PROGRESS IN THE THEORY OF COMPLEX ALGEBRAIC CURVES

DAVID EISENbud AND JOE HARRIS

Abstract. We describe the contemporary view of the theory of algebraic curves over the complex numbers, with emphasis on the moduli spaces of curves and linear series on them. We then give an exposition of some of the recent work on the question of the rationality of the moduli space.

Introduction. In the last twenty years there has been a major development in our understanding of algebraic curves. A number of the classical problems have been solved and new directions of investigation have been begun. We will describe some of the history of the theory, and how it led to the modern point of view, and we will sketch proofs of many of the main assertions. Then we will explain something of how the modern ideas have been used to solve some old problems.

Many features of the current wave of progress are closely connected with patterns that go back to the earliest period in the development of the theory, so it's best to start with ancient history. To talk of complex projective curves, you need the complex numbers and you need projective space, so “ancient history” for us will start when these things become available, between about 1800 (Gauss' proof of the fundamental theorem of algebra) and 1830, (the introduction by Plücker of homogeneous coordinates for the projective plane). It goes without saying that the history below is that of a Mathematician and not of an Historian—it should probably be described as “fictionalized.”

Naturally we will have to leave out parts of the theory of curves at least as rich as the parts we can put in. We beg pardon in advance from anyone whose favorite bit we’ve skipped. A more detailed development along the lines of this article may be found in lectures from the Bowdoin conference (Harris [1988]). A very beautiful survey covering a different range of topics is that of Mumford [1976].

A curious aspect of the history of algebraic curves, as with other parts of algebraic geometry, is the breakdown of rigor which deeply affected the theory, leading to its virtual stagnation in a mire of unproved assertions and incomplete proofs in the end of the first third of this century. The late nineteenth and early twentieth centuries were of course a period of enormous vitality in algebraic geometry, a period in which a large fraction of our current knowledge of curves and surfaces was obtained. In some cases the low rigor was simply the result of the fact that people lacked a
good language for expressing their new discoveries, or the fact that they expressed as theorems what might now be called "certain true conjectures," in the fashion of modern number theory. But in shockingly many instances, we see now that the "proofs" were based on the then traditional, but totally inadmissible, lack of concern with certain "exceptional" cases—cases which were in fact not exceptional at all, but were crucial to the truth of the main assertions even in the best and most centrally important circumstances. A number of the assertions whose original proofs were thus tainted will appear in their modern guise in our story.

It is interesting that mathematicians have in general handled the circumstances of these wrong and incomplete proofs rather charitably. The phrase "X's Theorem," when it is finally noticed that X gave no proof, usually becomes "X's Conjecture," and turns back into "X's Theorem," or sometimes "X-Y's Theorem" if and when Y finds a proof. (Perhaps a new term for this class of mathematical utterances should be coined. Why not speak of "X's Folly"—not in the pejorative sense of the word folly, but rather in the sense of enlightened madness?)

In any case, perhaps partly encouraged by the successful reexamination of general position and transversality arguments in topology and differential geometry, these ideas have been broadly studied by algebraic geometers, and many of the applications to the theory of curves have been put right. It is one stream of the work in this direction, and the new results it has brought, to which we will address ourselves in the second part of this article. To prepare the ground, the first part will be taken up with a sketch of some of the fundamental ideas and results of the theory of curves.

There are, as we indicated, many recent developments of great interest which we will not discuss below. We will at least give references to a few of these, which include the theories of nodal plane curves (Harris [1986]), of curves in $\mathbb{P}^3$ (Gruson-Peskine [1978, 1982], see also Hartshorne [1982]), and of Weierstrass points (see for example Eisenbud-Harris [1987] and Morrison-Pinkham [1986], where there are also many further references); and the Schottky problem (Arbarello, DeConcini, Donagi, Fay, Gunning, Shiota, Welters,...—see for example the survey [1988] of Arbarello).

We are grateful to Scott Wolpert who read a preliminary version of this report, helped us with our history, and suggested many useful references.

1. Basic notions of the theory of curves.

In the beginning... curves were in the plane; that is, a curve was an equation in two variables (or a homogeneous equation in 3 variables, which defines a curve in $\mathbb{P}^2 = \{(x : y : z) = (x, y, z)/\text{scalars}\}$, a compactification of $\mathbb{C}^2$. Two features of this "world-view," in particular, had very significant consequences for the future development of the theory:

Families of curves. It was natural from the first to consider families of curves: having written an equation, why not let the coefficients vary? At least in the homogeneous case, this variation does not change the degree of the equation. Accordingly, nothing could have been more natural than to study the family of all curves of given degree as itself an algebraic variety—it is just the projective space of homogeneous equations of that
degree. (The use of projective space here is natural since two equations that differ by a scalar vanish on the same points and thus define the same curve.)

