to the other by an automorphism close to the identity. Thus the action is weakly proper, which completes the proof of Lemma 5.4, and hence proves Theorem 5.3 in the complex case.

In the real case, we need only note that the curve with a self-crossing point, as illustrated in Figure 18, has stabilizer generated by the reflection $(x, y) \mapsto (x, -y)$, while every smooth curve has stabilizer of order six. (See the more detailed exposition below.) Therefore, the action is not locally proper at this point. The proof that it is weakly proper proceeds just as in the complex case. This completes the proof of Theorem 5.3. \square

The Circle of Real Cubic Curves. Although $\mathbb{M}_3(\mathbb{C})$ is isomorphic to $\mathfrak{M}_4(\mathbb{C})$, the real analog $\mathbb{M}_3(\mathbb{R})$ is quite different from $\mathfrak{M}_4(\mathbb{R})$. We have shown that the space $\mathbb{M}_3(\mathbb{R})$ of projective equivalence classes of cusp-free irreducible real cubic curves is diffeomorphic to the unit circle. In fact each such curve-class has a unique representative with equation of the form

$$y^2 = x^3 + ax + b \quad \text{with} \quad a^2 + b^2 = 1 , \quad (15)$$

However, this “unit circle” normal form seems somewhat arbitrary. For example, if we used a circle of different radius, then we would get a quite different parametrization.

Note that the natural map from $\mathbb{M}_3(\mathbb{R})$ onto the real part of $\mathbb{M}_3(\mathbb{C})$ is definitely not one-to-one. In fact it is precisely two-to-one. Two real curves of the form $y^2 = x^3 + ax + b$ and $y^2 = x^3 + ax - b$ with $b \neq 0$ are not real projectively equivalent; yet they have the same ratio $(a^3 : b^2)$ and hence are complex projectively equivalent.

We know of two quite natural ways of mapping $\mathbb{M}_3(\mathbb{R})$ homeomorphically (but not diffeomorphically) onto the real projective line $\mathbb{P}^1(\mathbb{R})$. One is given by representing each smooth curve-class by its unique Hesse normal form

$$x^3 + y^3 + z^3 = 3kxyz , \quad (16)$$

with $k \neq 1$. (Compare [BM, Theorem 6.3 and Figure 10].) This is very convenient, since it displays the stabilizer as the group of permutations of the three coordinates, with order six. It also shows that each smooth cubic curve \mathcal{C} has a unique **center of symmetry** in $\mathbb{P}^2 \setminus |\mathcal{C}|$ which is invariant under all projective automorphisms, and which corresponds to the point $(1 : 1 : 1)$ in Hesse normal form.²¹

²¹Using the standard normal form, the center of symmetry \mathbf{c} can be computed as follows. If \mathbf{m} is the midpoint of the two finite flex points, then the tangent lines at these two flex points intersect at the point

The Hesse normal form works beautifully for smooth curve-classes. However, the limits as k tends to $+1$ or $\pm\infty$ are badly behaved, and yield reducible curves.²²

Another possible choice is the **flex-slope normal form**

$$y^2 = x^3 + (sx + 1)^2, \quad (17)$$

in affine²³ coordinates, where s can be any real number. Here the two finite flex points are located at $(x, y) = (0, \pm 1)$, and s is equal to the slope dy/dx at the upper flex point.

Lemma 5.5. *For any irreducible real cubic curve $\mathcal{C} \subset \mathbb{P}^2(\mathbb{R})$, the following three properties are equivalent²⁴*

- (a) \mathcal{C} is either smooth, or else smooth except at one isolated point.
- (b) \mathcal{C} contains three flex points.
- (c) \mathcal{C} is equivalent, under a real projective change of coordinates, to one and only one curve $\mathcal{F}(s)$ in the flex-slope normal form.

Note that every curve in the form (17) has a flex point of slope s at $(0, 1)$, as well as a flex point of slope $-s$ at $(0, -1)$. In fact it follows easily from (17) that

$$y = \pm(1 + sx) + O(x^3) \quad \text{as } x \rightarrow 0,$$

so that the points $(0, \pm 1)$ satisfy $dy/dx = \pm s$ and $d^2y/dx^2 = 0$.

See Figure 17 for some typical examples. Note that the first four curves in this figure are connected, while the last three have two components (although one component in Figure 17(ε) is extremely small). The first and last curves would be much larger if drawn to scale: they have been shrunk to fit in the picture. If we rescale by setting $x = s^2X$ and $y = s^3Y$ so that $Y^2 = X^3 + (X + 1/s^3)^2$, then the curve would converge to $Y^2 = X^3 + X^2$ as $s \rightarrow \pm\infty$, with a self-crossing singular point at $X = Y = 0$, as shown in Figure 18.

Proof of 5.5. The proof proceeds in three steps as follows.

Proof that (a) \implies (b). Every smooth real or complex cubic curve has a classical additive group structure. Choosing a flex point as identity element (compare Lemma 5.2), the flex points are the solutions to the equation $3\mathbf{x} = 0$. In the complex case the group is isomorphic to $\mathbb{R}/\mathbb{Z} \oplus \mathbb{R}/\mathbb{Z}$ with nine flex points; while in the real case the group is isomorphic to either \mathbb{R}/\mathbb{Z} or $\mathbb{R}/\mathbb{Z} \times (\mathbb{Z}/2)$, with a unique subgroup of order three, and hence exactly three flex points. (Compare [BM, §5].)

There are just two real cubic curve-classes made up of singular curves. (Compare Figure 17 ε and Figure 18.) For a curve with an isolated singular point, we must prove that there are still three flex points. Putting the curve in standard normal form, with one flex point on the line at infinity, note that the slope dy/dx of the upper branch $y > 0$ of this curve tends to $+\infty$, both as $y \rightarrow 0$ and as $x, y \rightarrow +\infty$. Therefore the slope must take on a minimum value, necessarily at a flex point, somewhere on this branch. It follows easily that there are three flex points altogether. \square

²² $\frac{2}{3}\mathbf{m} + \frac{1}{3}\mathbf{c}$. (If a sequence of smooth cubics converges to an irreducible singular cubic, one can check that the centers of symmetry converge to the singular point. Compare Figure 17(ε).)

²³We could get around this, and obtain the correct irreducible curve-class by taking the limit of carefully rescaled curves.

²⁴In homogeneous coordinates, the equation is $y^2z = x^3 + z(sx + z)^2$.

²⁵Another equivalent condition is that the group of automorphisms is the non-abelian group of order 6.

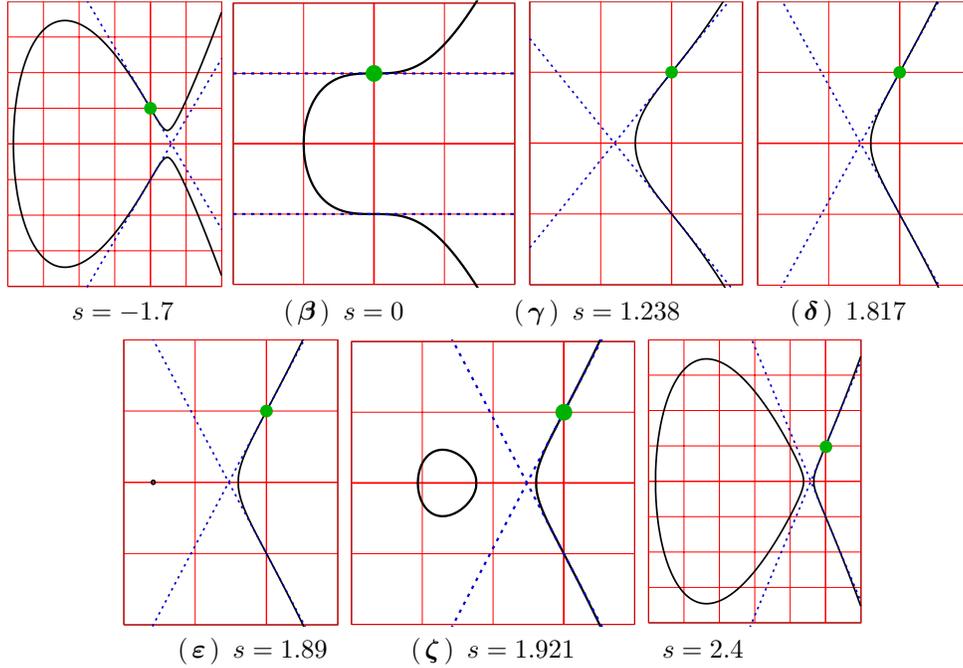


FIG. 17. Graphs of seven curves $\mathcal{F}(s)$ in flex-slope normal form, so that the finite flex points are at $(0, \pm 1)$. Here the slope s ranges from -1.7 to $+2.4$. The tangent lines at the flex points are indicated by dotted lines. The grid of points with an integer coordinate is also shown. Note the isolated singular point which appears at $s = 1.88988\dots$ and immediately expands to a circle. The figures blow up as $s \rightarrow \pm\infty$. (See Figure 18 for the limiting behavior.) The five middle curves, labeled as (β) through (ζ) , correspond to the points with the same labels in Figure 19.

(On the other hand, as we converge towards the rescaled curve of Figure 18, the equation $Y^2 = X^3 + (X + 1/s^3)^2$ converges to $Y^2 = X^3 + X^2$, and the two finite flex points at $(0, \pm 1/s^3)$ converge to the singular point at the origin.)

Proof that (b) \implies (c). Now assume that there are three flex points. According to Proposition 5.1, we can put the curve into standard normal form $y^2 = x^3 + ax + b$, with one flex point on the line at infinity. Let $(x_1, \pm y_1)$ be the two finite flex points. Translating the x -coordinate appropriately, and multiplying the y coordinate by a constant, we can move the flex points to $(0, \pm 1)$. The curve will then be defined by a monic equation of the form $y^2 = p(x)$ so that

$$\frac{dp}{dx} = 2y \frac{dy}{dx} \quad \text{and} \quad \frac{d^2p}{dx^2} = 2 \left(\frac{dy}{dx} \right)^2 + 2y \frac{d^2y}{dx^2}.$$

Now evaluate at $(x, y) = (0, 1)$. Since $y = 1$, and $dy/dx = s$, and $d^2y/dx^2 = 0$, it follows easily that $p(x) = 1 + 2sx + s^2x^2 + x^3$, as required. Furthermore, since the construction

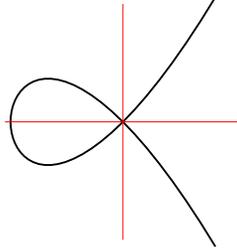


FIG. 18. Although the flex-slope normal form blows up as s tends to $\pm\infty$, a carefully rescaled version, with $Y^2 = X^3 + (X + 1/s^3)^2$, tends to the illustrated curve, with a simple self-crossing at the limit of the two finite flex points. This limit belongs to the leaf α in Figure 19.

is uniquely specified, it follows that the parameter s is uniquely determined by the curve-class. \square

Proof that (c) \implies (a). Finally, assuming that the curve is in flex-slope normal form (17), we must show that it is either smooth everywhere, or else has just one isolated point as singularity. It is not hard to show that any singularity must lie on the x -axis, and correspond to a double or triple root of the polynomial $x^3 + (sx + 1)^2$. Using the standard formula for the discriminant of a cubic polynomial (see for example [BMac]), we can check that the discriminant of this polynomial is given by $\Delta = 4s^3 - 27$. Therefore, the corresponding curve is singular if and only if

$$s = 3/\sqrt[3]{4} = 1.88988\dots$$

For this value of s , it is not hard to check that the associated polynomial factors as

$$x^3 + (sx + 1)^2 = (x + r)^2(x + r/4) \quad \text{where } r = \sqrt[3]{4},$$

with a double root at $-r$ and a simple root at $-r/4$. It follows easily that the associated curve has an isolated point at $(-r, 0)$. Thus we have proved that (a) \implies (b) \implies (c) \implies (a), completing the proof of Lemma 5.5. \square

Remark 5.6. More generally, for any curve of the form

$$Y^2 = X^3 + AX^2 + BX + C$$

the flex-slope invariant can be computed as

$$s = \frac{dY/dX}{\sqrt[3]{Y}},$$

to be evaluated at either of the finite flex points $(X_0, \pm Y_0)$. In fact, if we set $x = \lambda^2 X$ and $y = \lambda^3 Y$, then the slope will be $dy/dx = \lambda dY/dX$. Choosing $\lambda = 1/\sqrt[3]{Y_0}$, the y coordinate at the finite flex points will be ± 1 , and we can translate the x coordinate so that it will be zero at these points.

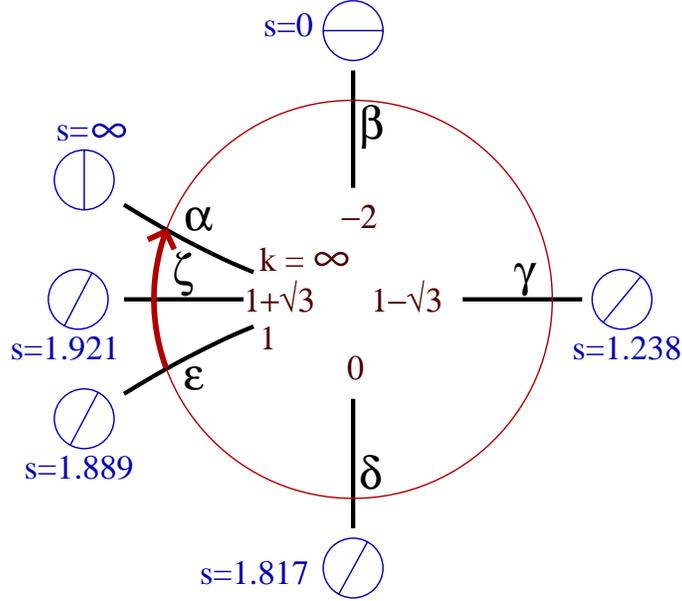


FIG. 19. Showing the unit circle in the (a, b) -plane together with six leaves from the foliation of Figure 16. The labels α through ζ on these leaves correspond to the labels on the cubic curves of Figures 17 and 18. The numbers outside the circle in this figure give the flex-slope invariant s for the corresponding curves, and the numbers inside the circle give the Hesse invariant k . Note that both s and k increase monotonically from $-\infty$ to $+\infty$ as we follow the circle clockwise from α back around to α . (Thus s takes negative values between α and β .) The curve associated with any point of this circle has two connected components if and only if $k \geq 1 \iff s \geq 1.88988\dots$; corresponding to the emphasized arc on the left.

Thus we have three different possible normal forms for real cubic curves: the unit circle normal form (15), the Hesse normal form (16), and the flex-slope normal form (17). These are compared in Figure 19. Two of the leaves in this figure correspond to singular curves, and separate the connected cubic curves from curves with two components: The α -leaf is the set of curves with a self-crossing point, giving rise to improper group action; while the ε -leaf is the set of curves with an isolated (necessarily singular) point. A cubic curve has two components if and only if $1 \leq k < \infty \iff 1.889\dots \leq s < \infty$.

The β leaf also corresponds to cubic curves with a distinctive geometry. These are the only real cubics such that the tangent lines at the three flex points all pass through a common point. Note that the β leaf and the δ leaf both lie in the coordinate line $a = 0$, with shape invariant $\mathbf{J} = 0$. These correspond to complex curves with six-fold rotational symmetry. Similarly both the γ -leaf and the ζ -leaf lie in the coordinate line $b = 0$ with $\mathbf{J} = 1$, corresponding to complex cubics with four-fold rotational symmetry.

It is noteworthy that the one singular curve-class in $\mathbb{M}_3(\mathbb{C})$ splits into two distinct singular curve-classes α and ε in $\mathbb{M}_3(\mathbb{R})$. However, only the singularity of type α , corresponding to curves with a real self-crossing point, is improper. All other curves in $\mathfrak{C}_3^{\text{fs}}(\mathbb{R})$ have

an automorphism group \mathfrak{S}_3 of order six; but real curves of class α have an automorphism group of order two.

6. COMPLEX CURVES OF DEGREE AT LEAST FOUR

Recall from Section 4 that the moduli space $\mathbb{M}_n = \mathbb{M}_n(\mathbb{C})$ is defined to be the quotient space $\widehat{\mathfrak{C}}_n^{\text{fs}}/\mathbf{G}$, where $\widehat{\mathfrak{C}}_n^{\text{fs}} = \widehat{\mathfrak{C}}_n^{\text{fs}}(\mathbb{C})$ is the open subset consisting of 1-cycles with finite stabilizer in the complex projective space $\widehat{\mathfrak{C}}_n$ consisting of all 1-cycles of degree n , and where \mathbf{G} is the projective linear group $\text{PGL}_3(\mathbb{C})$. Here is a preliminary statement.

Theorem 6.1. *For $n \geq 5$ the moduli space \mathbb{M}_n is not a Hausdorff space.*

Proof. Consider the subspace of \mathbb{M}_n consisting of formal sums

$$m_1 \cdot L_1 + \cdots + m_k \cdot L_k \quad \text{with} \quad n = \sum m_j ,$$

where the L_j are lines. Each line L_j in the plane \mathbb{P}^2 is dual to a point \mathbf{p}_j in the dual plane \mathbb{P}^{2*} , yielding an associated zero-cycle $m_1 \langle \mathbf{p}_1 \rangle + \cdots + m_k \langle \mathbf{p}_k \rangle$ in the dual plane. The argument is now similar to the proof of Lemma 3.12, but with suitable modification since we are now working in \mathbb{P}^2 rather than \mathbb{P}^1 .

By definition, four points of \mathbb{P}^2 are in *general position* if no three are contained in a common line. Note that the action of $\mathbf{G} = \text{PGL}_3$ on \mathbb{P}^2 is simply transitive on ordered 4-tuples $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4)$ which are in general position. In fact there is one and only one group element \mathbf{g} such that

$$\mathbf{g}(\mathbf{p}_1) = (1 : 0 : 0), \quad \mathbf{g}(\mathbf{p}_2) = (0 : 1 : 0), \quad \mathbf{g}(\mathbf{p}_3) = (0 : 0 : 1), \quad \text{and} \quad \mathbf{g}(\mathbf{p}_4) = (1 : 1 : 1) .$$

It follows that any zero-cycle which includes four points in general position will have finite stabilizer (of order at most $k!$ where k is the number of distinct points in the cycle).

First consider the case $n = 5$. For any $\varepsilon > 0$, let

$$\mathbf{g}_\varepsilon(x : y : z) = (x : y : \varepsilon z) .$$

The zero-cycle

$$\mathcal{D}_0 = (1 : 0 : 0) + (0 : 1 : 0) + (0 : 0 : 1)$$

is invariant under \mathbf{g}_ε . Let

$$\mathcal{D}_\varepsilon = \mathcal{D}_0 + (1 : 1 : 1) + (\varepsilon : \varepsilon : 1)$$

so that

$$\mathbf{g}_\varepsilon(\mathcal{D}_\varepsilon) = \mathcal{D}_0 + (1 : 1 : \varepsilon) + (1 : 1 : 1) .$$

Thus as $\varepsilon \rightarrow 0$ we have

$$\lim_{\varepsilon \rightarrow 0} \mathcal{D}_\varepsilon = (1 : 0 : 0) + (0 : 1 : 0) + 2 \cdot (0 : 0 : 1) + (1 : 1 : 1) ;$$

but

$$\lim_{\varepsilon \rightarrow 0} \mathbf{g}_\varepsilon(\mathcal{D}_\varepsilon) = (1 : 0 : 0) + (0 : 1 : 0) + (0 : 0 : 1) + (1 : 1 : 0) + (1 : 1 : 1) .$$

These two limits are clearly not \mathbf{G} -equivalent since the first limit has a point $(0 : 0 : 1)$ of multiplicity two, while the second limit is a zero-cycle with five distinct points of multiplicity one. This proves that the quotient space is not Hausdorff.

The proof for $n > 5$ can now be obtained by adding extra \mathbf{G} -invariant terms of the form $(x : y : 0)$ to \mathcal{D}_ε . \square

In this section and the next one we will describe moderately large open subsets of \mathbb{M}_n which are Hausdorff. We first prove that \mathbb{M}_n is a T_1 -space.

Lemma 6.2. *For any \mathbf{G} -orbit $((\mathcal{C})) \subset \widehat{\mathfrak{C}}_n$, every limit point in the complement $\overline{((\mathcal{C}))} \setminus ((\mathcal{C}))$ has infinite stabilizer.*

(Compare [AF2, p. 527].)

Proof. More generally, the corresponding statement holds for any algebraic action of an algebraic group \mathbf{G} on a projective space \mathbb{P}^N . (See [Br, Prop. 1.11].) The proof can be outlined very briefly as follows. The orbit of any $\mathbf{p} \in \mathbb{P}^N$ can be described as the image of a map $\mathbf{g} \mapsto \mathbf{g}(\mathbf{p})$ from \mathbf{G} to \mathbb{P}^N . Denoting this image by $\mathbf{G}\mathbf{p}$, it follows from a theorem of Chevalley that $\mathbf{G}\mathbf{p}$ is a “constructible set”, and hence contains a subset U which is dense and open as a subset of its closure $\overline{\mathbf{G}\mathbf{p}} \subset \mathbb{P}^N$, using the Zariski topology. (Compare [Gor] or the proof of [Harr, Thm. 3.16].) Thus the complement $\overline{\mathbf{G}\mathbf{p}} \setminus \mathbf{G}\mathbf{p}$ is contained in an algebraic set $\overline{\mathbf{G}\mathbf{p}} \setminus U$ with dimension strictly smaller than the dimension of \mathbf{G} . It follows, using equation (10), that every point in this complement has infinite stabilizer. \square

As an immediate Corollary, it follows that \mathbb{M}_n is a T_1 space:

Corollary 6.3. *Every \mathbf{G} -orbit which is contained in $\widehat{\mathfrak{C}}_n^{\text{fs}}$ is a closed subset of $\widehat{\mathfrak{C}}_n^{\text{fs}}$. In other words, every point in the quotient space $\mathbb{M}_n = \widehat{\mathfrak{C}}_n^{\text{fs}}/\mathbf{G}$ is closed.*

In the rest of this section we will use the concept of “virtual flex point” in order to prove that certain large subsets of \mathbb{M}_n are Hausdorff.

Definition 6.4. Let $\mathfrak{U}_n \subset \mathfrak{C}_n$ be the open set consisting of all *line-free curves* \mathcal{C} of degree n . In other words, for $\mathcal{C} \in \mathfrak{U}_n$ we assume:

- (1) that \mathcal{C} contains no line, and
- (2) that every irreducible component of \mathcal{C} has multiplicity one.

For $\mathcal{C} \in \mathfrak{U}_n$, a point $\mathbf{p} \in |\mathcal{C}|$ will be called a *virtual flex point* if it is either a singular point or a flex point.

Lemma 6.5. *Every virtual flex point \mathbf{p} for a curve in \mathfrak{U}_n can be assigned a **flex-multiplicity** $\varphi(\mathbf{p}) \geq 1$ with the following two properties:*

- (1) *The sum of $\varphi(\mathbf{p})$ over all virtual flex points is equal to $3n(n-2)$.*
- (2) *Under a generic small perturbation, each virtual flex point \mathbf{p} splits into $\varphi(\mathbf{p})$ distinct nearby simple flex points.*

Proof of Lemma 6.5. Define $\varphi(\mathbf{p})$ as the local intersection multiplicity between the curve \mathcal{C} of degree n and its associated Hessian curve $\mathcal{H}_{\mathcal{C}}$ of degree $3(n-2)$. (See for example [Gr], [Kir], [Kun], or [Sha].) The statement then follows from Bézout’s Theorem. In order to apply Bézout, it is first necessary to check that \mathcal{C} and $\mathcal{H}_{\mathcal{C}}$ have no common sub-curve. But such a sub-curve would have to be either a line or a component of multiplicity ≥ 2 ; and both possibilities have been excluded. To prove that $\varphi(\mathbf{p}) \geq 1$ at every singular point, we proceed as follows. Taking the singular point of a curve of degree n to be $(0:0:1)$ the defining equation must have the form

$$\Phi(x, y, z) = \sum_{j=2}^n \Phi_j(x, y) z^{n-j}$$

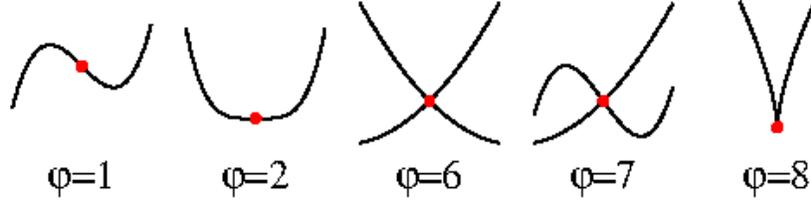


FIG. 20. Examples of virtual flex points with small flex-multiplicity. The first is a simple flex point, the second is a double flex point ($y = x^4$), and the remaining three are singular points.

where the Φ_j are homogeneous of degree j . It is then easy to check that the last row $(\Phi_{xz}, \Phi_{yz}, \Phi_{zz})$ of the Hessian matrix is identically zero at $(0, 0, 1)$; so that this point is a common zero of Φ and the Hessian determinant. \square

Example 6.6. Suppose that there are k smooth local branches (= curve germs) $\mathcal{B}_1, \dots, \mathcal{B}_k$ of the curve \mathcal{C} passing through \mathbf{p} . (Compare the cases $\varphi = 6, 7$ in Figure 20 and for other examples see Figure 22.) Then

$$\varphi_{\mathcal{C}}(\mathbf{p}) = \sum_j \varphi_{\mathcal{B}_j}(\mathbf{p}) + 6 \sum_{i < j} \mathcal{B}_i \cdot \mathcal{B}_j,$$

where $\mathcal{B}_i \cdot \mathcal{B}_j$ is the local intersection multiplicity. (Here $\varphi_{\mathcal{B}_j}(\mathbf{p})$ makes sense, since φ can be defined as a local analytic invariant.) This equation can be proved by choosing small generic translations of the \mathcal{B}_j so that they intersect transversally, and then noting that a simple intersection has flex-multiplicity six.²⁵

It will be convenient to consider the probability measure on \mathbb{P}^2 defined by

$$\widehat{\varphi}(S) = \frac{\sum_{\mathbf{p} \in S} \varphi(\mathbf{p})}{3n(n-2)} \in [0, 1] \quad \text{for every set } S \subset \mathbb{P}^2.$$

Here it is understood that $\varphi(\mathbf{p}) = 0$ unless \mathbf{p} is a virtual flex point in $|\mathcal{C}|$.

Definition 6.7. To every $\mathcal{C} \in \mathcal{U}_n$, we can assign the two rational numbers

$$\mathbf{p}_{\max} = \mathbf{p}_{\max}(\mathcal{C}) = \max_{\mathbf{p}} \widehat{\varphi}_{\mathcal{C}}(\mathbf{p}) \quad \text{and} \quad \mathbf{L}_{\max} = \mathbf{L}_{\max}(\mathcal{C}) = \max_L \widehat{\varphi}_{\mathcal{C}}(L),$$

where \mathbf{p} ranges over all points in $|\mathcal{C}| \subset \mathbb{P}^2$, and where L ranges over all lines in \mathbb{P}^2 . Evidently, since there are at most n points of \mathcal{C} on L

$$0 < \mathbf{p}_{\max} \leq \mathbf{L}_{\max} \leq 1, \quad \text{and} \quad \mathbf{L}_{\max} \leq n \mathbf{p}_{\max}. \quad (18)$$

(The large emphasized triangle in Figure 21 encloses all points satisfying these inequalities.)

²⁵Consider for example a cubic curve with a single double point, it follows from the proof of Lemma 5.2 that there are three flex points, which must certainly have multiplicity $\varphi = 1$. Since $\sum \varphi = 9$, it follows that $\varphi = 6$ at the double point. (Alternatively, see Figure 22(4) for an example of degree $n = 4$ with four simple double points and no flexes, and with $\sum \varphi = 3n(n-2) = 24$.)

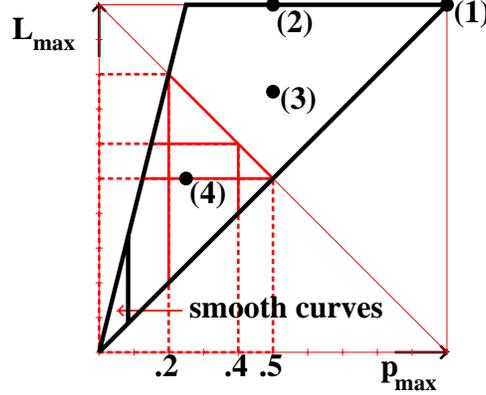


FIG. 21. The large black triangle encloses all possible pairs $(\mathbf{p}_{\max}(\mathcal{C}), \mathbf{L}_{\max}(\mathcal{C}))$ for curves of degree $n = 4$, while the small black subtriangle encloses the possible pairs for smooth degree four curves. (See the proof of Corollary 6.9.) By Theorem 6.8, the \mathbf{G} -action is locally proper for curves below the diagonal line $\mathbf{p}_{\max} + \mathbf{L}_{\max} = 1$. The boundaries of three typical rectangles $\mathfrak{U}_4(1/5)$, $\mathfrak{U}_4(2/5)$, and $\mathfrak{U}_4(1/2)$ of provably proper action are also shown. The heavy dots correspond to the four examples shown in Figure 22.

Theorem 6.8. *If $\mathbf{p}_{\max}(\mathcal{C}_0) + \mathbf{L}_{\max}(\mathcal{C}_0) < 1$ for a curve $\mathcal{C}_0 \in \mathfrak{U}_n$, then the action of \mathbf{G} on \mathfrak{U}_n is locally proper at \mathcal{C}_0 , hence the moduli space \mathbb{M}_n is locally Hausdorff at the corresponding point $\pi(\mathcal{C}_0) \in \mathbb{M}_n$. In fact, choosing a real number κ so that $\mathbf{p}_{\max}(\mathcal{C}_0) < \kappa$ and $\mathbf{L}_{\max}(\mathcal{C}_0) < 1 - \kappa$, the action of \mathbf{G} is proper throughout the entire open subset $\mathfrak{U}_n(\kappa) \subset \mathfrak{U}_n$ consisting of curves which satisfy $\mathbf{p}_{\max}(\mathcal{C}) < \kappa$ and $\mathbf{L}_{\max}(\mathcal{C}) < 1 - \kappa$.*

Corollary 6.9. *The action of \mathbf{G} is locally proper at every smooth curve in \mathfrak{C}_n . In fact, every smooth curve belongs to $\mathfrak{U}_n(\kappa)$ for every κ between $1/(n+1)$ and $1/2$. Furthermore, if \mathcal{C} is any curve in \mathfrak{U}_n which satisfies $\mathbf{p}_{\max} + \mathbf{L}_{\max} < 1$, then \mathcal{C} belongs to $\mathfrak{U}_n(\kappa)$ for some κ between $1/(n+1)$ and $1/2$. Thus every curve-class with $\mathbf{p}_{\max} + \mathbf{L}_{\max} < 1$ belongs to an open subset of \mathbb{M}_n which is a Hausdorff orbifold, and which contains every smooth curve-class.*

Proof of Corollary 6.9 (assuming Theorem 6.8). If \mathcal{C} is a smooth curve, then every virtual flex point is an actual flex point, with flex-multiplicity

$$\varphi \leq n - 2, \quad \text{hence} \quad \widehat{\varphi} \leq 1/3n.$$

Thus $\mathbf{p}_{\max} \leq 1/3n$; and therefore $\mathbf{L}_{\max} \leq 1/3$ since there are at most n points of \mathcal{C} on any line. It follows easily that $\mathcal{C} \in \mathfrak{U}_n(\kappa)$ for every κ between $1/(n+1)$ and $1/2$.

Note that if we parametrize the line $\mathbf{p}_{\max} + \mathbf{L}_{\max} = 1$ by setting

$$(\mathbf{p}_{\max}, \mathbf{L}_{\max}) = (\kappa, 1 - \kappa),$$

then this line intersects the triangle defined by the inequalities (18) precisely in the interval $1/(n+1) \leq \kappa \leq 1/2$. (Compare Figure 21.) Further details are easily supplied. \square

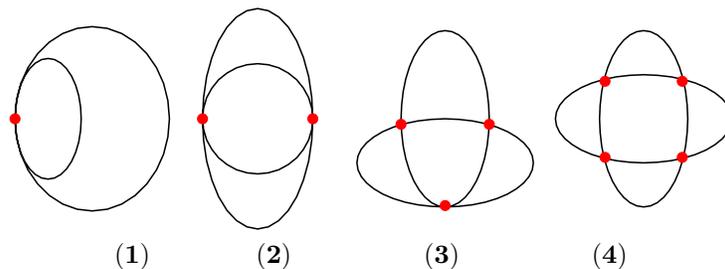


FIG. 22. Representatives for the four equivalence classes of curves which are unions of two smooth quadratic curves. The first two are W-curves. (See §8.) In each case there are four intersection points, counted with multiplicity.

Example 6.10. Let \mathcal{C} be a curve of degree four which is the union of two smooth curves of degree two. Then \mathcal{C} has no flex points, but may have either one, two, three, or four singular points, as illustrated in Figure 22. The corresponding values of \mathbf{p}_{\max} and \mathbf{L}_{\max} can be tabulated as follows. (Compare the four labeled points in Figure 21.)

| | | | | |
|---|-------------|----------------|-------------------|------------------|
| #singular points : | (1) | (2) | (3) | (4) |
| $\mathbf{p}_{\max} + \mathbf{L}_{\max}$: | $1 + 1 = 2$ | $.5 + 1 = 1.5$ | $.5 + .75 = 1.25$ | $.25 + .5 = .75$ |

Thus Theorem 6.8 implies that the last curve represents a point in moduli space which is proper, and hence locally Hausdorff. On the other hand, the first two are W-curves, and hence do not represent any point of moduli space. (Compare Figure 28 in Section 8.) The point of \mathbb{M}_4 corresponding to the third curve is more interesting:

Proposition 6.11. *Let \mathcal{C}_3 be a curve in $\mathbb{P}^2(\mathbb{C})$ which is the union of two smooth quadratic curves which have three intersection points (as in Figure 22(3)). Then the action of \mathbf{G} is not even weakly proper at \mathcal{C}_3 .*

Proof. We will make use of the statement that the automorphism group (= stabilizer) of a generic curve of degree four is trivial. (See Theorem 9.3 below.)

Note that any curve \mathcal{C} which is the union of two smooth quadratic curves with four distinct intersection points has a group of projective automorphisms which is transitive on these four points.

To see this, note that the four points must be in general position, since a line can intersect a smooth quadratic in at most two points. Thus, after a projective transformation, we can put the intersection points at $(\pm 1, \pm 1)$. The general quadratic equation in affine coordinates can be written as $Q(x, y) + L(x, y) = c$; where Q is homogeneous quadratic and L is linear. If the equation is to hold at all of the points $(\pm 1, \pm 1)$, then it is easy to check that the linear term $L(x, y)$ must be zero, and that the coefficient of xy in the quadratic term must be zero. Thus we are reduced to an equation of the form $ax^2 + by^2 = c$; where evidently c must equal $a + b$.

Any curve defined by an equation of the form $ax^2 + by^2 = a + b$ is clearly invariant under the four element group

$$(x, y) \mapsto (\pm x, \pm y) .$$

It follows that any union \mathcal{C} of two such curves is also invariant under this four element group. Let $\mathbf{g}_{\mathcal{C}}$ be the element of the stabilizer $\mathbf{G}_{\mathcal{C}}$ corresponding to the involution $(x, y) \leftrightarrow (-x, -y)$. We can choose a curve \mathcal{C}_0 arbitrarily close to \mathcal{C} which has no non-trivial automorphism. It

follows that the curves \mathcal{C}_0 and $\mathbf{g}_\mathcal{C}(\mathcal{C}_0)$ represent the same point of \mathbb{M}_4 ; but that $\mathbf{g}_\mathcal{C}$ is the only group element carrying one to the other.

Now, as we move two of the intersection points together to obtain the third curve **(3)** in Figure 22, the corresponding sequence of involutions clearly cannot lie in any compact subset of \mathbf{G} . Thus the action of \mathbf{G} is not weakly proper at this curve. \square

We don't know whether the moduli space \mathbb{M}_4 is locally Hausdorff near the corresponding point.

The proof of Theorem 6.8 will be based on the following.

Lemma 6.12 (Distortion Lemma for \mathbb{P}^2). *For any $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset \mathbf{G} = \mathrm{PGL}_3(\mathbb{C})$ such that, for any $\mathbf{g} \in \mathbf{G} \setminus K_\varepsilon$, one or both of the following two conditions is satisfied. Either:*

(1) *there exists a line $L^+ \subset \mathbb{P}^2$ and a point $\mathbf{p}^- \in \mathbb{P}^2$ such that*

$$\mathbf{g}(N_\varepsilon(L^+)) \cup N_\varepsilon(\mathbf{p}^-) = \mathbb{P}^2, \text{ or}$$

(2) *there exists a point $\mathbf{p}^+ \in \mathbb{P}^2$ and a line $L^- \subset \mathbb{P}^2$ such that*

$$\mathbf{g}(N_\varepsilon(\mathbf{p}^+)) \cup N_\varepsilon(L^-) = \mathbb{P}^2.$$

Note that we can interchange the two cases simply by replacing \mathbf{g} by \mathbf{g}^{-1} . It is also important to note that both **(1)** and **(2)** are open conditions, which will remain valid under a small perturbation of ε .

Proof of Lemma 6.12. In order to prove this Lemma, we will discuss first the special situation in which the group action is diagonalizable of the form

$$\mathbf{g}(x : y : z) = (x' : y' : z') = (rx : sy : tz) \quad \text{with} \quad |r| \geq |s| \geq |t| > 0.$$

The ratio $\delta = |r/t| \geq 1$ can be thought of as a measure of the distortion of \mathbf{g} . Evidently δ is bounded on any compact set of diagonal matrices. We want to study the limiting behavior as $\delta \rightarrow \infty$.

Note that either $|r/s| \geq \sqrt{\delta}$, or $|s/t| \geq \sqrt{\delta}$ (or both if $|s|$ is precisely equal to $\sqrt{|rt|}$).

Case(1). Suppose that $|r/s| \geq \sqrt{\delta}$. It will be convenient to choose a representative (x, y, z) for $(x : y : z)$ so that $|x| + |y| + |z| = 1$, and also a representative (r, s, t) for $(r : s : t)$ (depending on x, y, z) so that the triple

$$(x', y', z') = (rx, sy, tz) \quad \text{satisfies} \quad |x'| + |y'| + |z'| = 1.$$

Then

$$\frac{|x'|}{|y'|} = \left| \frac{rx}{sy} \right| \geq \sqrt{\delta} \left| \frac{x}{y} \right| \quad \text{and similarly} \quad \left| \frac{x'}{z'} \right| = \left| \frac{sx}{tz} \right| \geq \delta \left| \frac{x}{z} \right|.$$

It follows that

$$|y'| \leq \frac{|x'y|}{|x|\sqrt{\delta}} \leq \frac{1}{|x|\sqrt{\delta}} \quad \text{and similarly} \quad |z'| \leq \frac{1}{|x|\delta}.$$

Thus if $|x|$ is bounded away from zero, and δ tends to infinity, then $|y'|$ and $|z'|$ must tend to zero; hence $(x' : y' : z')$ must tend to the point $(1 : 0 : 0)$. More explicitly, if $|x| > \varepsilon$ and if δ is large enough, then $(x : y : z)$ must map into the ε neighborhood of this point.

But according to Lemma 2.30, any element of $\mathbf{G} = \mathrm{PGL}_3$ can be written uniquely as a product $\mathbf{r} \circ \mathbf{g} \circ \mathbf{r}'$ where \mathbf{r} and \mathbf{r}' are unitary rotations, and where \mathbf{g} is diagonal. Suppose as

above that the diagonal transformation \mathbf{g} maps everything outside of a small neighborhood of the line $x = 0$ into a small neighborhood of the point $y = z = 0$. Then setting

$$L^+ = \mathbf{r}'^{-1}(\{x = 0\}) \quad \text{and} \quad \mathbf{p}^- = \mathbf{r}(\{y = z = 0\}),$$

we obtain the required line and point, with

$$\begin{array}{ccccccc} \mathbb{P}^2 & \xrightarrow{\mathbf{r}'} & \mathbb{P}^2 & \xrightarrow{\mathbf{g}} & \mathbb{P}^2 & \xrightarrow{\mathbf{r}} & \mathbb{P}^2 \\ L^+ & \xrightarrow{\cong} & \{x = 0\} & & \{y = z = 0\} & \xrightarrow{\cong} & \mathbf{p}^- . \end{array}$$

Case(2). If $|s/t| \geq \sqrt{\delta}$, then a similar argument shows that we have

$$|z'| \leq \frac{1}{|x|\delta} \quad \text{as above,} \quad \text{but now} \quad |z'| \leq \frac{1}{|y|\sqrt{\delta}}.$$

Thus if $(x : y : z)$ is bounded away from the point $(0 : 0 : 1)$, then either $|x| \geq \varepsilon$ or $|y| \geq \varepsilon$. Therefore, as δ tends to infinity, $|z'|$ must tend to zero, so that we converge to the line $z' = 0$. Further details of the proof are straightforward. \square

Proof of Theorem 6.8. Recall that $\mathfrak{U}_n(\kappa)$ is the set of line-free curves $\mathcal{C} \in \mathfrak{C}_n$ satisfying $\mathbf{p}_{\max}(\mathcal{C}) < \kappa$ and $\mathbf{L}_{\max}(\mathcal{C}) < 1 - \kappa$. Given two curves \mathcal{C}_0 and \mathcal{C}'_0 in $\mathfrak{U}_n(\kappa)$, first choose ε small enough so that the ε -neighborhoods $N_\varepsilon(\mathbf{p})$ of the various virtual flex points of \mathcal{C}_0 are disjoint, and so that there exists a line intersecting any three of the neighborhoods $N_{2\varepsilon}(\mathbf{p}_j)$ only if the center points \mathbf{p}_j are collinear. (Compare Figure 23. Such an ε must exist since, if there were such a line for arbitrarily small ε , then the center points would have to be collinear.) Furthermore, we also require the corresponding conditions for \mathcal{C}'_0 .

Now let \mathfrak{N} be the neighborhood of \mathcal{C}_0 consisting of all curves $\mathcal{C} \in \mathfrak{U}_n$ such that, for every virtual flex point \mathbf{p} of \mathcal{C}_0 , the number of virtual flex points of \mathcal{C} in $N_\varepsilon(\mathbf{p})$, counted with flex-multiplicity, is exactly the flex-multiplicity of $\mathbf{p} \in \mathcal{C}_0$. Construct the neighborhood \mathfrak{N}' of \mathcal{C}'_0 in the analogous way.

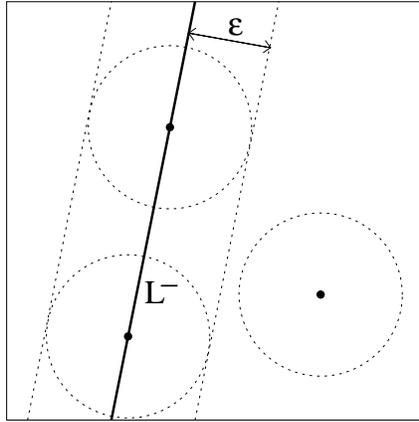


FIG. 23. Illustrating the proof of Theorem 6.8. For $\mathcal{C}' \in \mathfrak{N}'$, every virtual flex point of \mathcal{C}' must have distance less than ε from the corresponding virtual flex point of \mathcal{C}'_0 .

Choosing K_ε as in Lemma 6.12, we must show that there cannot be any $\mathcal{C} \in \mathfrak{N}$ and any $\mathbf{g} \notin K_\varepsilon$ such that $\mathbf{g}(\mathcal{C}) = \mathcal{C}'$ belongs to \mathfrak{N}' . Replace \mathbf{g} by \mathbf{g}^{-1} if necessary, so that we are in Case (2) of the Distortion Lemma 6.12. Then $\widehat{\varphi}_{\mathcal{C}}(N_\varepsilon(\mathbf{p}^+)) < \kappa$, and it follows from the Distortion Lemma that the ε -neighborhood of L^- contains more than $1 - \kappa$ virtual flex points of $\mathbf{g}(\mathcal{C}) = \mathcal{C}'$. This contradiction completes the proof of Theorem 6.8. \square

Remark 6.13 (The Classical Moduli Space $\mathcal{M}_{\mathbf{g}}$). Since a smooth curve of degree n in $\mathbb{P}^2(\mathbb{C})$ has genus $\mathbf{g}(n) = \binom{n-1}{2}$, it is natural to compare the moduli space $\mathbb{M}_n^{\text{sm}}(\mathbb{C})$ for smooth curves in \mathbb{P}^2 with the classical moduli space $\mathcal{M}_{\mathbf{g}(n)}$, consisting of conformal isomorphism classes of closed Riemann surfaces of genus $\mathbf{g}(n)$. The dimension of this classical moduli space is given by²⁶

$$\dim(\mathcal{M}_{\mathbf{g}}) = 3\mathbf{g} - 3 \quad \text{for } \mathbf{g} \geq 2; \quad \text{but} \quad \dim(\mathcal{M}_1) = 1 .$$

Compare [ACGH, p.28] or [Mu, Ch. 5].

For every $n \geq 3$ there is a natural inclusion

$$\mathbb{M}_n^{\text{sm}}(\mathbb{C}) \hookrightarrow \mathcal{M}_{\mathbf{g}(n)} .$$

(Compare Remark 9.10 below.) The case $n = 3$ is exceptional, since in this case, we obtain an isomorphism

$$\mathbb{M}_3^{\text{sm}}(\mathbb{C}) \xrightarrow{\cong} \mathcal{M}_1 ,$$

where both spaces are isomorphic to \mathbb{C} with two ramified points, using the shape invariant \mathbf{J} of §5 (which is just the classical j -invariant, up to a multiplicative constant: compare [BM]).

For $n = 4$, the inclusion $\mathbb{M}_4^{\text{sm}} \hookrightarrow \mathcal{M}_3$ can be understood as follows. Any closed Riemann surface \mathcal{C} of genus \mathbf{g} has \mathbf{g} linearly independent holomorphic 1-forms, say $\omega_1, \dots, \omega_{\mathbf{g}}$. For any point $\mathbf{p} \in \mathcal{C}$, the ratio

$$(\omega_1(\mathbf{p}) : \dots : \omega_{\mathbf{g}}(\mathbf{p}))$$

can be interpreted as a point in the projective space $\mathbb{P}^{\mathbf{g}-1}(\mathbb{C})$. Thus there is a canonical map from any Riemann surface of genus $\mathbf{g} > 1$ into $\mathbb{P}^{\mathbf{g}-1}(\mathbb{C})$, well defined up to projective automorphisms of $\mathbb{P}^{\mathbf{g}-1}(\mathbb{C})$. Furthermore any conformal automorphism of \mathcal{C} corresponds to a change of basis for the vector space of 1-forms, and hence to a projective automorphism of its image in the projective $(\mathbf{g} - 1)$ -space.

By definition, a smooth complex curve \mathcal{C} of genus $\mathbf{g} > 1$ is called *hyperelliptic* if it admits a holomorphic function $\mathcal{C} \rightarrow \mathbb{P}^1(\mathbb{C})$ of degree two (or equivalently if it is a 2-fold branched covering of the sphere). The following is proved for example in [Be] or [Gr].