**Singular curves.** The family of plane curves of a given degree naturally included singular curves, so that even if these did not have precisely equal rights, they were at least present to be reckoned with. As one followed paths around in the family, say

\[ y^2 = x(x - 1)(x - t) \]

as \( t \) varies, say \( t \) goes to 0, one saw nonsingular curves "degenerating" into a singular curve—in this case a "nodal" curve. Indeed, many degenerate forms arise, as indicated in the diagram

**Figure 1**
Even "reducible" curves (that is, curves that may be described as the union of several curves) and "nonreduced" curves are included here. One may think of a nonreduced curve as a curve "with multiplicity," such as for example the "double line" which is the limit of the family of ellipses \( x^2 + t \cdot y^2 = t \) as \( t \) approaches 0:

![Figure 2]

**Figure 2**

**Abstract curves.** Curves were finally freed from the plane about 1860 through the work of Riemann (their liberation was not universally recognized until at least the time of Weyl, however). At first curves appeared in the new "Function Theory" as objects which are (ramified) coverings of \( \mathbb{P}^1 \), renamed the "Riemann Sphere." This was perhaps not so different from considering them as objects with embeddings into \( \mathbb{P}^2 \)—a link to a projective space still existed—but the bond to a particular projective space had been shaken.

This can be seen as one instance of a global trend in mathematics, in which objects that had initially been defined as subsets of a standard object were replaced by abstract versions. For example, in differential geometry a "manifold" had been a subset of \( \mathbb{R}^n \) defined locally by \( \mathbb{C}^\infty \) functions with independent differentials; now it became a set with the additional structure of a topology, coordinate charts and possibly a Riemannian metric. Similarly, in group theory a "group" had been a subset of \( GL_n \) or \( \mathfrak{S}_n \) closed under inverse and multiplication; it became a set with binary operation satisfying the now-standard axioms. Analogously, through the contributions of Poincaré, Klein, ..., Weyl, from the 1880s to around 1920, (and of course very much through the rise of topology), curves finally shook off the need for attachment to a projective space. The term "curve," by which had previously meant a subset of projective space (and most commonly the loci of single polynomials in the projective plane), now came to mean a compact complex manifold of dimension 1.

But how is one to study an object that simply floats by itself? A curve in \( \mathbb{P}^2 \) has many "features;" for example, points of inflection, or points where the tangent line is again tangent at some other point of the curve, to name just two constructs that were important in the early theory. Similarly, if one studies a curve as a branched cover of \( \mathbb{P}^1 \), the ramification points of the covering provide convenient handholds for study of the curve. But an abstract curve seems at first rather featureless.

Another deficiency of the abstract viewpoint is that it is at first difficult to see how to define families and degenerations of abstract curves. The very first problem was to decide which curves should fit together in a family. Whereas the coefficients of a plane curve can be varied, giving
the family of curves of given degree, or the branch points of a ramified
cover can be varied, giving a family of curves admitting maps of a certain
degree to $\mathbb{P}^1$, the “obvious” invariant in the new point of view is rather its
topological structure, so one should expect to see families of curves with
given topology. A complex manifold is automatically orientable, so a com-
 pact complex 1-manifold is a real compact orientable surface—a sphere
to which a certain number of “handles” have been added, and as such it is
topologically classified by the number of handles, called its genus. How
should one think of varying the complex structure on a given topological
surface, and how would one ever know, for example, if that variation were
“algebraic?”

The repair of these deficiencies was of course among the earliest projects
of the developers of the new views, but it took a long time to achieve
satisfactory results. We will describe the modern view.

Maps to projective space, and linear series. To recover the rich-
ness of a projective embedding for an abstract curve, we must understand
the data of such an embedding. It cannot be defined by functions in the
usual sense, since any holomorphic function on a compact space must be
constant by the maximum modulus principle. Functions would be (global)
sections of the trivial line bundle. The way out of the problem of nonex-
istence of functions is to look for sections of a nontrivial line bundle: if
$s_0, \ldots, s_r$ are sections of a line bundle $L$ on a curve $C$, and if, say, $s_0$
does not vanish at a point $p$ of $C$, then the ratios $s_i(p)/s_0(p)$ are well-defined
complex numbers, so we may define a map, from at least that part of $C$
where not all the $s_j$ vanish, to $\mathbb{P}^r$ by sending $p$ to $(s_0(p) : \cdots : s_r(p))$. (From
a higher-brow point of view, $\mathbb{P}^r$ is the classifying space for line bundles
generated by $r + 1$ global sections.) To get a well-defined map to projec-
tive space one needs a collection of sections $s_i$ that do not all vanish at
any point of $C$. Note that if we replace the $s_i$ by a different collection
of sections generating the same vector space of sections, then we simply
compose the corresponding map with an automorphism of $\mathbb{P}^r$, so we may
replace the collection of sections $s_0, \ldots, s_r$ by the vector space of sections
that they generate.

Now it is rather a nuisance to be always checking whether there is a
point at which all elements of a certain collection of sections of a given
line bundle vanish, and for many purposes it is unnecessary, so we will drop
the condition and define a linear series on $C$ to be a pair $(L, V)$ consisting
of a line bundle $L$ and a space $V$ of global sections of $L$. If $(L, V)$ actually
corresponds to an embedding, we say that it is very ample. If $L$ together
with the vector space, written $H^0(L)$, of all its sections corresponds to an
embedding, then $L$ is called very ample.