Proposition 6.14. *If \mathcal{S} is a Riemann surface of genus three which is not hyperelliptic, then the canonical map $\mathcal{S} \rightarrow \mathbb{P}^2(\mathbb{C})$ is a smooth embedding. Furthermore, every embedding of a Riemann surface of genus 3 into $\mathbb{P}^2(\mathbb{C})$ can be obtained by this construction; and every conformal automorphism of the Riemann surface corresponds to a projective automorphism*

²⁶For $\mathbf{g} \geq 2$ the moduli space $\mathcal{M}_{\mathbf{g}}$ can be considered as a quotient space $\mathcal{T}_{\mathbf{g}}/\text{MCG}_{\mathbf{g}}$, with the associated orbifold structure. Here $\mathcal{T}_{\mathbf{g}}$ is the $(3\mathbf{g} - 3)$ -dimensional Teichmüller space, and the mapping class group $\text{MCG}_{\mathbf{g}}$ is a discrete group which acts on $\mathcal{T}_{\mathbf{g}}$. (See for example [Hub].) For $\mathbf{g} = 1$ the corresponding description would be that \mathcal{M}_1 is the quotient of the upper half-plane in \mathbb{C} by the group $\text{PSL}(2, \mathbb{Z})$ of fractional linear transformations.

of the embedded curve. On the other hand, a hyperelliptic Riemann surface of genus three cannot be embedded in $\mathbb{P}^2(\mathbb{C})$.

Corollary 6.15. *The moduli space $\mathbb{M}_4^{\text{sm}}(\mathbb{C})$ for smooth projective curves of degree four maps bijectively onto the open subset of \mathcal{M}_3 consisting of conformal equivalence classes of non-hyperelliptic Riemann surfaces of genus three. Furthermore, any conformal automorphism of a smooth projective curve of degree four extends to a projective automorphism of $\mathbb{P}^2(\mathbb{C})$.*

(See Remark 9.10 for a more general statement.)

Proof of Corollary. This follows since a smooth curve of genus three in $\mathbb{P}^2(\mathbb{C})$ necessarily has degree four; and since any conformal automorphism of a Riemann surface corresponds to a projective automorphism of its image in $\mathbb{P}^{g-1}(\mathbb{C})$. \square

Note that there exists a smooth curve of genus \mathfrak{g} in \mathbb{P}^2 only if \mathfrak{g} is a number of the form $\mathfrak{g}(n) = \binom{n-1}{2}$. For degrees $n \geq 5$ and hence $\mathfrak{g}(n) \geq 6$, the dimension

$$\dim \mathcal{M}_{\mathfrak{g}(n)} = 3\mathfrak{g}(n) - 3 = 3(n^2 - 3n)/2$$

is strictly larger than

$$\dim \mathbb{M}_n = \dim \widehat{\mathcal{C}}_n - 8 = (n^2 + 3n - 16)/2 .$$

Therefore only a very thin set of Riemann surfaces of genus $\mathfrak{g} \geq 6$ can be embedded smoothly into $\mathbb{P}^2(\mathbb{C})$. (However every Riemann surface can be immersed into $\mathbb{P}^2(\mathbb{C})$ with only simple double points. See for example [Hart, Corollaries 3.6 and 3.11, Chapter IV].)

7. SINGULARITY GENUS AND PROPER ACTION.

This Section will describe another criterion for proper action, based on the genus invariant for a singular point. (See Theorems 7.13 and 7.17.) However, we must first understand the concept of genus for surfaces which may be non-compact and non-algebraic.

The Topological Genus for More General Surfaces.

By a *surface* \mathcal{S} we will mean a C^1 -smooth,²⁷ oriented 2-dimensional Hausdorff manifold, possibly with C^1 -smooth boundary. (In particular, any Riemann surface is also a surface in this sense.) The genus $\mathfrak{g}(\mathcal{S}) \geq 0$ of such a surface is a fundamental topological invariant, taking integer values (or the value $+\infty$ in some non-compact cases).

It will be convenient to work with the homology group $H_1 = H_1(\mathcal{S}; \mathbb{Q})$, using rational coefficients.²⁸ Choosing an orientation for \mathcal{S} , any two homology classes $\alpha, \beta \in H_1$ have a well defined intersection number, yielding a skew-symmetric bilinear pairing

$$(\alpha, \beta) \mapsto \alpha \cdot \beta = -\beta \cdot \alpha \in \mathbb{Q} .$$

²⁷ Smoothness is not really necessary, but makes proofs easier.

²⁸ One could equally well use coefficients in any field. The field $\mathbb{Z}/2$ has the unique property that the definition works equally well for non-orientable surfaces. However the “genus” of a non-orientable surface, defined in this way, may be a half-integer. For example, the real projective plane has “genus” $1/2$, and the hyperbolic dodecahedron of Figure 13 has “genus” $5/2$.

Definition 7.1. The *genus* $\mathfrak{g}(\mathcal{S})$ is defined to be the rank of this intersection pairing, divided by two. In other words, $2\mathfrak{g}$ is the dimension of the “*reduced homology group*” H_1/N , where N is the null space, consisting of all α such that $\alpha \cdot \beta = 0$ for all β . As an example, if \mathcal{S} is a compact surface with boundary, then \mathfrak{g} is finite, and N is generated by the homology classes of the boundary circles. (It will be convenient in this section to use the word “circle” for any manifold which is homeomorphic to the standard circle.) Whenever \mathfrak{g} is finite, it is not difficult to choose a basis for H_1/N so that the matrix for this pairing consists of \mathfrak{g} blocks of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ along the diagonal, with zeros elsewhere.

Theorem 7.2. *Here are seven basic properties of the genus.*

- (1) **Additivity.** *If \mathcal{S} is the union of two disjoint open subsurfaces²⁹ \mathcal{S}_1 and \mathcal{S}_2 , then $\mathfrak{g}(\mathcal{S}) = \mathfrak{g}(\mathcal{S}_1) + \mathfrak{g}(\mathcal{S}_2)$.*
- (2) **Monotonicity.** *If $\mathcal{S} \subset \mathcal{S}'$, then $\mathfrak{g}(\mathcal{S}) \leq \mathfrak{g}(\mathcal{S}')$.*
- (3) **Puncture Tolerance.** *The genus of \mathcal{S} is unchanged if we remove any finite subset from \mathcal{S} . Similarly, it is unchanged if we remove the interior of a closed disk which is embedded in the interior of \mathcal{S} .*
- (4) **Compact versus Non-Compact.** *If \mathcal{S} is non-compact, then $\mathfrak{g}(\mathcal{S})$ is equal to the supremum of the genera of compact sub-surfaces. On the other hand, if \mathcal{S} is compact with boundary $\partial\mathcal{S}$, then the genus of \mathcal{S} is equal to the genus of the interior $\mathcal{S} \setminus \partial\mathcal{S}$.*
- (5) **The Closed Surface Case.** *For a compact surface \mathcal{S} with empty boundary (or more generally for one such that each connected component has at most one boundary component), the doubled genus is equal to the first Betti number:*

$$2\mathfrak{g}(\mathcal{S}) = \dim H_1(\mathcal{S}) . \tag{19}$$

- (6) **Cutting and Pasting.** *Let \mathcal{S} be a compact surface of genus \mathfrak{g} with ℓ connected components and with b boundary circles. Then the Euler characteristic can be computed as*

$$\chi(\mathcal{S}) = 2\ell - 2\mathfrak{g} - b . \tag{20}$$

If we form a new surface \mathcal{S}' by pasting together two boundary circles of \mathcal{S} , then $\chi(\mathcal{S}) = \chi(\mathcal{S}')$. (In fact the number b decreases by two, but either ℓ decreases by one, or \mathfrak{g} increases by one.)

- (7) **The Classical Basic Example.** *If $\mathcal{C} \subset \mathbb{P}^2$ is a smooth complex curve of degree n , then (as noted earlier) the genus is given by*

$$\mathfrak{g}(\mathcal{C}) = \binom{n-1}{2} . \tag{21}$$

Remark 7.3. Felix Klein defined³⁰ the genus as the maximal number of disjoint non-separating circles which can be placed on the surface (that is, circles C_j such that both sides of any C_j belong to the same component of $\mathcal{S} \setminus \bigcup C_j$). In particular, a surface has genus zero if and only if has the Jordan property of being separated by any embedded circle.)

²⁹It is customary to define genus only for connected surfaces, but this extension to non-connected cases is often convenient.

³⁰See [Gra, p. 230].

It is a non-trivial exercise, using properties (4) and (6), to check that this agrees with our definition of genus in all cases.

Proof of Theorem 7.2. The first five statements follow easily from corresponding statements for the “reduced homology group” H_1/N ; which are not difficult to check.

For (1): If \mathcal{S} is the disjoint union of \mathcal{S}_1 and \mathcal{S}_2 , then evidently

$$H_1(\mathcal{S})/N(\mathcal{S}) = H_1(\mathcal{S}_1)/N(\mathcal{S}_1) \oplus H_1(\mathcal{S}_2)/N(\mathcal{S}_2).$$

For (2): If $\mathcal{S} \subset \mathcal{S}'$, then $H_1(\mathcal{S})/N(\mathcal{S})$ maps injectively into $H_1(\mathcal{S}')/N(\mathcal{S}')$. In fact, if two elements of $H_1(\mathcal{S})$ have non-zero intersection number, then the corresponding two elements of $H_1(\mathcal{S}')$ will have the same non-zero intersection number.

For (3): If $\mathcal{S}' = \mathcal{S} \setminus \{\mathbf{p}\}$, then $H_1(\mathcal{S})$ can be obtained from $H_1(\mathcal{S}')$ by setting the homology class of a small loop around the puncture point equal to zero. However, this doesn't change the quotient $H_1(\mathcal{S}')/N(\mathcal{S}')$ since the homology class of this loop belongs to $N(\mathcal{S}')$.

For (4): This follows since $H_1(\mathcal{S})$ is clearly isomorphic to the direct limit of the homology groups of the compact subsurfaces of \mathcal{S} .

For (5), the Poincaré duality theorem for a compact oriented n -manifold without boundary implies that the intersection pairing $(\alpha, \beta) \mapsto \alpha \cdot \beta$ from $H_j \times H_{n-j}$ to \mathbb{Q} is non-singular,³¹ giving rise to an isomorphism from H_j to the cohomology group $H^{n-j} \cong \text{Hom}(H_{n-j}, \mathbb{Q})$. In particular, for the special case $j = n - j = 1$, this implies that the null space N is trivial, so that $H_1/N = H_1$.

The proof of (6) will make use of properties of the Euler characteristic. Note first the additive property

$$\chi(X \cup Y) = \chi(X) + \chi(Y) - \chi(X \cap Y), \quad (22)$$

which clearly holds whenever X and Y are finite complexes with $X \cap Y$ as a subcomplex. In the special case where $X \cap Y$ is a finite union of circles, since the Euler characteristic of a circle is zero, this simplifies to $\chi(X \cup Y) = \chi(X) + \chi(Y)$. As an example, suppose that X is a compact connected surface of genus \mathbf{g} bounded by b circles. Then we can choose Y to be a union of b closed disks so that $X \cap Y$ is the union of these circles, and so that $X \cup Y$ is a closed surface of genus \mathbf{g} . Then

$$\chi(X \cup Y) = 2 - 2\mathbf{g} = \chi(X) + \chi(Y) = \chi(X) + b,$$

yielding the standard formula $\chi(X) = 2 - 2\mathbf{g} - b$. Now if \mathcal{S} is the disjoint union of ℓ such manifolds, since both \mathbf{g} and the number of boundary circles are additive, we obtain the required identity

$$\chi(\mathcal{S}) = 2\ell(\mathcal{S}) - 2\mathbf{g}(\mathcal{S}) - b(\mathcal{S})$$

for any compact surface. (Where $\ell(\mathcal{S})$ is the number of components of \mathcal{S} .)

For the proof of (7), see for example [GH]. □

Here is a convenient consequence of property (6).

³¹See for example [GH]. The intersection pairing in homology corresponds to the cup product pairing in the dual cohomology groups.

Corollary 7.4. *Suppose that a compact connected surface \mathcal{S} of genus \mathfrak{g} can be obtained from a disjoint union of ℓ connected surfaces \mathcal{S}_j of genera $\mathfrak{g}_1, \dots, \mathfrak{g}_\ell$ by pasting together k pairs of boundary circles. Then*

$$\mathfrak{g} = k + 1 - \ell + \sum_{j=1}^{\ell} \mathfrak{g}_j . \quad (23)$$

In fact it follows from (20) that

$$\chi(\mathcal{S}_j) = 2 - 2\mathfrak{g}_j - b_j , \quad \text{and that} \quad \chi(\mathcal{S}) = 2 - 2\mathfrak{g} - \left(\sum b_j - 2k \right) ;$$

where the expression in parentheses is the number of boundary circles of \mathcal{S} . Equation (23) then follows easily since $\chi(\mathcal{S}) = \sum \chi(\mathcal{S}_j)$.

The Genus of a Singularity. Let \mathbf{p} be a (necessarily isolated) singular point of a complex curve $\mathcal{C} \subset \mathbb{P}^2$, and let $N_{\mathbf{p}}$ the ε -ball centered at \mathbf{p} , using the standard Study-Fubini metric (Definition 2.31). If ε is small enough, then for every smooth curve \mathcal{C}' of the same degree which approximates \mathcal{C} closely enough (depending on ε), the intersection $\mathcal{S}_{\mathbf{p}} = \mathcal{C}' \cap \overline{N_{\mathbf{p}}}$ will be a smooth³² compact connected surface with b boundary components, where $b \geq 1$ is the number of local branches of \mathcal{C} through the point \mathbf{p} ; and where the genus of $\mathcal{S}_{\mathbf{p}}$ is independent of the choice of \mathcal{C}' . This is proved for example in [Mi1] or [Wa].

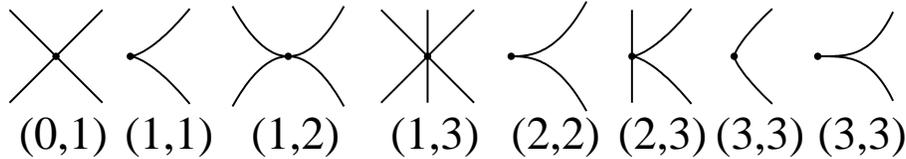


FIG. 24. Showing the pair of invariants $(\mathfrak{g}, \mathfrak{g}^+)$ for the eight possible singularities with $\mathfrak{g}^+ \leq 3$. These are respectively a simple node, an $x^3 = y^2$ cusp, a tacnode, a triple crossing, an $x^5 = y^2$ cusp, an $x^3 = y^2$ cusp together with a non-tangent line, an $x^3 = y^4$ cusp, and an $x^7 = y^2$ cusp. (For computation of these invariants, see Example 7.10 as well as Lemma 7.12.)

Definition 7.5. By the *genus* $\mathfrak{g}_{\mathcal{C}}(\mathbf{p})$ of the singularity we will mean the genus of this surface $\mathcal{S}_{\mathbf{p}}$. We also define the *augmented genus* (or *δ -invariant*³³) as

$$\mathfrak{g}_{\mathcal{C}}^+(\mathbf{p}) = \mathfrak{g}_{\mathcal{C}}(\mathbf{p}) + b - 1 .$$

Example 7.6. Here is an example to illustrate these ideas. Let \mathcal{C} be a singular projective curve with singular points $\mathbf{p}_1, \dots, \mathbf{p}_m$. The topological genus of the non-singular open subset

$$\mathcal{C} \setminus \{\mathbf{p}_1, \dots, \mathbf{p}_m\} .$$

is called³⁴ the *geometric genus* $\mathfrak{g}_{\text{geom}}(\mathcal{C})$.

³²To see this, note that the set of ε' for which the intersection is not transverse has measure zero by Sard's Theorem. In fact, it is a semi-algebraic set (see [BCR]), and hence must be finite.

³³See for example [Ser1, pp. 59-65] or [Nam, p.126]. We have preferred the \mathfrak{g}^+ notation since it makes the connection with genus clearer.

³⁴This is equivalent to the standard definition in the case of an irreducible curve. Our $\mathfrak{g}_{\text{geom}}(\mathcal{C})$ is just the sum of the geometric genera of the irreducible components of \mathcal{C} .

Lemma 7.7 (Degree and Geometric Genus). *With \mathcal{C} as above, we have*

$$\mathfrak{g}_{\text{geom}}(\mathcal{C}) + \sum_{j=1}^m \mathfrak{g}_{\mathcal{C}}^+(\mathbf{p}_j) = \binom{n-1}{2} + \ell - 1, \quad (24)$$

where n is the degree of \mathcal{C} and $\ell \geq 1$ is its number of irreducible components.

Proof. Choose a small closed ball around each \mathbf{p}_j and let \mathcal{C}' be a smooth degree n curve which closely approximates \mathcal{C} . Let \mathcal{S}_j be the intersection of \mathcal{C}' with the ball around \mathbf{p}_j and let \mathcal{S}' be the closure of $\mathcal{C}' \setminus (\mathcal{S}_1 \cup \dots \cup \mathcal{S}_m)$. If the balls are small enough and the approximation is close enough, then each \mathcal{S}_j will have genus $\mathfrak{g}(\mathbf{p}_j)$, and will have b_j boundary circles, where b_j is the number of local branches of \mathcal{C} at \mathbf{p}_j . Furthermore, \mathcal{S}' will be a smooth curve with $\sum_j b_j$ boundary circles, and with ℓ connected components \mathcal{S}'_k , where ℓ is the number of irreducible components of \mathcal{C} ; and with $\mathfrak{g}(\mathcal{S}')$ equal to the geometric genus $\mathfrak{g}_{\text{geom}}(\mathcal{C})$. Now applying Equation (23) to the surface \mathcal{C}' , which is the union of the \mathcal{S}_j together with the ℓ components \mathcal{S}'_k of \mathcal{S}' , pasted together along $\sum b_j$ boundary circles, we see that

$$\mathfrak{g}(\mathcal{C}') = \sum_{j=1}^m b_j + 1 - (m + \ell) + \left(\sum_{j=1}^m \mathfrak{g}(\mathcal{S}_j) + \sum_{k=1}^{\ell} \mathfrak{g}(\mathcal{S}'_k) \right).$$

Here the left side is equal to $\binom{n-1}{2}$, while the right side can be rearranged as

$$\sum_1^m \left(\mathfrak{g}(\mathcal{S}_j) + b_j - 1 \right) + \mathfrak{g}_{\text{geom}}(\mathcal{C}) + 1 - \ell = \sum_1^m \mathfrak{g}^+(\mathbf{p}_j) + \mathfrak{g}_{\text{geom}}(\mathcal{C}) + 1 - \ell.$$

The required equation (24) now follows easily. \square

Remark 7.8. The numbers $\mathfrak{g}_{\mathcal{C}}$ and $\mathfrak{g}_{\mathcal{C}}^+$ are closely related to the ‘‘Milnor number’’ μ . (See for example [Mi1], [Wa], [Ghy], [Sea].) Using affine coordinates (x, y) , this number μ for the curve $F(x, y) = 0$ at a point \mathbf{p} can be defined as the intersection multiplicity between the curves $F_x = 0$ and $F_y = 0$ at \mathbf{p} , where the subscripts indicate partial derivatives.

If $\mathbf{p} = (0, 0)$, then μ can be computed as the dimension of the quotient algebra $\mathbb{C}[[x, y]]/(F_x, F_y)$, where $\mathbb{C}[[x, y]]$ is the ring of formal power series in two variables and (F_x, F_y) stands for the ideal generated by these two partial derivatives. (Compare [Fu, p.9].) It follows easily that $\mu > 0$ if and only if \mathbf{p} is a singular point of \mathcal{C} .

Lemma 7.9. *The Milnor number μ of a singular point is the sum of the genus \mathfrak{g} and the augmented genus \mathfrak{g}^+ . That is,*

$$\mu = \mathfrak{g} + \mathfrak{g}^+. \quad (25)$$

Since $0 \leq \mathfrak{g}^+ \leq \mu \leq 2\mathfrak{g}^+$, it follows that $\mathfrak{g}^+ > 0$ if and only if \mathbf{p} is a singular point.

Proof. According to [Mi1, Theorem 7.2], μ is equal to the first Betti number $\dim(H_1(\mathcal{S}_{\mathbf{p}}))$ of the surface $\mathcal{S}_{\mathbf{p}}$. We must show that the sum

$$\mathfrak{g} + \mathfrak{g}^+ = 2\mathfrak{g} + b - 1$$

is equal to this Betti number $\dim(H_1(\mathcal{S}_{\mathbf{p}}))$. Recall from Theorem 7.2(6) that the Euler characteristic of a connected surface of genus \mathbf{g} with b boundary components is $2 - 2\mathbf{g} - b$. Comparing this with the standard expression

$$\dim(H_0) - \dim(H_1) + \dim(H_2)$$

for the Euler characteristic (where $H_2 = 0$ in our case), we obtain

$$2 - 2\mathbf{g} - b = 1 - \dim(H_1) + 0,$$

and hence $2\mathbf{g} + b - 1 = \dim(H_1)$. The equation (25) follows. \square

Example 7.10. For the curve $F(x, y) = x^p - y^q = 0$, the quotient algebra

$$\mathbb{C}[[x, y]]/(F_x, F_y) = \mathbb{C}[[x, y]]/(x^{p-1}, y^{q-1})$$

has an additive basis consisting of the $(p-1)(q-1)$ monomials $x^j y^k$ with $0 \leq j < p-1$ and $0 \leq k < q-1$. Therefore

$$\boldsymbol{\mu} = (p-1)(q-1).$$

Let δ be the greatest common divisor of p and q . Then it is not hard to check that there are δ branches of the curve through the point $(0, 0)$. From the equations $\mathbf{g} + \mathbf{g}^+ = \boldsymbol{\mu}$ and $\mathbf{g}^+ - \mathbf{g} = \delta - 1$ we see that

$$\mathbf{g} = (\boldsymbol{\mu} + 1 - \delta)/2 \quad \text{and} \quad \mathbf{g}^+ = (\boldsymbol{\mu} + \delta - 1)/2. \quad (26)$$

In the case of a cusp curve, with p and q relatively prime, this reduces to

$$\mathbf{g} = \mathbf{g}^+ = \boldsymbol{\mu}/2 = (p-1)(q-1)/2.$$

(Compare the four cusps shown in Figure 24.) On the other hand for $p = q = 2$ the curve

$$x^2 - y^2 = (x+y)(x-y) = 0$$

has a simple self-crossing, and we get $\boldsymbol{\mu} = 2$ and $(\mathbf{g}, \mathbf{g}^+) = (0, 1)$. For $p = 4, q = 2$ the curve $x^4 - y^2 = (x^2 + y)(x^2 - y)$ has a ‘‘tacnode’’ singularity, with two smooth branches intersecting tangentially. In this case $\boldsymbol{\mu} = 3$ and $(\mathbf{g}, \mathbf{g}^+) = (1, 2)$, again as in Figure 24.

Remark 7.11 (Erratum). In [Mi1, p. 60], it was stated incorrectly that the invariant $\boldsymbol{\mu}$ is equal to the classical multiplicity of the singularity. In fact the classical *multiplicity* \mathbf{m} of a singular point $\mathbf{p} \in \mathcal{C}$ is defined to be the intersection multiplicity at \mathbf{p} between \mathcal{C} and a generic line through \mathbf{p} . The following examples show that neither of these two invariants at the point $x = y = 0$ determines the other.

| curve | \mathbf{m} | $\boldsymbol{\mu}$ |
|-------------|--------------|--------------------|
| $x^3 = y^5$ | 3 | 8 |
| $x^3 = y^7$ | 3 | 12 |
| $x^4 = y^5$ | 4 | 12 |

Note that the multiplicity \mathbf{m} for a singular point of a curve of degree n satisfies

$$2 \leq \mathbf{m} \leq n.$$

The set of all curves in \mathfrak{C}_n which have a singularity of multiplicity \mathbf{m} or larger forms a closed algebraic subset of codimension $\binom{\mathbf{m}+1}{2} - 2$ in \mathfrak{C}_n . The proof is similar to the proof of Proposition 10.1 in §10.

Lemma 7.12 (Multi-Branch Lemma). *The augmented genus of a singularity with k local branches $\mathcal{B}_1 \dots, \mathcal{B}_k$ can be computed as follows*

$$\mathfrak{g}_{\mathcal{C}}^+(\mathbf{p}) = \sum_j \mathfrak{g}_{\mathcal{B}_j}(\mathbf{p}) + \sum_{i < j} \mathcal{B}_i \cdot \mathcal{B}_j ,$$

where $\mathcal{B}_i \cdot \mathcal{B}_j$ is the intersection number between the two branches.

(Compare the analogous formula for flex-multiplicity in Example 6.6.) As an example, if there are k smooth branches intersecting pairwise transversally at a common point \mathbf{p} , then $\mathfrak{g}_{\mathcal{C}}^+(\mathbf{p}) = \binom{k}{2}$, and it follows easily that $\mathfrak{g}_{\mathcal{C}}(\mathbf{p}) = \binom{k-1}{2}$.

Outline Proof. ³⁵ First choose a fixed small round neighborhood N of \mathbf{p} , and choose generic small translations $\mathcal{B}_j + \mathbf{v}_j$ of the various branches so that each one still intersects ∂N transversally, and so that any two translated branches intersect transversally in $\mathcal{B}_i \cdot \mathcal{B}_j \geq 1$ distinct points. Then approximate each translated branch very closely by a smooth curve. Thus we are reduced to the case of smooth curves intersecting transversally. The disjoint union of the resulting smooth curves will have k components, each with one boundary curve, and will have genus $\sum \mathfrak{g}_{\mathcal{B}_i}(\mathbf{p})$. A smooth curve which is close to the actual union of these transversally intersecting curves will be homeomorphic to the object obtained by removing a small round neighborhood of each transverse intersection point, and then gluing the $2 \sum \mathcal{B}_i \cdot \mathcal{B}_j$ resulting boundary circles together in pairs. By Theorem 7.2(6), each such pasting must either increase the genus by one or decrease the number of components by one. Since the total effect is to decrease the number of components from k to one, the final genus must be

$$\mathfrak{g}_{\mathcal{C}}(\mathbf{p}) = 1 - k + \sum_i \mathfrak{g}_{\mathcal{B}_i}(\mathbf{p}) + \sum_{i < j} \mathcal{B}_i \cdot \mathcal{B}_j .$$

Adding $k - 1$ to both sides, the conclusion follows. \square

Genus and Proper Action. Using the genus invariant for singular points, we can provide criteria for proper action of $\mathbf{G} = \mathrm{PGL}_3$ on subsets of \mathfrak{C}_n which are different from those in Section 6, and which involve only the complexity of the singularities. First, as in Section 6, we consider only line-free curves. Let $\mathfrak{U} \subset \mathfrak{C}_n$ be some \mathbf{G} -invariant open set of curves.

Theorem 7.13. *Suppose that the following three conditions are satisfied:*

- (1) *No curve in \mathfrak{U} contains a line.*
- (2) *No curve in \mathfrak{U} is separated by a single point.*
- (3) *The following inequality is satisfied*

$$\max_{\mathbf{p} \in \mathfrak{C} \in \mathfrak{U}} \mathfrak{g}_{\mathcal{C}}(\mathbf{p}) + \max_{\mathbf{p} \in \mathfrak{C} \in \mathfrak{U}} \mathfrak{g}_{\mathcal{C}}^+(\mathbf{p}) < \binom{n-1}{2} .$$

Then the action of \mathbf{G} on \mathfrak{U} is proper, and hence the open set $\mathfrak{U}/\mathbf{G} \subset \mathbb{M}_n$ is a Hausdorff orbifold.

Here Condition (2) is very weak, and fails to be satisfied only in the exceptional case that \mathcal{C} is the union of two sub-curves which intersect in only one point. (Using Lemma 7.12, one can check that this cannot happen for $n < 9$ when Conditions (1) and (3) are satisfied.)

³⁵For a detailed proof of an equivalent statement, see [Wa, Th. 6.5.1].

For degree $n = 4$, Condition **(3)** allows only simple nodes and cubic cusp singularities. However, for $n = 5$ the first six singularities of Figure 24 are allowed, and for larger degrees much wilder singularities are allowed.

The proof of Theorem 7.13 will require two lemmas.

Lemma 7.14. *Given any curve \mathcal{C} in \mathbb{P}^2 there exists an $\varepsilon_{\mathcal{C}} > 0$ with the following property. Any open ball of radius $\varepsilon_{\mathcal{C}}$ in \mathbb{P}^2 is contained in some open ball B which either*

- (a) *is disjoint from \mathcal{C} or intersects \mathcal{C} in a smooth topological disk, or*
- (b) *is centered at a singular point $\mathbf{p} \in \mathcal{C}$ and is small enough so that*

$$\mathbf{g}_{\mathcal{C}}^+(\mathbf{p}) = \mathbf{g}(\mathcal{C} \cap B) + b - 1$$

where b is the number of boundary components of $\mathcal{C} \cap B$.

Proof. Clearly every point of \mathbb{P}^2 is contained in at least one open ball which satisfies either (a) or (b). Since \mathbb{P}^2 is compact, it is contained in a union $B_1 \cup \cdots \cup B_m$ of finitely many balls, each of which satisfies (a) or (b). Now consider the function

$$\mathbf{p} \mapsto \max_j \text{dist}(\mathbf{p}, \mathbb{P}^2 \setminus B_j)$$

from \mathbb{P}^2 to the real numbers. This is continuous, and strictly positive everywhere. Hence it has a minimum value, which is the required number $\varepsilon_{\mathcal{C}} > 0$. \square

Lemma 7.15. *If the curve \mathcal{C} is line-free, then for every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: For any line $L \subset \mathbb{P}^2$, each connected component of the intersection $\mathcal{C} \cap N_{\delta}(L)$ has diameter less than ε .*

In particular, if we take $\varepsilon = \varepsilon_{\mathcal{C}}$ then the corresponding δ will be called $\delta_{\mathcal{C}}$.

Proof of Lemma 7.15. Otherwise, for some fixed $\varepsilon_0 > 0$, we could choose a sequence $\{\delta_j\}$ converging to zero, and an associated sequence of lines L_j , such that for each j some component of $\mathcal{C}_0 \cap N_{\delta_j}(L_j)$ has diameter $\geq \varepsilon_0$. After passing to an infinite subsequence, we may assume that $\{L_j\}$ converges to a limit line L' . It then follows that the intersection of any neighborhood of L' with \mathcal{C}_0 has one or more components of diameter $\geq \varepsilon_0$. Since any nested intersection of compact connected sets is again connected, it would follow that $L' \cap \mathcal{C}_0$ has a component of length $\geq \varepsilon_0$, which is impossible since \mathcal{C}_0 is line-free. \square

Proof of Theorem 7.13. If the action of \mathbf{G} on the open set $\mathfrak{U} \subset \mathfrak{C}_n$ were not proper then we could find:

- (1) a sequence of curves $\mathcal{C}_1(k)$ converging to some $\mathcal{C}_1 \in \mathfrak{U}$, and
- (2) a sequence of curves $\mathcal{C}_2(k)$ converging to some $\mathcal{C}_2 \in \mathfrak{U}$, and
- (3) group elements \mathbf{g}_k diverging to infinity in \mathbf{G} with $\mathbf{g}_k(\mathcal{C}_1(k)) = \mathcal{C}_2(k)$.

Choose ε smaller than the numbers $\varepsilon_{\mathcal{C}_1}$ and $\varepsilon_{\mathcal{C}_2}$ of Lemma 7.14, and also smaller than the numbers $\delta_{\mathcal{C}_1}$ and $\delta_{\mathcal{C}_2}$ of Lemma 7.15. We will make use of the Distortion Lemma 6.12. To fix our ideas, suppose that we are in the case where, for infinitely many k , there exists a point \mathbf{p}^+ and a line L^- , so that \mathbf{g}_k maps everything outside of $N_{\varepsilon}(\mathbf{p}^+)$ into $N_{\varepsilon}(L^-)$. (In some cases, it will be necessary to interchange the roles of \mathcal{C}_1 and \mathcal{C}_2 , and of \mathbf{g}_k and \mathbf{g}_k^{-1} , in order to reach this situation.)

Since our curves are not disconnected by any single point, it follows that $N_{\varepsilon}(\mathbf{p}^+)$ will be contained in an open ball B of type (a) or (b) such that the complement $\mathcal{C}_1 \setminus B$ is connected. Therefore this complement must map into a single component of $\mathcal{C}_2 \cap N_{\varepsilon}(L^-)$,

which necessarily has diameter less than $\varepsilon_{\mathcal{C}_2}$. Now approximate \mathcal{C}_1 closely by a smooth curve \mathcal{C}'_1 . We can compute the genus of $\mathcal{C}'_1 \setminus B$ as follows. Assuming that B is of type **(b)** centered at \mathbf{p} , so that $\mathfrak{g}_{\mathcal{C}'_1}^+(\mathbf{p}) = \mathfrak{g}(\mathcal{C}'_1 \cap B) + b - 1$, it follows from Corollary 7.4 that

$$\mathfrak{g}_{\mathcal{C}'_1}^+(\mathbf{p}) + \mathfrak{g}(\mathcal{C}'_1 \setminus B) = \binom{n-1}{2}.$$

In fact, if we cut \mathcal{C}'_1 along the boundary of B , then the part inside B will have genus $\mathfrak{g}_{\mathcal{C}'_1}(\mathbf{p})$, with $b \geq 1$ boundary circles. The union \mathcal{C}'_1 has genus $\binom{n-1}{2}$. According to Corollary 7.4 this must be equal to the sum $b + 1 - \ell + \mathfrak{g}_{\mathcal{C}'_1}(\mathbf{p}) + \mathfrak{g}(\mathcal{C}'_1 \setminus B)$, where $\ell = 2$, and where the first four summands add up to $\mathfrak{g}_{\mathcal{C}'_1}^+(\mathbf{p})$.

(On the other hand if B is of Type **(a)**, then we can simply replace $\mathfrak{g}_{\mathcal{C}'_1}^+(\mathbf{p})$ by zero.) Since $\mathcal{C}'_1 \setminus B$ must map bijectively into a neighborhood of the form $N_{\varepsilon_{\mathcal{C}_2}}(\mathbf{p}')$, it follows that $\mathfrak{g}_{\mathcal{C}'_1}^+(\mathbf{p}) + \mathfrak{g}_{\mathcal{C}_2}(\mathbf{p}') \geq \binom{n-1}{2}$; contradicting our hypothesis and completing the proof of Theorem 7.13. \square

Curves which contain lines. If we want to allow curves which contain lines, then we have to work a little harder. Let \mathcal{C} be a curve in \mathbb{P}^2 and let L be a line in \mathbb{P}^2 which may be contained in \mathcal{C} or may have finite intersection with \mathcal{C} . Then we define an invariant $\mathfrak{g}_{\max}(\mathcal{C}, L) \geq 0$ as follows.

Definition 7.16. If $\mathcal{C} \cap L$ is finite, let $\mathfrak{g}_{\max}(\mathcal{C}, L) \geq 0$ be the *maximum* of $\mathfrak{g}_{\mathcal{C}}(\mathbf{p})$ over all singular points of \mathcal{C} which are contained in L . On the other hand, if $L \subset \mathcal{C}$ define $\mathfrak{g}_{\max}(\mathcal{C}, L)$ to be the *sum* of $\mathfrak{g}_{\mathcal{C}}(\mathbf{p})$ over all singular points of \mathcal{C} in L .

Theorem 7.17. Let $\mathcal{U}' \subset \mathfrak{C}_n$ be some \mathbf{G} -invariant open set consisting of curves which are not separated by any point (as in Theorem 7.13); but which may contain lines. If the following condition is satisfied

$$(3') \quad \max_{\mathcal{C} \in \mathcal{U}'} \max_L \mathfrak{g}_{\max}(\mathcal{C}, L) + \max_{\mathbf{p} \in \mathcal{C} \in \mathcal{U}'} \mathfrak{g}^+(\mathbf{p}) < \binom{n-1}{2},$$

then the action of $\mathbf{G} = \mathrm{PGL}_3$ on \mathcal{U}' is proper, so that \mathcal{U}'/\mathbf{G} is a Hausdorff orbifold.

(If the curves in \mathcal{U}' do not contain any lines, then it is easy to check that this statement is completely equivalent to Theorem 7.13.)

Note that the property of being separated by a point is much easier to satisfy for curves which contain lines. For example, if $p(x)$ is a polynomial of degree $n-1$, then the curve which is described in affine coordinates by $y^2 = p(x)$ has only one point $\mathbf{p} = (0 : 1 : 0)$ on the line at infinity. Thus the union of this curve with the line at infinity is an example of a curve of degree n which is separated by a single point \mathbf{p} . A brief computation using Lemma 7.12 shows that $\mathfrak{g}^+(\mathbf{p}) = n-1$, and hence $\mathfrak{g}(\mathbf{p}) = n-2$. For $n \geq 6$, this curve satisfies Condition (3'); and yet the Theorem does not apply.

Here is a sample consequence.

Corollary 7.18. Let $\mathcal{U}' \subset \mathfrak{C}_n$ be the open set consisting of curves with no singularities other than simple double points and cubic cusps (or equivalently with $\mathfrak{g}^+(\mathbf{p}) \leq 1$ for all singular points). If $n \geq 4$, then the action of \mathbf{G} on \mathcal{U}' is proper, and hence the open set $\mathcal{U}'/\mathbf{G} \subset \mathbb{M}_n$ is a Hausdorff orbifold.

Note that the condition $n \geq 4$ is essential. For a cubic curve with a cusp point, the quotient is not even a T_1 -space; while for a cubic curve with a double point (Figure 18), the action is not proper (although the quotient is Hausdorff). This Corollary applies to a

union of four lines in general position; but not to a union of four lines where three pass through a common point. (Compare Figure 28.)

Proof of Corollary 7.18 (assuming Theorem 7.17). Since only simple double points are allowed, any singularity of \mathcal{C} within a line $L \subset \mathcal{C}$ must have genus zero. The conclusion then follows easily. (More generally, we can allow (p, q) -cusps with p and q relatively prime provided that $(p-1)(q-1) < \binom{n-1}{2}$.) \square

Proof of Theorem 7.17. The argument proceeds much as in the proof of Theorem 7.13; but now we must allow for the extra possibility that the line L^- is contained in the curve \mathcal{C}_2 . In that case, a small neighborhood $N_\varepsilon(L^-)$ will contain no singular points except those in L^- . Choose a small ball around each of these, and then approximate \mathcal{C}_2 very closely by a smooth curve \mathcal{C}'_2 . Then we see from Corollary 7.4 that the genus of $\mathcal{C}'_2 \cap N_\varepsilon(L^-)$ is precisely equal to the sum $\mathfrak{g}_{\mathcal{C}_2}(\mathbf{p}_1) + \cdots + \mathfrak{g}_{\mathcal{C}_2}(\mathbf{p}_k)$ of the genera of these singularities. In fact this intersection can be obtained from a surface of genus zero with k holes by pasting a surface of genus $\mathfrak{g}_{\mathcal{C}_2}(\mathbf{p}_j)$ into the j -th hole for each j between 1 and k .

We now need a modified version of Lemma 7.15. If a line L' is contained in one of the neighborhoods $N_\varepsilon(L)$ for a line $L \subset \mathcal{C}_2$, then we may use $N_\varepsilon(L)$ in the proof of the theorem. But if we exclude all such L' then the proof of Lemma 7.15 proceeds as before. Further details will be left to the reader. \square

Following is a special case where we can push this argument a little further.

Remark 7.19 (Extension to 1-cycles). If we consider 1-cycles rather than curves, with multiplicities allowed, then the arguments become more difficult since every point of a curve of multiplicity two or more is singular. As a consequence, in the definition of the genus associated with a point of \mathcal{C} or a line through \mathcal{C} we must take the lim-sup over all possible smooth approximating curves.

As a simplest example, consider a 1-cycle of the form $\mathcal{C} = \mathcal{C}_{n-2} + 2 \cdot L$ where \mathcal{C}_{n-2} is a generic smooth curve of degree $n-2$ and L is a generic line counted twice. Then one can check that the largest value of \mathfrak{g}^+ at a point is $\mathfrak{g}^+(\mathbf{p}) = 3$, corresponding to an intersection point in $\mathcal{C}_{n-2} \cap L$. Using Corollary 7.4, one can check that the largest value of \mathfrak{g} on a line (when $n \geq 4$) is $\mathfrak{g}(L) = n-2$ for the doubled line L . The required inequality

$$\max \mathfrak{g}^+(\text{point}) + \max \mathfrak{g}(\text{line}) < \binom{n-1}{2}$$

then reduces to $3 + (n-2) < (n-1)(n-2)/2$; which is satisfied if and only if $n \geq 6$. Thus we can conclude that \mathbb{M}_n is locally Hausdorff at $((\mathcal{C}_{n-2} + 2 \cdot L))$ whenever $n \geq 6$. Details of the argument will be omitted.

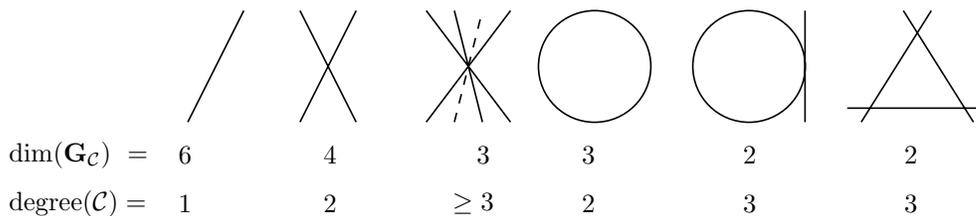


FIG. 25. The six highly symmetric curves.

8. CURVES WITH INFINITE STABILIZER

In this section we will always work over the complex numbers. Following Klein and Lie, a curve³⁶ $\mathcal{C} \in \mathfrak{C}_n$ is called a *W-curve* if it is invariant under a one-parameter group of projective transformations. Recall from Remark 4.5 that a curve or cycle \mathcal{C} has infinite stabilizer if and only if this stabilizer is a Lie group, necessarily containing a one-parameter subgroup. (It will be enough to study curves, since clearly a 1-cycle \mathcal{C} has infinite stabilizer if and only if its support $|\mathcal{C}|$ also has infinite stabilizer.) We use the notation \mathfrak{W}_n for the closed subset of $\widehat{\mathfrak{C}}_n$ consisting of all curves or cycles with infinite stabilizer.

A detailed classification of curves with infinite stabilizer has been provided by [AF1]. (See also [Ghi], [Pop].) We will provide a different version of the classification. Since many different W-curves may be invariant under the same group, it is convenient to first list the possible connected subgroups of $\text{PGL}_3(\mathbb{C})$ which can serve as the identity component $\mathbf{G}_{\mathcal{C}}^0$ of some stabilizer.

The most symmetric W-curves are those with a stabilizer of dimension two or more. These are relatively easy to describe:

Theorem 8.1. *There are only six connected Lie groups of dimension two or more which can occur as the component of the identity $\mathbf{G}_{\mathcal{C}}^0$ for some curve in \mathbb{P}^2 . The corresponding curves can be listed as follows. (Compare Figure 25.)*

One Line. *If \mathcal{C} is a line, the stabilizer $\mathbf{G}_{\mathcal{C}}$ has dimension six.³⁷ Putting this line at infinity, $\mathbf{G}_{\mathcal{C}} = \mathbf{G}_{\mathcal{C}}^0$ can be identified with the group consisting of all non-singular affine transformations*

$$(x, y) \mapsto (\alpha x + \beta y + \sigma, \gamma x + \delta y + \tau) \quad \text{with} \quad \alpha\delta - \beta\gamma \neq 0. \quad (27)$$

Two Lines. *If \mathcal{C} is the union of two distinct lines, the group $\mathbf{G}_{\mathcal{C}}$ is four-dimensional, and $\mathbf{G}_{\mathcal{C}}^0$ can be identified with the solvable subgroup of (27) consisting of transformations $(x, y) \mapsto (\alpha x + \beta y + \sigma, \delta y)$ (preserving the line $y = 0$, as well as the line at infinity).*

Concurrent Lines. *If \mathcal{C} is the union of three or more lines passing through a common point, the group $\mathbf{G}_{\mathcal{C}}$ is three-dimensional, and $\mathbf{G}_{\mathcal{C}}^0$ can be identified with the subgroup of (27) consisting of affine transformations*

$$(x, y) \mapsto (\alpha x + \beta y + \sigma, y)$$

³⁶Klein and Lie [KL] also considered transcendental curves (such as the logarithmic spiral) which are invariant under a one-parameter group, but we consider only algebraic curves.

³⁷This is the unique example for which the action of the stabilizer $\mathbf{G}_{\mathcal{C}}$ on $|\mathcal{C}|$ is not effective, and hence is larger than the group of automorphisms. The group of automorphisms of the line is the 3-dimensional group PGL_2 , which is a quotient group of the six dimensional stabilizer $\mathbf{G}_{\mathcal{C}}$.

which preserve every line $y = \text{constant}$. (This is the only case which includes curves of every degree $n \geq 3$. For $n \geq 4$ note that it includes infinitely many \mathbf{G} -equivalence classes, since any four lines through a point have a \mathbf{G} -invariant cross-ratio.)

Degree Two. If \mathcal{C} is a smooth degree two curve, the group $\mathbf{G}_{\mathcal{C}}$ is a three-dimensional simple group which is isomorphic to PGL_2 , but with a non-standard embedding of PGL_2 into PGL_3 .

Degree Two plus Tangent Line. For the union of a smooth degree two curve with a tangent line, the stabilizer has dimension two, isomorphic to the group of affine automorphisms $z \mapsto \alpha z + \beta$ of \mathbb{C} .

Three Non-concurrent Lines. For a triple of lines in general position, the group $\mathbf{G}_{\mathcal{C}}$ has dimension two, and $\mathbf{G}_{\mathcal{C}}^0$ can be identified with the abelian group consisting of non-singular diagonal transformations

$$(x, y) \mapsto (\alpha x, \beta y).$$

The proof of these statements will depend on the following catalog of all possible one-dimensional stabilizers.

Theorem 8.2. A curve $\mathcal{C} \subset \mathbb{P}^2(\mathbb{C})$ has infinite stabilizer if and only if, after a projective change of coordinates, it is invariant under one (or more than one if it is listed in Theorem 8.1) of the following kinds of one-parameter subgroups of $\text{PGL}_3(\mathbb{C})$. We first list the diagonalizable subgroups, and then one non-diagonalizable subgroup.

(1) **Type $D(p, q, r)$: Diagonalizable**, with typical equation

$$x^p = y^q z^r. \tag{28}$$

Here the integers $p \geq q \geq r \geq 0$ should be pairwise relatively prime with $p = q + r \geq 1$. (Compare Figure 26.) The automorphism takes the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} t^q & 0 & 0 \\ 0 & t^p & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{29}$$

where t varies over the multiplicative group of all non-zero complex numbers. In this case, the invariant curve \mathcal{C} can be any union of finitely many irreducible curves of the form $x = 0$, or $y = 0$, or $z = 0$, or

$$x^p = a y^q z^r \quad \text{with } a \neq 0. \tag{30}$$

(2) **Non-Diagonalizable**,³⁸ of type **ND**, with automorphism

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{31}$$

where t varies over the additive group of all complex numbers. In this case \mathcal{C} can be any union of curves of the form $z = 0$ or

$$x z = y^2/2 + a z^2, \quad \text{with } a \text{ constant}. \tag{32}$$

³⁸There is also another non-diagonalizable group $(x : y : z) \mapsto (x + ty : y : z)$ which we will ignore, since it occurs only as a subgroup of the 3-dimensional group $\mathbf{G}_{\mathcal{C}}$ where \mathcal{C} is a union of concurrent lines. (Compare Figure 25.) Every such \mathcal{C} has already been included under type $D(1, 1, 0)$.