We can immediately associate two characteristic numbers to a linear
series $(L, V)$: First, all the sections of a line bundle $L$ on a curve $C$ vanish
at the same number of points (counted with multiplicities) of $C$, and this
number is called the degree of the line bundle, or linear series. (It is in
fact a complete topological invariant for the line bundle, and generalizes
the degree of a plane curve, as the terminology suggests.) Second, we have
the dimension: if the vector space dimension of $V$ is $r + 1$, we speak of a
linear series of dimension \( r \), since it may correspond to a map to \( \mathbb{P}^r \). In the classical terminology, a linear series of dimension \( r \) and degree \( d \) is called a \( g^r_d \), a “group” of \( d \) points “moving” in an \( r \)-dimensional family. This view comes from the identification of a section of \( L \) (up to scalars) with the set on which it vanishes (with multiplicities). Alternately, thinking of \( L, V \) as corresponding to a map to \( \mathbb{P}^r \), we may think of a linear series as being the family of sets in which the curve meets hyperplanes in \( \mathbb{P}^r \).

**What linear series are there?** One could interpret this question as a question about an arbitrary but fixed curve, or as a question about how to choose a linear series for each curve simultaneously in some uniform way. We begin with the second interpretation.

Surprisingly, there is an essentially unique answer. This is essentially the content of Franchetta’s theorem (or “folly,” in the language of the introduction), which was recently proved through work of Mumford, Arbarello, and Cornalba [1987], using the topological work of Harer [1983], and refined by Mestrano [1987]. It says roughly that the only line bundles that can be chosen uniformly are the (positive and negative) tensor powers of the complex tangent bundle. The tangent bundle and its positive powers are not so interesting from the point of view of linear series, since for curves of genus \( g > 1 \) they have no sections! But the dual of the tangent bundle, the cotangent bundle, which we will write as \( \omega_C \), always has \( g \) sections. This bundle plays such a central role in the theory of curves that it is called the **canonical bundle** of \( C \). One of the basic theorems in the subject says that for \( g \geq 1 \) these sections do not vanish simultaneously anywhere on the curve, so that the “canonical” linear series consisting of the canonical bundle together with all its sections always gives a map of the curve to \( \mathbb{P}^{g-1} \). As soon as \( g \geq 3 \), there is an open dense set of curves for which this “canonical map” is an embedding.

The study of the ideal defining the curve in the canonical embedding is an important part of the theory of curves that was begun (with partially incomplete proofs and statements that are sometimes simply wrong) by Petri in the early part of this century, and brought up to date by St.-Donat [1973] and others. It continues to be an active topic of research; for example, it is believed that the minimal free resolution of this ideal reflects rather detailed information about what other interesting linear series the curve possesses (Green-Lazarsfeld [1986] and Schreyer [1986] and [1988]).

We now turn to the linear series on an individual curve. We will first need to know about the line bundles on the curve, then about the space of global sections of a line bundle.

**Line bundles.** An algebraic line bundle \( L \) on a curve \( C \) is by definition a map of spaces \( L \to C \) whose fibers are all isomorphic to the complex line, \( \mathbb{C} \), and so that there is a cover \( C = \bigcup U_i \) of \( C \) by open sets such that the restriction of \( L \) to each set \( U_i \) is isomorphic to the trivial bundle \( \mathbb{C} \times U_i \), the isomorphism being linear on the fibers. It follows that the identification of \( L|_{U_i} \cong \mathbb{C} \times U_i \) with \( L|_{U_j} \cong \mathbb{C} \times U_j \) over \( U_i \cap U_j \) is by means of a function \( f_{ij} \) from \( U_i \cap U_j \) to the set of automorphisms of a 1-dimensional vector space, that is, to \( \mathbb{C}^* \). If we wish to define algebraic bundles, then we should properly deal with “algebraic open sets” and polynomial functions;
but in fact, since we are dealing only with projective curves, algebraic and analytic line bundles are the same, and we will work here in the somewhat simpler analytic category. Such a collection of analytic functions \( f_{ij} \) may be regarded as a Čech 1-cocycle with values in the sheaf \( \mathcal{O}^*_{\text{an}} \) of nowhere vanishing analytic functions on \( C \), and we associate to \( L \) the corresponding cohomology class, in \( H^1(C, \mathcal{O}^*_{\text{an}}) \). It is not hard to show that in this way the group of all line bundles (the group operation being tensor product), called the Picard group of \( C \), and written \( \text{Pic} C \), is actually isomorphic to \( H^1(C, \mathcal{O}^*_{\text{an}}) \).

This is useful because of the ease with which we can compute \( H^1(C, \mathcal{O}^*_{\text{an}}) \). If \( f \) is a function on some connected open set \( U \) of \( C \), then \( \exp f \), defined by \( (\exp f)(z) = e^{(2\pi i)f(z)} \), say, is a nowhere vanishing function on \( U \), identically equal to 1 iff \( f \) is a constant function equal to an integer. Conversely, if \( U \) is small enough to be simply connected, then we may take the logarithm of a nowhere vanishing function on \( U \), so that the “exponential sequence”

\[
0 \to \mathbb{Z} \to \mathcal{O}_{\text{an}} \to \mathcal{O}^*_{\text{an}} \to 0,
\]

where we have written \( \mathcal{O}_{\text{an}} \) for the sheaf of analytic functions on \( C \), is an exact sequence of sheaves on \( C \). Applying cohomology, and using the fact that any globally defined analytic function is a constant by the maximum modulus principle, we get the exact sequence

\[
0 \to H^1(C, \mathbb{Z}) \to H^1(C, \mathcal{O}_{\text{an}}) \to H^1(C, \mathcal{O}^*_{\text{an}}) \to H^2(C, \mathbb{Z}).
\]

The first and last terms in this sequence are isomorphic to the ordinary simplicial or singular cohomology groups of \( C \); since \( C \) is a compact orientable surface of genus \( g \) they are \( \mathbb{Z}^{2g} \) and \( \mathbb{Z} \) respectively. The second term in the sequence is a complex vector space of dimension \( g \) by the Riemann-Roch Theorem, to be discussed below. The third term is by our discussion above the sought for group of line bundles, \( \text{Pic} C \). Furthermore, it is not hard to show that the map in the sequence \( \text{Pic} C \to H^1(C, \mathbb{Z}) \) is given by sending a line bundle to its degree. Since every degree is possible (for example, it is easy to see that if \( p \) is a point of \( C \) then the line bundle whose sections, locally, give the functions vanishing to some given order \( d \) at \( p \) has degree \(-d\)), so the map is onto. Thus for any integer \( d \), the set of line bundles of degree \( d \) on \( C \) is isomorphic to

\[
H^1(C, \mathcal{O}_{\text{an}})/H^1(C, \mathbb{Z}) \cong \mathbb{C}^g/\mathbb{Z}^{2g}.
\]

It turns out that \( \mathbb{Z}^{2g} \) is here embedded in \( \mathbb{C}^g \) as a lattice (called the period lattice of \( C \)—it actually determines \( C \), a result called the Torelli Theorem). Moreover, it turns out that this complex torus \( \mathbb{C}^g/\mathbb{Z}^{2g} \) is actually a projective algebraic variety, called the Jacobian variety of \( C \).

In sum, the family \( \text{Pic}^d C \) of line bundles of degree \( d \) on \( C \) forms a compact, even projective algebraic, variety of dimension equal to the genus of the curve \( C \). In particular, any curve of genus \( \geq 1 \) possesses lots of line bundles of each degree.

Of course we do not immediately get linear series from this—we need to know about the sections of these line bundles as well. This brings us to one of the most useful of all results in algebraic geometry:
Sections: the Riemann-Roch Theorem. The Riemann-Roch Theorem tells us how many sections a line bundle on a curve has. (There are versions for higher dimensional varieties, but they give progressively less complete information as the dimension goes up.) In the case of line bundles whose degree is large compared to the genus of the curve, the answer is in terms of the genus $g$ and the degree $d$ of the bundle—the topology of the situation—alone: it says that there are $d - g + 1$ independent sections. That is, we get a

$$g_d^{d-g}.$$  

With this information it is not difficult to deduce the existence of many embeddings of the curve, and to gain considerable insight into their geometry. (Many mathematicians—including both the authors of this article—hear for the first time about the theory of curves in a course where the Riemann-Roch Theorem is the last thing done. For years afterwards, they wonder what, if anything, the Theorem is really good for. We feel that it should rather be among the first things done in a course on curves—the proof is not really very difficult, and it makes a great deal else possible.)

The Brill-Noether computation. It turns out that in practice one is often interested in linear series of lower degree, where Riemann-Roch tells one less. In general, writing $h^0 L$ for the dimension of the vector space $\mathcal{H}^0(L) = H^0(C, L)$ of all global sections of $L$, and $h^1 L$ for the dimension of the (Čech) cohomology group $\mathcal{H}^1(L) = H^1(C, L)$, the statement of Riemann-Roch is

$$h^0 L = d - g + 1 + h^1 L,$$

and $\mathcal{H}^1(L)$ is isomorphic to the dual of the vector space $H^0(L^{-1} \otimes \omega_C)$ (this is called Serre Duality), so that

$$h^1 L = h^0(L^{-1} \otimes \omega_C).$$  

The reason that the formula simplifies for $d > 2g - 2 = \deg \omega_C$ is that then $L^{-1} \otimes \omega_C$ has negative degree, and thus no sections, so $h^1 L = 0$. But for bundles $L$ of lower degree, $h^1(L^{-1} \otimes \omega_C)$ may be fairly large, giving rise to interesting "special" linear series, and sometimes embeddings. There is a general result that gives us at least a lower bound, sharp for the general curve, on the dimension of the family of these special linear series. This is the Brill-Noether "Theorem" (more properly, "folly"), stated with an incomplete proof by Brill and Noether in the 1880s, and completely proved about 5 years ago. (Brill-Noether [1874], Kempf [1971], Kleiman-Laksov [1972] and [1974], Griffiths-Harris [1980], Gieseker [1982], Eisenbud-Harris [1983], Fulton-Lazarsfeld [1981], Lazarsfeld [1986]). In one of its sharp forms, it says that the family $G_d'(C)$ of $g_d'$'s has dimension

$$\dim G_d'(C) \geq \rho := g - (r + 1)(g - d + r),$$  

and that for most $C$ the inequality is an equality, with $G_d'(C)$ smooth and irreducible.