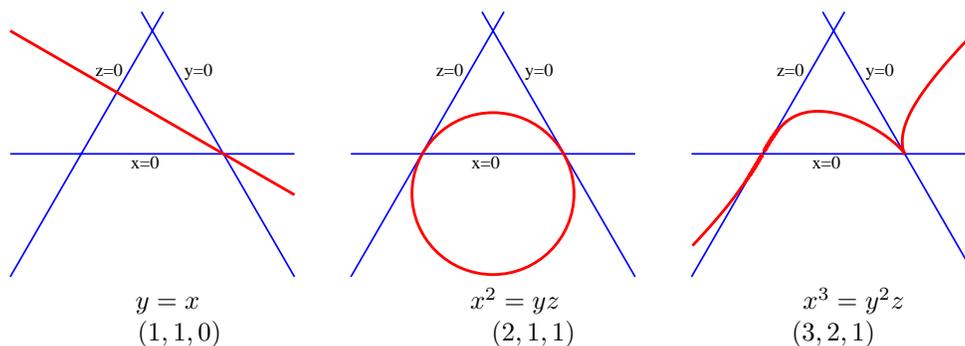


FIG. 26. Graphs of curves of the form (28) of types $(1, 1, 0)$, $(2, 1, 1)$ and $(3, 2, 1)$, shown in the real affine plane with normalization

$$x + y + z = 1 .$$

For arbitrary $a \neq 0$ the curves (30) of type $(1, 1, 0)$ are straight lines through the point $(0, 0, 1)$; curves of type $(2, 2, 1)$ are degree two curves tangent to the coordinate lines at the points $(0, 1, 0)$ and $(0, 0, 1)$; while curves of type $(3, 2, 1)$ have a flex at $(0, 1, 0)$ and a 3-2 cusp at $(0, 0, 1)$.

(In the affine plane with $z = 1$, these are just “parallel” parabolas $x = y^2/2 + a$.)

Remark 8.3 (The catalog of curves in \mathfrak{W}_n). Before proving Theorem 8.2, we will describe these curves in more detail. At the same time, we will compute the dimension of the space \mathfrak{W}_n .

- **Type $D(1, 1, 0)$.** The curves of type $D(1, 1, 0)$ are the easiest to describe. To be invariant under the action $(x, y, z) \mapsto (tx, ty, z)$ a curve must be a union of lines $(x : y) = \text{constant}$ through the point $(0 : 0 : 1)$, possibly together with the “line at infinity” $z = 0$. In other words, a curve \mathcal{C} of degree n has type $D(1, 1, 0)$ if and only if it is a union of n lines, at least $n - 1$ of which pass through a common point. To compute the dimension of the corresponding subset of \mathfrak{W}_n , note that we need two parameters in order to specify the intersection point, two parameters to specify the free line, and then one-parameter for each additional line. Hence the dimension of the corresponding irreducible subset of \mathfrak{W}_n is $n + 3$ (provided that $n \geq 3$). If $n \geq 4$ then this component contains infinitely many different projective equivalence classes. In fact, for $n > 4$ there are $n - 4$ independent cross-ratio invariants. The algebraic subset consisting of lines through a common point has $n - 3$ cross-ratio invariants.

- **Type $D(2, 1, 1)$.** By definition each irreducible non-linear curve of type $D(2, 1, 1)$ can be put in the form $x^2 = ayz$ with $a \neq 0$. Any two curves in this form intersect in the two points $(0 : 0 : 1)$ and $(0 : 1 : 0)$. For example, in the region $z \neq 0$ we can use affine coordinates with $z = 1$. The curves are then parabolas $x^2 = ay$ which are tangent to each other at the origin. Thus any automorphism which maps each curve to itself and fixes the origin must also map the tangent line $y = 0$ to itself. Similarly the tangent line $z = 0$ at the point $(0 : 1 : 0)$ must map to itself, and the line $x = 0$ joining the two intersection points must map to itself. A union of k such curves, with $k \geq 2$, can be determined by $k + 6$ independent parameters: namely 6 parameters to determine the three coordinate lines, and one more for each curve. Thus the corresponding irreducible variety in \mathfrak{W}_{2k} has dimension

$6 + k$. Note that we can obtain varieties of higher degree, but the same dimension, by adjoining one or more of the three coordinate lines to the curve.

Whenever $p \geq 2$, the orbits under this one-parameter group of automorphisms form a smooth foliation of the thrice punctured plane. Here the three coordinate points $(1 : 0 : 0)$, $(0 : 1 : 0)$, and $(0 : 0 : 1)$ are singularities of the foliation, since they are fixed points of this automorphism group. (For type $(1, 1, 0)$ with $p = 1$, there is also an entire line $z = 0$ of fixed points.)

Caution. The real pictures which we use as illustrations can never completely describe a complex curve. As one example, any finite union of concentric circles

$$u^2 + v^2 = \rho^2 w^2$$

is clearly a W-curve; yet superficially it doesn't look like one of our examples. But in fact it can easily be put in the $D(2, 1, 1)$ form $x^2 = ayz$ by setting

$$x = w, \quad y = u + iv, \quad z = u - iv, \quad \text{and} \quad a = 1/\rho^2.$$

- **Type $D(p, q, r)$ with $p > 2$.** For a curve of the form

$$x^p = ay^qz^r, \quad \text{with } p > 2 \text{ hence } q \geq 2,$$

the point $(x : y : z) = (0 : 0 : 1)$ is a cusp-point of the form $x^p = ay^q$, using affine coordinates with $z = 1$. On the other hand, using affine coordinates with $y = 1$, the point $(0 : 1 : 0)$ is either a cusp point of the form $x^p = az^r$ if $r > 1$, or a flex-point of the form $x^p = az$ if $r = 1$. In either case these two points are distinguished. Hence, as in Case $D(2, 1, 1)$ it follows that one, two, or all three of the coordinate lines $x = 0$, $y = 0$ and $z = 0$ can be adjoined to the curve without increasing the dimension of the associated irreducible components. (Compare the last three curves on the top line of Figure 28.) As in Case $D(2, 1, 1)$, this dimension is $k + 6$ where k is the number of non-linear components; but now we need only require that $k \geq 1$.

- **Non-Diagonalizable Type.** Using affine coordinates with $z = 1$, the transformation (31) takes the form

$$(x, y) \mapsto (x + ty + t^2/2, y + t).$$

It is not hard to check that each parabola of the form $2x = y^2 + k$ maps onto itself under this transformation. Evidently these parabolas form a smooth foliation of the affine plane. However, they must intersect somewhere on the line at infinity. Let us write this equation in projective coordinates as $2xz = y^2 + kz^2$, and then rename the coordinates as $x = Z$, $y = X$ and $z = Y$, so that the equation now becomes

$$2ZY = X^2 + kY^2.$$

Evidently any two of these curves intersect at the point $(X : Y : Z) = (0 : 0 : 1)$. If we again pass to affine coordinates, but now with $Z = 1$, this yields the family of quadratic curves

$$2Y = X^2 + kY^2 \tag{33}$$

in the (X, Y) -plane, as illustrated in Figure 27. These curves are all tangent to the X -axis at the origin. The limit of these curves as $k \rightarrow \infty$ is the 1-cycle defined by the equation $Y^2 = 0$, or in other words the X -axis counted twice.

More generally, it is not hard to check that a union of $k \geq 2$ smooth curves of degree two can be put simultaneously into the non-diagonalizable normal form (32) if and only if these curves all are mutually tangent at a common point of intersection and have no

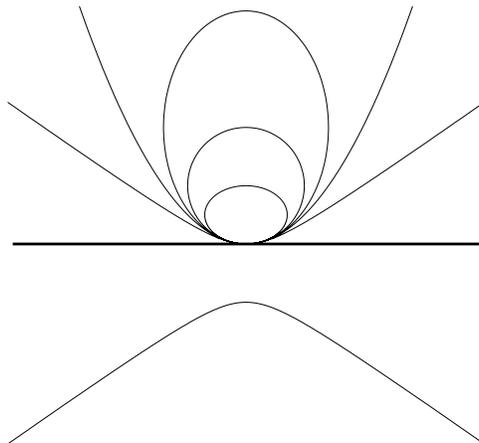


FIG. 27. Six curves from the non-diagonalizable family (32), shown in the real affine plane after a projective transformation which moves the line at infinity $z = 0$ to the x -axis. (See (33) below.) Here we obtain ellipses for $k > 0$ (shown for $k = 1/2, 1, 2$), tending to the origin as $k \rightarrow +\infty$, and tending to a parabola, which is shown, as $k \rightarrow 0$. For $k < 0$ we obtain a hyperbola (shown for $k = -2$), which converges to a single line counted twice as $k \rightarrow -\infty$. Any two of these curves intersect only at the origin, with intersection multiplicity four.

other intersection. (Thus the pairwise intersection multiplicity at this point must be four.) Equivalently, these curves must belong to the pencil consisting of all sums

$$\{\alpha\Phi_1 + \beta\Phi_2\}$$

where $\Phi_1 = 0$ defines a smooth degree two curve and $\Phi_2 = 0$ is one of its tangent lines, counted with multiplicity two. The corresponding irreducible component of \mathfrak{W}_{2k} has dimension $5 + k$, assuming that $k \geq 2$. (It takes six parameters to specify a quadratic curve plus distinguished point, and one more for each additional curve.) We can also adjoin the common tangent line without increasing the dimension of the locus in the appropriate space \mathfrak{C}_n .

Examples. For $n = 3$, it is not hard to check that the algebraic set $\mathfrak{W}_3 \subset \mathbb{C}_3$ is the union of two maximal irreducible subvarieties, both of dimension seven. One consists of the \mathbf{G} -equivalence class of the cusp curve $x^3 = y^2z$, together with the classes of $x^3 = 0$ and $y^2z = 0$. A generic curve in the other is a smooth quadratic curve together with a line which intersects it transversally. There are also five non-maximal subvarieties having the following as generic elements: (1) a smooth degree two curve plus tangent line, (2) three lines in general position, (3) three distinct lines through a common point, (4) two lines, one with multiplicity two, and (5) one line with multiplicity three.

For $n = 4$, there are eight different maximal irreducible subvarieties, with generic representatives as illustrated in Figure 28. (Again there are also many non-maximal subvarieties. Compare Remark 8.5.) The dimension of the algebraic set \mathfrak{W}_4 is equal to eight, the maximum of the dimensions of its irreducible components.

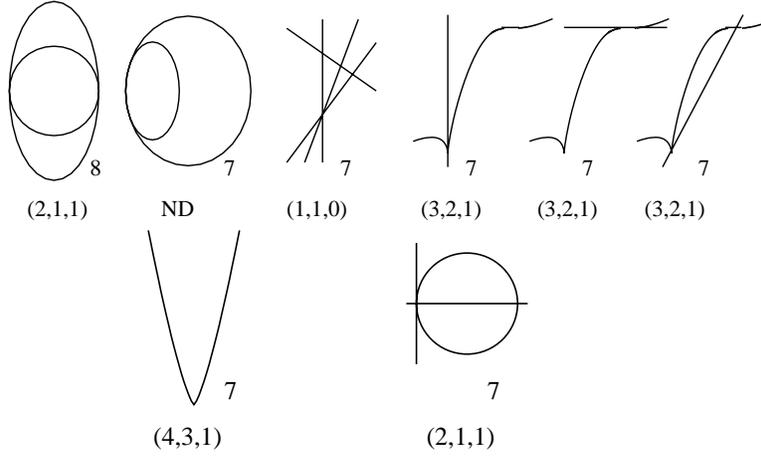


FIG. 28. The algebraic set $\mathfrak{W}_4 \subset \widehat{\mathfrak{C}}_4$ consisting of curves or cycles of degree four with infinite stabilizer is the union of eight maximal irreducible subvarieties. Representative generic curves from each of these subvarieties are shown. In each case, the dimension of the algebraic subset is listed, as well as the automorphism type indicated by the appropriate indices $p \geq q \geq r$, or by ND for non-diagonalizable.

Remark 8.4 (Dimension of \mathfrak{W}_n). From the discussion above, for n equal to three or four, the dimension of the complex algebraic set \mathfrak{W}_n is $n + 4$. On the other hand, for $n > 4$ the dimension of \mathfrak{W}_n is $n + 3$. In fact the maximal irreducible component of type $D(1, 1, 0)$ has dimension $n + 3$ for all $n \geq 3$. It is not hard to check that the dimension of any other kind of maximal irreducible component increases more slowly as a function of n , so that $n + 3$ is the maximal dimension for all $n > 4$.

Remark 8.5. Of course not every curve in a maximal irreducible component is generic. For example, there are many other curves in \mathfrak{W}_4 which are not shown in Figure 28, but which can be obtained as degenerations of the generic curves. Thus, in the $(2, 1, 1)$ example, the outer ellipse can expand and converge to a union of two vertical tangent lines. Similarly the inner ellipse can shrink and converge to a horizontal line counted with multiplicity two.

Remark 8.6. It is clear from these examples that the compact sets $\mathfrak{W}_n \subset \widehat{\mathfrak{C}}_n$ have rather complicated topologies. However, the \mathfrak{W}_n have fairly high codimension in $\widehat{\mathfrak{C}}_n$, so only the high dimensional homology of the complement $\widehat{\mathfrak{C}}_n^{\text{fs}} = \widehat{\mathfrak{C}}_n \setminus \mathfrak{W}_n$ is affected. Suppose for example that $n > 4$, so that the real dimension $\dim_{\mathbb{R}}(\mathfrak{W}_n)$ is $2(n + 3)$. Using standard tools of algebraic topology (see for example [Sp]), there is an exact sequence

$$H^{k-1}(\mathfrak{W}_n) \rightarrow H^k(\widehat{\mathfrak{C}}_n, \mathfrak{W}_n) \rightarrow H^k(\widehat{\mathfrak{C}}_n) \rightarrow H^k(\mathfrak{W}_n)$$

of cohomology groups with integer coefficients. Since $H^k(\mathfrak{W}_n) = 0$ for $k > 2(n + 3)$, it follows that $H^k(\widehat{\mathfrak{C}}_n, \mathfrak{W}_n)$ maps isomorphically onto $H^k(\widehat{\mathfrak{C}}_n)$ when $k \geq 2(n + 4)$. Now

using the duality theorem

$$\begin{array}{ccc} H^k(\widehat{\mathcal{C}}_n, \mathfrak{W}_n) & \xrightarrow{\cong} & H^k(\widehat{\mathcal{C}}_n) \\ \downarrow \cong & & \downarrow \cong \\ H_{d-k}(\widehat{\mathcal{C}}_n, \mathfrak{W}_n) & \longrightarrow & H_{d-k}(\widehat{\mathcal{C}}_n), \end{array}$$

where $d = n(n+3) = \dim_{\mathbb{R}}(\widehat{\mathcal{C}}_n)$, we see that $H_j(\widehat{\mathcal{C}}_n^{\text{fs}})$ maps isomorphically onto

$$H_j(\widehat{\mathcal{C}}_n) \cong \begin{cases} \mathbb{Z} & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd} \end{cases}$$

whenever $j \leq d - 2(n+4) = n^2 + n - 8$. (Here we are assuming that $n > 4$. For n equal to 3 or 4 the corresponding condition would be $j \leq n^2 + n - 10$.)

Proof of Theorem 8.2. We know that every infinite stabilizer must contain one or more one-parameter Lie groups. Every one parameter subgroup of $\text{PGL}_3(\mathbb{C})$ can be parametrized as

$$t \mapsto \exp(tA) = I + tA + (tA)^2/2! + (tA)^3/3! + \dots,$$

where A is a 3×3 matrix. We can simplify this matrix in three different ways:

- We can put A into Jordan normal form by a linear change of coordinates.
- We can add a constant multiple of the identity matrix to A , or in other words multiply $\exp(tA)$ by a non-zero constant, since this will not affect the image in PGL_3 .
- We can multiply the matrix A itself by a non-zero constant; this is just equivalent to multiplying the parameter t by a constant.

We will first show, using these three transformations, that the matrix A can be reduced to one of the following, which we will refer to as Cases 1 through 4.

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In fact if A has three linearly independent eigenvectors, then we are in Case 1. In particular, if the eigenvalues a, b, c are all distinct, then we are always in Case 1. At the opposite extreme, if the eigenvalues are all equal then subtracting a multiple of the identity matrix we may assume that they are all zero. The Jordan normal form will then correspond to either Case 2 or 3. (Evidently A cannot be the zero matrix.) Finally, suppose that just two of the eigenvalues are equal. Then we can assume that the two are zero and the third is one, so that the Jordan normal form is either a diagonal matrix, so that we are in Case 1, or corresponds to the matrix of Case 4. The four corresponding matrices $\exp(tA)$ can now be listed as follows.

$$\begin{pmatrix} e^{ta} & 0 & 0 \\ 0 & e^{tb} & 0 \\ 0 & 0 & e^{tc} \end{pmatrix}, \quad \begin{pmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^t \end{pmatrix}.$$

Case 1. Suppose that a curve is invariant under the transformation

$$(x : y : z) \mapsto (e^{at}x : e^{bt}y : e^{ct}z) .$$

Clearly this maps each of the three coordinate axes to itself. In affine coordinates with $z = 1$, we can write this as

$$(x_0, y_0) \mapsto (x, y) = (e^{a't}x_0, e^{b't}y_0) \quad \text{where } a' = a - c, \quad b' = b - c .$$

In other words, if x_0 and y_0 are non-zero, we can write

$$x/x_0 = e^{a't}, \quad y/y_0 = e^{b't} . \tag{34}$$

Since a, b, c cannot all be equal, we may assume (after permuting the coordinates if necessary) that a' and b' are non-zero.

First suppose that the ratio b'/a' is a rational number, which we can write as a fraction in lowest terms as $\pm\ell/m$ with $\ell, m > 0$. Then $mb' = \pm\ell a'$, so that

$$(y/y_0)^m = e^{mb't} = e^{\pm\ell a't} = (x/x_0)^{\pm\ell} .$$

In other words, we have an equation of the form either

$$y^m = ax^\ell \quad \text{or} \quad y^m x^\ell = a$$

for a suitable constant a . After permuting the coordinates appropriately, this takes the required form (30).

On the other hand, if b'/a' is irrational or imaginary, then the invariant curve cannot be algebraic. Choosing t so that $a't$ is an integral multiple of $2\pi i$ in the equation (34), we see that $x = x_0$ but that y takes a countably infinite number of distinct values, which is impossible for any algebraic curve.

Case 2. Using affine coordinates $(x : y : 1)$, the automorphism will take the form

$$(x_0, y_0) \mapsto (x, y) = (x_0 + ty_0 + t^2/2, y_0 + t) .$$

Eliminating $t = y - y_0$ from this equation, the curve through (x_0, y_0) takes the form $x = y^2/2 + (x_0 - y_0^2/2)$, which agrees with the required normal form (32).

Case 3. In this case the transformation takes the simpler form

$$(x_0 : y_0 : z_0) \mapsto (x_0 + ty_0 : y_0 : z_0) ,$$

so that the invariant curves are just the parallel lines $y = y_0$ and $z = z_0$, or in other words lines which pass through the point $(1 : 0 : 0)$ on the line at infinity. These can easily be put in the Case 1 normal form, of type $D(1, 1, 0)$.

Case 4. Here the transformation takes the more ominous form

$$(x_0, y_0, z_0) \mapsto (x, y, z) = (x_0 + ty_0, y_0, e^t z_0) .$$

Thus $y = y_0$ is constant. If $y_0 = 0$, then the invariant lines are just parallel lines with $x = x_0$, again of type $D(1, 1, 0)$. But if $y_0 \neq 0$, then we can solve for $t = (x - x_0)/y_0$, so that the invariant curves have the form

$$z = z_0 \exp((x - x_0)/y_0) .$$

Since this cannot be the equation of any algebraic curve, this completes the proof of Theorem 8.2. \square

Proof of Theorem 8.1. We must show that every curve with stabilizer of dimension two or more is contained in the list which is illustrated in Figure 25.

Let $\mathcal{C} = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_k$ be a curve with irreducible components \mathcal{C}_j . Then any automorphism of $\mathbf{G}_{\mathcal{C}}$ must permute the \mathcal{C}_j . Thus the intersection $\bigcap_j \mathbf{G}_{\mathcal{C}_j}$, consisting of projective automorphisms which map each \mathcal{C}_j to itself, is a subgroup of finite index (at most $k!$) in $\mathbf{G}_{\mathcal{C}}$. If $\dim \mathbf{G}_{\mathcal{C}} \geq 2$, then it follows that every irreducible component of \mathcal{C} must have $\dim \mathbf{G}_{\mathcal{C}_j} \geq 2$. Therefore every irreducible component must be a curve of degree one or two. In fact, any irreducible curve of degree three or more is either a cusp curve, with $\dim \mathbf{G}_{\mathcal{C}} = 1$, or else has finite stabilizer.

Thus we need only consider unions of lines and smooth quadratic curves. In particular, we can ignore curves of type $D(p, q, r)$ with $p > 2$, since they either contain a cusp curve, or consist of at most three lines. (Note that any union of at most three lines is already included in the list represented by Figure 25.) We can also ignore curves of type ND with two or more components, since it is easy to check that they have a one-dimensional stabilizer. Thus we only need to consider curves of type $D(2, 1, 1)$ or $D(1, 1, 0)$.

It is easy to check that a curve of type $D(1, 1, 0)$ has stabilizer of dimension two or more if and only if it either consists of concurrent lines, or consists of exactly three lines. Since these are already included, it suffices to consider curves which have at least one component of degree two.

Consider first a single smooth curve of degree two. Since such a curve has genus zero, it is conformally isomorphic to \mathbb{P}^1 , and its group of conformal isomorphisms is isomorphic to the group PGL_2 , of dimension three. The subgroup of isomorphisms fixing two points is one dimensional, and is a group of type $D(2, 1, 1)$. (Compare Figure 26.) In particular, it extends to a projective automorphism of \mathbb{P}^2 . Since the group of all conformal automorphisms is clearly generated by automorphisms with two fixed points, it follows that every automorphism of a degree two curve extends to an automorphism of \mathbb{P}^2 . This yields the required non-standard embedding of PGL_2 into PGL_3 . The discussion of automorphisms of a genus zero curve with one fixed point is similar. Since any line must intersect the degree two curve in one or two points, the other possible cases are easily excluded. \square

9. AUTOMORPHISMS OF SMOOTH CURVES AND OF RIEMANN SURFACES.

This section will discuss the projective automorphism groups of smooth curves in $\mathbb{P}^2(\mathbb{C})$, first answering the following question:

For which degrees n and which primes p does there exist a smooth plane curve of degree n which admits a projective automorphism of period p ?

It will also show, following [Ch], that the stabilizer (= projective automorphism group) of a generic curve of degree 4 is trivial (Theorem 9.3); and that every finite subgroup of $\mathrm{PGL}_3(\mathbb{C})$ can occur as the stabilizer for some smooth curve (Theorem 9.4). The second half of the section will discuss the corresponding questions for conformal automorphisms of arbitrary compact Riemann surfaces.

Theorem 9.1. *Let n be a positive integer, and let p be prime. There exists a smooth curve of degree n in $\mathbb{P}^2(\mathbb{C})$ with a projective automorphism of period p if and only if either*

- (1) n is congruent to either 0, 1, or 2 modulo p , or

(2) the integer

$$2g + 1 = 2 \binom{n-1}{2} + 1 = n^2 - 3n + 3$$

is congruent to zero modulo p .

In Case (2), this congruence is satisfied for some n if and only if p is either equal to 3 or else has the form $p = 6m + 1$ with $m \in \mathbb{Z}$.

(Compare Remark 9.5; and see [Harui] which provides a detailed classification of automorphism groups of smooth plane curves.)

For n equal to 1 or 2, all primes can occur. Table 1 lists the possible primes for other small n , as well as the associated genus. Note that the largest prime is $2g + 1$ whenever $2g + 1$ is a prime, and is strictly less than $2g + 1$ otherwise. (Compare Theorem 9.9.)

| | | | | | | | | |
|--------|-------------|----------------|--------------------|---------------------------|-----------------------|--------------------|----------------------------|--------------------|
| n | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| g | 1 | 3 | 6 | 10 | 15 | 21 | 28 | 36 |
| primes | 2, 3 | 2, 3, 7 | 2, 3, 5, 13 | 2, 3 , 5, 7 | 2, 3, 5, 7, 31 | 2, 3, 7, 43 | 2, 3 , 7, 19 | 2, 3, 5, 73 |

TABLE 1. Degree, genus, and possible prime periods for smooth plane curves.

Here the primes coming from Case (2) have been shown in boldface. Evidently these can be much larger than those from Case (1). (Automorphisms of period $p = 3$ can be obtained from Case (1) in all cases, but also from Case (2) when n is divisible by 3.) As one example, for $n = 4$ and $p = 7$ the corresponding curve is known as the *Klein quartic*. The full automorphism group is a simple group of order $168 = 2^3 \cdot 3 \cdot 7$.

Table 2 shows the prime factorization of $2g + 1$ for small n (or a dash if $2g + 1 = 1$). These prime factors are exactly the primes which can occur in Case (2).

| | | | | | | | | | | | | | | |
|----------|---|---|---|---|----|-----|----|----|------|----|------|------|------|-----|
| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| $2g + 1$ | - | - | 3 | 7 | 13 | 3·7 | 31 | 43 | 3·19 | 73 | 7·13 | 3·37 | 7·19 | 157 |

TABLE 2. Showing $2g + 1$ as a product of primes.

Proof of Theorem 9.1. To show that these conditions are sufficient, it suffices to provide a list of specific examples, as follows. (Of course we do not claim that these are the only such examples.)

| | | |
|----------------------------|---|--|
| $n \equiv 0 \pmod{p}$ | $\Phi = x^n + y^n + z^n$ | $(x : y : z) \mapsto (x : y : \xi z)$ |
| $n \equiv 1 \pmod{p}$ | $\Phi = x^n + y^n + z^{n-1}y$ | $(x : y : z) \mapsto (x : y : \xi z)$ |
| $n \equiv 2 \pmod{p}$ | $\Phi = x^n + y^{n-1}z + z^{n-1}y$ | $(x : y : z) \mapsto (x : \xi y : \xi^{-1}z)$ |
| $2g + 1 \equiv 0 \pmod{p}$ | $\Phi = x^{n-1}y + y^{n-1}z + z^{n-1}x$ | $(x : y : z) \mapsto (x : \xi y : \xi^{2-n}z)$ |

TABLE 3. Here ξ is to be a primitive p -th root of unity.

Each of these four curves is smooth, since in each case it is not difficult to check that the only solution to the equations $\Phi_x = \Phi_y = \Phi_z = 0$ is $x = y = z = 0$. Furthermore, the indicated mappings are clearly period p automorphisms of $\mathbb{P}^2(\mathbb{C})$, and it is not difficult to check that each one maps the corresponding locus $\Phi = 0$ to itself.

The last case (corresponding to Case (2) above) is the most interesting. Note that Φ is the sum of three monomials, which are multiplied respectively by ξ , by $\xi^{n-1+2-n} = \xi$, and by $\xi^{(n-1)(2-n)}$. Thus, in order to map the curve $\Phi = 0$ to itself, we must have

$$(n-1)(2-n) \equiv 1 \pmod{p}.$$

But this is equivalent to our hypothesis that $n^2 - 3n + 3 \equiv 0 \pmod{p}$.

We must also check that this congruence

$$n^2 - 3n + 3 \equiv 0 \pmod{p}$$

has a solution n for a given prime $p > 3$ if and only if p is of the form $6m+1$. Multiplying this congruence by 4 and then making the substitution $2n = b+3$ we obtain the equation $b^2 + 3 \equiv 0 \pmod{p}$. In other words there is a solution if and only if the quadratic residue symbol $\left(\frac{-3}{p}\right)$ is equal to one. By quadratic reciprocity (see for example [L]), we have

$$\left(\frac{3}{p}\right) = (-1)^{(p-1)/2} \left(\frac{p}{3}\right) \quad \text{and} \quad \left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} \quad \text{hence} \quad \left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right).$$

Thus there is a solution if and only if $p \equiv 1 \pmod{3}$. Since p is odd, this means that $p \equiv 1 \pmod{6}$.

To prove that the conditions of Theorem 9.1 are necessary, we will make use of the following.

Lemma 9.2. *Any automorphism of $\mathbb{P}^2(\mathbb{C})$ of prime order p can be put into the normal form*

$$(x : y : z) \mapsto (x : \xi y : \xi^k z) \tag{35}$$

by a linear change of coordinates in \mathbb{P}^2 , where ξ is a primitive p -th root of unity and k is an integer modulo p .

Proof. Any automorphism of $\mathbb{P}^2(\mathbb{C})$ lifts to an automorphism of \mathbb{C}^3 . If the given automorphism is non-trivial with finite order, then using the Jordan normal form we see that the lifted automorphism is diagonalizable, with either two or three distinct eigenvalues. After dividing by a suitable constant and permuting the variables, we can always put the map into the form $(x : y : z) \mapsto (x : \xi y : \xi^k z)$. If the automorphism has prime order p then we can choose ξ to be a primitive p th root of unity, as required. \square

The proof that the conditions of Theorem 9.1 are necessary now proceeds as follows. Every homogeneous polynomial $\Phi(x, y, z)$ can be expressed as a sum of monomials $a_{h,i,j} x^h y^i z^j$ with $h+i+j$ equal to the degree n . If the curve is smooth, there must be at least one such non-zero monomial with $h \geq n-1$, since otherwise the curve would be singular at the point $(1 : 0 : 0)$. Similarly, there must be one such monomial with $i \geq n-1$ and one with $j \geq n-1$. There may well be other non-zero monomials in the defining equation; but we will ignore them and focus on the nine monomials which satisfy these conditions.

Under the linear map $(x, y, z) \mapsto (x, \xi y, \xi^k z)$ from \mathbb{C}^3 to itself which corresponds to (35), each monomial $x^h y^i z^j$ will be multiplied by the power ξ^ℓ , where $\ell \equiv i + kj \pmod{p}$. The exponents ℓ corresponding to our nine chosen monomials can be tabulated as shown in Table 4.

| | A | | | B | | | C | | |
|---------------|-------|------------|------------|-------|------------|------------|-------|------------|------------|
| $x^h y^i z^j$ | x^n | $x^{n-1}y$ | $x^{n-1}z$ | y^n | $y^{n-1}z$ | $y^{n-1}x$ | z^n | $z^{n-1}y$ | $z^{n-1}x$ |
| ℓ | 0 | 1 | k | n | $n-1+k$ | $n-1$ | nk | $1+(n-1)k$ | $(n-1)k$ |

TABLE 4. Special monomials and the exponent of ξ

Here the chosen monomials have been separated into three groups, which we will refer to as A, B and C. If we choose one monomial from each of these three groups, the sum will be a homogeneous polynomial defining a (not necessarily smooth) plane curve. But in order for the resulting curve to admit (35) as an automorphism of period p , the following requirement must be satisfied:

Condition (*). The exponents ℓ corresponding to the three chosen monomials must be congruent to each other modulo p .

If n is congruent to 0, 1, or 2 mod p , then we know from Table 3 that this condition can be satisfied. Suppose then that n is not congruent to 0, 1, or 2 mod p . For the rest of the proof, “congruence” will always mean congruence mod p .

First suppose that we try to choose the first monomial in Group A, with exponent $\ell = 0$. Then since $n-1 \neq 0$ and $n \neq 0$, must choose $n-1+k$ as the exponent in Group B, so that $k \equiv 1-n \neq 0$. We can then check that none of the three choices in Group C will work. (As an example, to show that $1+(n-1)k$ cannot be congruent to zero, note that $k \equiv 1-n$ and hence $1+(n-1)k \equiv 1-(n-1)^2 = n(2-n)$, which cannot be congruent to zero.)

However, if we choose 1 as exponent in Group A then we must choose the middle entry $n-1+k$ in Group B, where now k is congruent to $2-n \neq 0$. A straightforward computation shows that the first and last choices in Group C will not work; but the middle choice gives us the required polynomial of Case (2). If we choose the last entry in Group A, then a similar argument shows that we must choose the last entry in Group B and in Group C, yielding a polynomial which is equivalent to Case (2). This proves Theorem 9.1. \square

Caution: There is no claim that these are the only possible curves. We can form many other curves by adding other monomials to the defining equation. The only claim is that these are the only pairs (n, p) which can be realized.

Now consider the following.

Theorem 9.3 (Chang). *For $n \geq 4$, a generic real or complex curve of degree n in \mathbb{P}^2 has no projective automorphisms other than the identity map.*

On the other hand, for degree $n = 3$ the projective automorphism group of a generic curve has order six in the real case, and eighteen in the complex case. (See [BM, §3]. In terms of the additive group structure on an elliptic curve, taking a flex point as zero element, the group of automorphisms is generated by the translations $\mathbf{p} \mapsto \mathbf{p} + \mathbf{p}_0$, where \mathbf{p}_0 can be any one of the flex points, together with the reflection $\mathbf{p} \mapsto -\mathbf{p}$. There are three flex points in the real case, and nine in the complex case.)

To give the idea of the argument without too many complicated estimates, we will consider only the degree 4 case.

Proof of Theorem 9.3 for $n = 4$. Clearly it suffices to prove that a generic curve of degree 4 has no automorphism of prime order. The space of smooth curves of degree $n = 4$

has dimension $n(n+3)/2 = 14$. We must show that the algebraic subspace consisting of curves with an automorphism of period p has dimension strictly less than 14 for each p . In fact we know from Table 1 that we need only to consider the primes 2, 3, and 7.

Any automorphism of \mathbb{P}^2 of prime period can be put into diagonal form by a linear change of variables. First suppose that there are only two distinct eigenvalues, so that it can be put into the form

$$(x : y : z) \mapsto (x : y : \xi z) \quad \text{so that} \quad x^h y^i z^j \mapsto \xi^j x^h y^i z^j, \quad (36)$$

where ξ is a primitive p -root of unity. First consider the prime $p = 2$. It is not hard to check that there are nine monomials of degree four with j even and only six with j odd. Thus, if we consider only linear combinations of monomials with j even we get an 8 dimensional space of curves with an automorphism of period 2. Since these include the Fermat curve $x^4 + y^4 + z^4$ which is certainly smooth, it follows that a generic curve of this form is smooth. We could also consider the case j odd, but this would yield a 5 dimensional space, which we can ignore.

Now note that the normal form (36) corresponds to a splitting of \mathbb{C}^3 as the direct sum of a two-dimensional eigenspace (for the x, y variables), and a one dimensional eigenspace (for the z variable). It takes two parameters to specify the first eigenspace, and two more parameters to specify the second eigenspace, and finally at most 8 more parameters to specify the curve. Thus the dimension of the space of all such curves is $2 + 2 + 8 = 12$, which is less than 14, as required.

For period $p = 3$, a similar argument shows that there are 7 monomials with $j \equiv 0 \pmod{3}$, and 5 with $j \equiv 1$, leaving 3 with $j \equiv 2$. Thus the space of all such smooth curves has dimension at most $2 + 2 + (7 - 1) = 10$. (We say ‘‘at most’’ 10, since we have not checked for smoothness.)

For $p \geq 5$, there are no smooth curves of degree 4 with an automorphism having only two distinct eigenvalues. In fact every such curve must either contain the line $z = 0$ (if $j > 0$), or else be a cone over the point $(0, 0, 1)$ (if $j = 0$).

Now consider an automorphism of period 3 with three distinct eigenvectors, so that we can use the normal form $(x, y, z) \mapsto (x, \xi y, \xi^2 z)$. Then we must consider the value of $i + 2j$ modulo 3. It turns out that there are 6 monomials with $i + 2j \equiv 0$, and 4 with $i + 2j \equiv 1$, and hence 5 with $i + 2j \equiv 2$. Since it takes 6 parameters to specify the three eigenspaces, we get an upper bound of $6 + (6 - 1) = 11$, which is again less than 14.

The case of period 7 with three distinct eigenvalues is more complicated, since there are several possible choices of ξ' . However one can check that it is always possible to reduce to the case $\xi' = \xi^2$ or $\xi' = \xi^3$. In the first case the most frequent value is $1 + 2j \equiv 4 \pmod{7}$ which occurs three times, while in the second case it is $1 + 3j \equiv 3$, which occurs three times. Thus the dimension is at most $6 + (3 - 1) = 8$. This completes the proof for curves of degree 4. \square

It follows that in the space of smooth degree four curves, the codimension of the subset consisting of curves with non-trivial stabilizer is precisely $14 - 12 = 2$.

Theorem 9.4. *Any finite subgroup of $\mathbf{G} = \mathrm{PGL}_3(\mathbb{C})$ can occur as a stabilizer. More explicitly, for any subgroup $\Gamma \subset \mathbf{G}$ of finite order m , there exists a smooth projective curve \mathcal{C} of degree $4m$ with stabilizer $G_{\mathcal{C}}$ equal to Γ .*

See [HMO] and [Harui] for discussion of all stabilizers of smooth plane curves.

Remark 9.5. A catalog of all possible finite subgroups of $\mathbf{G} = \mathrm{PGL}_3(\mathbb{C})$ has been provided by Miller, Blichfeld, and Dickson [MBD, Part II]. (See also Hambleton and Lee [HL].) Without giving the complete list, here are some examples: The group can have arbitrary order, since any abelian group with two generators is contained in the stabilizer for three lines in general position, as described at the beginning of Section 8. Any finite subgroup of the rotation group SO_3 can occur, since $\mathrm{SO}_3 \subset \mathrm{PGL}_2 \subset \mathrm{PGL}_3$. There are only three simple groups:

- The group $\mathrm{PSL}_2(\mathbb{F}_7)$ of order 168 can be realized as the stabilizer of the Klein quartic. (See the discussion following Table 1 above.)
- The alternating group \mathfrak{A}_6 of order 360 is the stabilizer of a smooth degree six curve which was described by Wiman [Wim] in 1896.
- The alternating group \mathfrak{A}_5 of order 60 (= icosahedral group) also occurs as the stabilizer of a degree six curve. Dolgachev, Farb and Looijenga [DFL], following Edge, have described a pencil of degree six curves, first studied by Wiman, such that the generic element is smooth, with \mathfrak{A}_5 as stabilizer.

Another noteworthy example, of order 216, is the automorphism group of the Hesse configuration, which can be described as a singular curve consisting of twelve lines which intersect in the nine flex points of an elliptic curve. (Compare [BM, Figure 7].)

Proof of Theorem 9.4. Let Γ be a finite subgroup of $\mathbf{G} = \mathrm{PGL}_3(\mathbb{C})$ with m elements. Since the case $m = 1$ is taken care of by Theorem 9.3, we may assume that $m > 1$. It is easy to construct a singular curve in \mathfrak{C}_{4m} which has Γ as stabilizer: According to Theorem 9.3, a generic curve $\mathcal{C}_1 \in \mathfrak{C}_4$ has trivial stabilizer. Let $\mathcal{C}_1^\Gamma \in \mathfrak{C}_{4m}$ be the union of the translates $\mathbf{g}(\mathcal{C}_1)$ by the elements $\mathbf{g} \in \Gamma$. Then it is not hard to see that the stabilizer of \mathcal{C}_1^Γ is precisely the group Γ .

In order to find a smooth example, we will use Bertini's Theorem,³⁹ which asserts that a locus of the form $\alpha_1\Phi_1 + \cdots + \alpha_k\Phi_k = 0$, where the Φ_j are homogeneous polynomials of the same degree, is non-singular for a generic choice of the coefficients α_j , provided that the common zero locus

$$\Phi_1 = \cdots = \Phi_k = 0$$

is empty. To apply this theorem, choose three curves $\mathcal{C}_i \in \mathfrak{C}_4$ which are generic in the sense that the triple $(\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3)$ is a generic point of $\mathfrak{C}_4 \times \mathfrak{C}_4 \times \mathfrak{C}_4$. Then each pair \mathcal{C}_i^Γ and \mathcal{C}_j^Γ will intersect transversally in $(4m)^2$ distinct points, but the 3-fold intersection $\mathcal{C}_1^\Gamma \cap \mathcal{C}_2^\Gamma \cap \mathcal{C}_3^\Gamma$ will be empty. Now let $\Phi_j = 0$ be the equation of \mathcal{C}_j^Γ . Then for a generic choice of coefficients α_j the locus $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ defined by the equation

$$\alpha_1\Phi_1 + \alpha_2\Phi_2 + \alpha_3\Phi_3 = 0 \tag{37}$$

will be a smooth Γ -invariant curve. We will prove the following.

Lemma 9.6. *For $(\alpha_1 : \alpha_2 : \alpha_3)$ in a small neighborhood of $(1 : 0 : 0)$ the curve $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ has stabilizer precisely equal to Γ .*

Proof. Since \mathcal{C}_1 is a generic curve, the intersections between \mathcal{C}_1 and its translates by elements of Γ must all be transverse. It follows that \mathcal{C}_1^Γ , which is a union of m translates of \mathcal{C}_1 , is smooth except for $4^2 \binom{m}{2} = 8m(m-1)$ simple self-crossing points.

Since \mathcal{C}_1^Γ has only simple self-crossing singularities, we know by Theorem 7.13 or by Corollary 7.18 that there exists a \mathbf{G} -invariant neighborhood U of \mathcal{C}_1^Γ in \mathfrak{C}_{4m} on which

³⁹See for example [Harr] or [Nam]. We thank Robert Lazarsfeld for suggesting this argument.

the action is proper. Therefore we know by Lemma 2.5 that the order of the stabilizer is upper semi-continuous as a function of the parameters, and hence is less than or equal to its value at the point $(1 : 0 : 0)$ for all points sufficiently close to $(1 : 0 : 0)$. Since the stabilizer is equal to Γ at $(1 : 0 : 0)$ and contains Γ for all parameter values, this shows the stabilizer is equal to Γ throughout some neighborhood of $(1 : 0 : 0)$. This proves the Lemma. Since a generic $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ in this neighborhood is smooth, it also completes the proof of Theorem 9.4. \square

Remark 9.7. With a little more work we can prove that the curve $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ has stabilizer Γ for a generic choice of $(\alpha_1 : \alpha_2 : \alpha_3)$. The proof will be based on the following.

Let H be an arbitrary finite group of order $h > m$, and let $\text{Hom}(H, \mathbf{G})$ be the set of all group homomorphisms (or representations) $\rho : H \rightarrow \mathbf{G} = \text{PGL}_3$. Then $\text{Hom}(H, \mathbf{G})$ is an algebraic subset of the h -fold cartesian product \mathbf{G}^H consisting of *all* functions from H to \mathbf{G} . (Of course this set is not compact, since \mathbf{G} is not compact.)

The subset $\text{Hom}_{\text{inj}}(H, \mathbf{G})$ consisting of injective homomorphisms (= faithful representations) is a closed algebraic subset. In fact it is both open and closed in the classical topology, as a subset of $\text{Hom}(H, \mathbf{G})$. (If a sequence of injective homomorphisms ρ_j converged to a non-injective homomorphism, then the images $\rho_j(h_j)$ of some non-identity elements of H would have to converge to the identity element of \mathbf{G} . But this would imply the the order of $\rho_j(h_j)$ tends to infinity, which is impossible. Compare the proof of Lemma 2.25.) Since every irreducible set is connected in the classical topology, it follows that $\text{Hom}_{\text{inj}}(H, \mathbf{G})$ is also a Zariski closed algebraic subset of $\text{Hom}(H, \mathbf{G})$.

Let $\mathbb{P}_{\text{par}}^2$ be the parameter plane consisting of all $(\alpha_1 : \alpha_2 : \alpha_3)$, and let

$$X(H) \subset \mathbb{P}_{\text{par}}^2 \times \text{Hom}_{\text{inj}}(H, \mathbf{G})$$

be the algebraic set consisting of all pairs $((\alpha_1 : \alpha_2 : \alpha_3), \rho)$ in $\mathbb{P}_{\text{par}}^2 \times \text{Hom}_{\text{inj}}(H, \mathbf{G})$ such that the curve of Equation (37) is invariant under $\rho(H)$. Finally, let $Y(H) \subset \mathbb{P}_{\text{par}}^2$ be the image of $X(H)$ under projection to the first factor. By a Theorem of Chevalley, this image is “constructible”, (meaning that it can be obtained from Zariski open and closed sets by taking finite unions and finite intersections). In particular, if $Y(H)$ is Zariski dense in $\mathbb{P}_{\text{par}}^2$, then it must contain a Zariski open subset of $\mathbb{P}_{\text{par}}^2$, (Compare [Gor], or [Harr, Lem. 3.17]; as well as the proof of Lemma 6.2 above.)

It follows that this image $Y(H)$ has dimension at most one. For if it had dimension two then there would be a Zariski dense and open subset of $\mathbb{P}_{\text{par}}^2$ for which the corresponding stabilizer contains a copy of H . In particular, this would be true for almost all points in a small neighborhood of the point $(1 : 0 : 0)$. But Lemma 9.6 shows that this is not possible. Therefore the dimension of $Y(H)$ is at most one; so that a generic curve $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ does *not* admit H as an automorphism group.

Finally, note that only finitely many groups H need be considered, since according to Hurwitz there is an upper bound on the orders of possible automorphism groups of smooth curves. (Compare Remark 9.8). This proves that a generic $\mathcal{C}(\alpha_1 : \alpha_2 : \alpha_3)$ must have stabilizer precisely equal to Γ .

Remark 9.8 (The Hurwitz Theorem: Conformal Automorphisms of Riemann Surfaces). The group of all conformal automorphisms of a compact Riemann surface of genus $g \geq 2$ has been much studied since the time of Hurwitz [Hur], who proved that such a group has at most $84(g - 1)$ elements. The cases in which the group has exactly $84(g - 1)$ elements are particularly interesting since they correspond to tessellations of the

curve by hyperbolic triangles, each with angles $\pi/2$, $\pi/3$ and $\pi/7$ and with area $2\pi/84$. (For more on this see [Co].) There are now many examples of groups which realize this Hurwitz maximum. As an extreme example, Wilson [Wi] has shown that the “monster group” of order roughly 8×10^{53} is one such group.

In Theorem 9.3 we proved that for $n \geq 4$, a generic real or complex curve of degree n in \mathbb{P}^2 has no projective automorphisms other than the identity map. The corresponding statement for arbitrary Riemann surfaces is that a generic Riemann surface of genus $\mathfrak{g} \geq 3$ has no non-trivial conformal automorphism. (See [Ba] as well as [Po].) However, every Riemann surface of genus two is hyperelliptic, and hence has an automorphism of period two. Of course any surface of genus zero or one has automorphisms of all orders.

Theorem 9.9. *Given a genus $\mathfrak{g} \geq 2$ and a prime p there exists a compact Riemann surface with an automorphism of period p if and only if p has the form*

$$p = 1 + \frac{2(\mathfrak{g} - \mathfrak{g}_0)}{\Delta}, \quad (38)$$

with $\mathfrak{g}_0 \geq 0$ and $\Delta \geq 1$, where the expression

$$\mathfrak{f} = \Delta + 2 - 2\mathfrak{g}_0$$

must satisfy $\mathfrak{f} \geq 0$ with $\mathfrak{f} \neq 1$.

Here \mathfrak{g}_0 is the genus of the quotient surface in which each orbit of the automorphism has been collapsed to a point. (Compare [HKKO].) Evidently it follows from (38) that $\mathfrak{g} > \mathfrak{g}_0$ in all cases. The denominator Δ has no direct geometric interpretation; however \mathfrak{f} can be described as the number of fixed points of the automorphism.