Though the full proof of the Brill-Noether Theorem remains somewhat complex, the idea behind it is very simple: Most $g_d'$'s in the interesting
ranges of $d$ and $r$ will be obtained by taking all the sections of a line bundle of degree $d$, so that to prove the Brill-Noether Theorem we should compute the codimension, in the $g$-dimensional family of line bundles $\text{Pic}^d(C)$, of the set of those with (at least) $r + 1$ sections, and show that this codimension is $(r + 1)(g - d + r)$. If $L_0$ is a fixed line-bundle of some arbitrary but high degree $e$, and $\mathcal{P} \subset C$ is the set of points at which some section of $L_0$ vanishes, then for all $L$ of degree $d$ we will have an exact sequence

$$0 \to H^0(L) \to H^0(L \otimes L_0) \xrightarrow{\varphi_L} \mathbb{C}^\mathcal{P},$$

where we have identified $H^0(L \otimes L_0)$ with the functions $\mathbb{C}^\mathcal{P}$ from $\mathcal{P}$ to $\mathbb{C}$, and $\varphi_L$ is the restriction map.

Since the degree of $L_0$ was chosen “large,” we may compute the dimension of $H^0(L \otimes L_0)$ by Riemann-Roch; it is $d + e - g + 1$, while the dimension of $\mathbb{C}^\mathcal{P}$, the set of functions from $\mathcal{P}$ to $\mathbb{C}$, is $e$, the number of points in $\mathcal{P}$. Thus the locus of $L$ with $h^0L \geq r + 1$ is the locus, in the family of those $(d + e - g + 1) \times e$ matrices of the form $\varphi_L$, of the matrices with rank at most $d + e - g + 1 - (r + 1)$. (This argument is only “locally” valid, since it is not possible to choose a family of bases for the $H^0(L \otimes L_0)$ which varies analytically with $L$; these spaces are the fibers of a nontrivial, but topologically well-understood vector bundle.)

Now the locus of $a \times b$ matrices of rank $\leq k$ is well known to have codimension $(a - k)(b - k)$ in the space of all $a \times b$ matrices. After a little arithmetic one sees that if the family of matrices $\varphi_L$ met the locus of matrices of rank $\leq d + e - g + 1 - (r + 1)$ transversely, then the dimension statement (and in fact all) of the Brill-Noether theorem would follow. This remark, together with some examples with low genus, is essentially all the proof given by Brill and Noether; it actually only proves that any (nonempty) component of $G^r_d(C)$ will have at least dimension $p$, and it does nothing to establish the existence of any. (These deficiencies, along with a possible program for correcting them, were first pointed out by Severi in the 1920s, after the Theorem had been accepted for 40 years!)

We have now touched on the fundamentals of the modern approach to the “projective geometry” of an abstract curve. We will next turn to the notions that are necessary to replace the family of plane curves, and of degenerations in that family.

Families of curves and degenerations: Spaces of moduli and their compactification. First, what is a family of abstract curves, and when is it algebraic? It turns out that all the constructions one can think of fit neatly into a single general framework: A family of smooth projective curves is simply a map $f: X \to B$ whose fibers are smooth curves. The family deserves to be called algebraic (or analytic) if $f$, $X$ and $B$ are all algebraic (analytic). If $X$ and $B$ are smooth, a case sufficient for most of our needs, no more conditions on $f$ are necessary; in the most general case, and in particular if the fibers are singular, one wants to require a weak local triviality condition on the family—“flatness” is the one most often chosen, since it allows many desirable families and is easy to manipulate.
It may take a little experience to realize that one’s favorite families of curves fit this pattern. For example, to make the family of all plane curves of a given degree into a map as above, we let $B$ be the projective space of all homogeneous forms in 3 variables of degree $d$—that is, the projective space of all equations of plane curves of degree $d$. Then let $X \subset \mathbb{P}^2 \times B$ be the set of all pairs $(p, F)$ such that $F(p) = 0$, and let $f: X \to B$ be the projection on the second factor. The fiber over a point $F \in B$ is the curve defined by $F$.

This example is not just “some” family of plane curves of degree $d$. It is in an obvious sense the universal family—every other family $f': X' \to B'$ of plane curves may be deduced from it by “pulling back” this family $f: X \to B$ along some map $\varphi: B' \to B$. Thus, in principle, we can study all possible families of plane curves by studying this one.