To understand this result, first consider the case $\Delta = 1$. The conditions $\mathfrak{f} \geq 0$ and $\mathfrak{f} \neq 1$ then require that $\mathfrak{g}_0 = 0$, yielding $p = 2\mathfrak{g} + 1$. Thus whenever $2\mathfrak{g} + 1$ is prime, there exists an automorphism of period $2\mathfrak{g} + 1$. In all cases, since $\Delta \geq 1$ and $\mathfrak{g}_0 \geq 0$, it follows easily that $p \leq 2\mathfrak{g} + 1$ for any p which satisfies (38). (Compare [Bars].)

Next consider the case $\Delta = 2$. Then \mathfrak{g}_0 can be either 0, 1 or 2. Correspondingly the expression $1 + 2(\mathfrak{g} - \mathfrak{g}_0)/2$ will be either $\mathfrak{g} - 1$, \mathfrak{g} , or $\mathfrak{g} + 1$. Thus whenever one of these numbers is prime, there is a corresponding automorphism of this prime order. In some extreme cases such as $\mathfrak{g} = 6$ or 18 or 30, all three of the numbers $2\mathfrak{g} + 1$ and $\mathfrak{g} \pm 1$ will be primes. At the other extreme, for $\mathfrak{g} = 25$ none of the numbers \mathfrak{g} , $\mathfrak{g} \pm 1$ and $2\mathfrak{g} + 1$ is prime.

In all of the remaining cases, with $\Delta \geq 3$, it is easy to check that $p \leq 1 + 2\mathfrak{g}/3$, which is strictly smaller than $\mathfrak{g} - 1$ when $\mathfrak{g} > 6$.

Now consider small primes. The primes 2 and 3 can occur in any genus:

For $p = 2$ we can choose

$$\mathfrak{g}_0 = 0 \quad \text{and} \quad \Delta = 2\mathfrak{g} \quad \text{so that} \quad \mathfrak{f} = 2 + 2\mathfrak{g}.$$

For $p = 3$,

$$\mathfrak{g}_0 = 0 \quad \text{and} \quad \Delta = \mathfrak{g} \quad \text{so that} \quad \mathfrak{f} = 2 + \mathfrak{g}.$$

The case $p = 5$ is more delicate. Equation (38) implies that $\mathfrak{g} - \mathfrak{g}_0 = 2\Delta$. If we choose \mathfrak{g}_0 to be 0 or 1 according as \mathfrak{g} is even or odd, then the necessary inequality $\mathfrak{f} = \Delta + 2 - 2\mathfrak{g}_0 \geq 0$ is certainly satisfied; with $\mathfrak{f} > 1$ when $\mathfrak{g} \neq 3$. However, in the special case $\mathfrak{g} = 3$, the only possible choice is $\mathfrak{g}_0 = 1$ with $\Delta = 1$. But then $\mathfrak{f} = \Delta + 2 - 2\mathfrak{g}_0 = 1$, which is not allowed.

Thus there exists a surface of genus \mathfrak{g} with an automorphism of period 5 if and only if $\mathfrak{g} \neq 3$.

| | | | | | | | | | | | |
|----------------------|---|---|---|----|-------|---|-------|-------|-------|--------|--------|
| g | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $p \geq 7$ | | 7 | | 11 | 7, 13 | 7 | 7, 17 | 7, 19 | 7, 11 | 11, 23 | 11, 13 |
| g₀ | | 0 | | 0 | 0, 0 | 1 | 2, 0 | 0, 0 | 1, 0 | 1, 0 | 2, 0 |
| f | | 3 | | 3 | 4, 3 | 2 | 0, 3 | 5, 3 | 3, 4 | 2, 3 | 0, 4 |
| Δ | | 1 | | 1 | 2, 1 | 2 | 2, 1 | 3, 1 | 3, 2 | 2, 1 | 2, 2 |

TABLE 5. Showing the possible primes $p \geq 7$ which can occur for each genus between 2 and 12. The auxiliary numbers g_0 , f , and Δ are also shown. The primes 2, 3 and 5 are not listed here; but are discussed in detail above.

Table 5 shows the possible choices for curves of genus $2 \leq g \leq 12$ and for primes $p \geq 7$. (For these particular values of g and p , there is at most one choice of the quotient surface genus g_0 , and hence at most one choice for f and Δ . However this is certainly not true for all g and p . Even for the prime 5, there are two possible choices for g_0 when the genus is 6, 10, or 11.)

Remark 9.10. Chang [Ch] has proved the following.

Any conformal isomorphism between smooth plane projective curves of degree at least four is actually a projective isomorphism.

(In fact he proves the same even for birational isomorphisms.) In particular, the conformal automorphism group for such a curve can be identified with its projective automorphism group. It follows that the primes which occur in Table 1 for degree n must also occur in Table 5 for genus $g = \binom{n-1}{2}$. Any primes which occur in Table 5 but not in Table 1 must correspond to Riemann surfaces which cannot be smoothly embedded in \mathbb{P}^2 . As examples, a surface of genus 6 with an automorphism of period 7, or a surface of genus 10 with an automorphism of period 11, cannot be embedded into \mathbb{P}^2 .

Proof of Theorem 9.9. The basic tool is the Riemann-Hurwitz formula for a branched covering map $\mathcal{S} \rightarrow \mathcal{S}_0$ of degree p , which states that:

the number of branch points, counted with multiplicity, is equal to $p\chi(\mathcal{S}_0) - \chi(\mathcal{S})$,

where χ is the Euler characteristic. (See for example [Mi2, Theorem 7.2].) In the case of a cyclic p -fold branched covering $\mathcal{S} \mapsto \mathcal{S}_0$ with f fixed points there are just f branch points, each with multiplicity $p - 1$, so the formula takes the form

$$(p - 1)f = p\chi(\mathcal{S}_0) - \chi(\mathcal{S}) . \tag{39}$$

Solving for p we obtain

$$p = \frac{f - \chi(\mathcal{S})}{f - \chi(\mathcal{S}_0)} = \frac{f + 2g - 2}{f + 2g_0 - 2} ,$$

Here the numerator is positive if $g \geq 2$, so the denominator must also be positive. Denoting this denominator by

$$\Delta = f + 2g_0 - 2$$

we can solve for $f = \Delta + 2 - 2g_0$. Substituting this expression for f in the numerator, we obtain the required formula

$$p = \frac{(\Delta + 2 - 2g_0) + 2g - 2}{\Delta} = 1 + \frac{2(g - g_0)}{\Delta} .$$

Here $g_0 \geq 0$ and $f \geq 0$ by definition.

Lemma 9.11. *Such a branched covering exists except in the following special cases:*

- (1) *It cannot exist if $\mathbf{f} = 1$, or if $\mathbf{f} = \mathbf{g}_0 = 0$.*
- (2) *In the case $p = 2$, it exists only under the extra condition that \mathbf{f} is even.*

Proof. Choose a surface \mathcal{S}_0 of genus \mathbf{g}_0 , and choose a finite subset $K \subset \mathcal{S}_0$ consisting of \mathbf{f} points. A p -fold cyclic covering of $\mathcal{S}_0 \setminus K$ is determined by a homomorphism from the fundamental group $\pi_1(\mathcal{S}_0 \setminus K)$, or equivalently from the abelianized fundamental group $H_1(\mathcal{S}_0 \setminus K)$, onto the cyclic group \mathbb{Z}/p . (See [Li] or [Mun].) This covering will extend to a branched covering \mathcal{S} of \mathcal{S}_0 , branched over each of the points of K , if and only if a small loop around each one of the \mathbf{f} points maps nontrivially to \mathbb{Z}/p . If $\mathbf{f} = 0$ there is an additional requirement that $\mathbf{g}_0 > 0$ so that $\mathcal{S}_0 \setminus K$ is not simply-connected. If $\mathbf{f} = 1$, then a small loop around the single point of K is homologous to zero, so no such homomorphism can exist. If $p = 2$, since the sum of the elements of H_1 corresponding to such small loops is zero, we must have \mathbf{f} even. However in all other cases, it is not difficult to construct the required homomorphism. This completes the proof of Lemma 9.11; \square
and hence the proof of Theorem 9.9. \square

Whenever there exists a pair consisting of a Riemann surface \mathcal{S} of genus \mathbf{g} and an automorphism α of period p with \mathbf{f} fixed points, the proof shows that the space of all isomorphism classes of such pairs is a finite covering of the moduli space $\mathcal{M}_{\mathbf{g}_0, \mathbf{f}}$.

Remark 9.12. As a final question, we can ask which groups can occur as conformal automorphism groups. This has a very easy answer.

Every finite group Γ can occur as the group of conformal automorphisms of some compact Riemann surface.

This can be proved as follows. Choosing some set of generators $\mathbf{g}_1, \dots, \mathbf{g}_k$ for Γ , we can form the associated Cayley graph, with one vertex for each group element, and with one edge from \mathbf{g} to $\mathbf{g}_j \mathbf{g}$ for each $\mathbf{g} \in \Gamma$ and each generator \mathbf{g}_j . Now replace each edge by a cylinder $[0, 1] \times \mathbb{R}/\mathbb{Z}$. Choose a smooth surface of genus zero with $2k$ boundary circles, and identify these boundary circles with the $2k$ cylinder ends which are closest to the vertex corresponding to the identity element. It is not hard to do this so that the result is C^∞ smooth. Paste an isomorphic copy of this genus zero surface around each other vertex of the graph, so that Γ acts transitively on the resulting compact C^∞ surface M . It is not hard to check that the resulting surface has genus

$$\mathbf{g} = 1 + (k - 1)m ,$$

where m is the order of Γ . We need to have $\mathbf{g} > 1$; but this can always be achieved by enlarging the chosen list of generators if necessary. Assuming that $\mathbf{g} > 1$, the Uniformization Theorem asserts that M has a uniquely determined complex structure, and hence a unique metric of constant curvature -1 .

This surface clearly admits Γ as a group of isometries. However, it may well have extra isometries. (For example this will be the case whenever there is an automorphism of Γ which preserves the list of generators.) For this reason, we need to make the construction a little less regular. Perhaps the easiest way to do this is to pinch each cylinder near one end, to create a very short geodesic. Furthermore, we can do this so that different generators correspond to short geodesics of different length. Since every automorphism must map a short geodesic to another short geodesic of the same length, we can guarantee in this way that the only automorphisms are the elements of Γ .

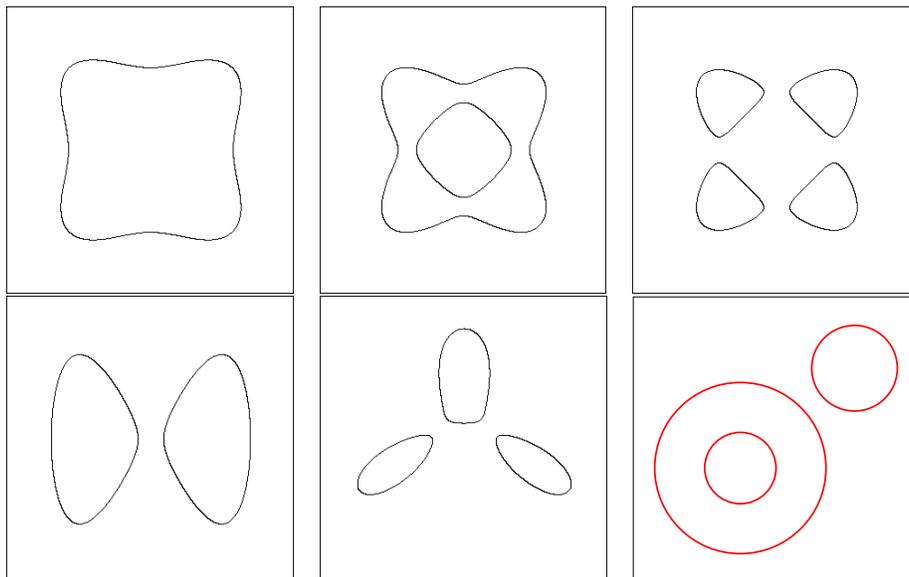


FIG. 29. Six examples. The first five panels above show representative curves for the five connected components⁴⁰ of $\mathcal{C}_4^{\text{cs}}(\mathbb{R})$ for which the real locus $|\mathcal{C}|_{\mathbb{R}}$ is non-empty. Note that the various components of $|\mathcal{C}|_{\mathbb{R}}$ always arrange themselves so that no line intersects more than two of them. The sixth panel shows a configuration of three components which cannot occur for any curve of degree less than six, since a line through the two smaller circles crosses all three circles, and hence has six intersection points.

10. REAL CURVES: THE HARNACK-HILBERT PROBLEM.

This section will be a digression, discussing a different kind of problem. Harnack in 1876 proved that⁴¹

The number of connected components of a smooth curve of degree n in the real projective plane is at most $g + 1$ where $g = \binom{n-1}{2}$ is the genus.

As examples, for degree three the curve $|\mathcal{C}|_{\mathbb{R}}$ has most two components, and for degree four at most four. (Compare Figure 29.)

The most famous question about such curves is the first part of Hilbert's 16th Problem (see [Hi1, Hi2]), which concerns real algebraic curves:

"The maximum number of closed and separate branches which a plane algebraic curve of the n th order can have has been determined by Harnack. There arises the further question as to the relative position of the branches in the plane. ..."

Hilbert goes on to discuss real varieties of higher dimension (see [Kh]), and also algebraic differential equations (see [Il]). However, we will discuss only his question about real curves.

⁴⁰The classification in Figure 29 depends on Zeuthen's proof that any two smooth curves of degree four which are topologically isotopic are necessarily isotopic through a smooth one-parameter family of projective curves, and hence belong to the same connected component of $\mathcal{C}_4^{\text{cs}}$. (See [Z] or [V3, p. 197].)

⁴¹See [Harn]. Klein, who was Harnack's advisor, later published his own proof, as described below.

For further information, see for example [Roh], [Pe], [Arn], [Gu], [R1, R2, R3], [W], [V1, V2, V3] (listed in chronological order).

Real-Smooth and Complex-Smooth Curves

A curve \mathcal{C} defined over \mathbb{R} will be called *real-smooth* if there are no singularities in the real zero-locus $|\mathcal{C}|_{\mathbb{R}}$, and *complex-smooth* if there are no singularities in the complex zero locus $|\mathcal{C}|_{\mathbb{C}}$. Thus there are open subsets

$$\mathfrak{C}_n^{\text{cs}} = \mathfrak{C}_n^{\text{cs}}(\mathbb{R}) \subset \mathfrak{C}_n^{\text{rs}} = \mathfrak{C}_n^{\text{rs}}(\mathbb{R}) \subset \mathfrak{C}_n(\mathbb{R})$$

consisting of complex-smooth and real-smooth curves. It is somewhat easier to construct examples if we work in the larger space $\mathfrak{C}_n^{\text{rs}}$. For example, any union of two or more disjoint circles or ellipses in $\mathbb{P}^2(\mathbb{R})$ is real-smooth but not complex-smooth. (Of course the number of components in such trivial examples is very much smaller than Harnack's upper bound.)

On the other hand, in the complex-smooth case we can obtain extra information by considering the way that the real locus $|\mathcal{C}|_{\mathbb{R}}$ is embedded in the Riemann surface $|\mathcal{C}|_{\mathbb{C}}$. As Klein noted, this leads to a quite easy proof of Harnack's Theorem, as follows (see [K]):

Any $g + 1$ disjoint embedded circles in a Riemann surface necessarily disconnect it. (Compare Remark 7.3.) If we choose a minimal set of such disconnecting circles, then there must be exactly two complementary components. Now suppose that each of these circles is pointwise invariant under complex conjugation. Then complex conjugation must interchange the two complementary components. Therefore no further embedded circles can be invariant under complex conjugation. It follows easily that there can never be more than $g + 1$ disjoint circles which are invariant under complex conjugation.⁴²

In fact, for Hilbert's problem it doesn't matter whether we work with real-smooth or complex-smooth curves:

Proposition 10.1. *Every real-smooth curve can be approximated arbitrarily closely by a complex-smooth curve. Furthermore, in the space $\mathfrak{C}_n^{\text{rs}}$ of real-smooth curves of degree $n \geq 3$, let $\mathfrak{C}_{\text{sing}}$ be the algebraic subset consisting of curves \mathcal{C} such that the complex zero-set $|\mathcal{C}|_{\mathbb{C}}$ is singular. Then $\mathfrak{C}_{\text{sing}}$ has codimension two. Therefore no connected component in $\mathfrak{C}_n^{\text{rs}}$ is disconnected by this algebraic subset. In other words, every connected component in $\mathfrak{C}_n^{\text{rs}}$ determines a unique connected component in the smaller set $\mathfrak{C}_n^{\text{cs}}$.*

Proof. Step 1. The space of all curves of degree n in $\mathbb{P}^2(\mathbb{C})$ has complex dimension $d(n) = \binom{n+2}{2} - 1 = n(n+3)/2$. Let V_n be the subvariety consisting of curves having singular points at $(0 : 0 : 1)$ and $(0 : 1 : 0)$. Then the dimension of V_n is $d(n) - 6$. In fact the curve defined by the equation⁴³

$$\Phi(x, y, z) = \sum_{i+j+k=n} a_{i,j,k} x^i y^j z^k = 0$$

⁴²The case of genus one provides an interesting elementary exercise. It is easy to describe an involution of a torus with two invariant circles; but harder to visualize an involution with only one invariant circle. (The two cases are topologically quite different, so it is impossible to pass smoothly from one to the other)

⁴³Here Φ should have no squared factor, so that this equation defines a curve rather than a 1-cycle.

will pass through these two points only if $a_{0,0,n} = a_{0,n,0} = 0$, and will be singular at these two points only if

$$a_{1,0,n-1} = a_{0,1,n-1} = a_{1,n-1,0} = a_{0,n-1,1} = 0 .$$

Step 2. Given two (not necessarily disjoint) small open sets $U_1, U_2 \subset \mathbb{P}^2(\mathbb{C})$, let W_{U_1, U_2} be the set of triples $(\mathcal{C}, \mathbf{p}, \mathbf{q})$ consisting of a degree n complex curve \mathcal{C} having a marked singular point $\mathbf{p} \in U_1$ and a marked singular point $\mathbf{q} \in U_2$, with $\mathbf{p} \neq \mathbf{q}$. It takes four parameters to specify \mathbf{p} and \mathbf{q} . We can choose a projective transformation $T_{\mathbf{p}, \mathbf{q}}$ depending on these four parameters which carries \mathbf{p} to $(0 : 0 : 1)$ and \mathbf{q} to $(0 : 1 : 0)$. The equation $\Phi(T_{\mathbf{p}, \mathbf{q}}(x, y, z)) = 0$ will then uniquely describe the most general curve of degree n with \mathbf{p} and \mathbf{q} as singular points. It follows that the dimension of W_{U_1, U_2} is

$$4 + (d(n) - 6) = d(n) - 2 .$$

Since a curve can have at most finitely many critical points, it follows that the projection map from W_{U_1, U_2} into $\mathfrak{C}_n(\mathbb{C})$ is finite-to-one; and hence must map to a $d(n) - 2$ dimensional set W'_{U_1, U_2} . Now cover $\mathbb{P}^2(\mathbb{C}) \times \mathbb{P}^2(\mathbb{C})$ by finitely many $U_1 \times U_2$ and let W_n be the union of the W'_{U_1, U_2} .

Step 3. Since the Zariski closure $\overline{W_n}$ is invariant under complex conjugation, it must be defined over the real numbers. Therefore its intersection with $\mathfrak{C}_n(\mathbb{R})$ has real codimension two in $\mathfrak{C}_n^{cs}(\mathbb{R})$. But every real-smooth curve with a complex singularity must also have a complex conjugate singularity, and hence must belong to the codimension two subset $\overline{W_n} \cap \mathfrak{C}_n(\mathbb{R})$. \square

Definition 10.2. A circle smoothly embedded in $\mathbb{P}^2 = \mathbb{P}^2(\mathbb{R})$ will be called an *oval* if it is two-sided, separating the projective plane into two components, and a *non-oval* if it is one-sided, not separating the plane. (Note than an oval in this sense need not be convex. For example, two of the ovals in Figure 29 are non-convex.)

Equivalently, an embedded circle is an oval if and only if the associated homology class in $H_1(\mathbb{P}^2; \mathbb{Z}/2)$ is zero. Every oval has a neighborhood which is an annulus. Furthermore, one of its two complementary components must be a topological disk, while the other must be a Möbius band. On the other hand, every non-oval has a Möbius band neighborhood, and an open topological disk as complement. It follows from this that any two non-ovals must intersect each other, since it is impossible to embed a Möbius band in a disk. Note that a generic line intersects an oval in an even number of points, and a non-oval in an odd number of points.

Now consider a curve $\mathcal{C} \in \mathfrak{C}_n^{cs}$. The number of intersections between $|\mathcal{C}|_{\mathbb{R}}$ and a generic line in $\mathbb{P}^2(\mathbb{C})$ is always congruent to $n \pmod{2}$. (In fact the complexified line intersects $|\mathcal{C}|_{\mathbb{C}}$ n times; but an even number of these intersection points belong to complex conjugate pairs.) Therefore the discussion above implies that the real locus $|\mathcal{C}|_{\mathbb{R}}$ is a union of ovals if the dimension is even; but that it contains exactly one non-oval if n is odd.

In order to distinguish between different configurations of topological circles, it is convenient to introduce the dual graph, which is a combinatorial description of the topological arrangement.

Definition 10.3. First consider a collection of N disjoint ovals O_1, \dots, O_N in $\mathbb{P}^2(\mathbb{R})$, as in Figure 30. The associated **dual graph** Γ is a rooted tree which has $N + 1$ vertices, one vertex v_k corresponding to each connected component U_k of the complementary region. The root point corresponds to the unique complementary region U_0 which is non-orientable. Two vertices are joined by an edge, which will be labeled e_j , if and only if the closures of the corresponding regions intersect in the common boundary curve O_j . (We should think of this dual graph as an abstract tree: It can be embedded in the plane for illustrative purposes, but the particular choice of embedding is arbitrary.)

Now suppose that we are given a configuration consisting of $N - 1$ ovals together with one non-oval O_N . The root point will now correspond to the complementary region U_N which surrounds O_N . Since we cross from U_N to itself as we cross O_N , it is natural to define the resulting dual graph to be the rooted tree as constructed above, but augmented by an extra edge e_N which is a loop, with both endpoints at the root point. Compare Figure 31, which shows five ovals plus one non-oval, indicated schematically by a line segment, together with the corresponding dual graph. (Of course, if we ignore the non-oval, then we get a rooted tree in all cases.)

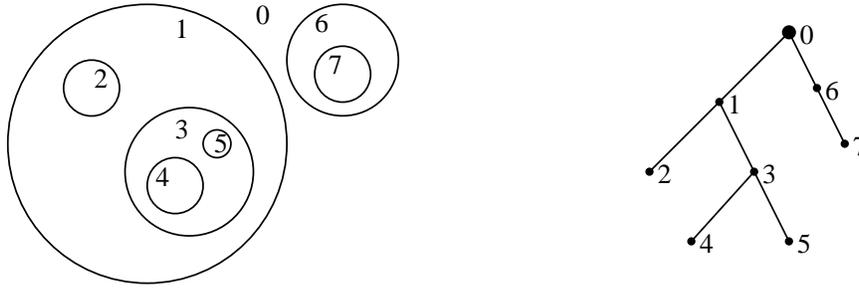


FIG. 30. A collection of seven ovals in the plane, and the associated dual graph. Each numbered vertex corresponds to the associated numbered complementary region, and each edge corresponds to the oval which separates two such regions.

It is not hard to check that one collection of embedded topological circles in $\mathbb{P}^2(\mathbb{R})$ can be deformed continuously into another if and only if they have isomorphic dual graphs, where the isomorphism is required to preserve the root point.