It is the urge to have a similar object to help in studying families of abstract curves that leads to the search for and study of the moduli space of curves. It is not hard to show that families of smooth curves defined as above, with $B$ connected, will consist entirely of curves of a single genus (indeed, the family will be locally $C^\infty$ trivial over $B$), so the most “universal” family that one could hope to find would be a universal family consisting of all curves of a given genus $g$. The base space $\mathcal{M}_g$ of such a family would be called the moduli space of curves of genus $g$. (The word moduli, hoary with age, denoted originally a collection of complex numbers that determine the isomorphism class of a curve.)

Already Riemann was interested in families of curves of given genus, and he even calculated the dimension that such a family could have. Perhaps on the basis of the example of curves of genus 1 (explained below), which was well understood, his contemporaries spoke of at least a space whose points would correspond to the isomorphism class of curves. However (as Brill and Noether point out, apparently to justify their hero Riemann before the skeptics) Riemann, at least, was careful not to assert the existence of a universal family, or even a space whose points corresponded to the isomorphism classes of all curves.

It is a sad truth that no universal family of smooth projective curves of genus $g$ exists. But the space whose points correspond to the isomorphism classes of all smooth curves of genus $g$—the space which “would be the base space of the universal family if that family existed”—does exist! (This shows, perhaps, that we have learned something in the 120 years since Riemann.) The peculiar phrase in quotation marks can even be turned into a valid definition:

**Definition of $\mathcal{M}_g$.** If a universal family of smooth curves existed, its base space $\mathcal{M}_g$ would have points in one to one correspondence with the isomorphism classes of such curves; this follows because the “family” of curves consisting of a single curve, with base space a single point, would have to be pulled back along a unique map of the point into $\mathcal{M}_g$. Thus the base space certainly exists, as a set—it is just the set of isomorphism classes of curves of the appropriate sort. Further, if there were a universal family of curves over $\mathcal{M}_g$, we could deduce the structure of $\mathcal{M}_g$ as a variety from its universal property: given any family $f: X \to B$ of smooth curves
of genus $g$, the set-theoretic map $\varphi_f: B \to \mathcal{M}_g$ sending a point $b \in B$ to the isomorphism class of the fiber $f^{-1}(b)$ would be a map of varieties—indeed, $\varphi_f$ would be the map of varieties necessary to obtain $f$ as a pullback of the universal family. The existence of such a map $\varphi_f$ for every family $f$, and the fact that if all these maps factor through some map of varieties $\psi: \mathcal{M} \to \mathcal{M}_g$, then $\psi$ is an isomorphism, are properties which define (at most!) a unique structure of algebraic variety $\mathcal{M}_g$ even without the existence of a universal family.

Something approximating the universal family of curves may be defined in a similar way: we can look at the set of isomorphism classes of pairs $(C, p)$ where $C$ is a smooth curve of genus $g$ and $p$ is a point of $C$, and define on it a structure of algebraic variety by a definition similar to the above. The resulting space (which really does exist) is called $\mathcal{C}_g$. There is a natural map $\pi: \mathcal{C}_g \to \mathcal{M}_g$ obtained by “forgetting” $p$, and the fiber of this map over a point corresponding to the isomorphism class of a curve $C$ is clearly $C/\text{Aut}(C)$. Since for $g \geq 2$ the automorphism group of any curve is finite, and for $g > 2$ most curves have no automorphisms at all, $\mathcal{C}_g$ will be the desired universal family of curves at least over an open subset of $\mathcal{M}_g$.

Of course these characterizations do nothing for the question of whether varieties with the properties ascribed to $\mathcal{M}_g$ and $\mathcal{C}_g$ exist. In fact, if the problem is changed slightly, say by including all reduced and irreducible curves instead of just the smooth ones, then again no such nice structure exists! The fact that things work for smooth curves is one of the beautiful miracles of the subject.

Constructions of the moduli space conform to one general pattern: they all consist of first constructing a moduli space for curves with some additional structure, and then realizing the moduli space $\mathcal{M}_g$ of curves alone as a quotient of this space. There are, broadly, three different approaches, associated with the names of three pioneers, depending on the extra structure attached. We will discuss only the case of $\mathcal{M}_g$ below, that of $\mathcal{C}_g$ being similar.

**Teichmüller.** The first successful attempt to construct $\mathcal{M}_g$ (as an analytic variety) was by Teichmüller [1940] and [1943]. He considered pairs $(C; \delta_1, \ldots, \delta_{2g})$ consisting of a curve $C$ and a set of generators $\delta_1, \ldots, \delta_{2g}$ for the fundamental group $\pi_1(C)$ satisfying certain relations, modulo the action of the inner automorphisms of the fundamental group, and he showed that the set of such pairs can be parametrized by an analytic variety, called *Teichmüller space* $\mathcal{T}_g$. Set-theoretically it is easy to see that the moduli space $\mathcal{M}_g$ is Teichmüller space modulo the *mapping class group*—the group of homeomorphisms modulo those isotopic to the identity—of a topological surface of genus $g$. Unfortunately, the analytic structure given by Teichmüller was the wrong one for this—it was not invariant under the mapping class group! The correct definitions and proofs were finally given by Ahlfors [1960], Bers [1960] and Weil [1958], after some partial results (giving the correct structure away from the hyperelliptic locus) of Rauch [1955]; see Bers [1972] for a survey.
Teichmüller’s construction has been especially important of late because of the handle it gives us on the topology of the moduli space. The structure of $\mathcal{F}_g$ itself is rather simple—as Bers proved in [1970], it is isomorphic to an open domain in $C^{3g−3}$, and homeomorphic to a ball, so it is in particular contractible. Since the mapping class group acts on $\mathcal{F}_g (g > 2)$ with only finite stabilizers of points, we may use this fact to identify the cohomology of the moduli space with $\mathbb{Q}$ coefficients with that of the mapping class group. On the other hand, the mapping class group is an essentially topological object, accessible to study, for example, by the methods of Thurston’s school. This is the basis of Harer’s theorems, discussed below.

Schottky. Another approach to constructing $\mathcal{M}_g$ is to look at pairs $(C; \delta_1, \ldots, \delta_{2g})$ consisting of a curve $C$ and a canonical basis $\delta_1, \ldots, \delta_{2g}$ for the first homology $H_1(C, \mathbb{Z})$, modulo suitable isomorphisms. To the pair $(C; \delta_1, \ldots, \delta_{2g})$ we can then associate the period matrix $\Omega$ consisting of the integrals, over the cycles $\delta_{2g+1}, \ldots, \delta_{2g}$, of the $g$ holomorphic differential forms $\omega_i$ on $C$ determined by the condition that for $i, j = 1, \ldots, g$, the integral of $\omega_i$ over the cycle $\delta_j$ is the Kronecker delta $\delta_{ij}$. The period matrix then varies in an analytic subvariety of the Siegel upper half space $\mathcal{H}_g$ of $g \times g$ symmetric matrices with positive definite imaginary part; and since by a basic theorem of Torelli a curve is characterized by its period matrix, we can in this way identify the space of pairs $(C; \delta_1, \ldots, \delta_{2g})$ with a subvariety $\mathcal{F}_g$ of $\mathcal{H}_g$. The moduli space is then realized as the quotient of $\mathcal{F}_g$ by the action of the group $\text{Sp}(2g, \mathbb{Z})$, corresponding to the choice of basis $\delta_1, \ldots, \delta_{2g}$.

This construction had, until recently, the defect that the space $\mathcal{F}_g$ was extremely mysterious; for example, no one knew how to tell whether a given period matrix $\Omega \in \mathcal{H}_g$ came from a Riemann surface. It still is pretty mysterious, but less so, thanks to the efforts of Arbarello-DeConcini [1984] and Shioita [1986] who have given explicit characterizations of period matrices arising from curves. Thanks to work of Satake, Bailey-Borel and others (see Baily [1962]), however, it is possible to see from this construction that $\mathcal{M}_g$ is indeed an algebraic variety. In fact, by compactifying the quotient of $\mathcal{H}_g$ by $\text{Sp}(2g, \mathbb{Z})$, Satake and Bailey-Borel [1966] give a compactification $\overline{\mathcal{M}_g}$ such that $\mathcal{M}_g$ is a projective variety and $\overline{\mathcal{M}_g} - \mathcal{M}_g$ has codimension 2 in $\overline{\mathcal{M}_g}$, showing in particular that there are no nonconstant holomorphic functions on $\mathcal{M}_g$.

Mumford. A third, and purely algebraic, construction of the moduli space of curves was given as an application by Mumford and Deligne of Mumford’s “Geometric Invariant Theory” (Mumford [1965], Deligne-Mumford [1969]). Here the extra structure tacked on to a curve is exactly that of a particular projective embedding: namely, we consider curves $C$ embedded in projective space $\mathbb{P}^{3g−6}$ by the linear system of sections of the third power of the canonical bundle $(\omega_C)^{\otimes 3}$ (the third power is chosen because it gives an embedding of any curve of genus $\geq 2$, and is the smallest power to do so). Just as plane curves may be parametrized in a fairly elementary way by the points of a projective space, such images—the set
of curves $C \subset \mathbb{P}^{g-6}$ of degree $6g-6$ and genus $g$ satisfying $\mathcal{H}_C(1) \cong \omega_C$—are naturally parametrized by a subvariety $\mathcal{H}$ of a projective variety, called the Hilbert scheme; the moduli space $\mathcal{M}_g$ is the quotient of this subvariety by the action of the group $\text{PGL}(5g-5, \mathbb{C})$ of automorphisms of $\mathbb{P}^{g-6}$.

This is the construction that most directly shows us the algebraic (as opposed to analytic) structure of $\mathcal{M}_g$, and in fact can be used to give $\mathcal{M}_g$ the structure of a variety over $\mathbb{Q}$. Unfortunately, it is hard to use this approach directly to investigate $\mathcal{M}_g$ because we have virtually no knowledge of the structure of $\mathcal{H}$. But there are features which make this construction very useful. In particular, because we are here taking the quotient of a quasiprojective variety by a reductive linear group, we may apply Geometric Invariant Theory to arrive at natural compactifications of $\mathcal{M}_g$ and $\mathcal{C}_g$. (This compactification is so natural that it was also discovered analytically by Earle-Marsden and Bers [1974].)