However, we are interested in smooth algebraic curves in $\mathbb{P}^2(\mathbb{R})$. If two curves of degree n represent the same connected component in the space \mathcal{C}_n^{rs} or \mathcal{C}_n^{cs} of smooth curves, then it follows that they have isomorphic rooted graphs. However the converse statement is false.

One can learn much about a curve $\mathcal{C} \in \mathcal{C}_n^{cs}$ by thinking of its real locus $|\mathcal{C}|_{\mathbb{R}}$ as a collection of topological circles embedded in the smooth Riemann surface $|\mathcal{C}|_{\mathbb{C}}$. Following Klein, a curve is said to be of **Type 1** if the Riemann surface $|\mathcal{C}|_{\mathbb{C}}$ is disconnected by $|\mathcal{C}|_{\mathbb{R}}$, and of **Type 2** if the difference set $|\mathcal{C}|_{\mathbb{C}} \setminus |\mathcal{C}|_{\mathbb{R}}$ is connected. Rokhlin [R3] described an example of two connected components in the space \mathcal{C}_5^{cs} such that the real loci $|\mathcal{C}|_{\mathbb{R}}$ for curves in one

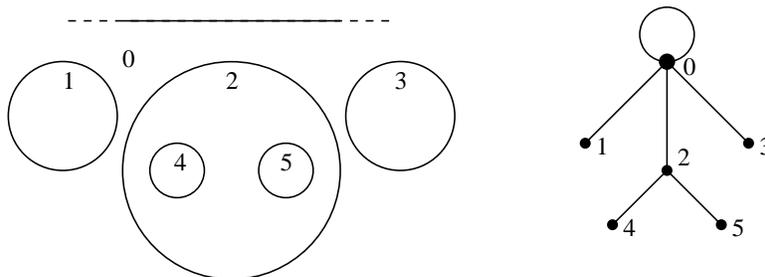


FIG. 31. A similar figure with five ovals plus one non-oval.

component can be deformed continuously to the real loci for curves in the other component, even though one of these components has Type 1, while the other has Type 2. (Compare [V3, Pg. 12].) For curves of degree six, Rokhlin [R3] and Nikulin [Ni] showed that the space $\mathfrak{C}_6^{\text{cs}}$ has 64 distinct connected components, although there are only 56 distinct real topological types. (Compare [KKPSS].)

Perhaps Hilbert's Problem should be reformulated in more modern terms as follows:

Is it possible to find an algorithm which, for any specified degree n and any rooted tree, will decide whether or not there is a curve in $\mathfrak{C}_n^{\text{cs}}(\mathbb{R})$ with the topological type corresponding to this tree? More precisely, can it decide how many components in $\mathfrak{C}_n^{\text{cs}}(\mathbb{R})$ have this topological type; and can it decide when two given curves belong to the same component? (Of course, to be really useful such an algorithm would have to run in polynomial time.)

Of course, as Hilbert indicated, the special case of curves with a maximal number of components would be of particular interest. These are often called *M-curves*.

It would also be interesting to find out what one can say about the topology of the various components of $\mathfrak{C}_n^{\text{cs}}(\mathbb{R})$. Perhaps, some components have a complicated fundamental group? For even n there is one component which is easy to understand: It is not hard to see that the component consisting of curves \mathcal{C} with no real points, so that $|\mathcal{C}|_{\mathbb{R}}$ is empty, is a convex subset of projective space.

Remark 10.4. One can also consider the *moduli space* $\mathbb{M}_n^{\text{cs}} = \mathfrak{C}_n^{\text{cs}}/\mathbf{G}$ for real curves which are complex-smooth, where $\mathbf{G} = \text{PGL}_3(\mathbb{R})$. Since $\mathfrak{C}_n^{\text{cs}}(\mathbb{R})$ is by definition a subset of $\mathfrak{C}_n^{\text{sm}}(\mathbb{C})$, it follows easily from Corollary 6.9 that the action of \mathbf{G} on $\mathfrak{C}_n^{\text{cs}}(\mathbb{R})$ is proper, and hence that the quotient space $\mathbb{M}_n^{\text{cs}}(\mathbb{R})$ is a Hausdorff orbifold. Since this group \mathbf{G} is connected, it follows easily that there is a one-to-one correspondence between connected components of $\mathfrak{C}_n^{\text{cs}}$ and connected components of \mathbb{M}_n^{cs} .

APPENDIX A. REMARKS ON THE LITERATURE.

The moduli space $\mathbb{M}_n^{\text{sm}}(\mathbb{C})$ for smooth curves of degree n has long been studied. In many cases it is known to be a rational variety. (Compare [S-B] and [BBK].) For the problem of compactifying $\mathbb{M}_n^{\text{sm}}(\mathbb{C})$, compare [Hac].

The Algebraic Geometer’s Bible for studying moduli spaces is Mumford’s “Geometric Invariant Theory” [Mu]. For other expositions of this theory, see for example [New], or [Sim]; and for the special case of PGL_{k+1} acting on hypersurfaces in \mathbb{P}^k see [Ne], as well as [Mu, Ch.4, §2]. The theory takes a simpler form in the very special case where the reductive Lie group \mathbf{G} acts on a variety $\mathbf{X} \subset \mathbb{P}^k$ *linearly*, that is by an embedding into PGL_{k+1} which lifts to an embedding into GL_{k+1} . A point of $\mathbf{x} \in \mathbf{X}$ is then called *stable* if the stabilizer $\mathbf{G}_{\mathbf{x}}$ of \mathbf{x} is finite, and if the orbit of a representative point $\widehat{\mathbf{x}} \in \mathbb{C}^{k+1} \setminus \{\mathbf{0}\}$ over \mathbf{x} is closed and bounded away from zero. If \mathbf{X}^s is the open subset consisting of stable points, then the quotient \mathbf{X}^s/\mathbf{G} is well behaved. The extension of this definition to more general group actions depends on a study of suitably linearized line bundles over \mathbf{X} . Particularly noteworthy are the Hilbert-Mumford numerical criterion for stability [Mu, Ch. 2], and the related Kempf-Ness criterion [KN].

Alternative definitions can be provided in a somewhat simpler way by introducing symplectic structures (Compare [GRS], or [MS2].) Here is a brief outline: Suppose that \mathbf{G} is a complex reductive group with maximal compact subgroup \mathbf{K} (for example $\mathbf{G} = \mathrm{PGL}_k(\mathbb{C})$, with $\mathbf{K} = \mathrm{PU}_k$), and suppose that \mathbf{G} acts on a manifold \mathbf{X} , which is provided with a \mathbf{K} -invariant symplectic structure. In good cases, there is an associated *moment map*

$$\mathfrak{m} : \mathbf{X} \rightarrow \mathcal{L}^*,$$

where $\mathcal{L}^* = \mathrm{Hom}_{\mathbb{R}}(\mathcal{L}, \mathbb{R})$ is the dual vector space to the Lie algebra $\mathcal{L} = \mathcal{L}(\mathbf{K})$, considered as a real vector space. (Compare [GGK].) This map \mathfrak{m} has two important properties: The given action of \mathbf{K} on \mathbf{X} corresponds to the adjoint action of \mathbf{K} on \mathcal{L}^* . Furthermore, for each vector $\mathbf{v} \in \mathcal{L}$, if we think of the map $\mathbf{x} \mapsto \mathfrak{m}(\mathbf{x})(\mathbf{v})$ from \mathbf{X} to \mathbb{R} as a Hamiltonian function, then the solution curves for the associated Hamiltonian differential equation on \mathbf{X} are just the orbits $t \mapsto \exp(t\mathbf{v})(\mathbf{x})$ under the one parameter subgroup $t \mapsto \exp(t\mathbf{v})$ of \mathbf{K} which is generated by \mathbf{v} . A point $\mathbf{x} \in \mathbf{X}$ is called *stable*, with respect to this choice of \mathfrak{m} , if the stabilizer $\mathbf{G}_{\mathbf{x}}$ is finite, and if the intersection of the set $\mathfrak{m}^{-1}(\mathbf{0})$ with the \mathbf{G} orbit of \mathbf{x} is non-empty.

In the case of interest, with $\mathbf{G} = \mathrm{PGL}_k(\mathbb{C})$ with $k \geq 2$, there is a unique choice of the moment map \mathfrak{m} ; hence the open set \mathbf{X}^s consisting of stable points is also uniquely defined. Furthermore, the quotient space \mathbf{X}^s/\mathbf{G} is a well defined orbifold with a Hausdorff topology.

In §6 and §7 we describe open subsets of \mathbf{X} with a Hausdorff orbifold quotient. Perhaps these are contained in Mumford’s set of stable points; but we don’t have a proof.

REFERENCES

- AF1. P. Aluffi and C. Faber, *Plane curves with small linear orbits I and II*, Ann. Inst. Fourier (Grenoble) **50** (2000) 151–196, and Internat. J. Math. **11** (2000) 591–608.
- AF2. P. Aluffi and C. Faber, *Limits of $\mathrm{PGL}(3)$ -translates of plane curves I*, J. Pure Appl. Algebra **214** (2010) 526–547.
- ACGH. E. Arbarello, M. Cornabla, P. Griffiths, and J. Harris, “Geometry of Algebraic Curves I”, Springer 1985.
- Ar1. M. Arfeux, “Dynamique holomorphe et arbres des sphères”, (Thesis) Université Toulouse III (Paul Sabatier) 2013.
- Ar2. M. Arfeux, *Berkovich spaces and Deligne-Mumford compactification*, ArXiv:1506.02552, 2017.
- Arn. V. I. Arnold, *On the arrangement of ovals of real algebraic plane curves, involutions on four-dimensional smooth manifolds, and the arithmetic of integral quadratic forms*, Functional Anal. Appl. **5** (1971) 169–176.
- Ba. W. Baily, *On the automorphism group of a generic curve of genus > 2* , J. Math. Kyoto Univ. **1** (1961/1962), 101–108; correction, 325.

- Bars. F. Bars, *Automorphisms groups of genus 3 curves*, Notes del Seminari de Teoria Nombres UB-UAB-UPC 2004/05 <http://mat.uab.es/francesc/>
- Bec. A. Becker. “What is Real”, Basic Books, 2018.
- Be. A. Bertram, *Complex Algebraic Geometry: Smooth Curves*, Lecture Notes, 2010.
<http://www.math.utah.edu/~bertram/6030/12Classification.pdf>
- BMac. G. Birkhoff and S MacLane, “A Survey of Modern Algebra”, Macmillan 1953.
- BCR. J. Bochnak, M. Coste and M.-F. Roy, “Real Algebraic Geometry”, Springer 1998.
- BBK. C. Böhning, H.-C. Graf v. Bothmer and J. Kröker, *Rationality of moduli spaces of plane curves of small degree*, Exp. Math. **18** (2009) 499–508.
- BMP. M. Boileau, S. Maillot and J. Porti, “Three-dimensional Orbifolds and Their Geometric Structures”, Société Mathématique De France, Paris, 2003.
- BM. A. Bonifant and J. Milnor, *On Real and Complex Cubic Curves*, J. L’Enseignement Mathématique (2) **63**, (2017), 21–61. DOI: 10.4171/LEM/63-1/2-2.
- Br. M. Brion, *Introduction to actions of algebraic groups*,
https://www-fourier.ujf-grenoble.fr/~mbrion/notes_luminy.pdf
- Ca. J. W. Cannon, *Shrinking cell-like decompositions of manifolds. Codimension three*, Annals of Mathematics. Second Series (1) **110** (1979) 83–112.
- Ch. H. C. Chang, *On plane algebraic curves*, Chinese Journal of Mathematics **6** (1978) 185-189. (Available through JSTOR)
- Co. M. Conder, *Hurwitz groups: A brief survey*, Bulletin (New Series) of the American Mathematical Society, **23** (1990) 359–370.
- DFL. I. Dolgachev, B. Farb, E. Looijenga, *Geometry of the Wiman-Edge pencil, I: algebro-geometric aspects*, Eur. J. Math. **4** (2018) 879–930.
- DK. J. Duistermaat and J. Kolk, “Lie Groups”, Springer 1999.
- Ed. R. D. Edwards, *Suspensions of homology spheres*, (2006) arXiv:math/0610573 [math.GT] (Reprint of private, unpublished manuscripts from the 1970’s.)
- EHKR. P. Etingof, A. Henriques, J. Kamnitzer and E. Rains. *The cohomology ring of the real locus of the moduli space of stable curves of genus 0 with marked points*. Annals of Math. **171** (2010) 731–777.
- Fu. W. Fulton. “Intersection Theory”, Springer, 1998.
- FMCP. W. Fulton and R. MacPherson, *A compactification of configuration spaces*, Ann. of Math. **139** (1994) 183–225.
- GRS. V. Georgoulas, J. Robbin, D. Salamon, “The moment-weight inequality and the Hilbert-Mumford criterion,” (2018).
<https://people.math.ethz.ch/~salamon/PREPRINTS/momentweight-book.pdf>
- Ghi. A. Ghizzetti, *Sulle curve limiti di un sistema continuo ∞^1 di curve piane omografiche*, Memorie R. Accad. Sci. Torino, **68** (1936) 123–141.
- Ghy. É. Ghys, “A Singular Mathematical Promenade”, Lyon, ENS éditions, 2017.
- Gor. E. Z. Goren, “Algebraic Groups: Part II” McGill University 2011.
<http://www.math.mcgill.ca/goren/MATH722.2010/AlgebraicGroups.2.pdf>
- Gra. J. Gray, “Henri Poincaré: A Scientific Biography”, Princeton U. Press 2012.
- Gr. P. Griffiths, “Introduction to Algebraic Curves”, A.M.S 1989.
- GH. P. Griffiths and J. Harris, “Principles of Algebraic Geometry”, Wiley & Sons Inc. 1978.
- GGK. V. Guillemin, V. Ginzburg, and Y. Karshon, “Moment Maps, Cobordisms, and Hamiltonian Group Actions”, A.M.S. 2002. (See Appendix B.)
- Gu. D. A. Gudkov, *Construction of a new series of M-curves*, Soviet Math. Dokl. **12** (1971) 1559–1563.
- Hac. P. Hacking, *Compact moduli of plane curves*, Duke Math. J. **124** (2004) 213–257.
- HL. I. Hambleton and R. Lee, *Finite group actions on $P^2(\mathbb{C})$* , J. Algebra **116** (1988) 227–242.
- Harn. A. Harnack, *Über Vieltheiligkeit der ebenen algebraischen Curven*, Math. Ann. **10** (1876) 189–199.
- Harr. J. Harris, “Algebraic Geometry, a first course”, Springer 1992.
- Hart. R. Hartshorne, “Algebraic Geometry”, Springer 1977.
- HKKO. T. Harui, T. Kato, J. Komeda and A. Ohbuchi, *Quotient Curves of Smooth Plane Curves with Automorphisms*, Kodai Math. J. **33** (2010) 164–172.
- Harui. T. Harui, *Automorphism groups of smooth plane curves*, Kodai Math. J. **42** (2019) 308331.
- HMO. T. Harui, K. Miura, A. Ohbuchi, *Automorphism group of plane curve computed by Galois points, II*, Proc. Japan Acad. Ser. A Math. Sci. **94** (2018) 59–63.
- Hi1. D. Hilbert, *Über die reellen Züge algebraischen Curven*, Mathematische Annalen **38** (1891) 115–138.

- Hi2. D. Hilbert, *Mathematical Problems*, Proc. Symposia in Pure Math. **28** A.M.S 1976, 1–34 (translated from Göttinger Nachr. 1900, 253–297).
- HNN. G. Higman, B.H. Neumann and H. Neumann, *Embedding theorems for groups*. J. London Math. Soc. **24** (1949) 247–254.
- Hub. J. H. Hubbard, “Teichmüller Theory, 1”, Matrix Editions 2006.
- Hur. A. Hurwitz, *Über algebraische Gebilde mit Eindeutigen Transformationen in sich*, Mathematische Annalen, **41** (1893) 403–442.
- Il. Y. Iyashenko, *Centennial History of Hilbert’s 16th Problem*, Bull. Amer. Math. Soc. **39** (2002) 301–354.
- JS. W. Jaco and P. B. Shalen, *A new decomposition theorem for irreducible sufficiently-large 3-manifolds*. Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2, pp. 71–84, Proc. Sympos. Pure Math., XXXII, Amer. Math. Soc., Providence, R.I., 1978.
- J. K. Johannson, “Homotopy equivalences of 3-manifolds with boundaries”, Lecture Notes in Mathematics, **761**. Springer, Berlin, 1979.
- KKPSS. N. Kaihnsa, M. Kummer, D. Plaumann, M. Sayyary Namin, and B. Sturmfels, *Sixty-Four Curves of Degree Six*, Experimental Mathematics, 2017. DOI: 10.1080/10586458.2017.1360808
- Ke. S. Keel, *Intersection theory of moduli spaces of n -pointed curves of genus zero*, Trans. Amer. Math. Soc. **330** (1992) 545–574.
- KN. G. Kempf and L. Ness. *The length of vectors in representation spaces*, Springer Lecture Notes Math. **732** (1979).
- Kh. V. M. Kharlamov, *Topological types of non-singular surfaces of degree 4 in $\mathbb{R}P^3$* , Functional Anal. Appl. **10** (1976) 295–305.
- Kir. F. Kirwan, “Complex Algebraic Curves”, Cambridge U. Press 1992.
- K. F. Klein, *Über eine neue Art von Riemannschen Flächen*, Math. Ann. **10** (1876) 398–416.
- KL. F. Klein and S. Lie, *Ueber diejenigen ebenen Curven, welche durch ein geschlossenes System von einfach unendlich vielen vertauschbaren linearen Transformationen in sich übergehen*, Math. Ann. **4** (1871) 50–84.
- Knu. F. F. Knudsen, *The projectivity of moduli spaces of stable curves II*, Funct. Anal. Appl. **19** (1983) 161–199.
- Kun. E. Kunz, “Introduction to Plane Algebraic Curves”, Birkhäuser, 2005.
- L. F. Lemmermeyer, “Reciprocity Laws: From Euler to Eisenstein”, Springer Monographs in Mathematics, Springer-Verlag Berlin Heidelberg 2000.
- Li. E. L. Lima, “Fundamental Groups and Covering Spaces”. A K Peters, Ltd. Natick, MA. 2003.
- MS1. D. McDuff and D. Salamon, “J-holomorphic Curves and Symplectic Topology.” Second Edition, A.M.S. 2012. (See Appendix D.)
- MS2. D. McDuff and D. Salamon, “Introduction to Symplectic Topology,” Third Edition, Oxford U. Press 2017. (See Section 5.7.)
- Mei. E. Meinrenken, *Group actions on manifolds*, Lecture Notes, U. Toronto 2003. (<http://www.math.toronto.edu/mein/teaching/LectureNotes/action.pdf>)
- MBD. G. Miller, H. Blichfeld, and L. Dickson, “Theory of Finite Groups”, Wiley, N.Y. 1916.
- Mi1. J. Milnor, “Singular Points of Complex Hypersurfaces”, Princeton U. Press 1968.
- Mi2. J. Milnor, “Dynamics in One Complex Variable”, Annals of Mathematics Studies **160**, Third Edition, Princeton University Press 2006.
- Mu. D. Mumford, “Geometric Invariant Theory”, Springer-Verlag 1965; or 3rd edition with Fogarty and Kirwan, 1994.
- Mun. J. R. Munkres, “Topology”, Second Edition, Prentice Hall, Inc., 2000.
- Nam. M. Namba, “Geometry of Projective Algebraic Curves”, Marcel Dekker, 1984.
- Ne. L. Ness, *Mumford’s numerical function and stable projective hypersurfaces*, Springer Lecture Notes **732** (1978).
- New. P. E. Newstead, “Introduction to Moduli Problems and Orbit Spaces”. Narosa Publishing House, available via the A.M.S. 1978, (2012).
- Ni. V. V. Nikulin, *Integer symmetric bilinear forms and some of their geometric applications*, USSR-Izv. **14** (1980) 103–167.
- Pe. I. Petrovsky, *On the topology of real plane algebraic curves*, Ann. of Math. **39** (1938) 189–209.
- Po. B. Poonen, *Varieties without extra automorphisms. I. Curves*, Math. Res. Letters **7** (2000) 77–82.

- Pop. V. L. Popov, *Algebraic curves with an infinite automorphism group*, Math. Notes **23** (1978) 102–108.
- Roh. K. Rohn, *Die Maximalzahl und Anordnung der Ovale bei der ebenen Kurve 6. Ordnung und bei der Fläche 4. Ordnung*, Mathematische Annalen **73** (1913) 177–229.
- R1. V. A. Rokhlin, *Congruences modulo 16 in Hilbert's sixteenth problem. I, II*, Functional Analysis and Its Applications, **6** (1972) 301–306. Part II, Functional Analysis and Its Applications, **7** (1973) 163–164.
- R2. V. A. Rokhlin, *Complex orientations of real algebraic curves*, Funct. Anal. Its Appl. **8**:331 (1974). <https://doi.org/10.1007/BF01075491>
- R3. V. A. Rokhlin, *Complex topological characteristics of real algebraic curves*, Russian Math. Surveys **33** (1978) 85–98.
- Sa1. I. Satake, *On a generalization of the notion of manifold*, Proc. Nat. Acad. Sci. USA **42** (1956) 359–363.
- Sa2. I. Satake, *The Gauss-Bonnet theorem for V-manifolds*, J. Math. Soc. Japan **9** 1957 464–492.
- Sea. J. Seade, *On Milnor's fibration theorem and its offspring after 50 years*, Bull. Amer. Math. Soc. **56** (2019) 281–348. DOI: <https://doi.org/10.1090/bull/1654>
- Ser1. J.-P. Serre, “Algebraic Groups and Class Fields”, Springer 1988.
- Ser2. J.-P. Serre, *Arbres, amalgams, SL_2* , Astérisque, **46** (1977).
- Sha. I. Shafarevich, “Basic Algebraic Geometry 1 (2nd edition)”, Springer 1994.
- S-B. N. I. Shepherd-Barron, *The rationality of some moduli spaces of plane curves*, Compositio Math. **67** (1988) 51–88.
- Sim. J. Simental, *Introduction to Geometric Invariant Theory*, mathserver.neu.edu/jose/GIT.pdf.
- Sp. E. H. Spanier, “Algebraic Topology”, Springer, 1981.
- Th. W. Thurston, *Geometry and topology of three-manifolds*, version 1 of March 2002. <http://library.msri.org/books/gt3m/PDF/13.pdf>
- V1. O. Viro, *Progress in the topology of real algebraic varieties over the last six years*, Russ. Math. Surveys (Uspeki), **41** (1986) 45–67.
- V2. O. Viro, *Real algebraic plane curves: Constructions with controlled topology*, Leningrad Math. J. **1**:5 (1990) 1059–1134.
- V3. O. Viro, *From the sixteenth Hilbert Problem to tropical geometry*, Japan. J. Math. **3** (2008) 185–214.
- Wa. C. T. C. Wall, “Singular Points of Plane Curves”. London Math. Soc. 2004.
- W. G. Wilson, *Hilbert's sixteenth problem*, Topology **17** (1978) 53–73.
- Wi. R. A. Wilson, *The monster is a Hurwitz group*, J. Group Theory **4** (2001) 367–374.
- Wim. A. Wiman, *Ueber eine einfache Gruppe von 360 ebenen Collineationen*, Math. Ann. **47** (1896) 531–556.
- Z. H. G. Zeuthen, *Almindelige Egenskaber ved Systemer af plane Kurver*, Kongelige Danske Videnskaberne Selskabs Skrifter – Naturvidenskabelig og Matematisk, **10** (1873) 285–393. Danish with French summary.

MATHEMATICS DEPARTMENT, UNIVERSITY OF RHODE ISLAND

Email address: bonifant@uri.edu

Current address: Institute for Mathematical Sciences, Stony Brook University

INSTITUTE FOR MATHEMATICAL SCIENCES, STONY BROOK UNIVERSITY

Email address: jack@math.stonybrook.edu