**Stable curves and the Mumford compactification.** Unfortunately, $\mathcal{M}_g$, the space of smooth curves, is not a compact variety. Smooth curves can degenerate to singular ones—as we have seen, they even do it in the plane. Fortunately, Mumford’s Geometric Invariant Theory tells us what to do: To compactify $\mathcal{M}_g$ for $g \geq 2$, Mumford’s theory adds points corresponding to the isomorphism classes of certain singular curves, called “stable curves.” These are curves with only ordinary double points as singularities and having no smooth rational components meeting other components in less than three points.

The last condition simply prevents the automorphism group from being large. The choice of allowable singularities is more significant. Something of its delicacy can be appreciated from the following example, occurring in genus one: It is easy to make a rational nodal curve degenerate to a cuspidal curve by “pulling the noose closed” little by little; algebraically, by considering the family $y^2 = x^3 + t \cdot x^5$ parametrized by $t$. However, all the nodal curves in this particular family are isomorphic to one another, since there is an automorphism of $\mathbb{P}^1$ moving any two points to any other two points. Thus if we allowed the rational cuspidal curve as a point
in our moduli space with the nodal curve, the space would become non-Hausdorff! We could perhaps get around this by simply identifying the point corresponding to the isomorphism class of the cuspidal curve with the point corresponding to the isomorphism class of the nodal curves, but more trouble lurks: for example, the curves in the family \( y^2 = x^3 + t \) corresponding to values of \( t \) other than 0 are all isomorphic (the map \((x,y) \mapsto (t^{-1/3}x, t^{-1/2}y)\) carries the curve \( C_t \) to \( C_1 \)); but the curve \( C_0 \) is our cuspidal curve again. Geometric invariant theory tells us exactly what singular curves to include in our moduli, and which, for reasons like the above, we have to leave out; in the examples above, we are told to simply exclude cuspidal curves. Of course, it is then far from clear that the moduli space we get is compact; it is one of the main results of the subject—the “semistable reduction theorem” (see for example the book of Kempf et al. [1973])—that given any family of stable curves, parametrized by a smooth curve and tending to an arbitrary limit curve, we may after base change and birational transformations affecting only the limit curve arrive at a family tending to a limit curve that is stable. This is enough to establish the compactness of the moduli space of stable curves, called \( \mathcal{M}_g \).

Similarly, if we take isomorphism classes of “stable pointed curves”—that is pairs \((C,p)\) with \( C \) a stable curve and \( p \) a smooth point on it (allowing \( C \) this time to have a smooth rational component meeting only two other components, so long as it contains the point), then we obtain a compactification \( \overline{\mathcal{M}}_g \) of \( \mathcal{M}_g \). The map \( \pi \) extends to a map \( \pi: \overline{\mathcal{M}}_g \to \mathcal{M}_g \) (which collapses the “excess” rational component caused by the difference in the definitions of stability, if any, to a point), and over the open set \( \mathcal{M}_g \subset \overline{\mathcal{M}}_g \) consisting of curves without automorphisms, \( \overline{\mathcal{M}}_g^0 = \pi^{-1}(\mathcal{M}_g^0) \) is the “universal” stable curve.

**Diagram of the theory of curves.** We have been trying to relate the spaces of curves in projective space to moduli spaces of abstract curves by describing the additional structure needed to specify the former, given the latter. We may put this all together in a diagram of spaces, which can be thought of as corresponding to curves with progressively more and more
“projective” data as one goes from left to right:

\[
\begin{array}{c}
\mathcal{C}_g \ni \mathcal{C}_g \\
\downarrow \\n\mathcal{M}_g \ni \mathcal{M}_g \\
\sim \quad \mathcal{Pic}^d \leftarrow \mathcal{G}_d \leftarrow \mathcal{H}_d.
\end{array}
\]

Here \(\mathcal{Pic}^d\) is the space of pairs \((C, L)\) where \(C\) is a smooth curve of genus \(g\) and \(L\) a line bundle of degree \(d\) on \(C\); \(\mathcal{H}_d\) is the space of triples \((C, L, V)\) where \((C, L) \in \mathcal{Pic}^d\) and \(V \subset H^0(C, L)\) is a linear series; and \(\mathcal{G}_d\) is the space of smooth curves of degree \(d\) and genus \(g\) in \(\mathbb{P}^r\), each of which corresponds, by what has been said above, to a triple \((C, L, V)\) in \(\mathcal{H}_d\) together with a choice of basis for the vector space \(V\).

There are at least two good reasons for including singular curves in this picture (in the compactified spaces at the left of the diagram), even if one is only interested in the geometry of smooth ones. First, since many of the techniques developed for studying varieties deal only with compact ones, we will often wish to have compactifications of the varieties given, and the nicest of these naturally involve singular curves. Second, it is often the case that singular curves are easier to understand than their smooth counterparts rather than harder (as is seen for example, in most of the proofs of the Brill-Noether theorem), so that in trying to understanding the geometry of the varieties above it makes sense to start at points representing singular curves.

As we have seen, Mumford’s approach to the moduli space \(\mathcal{M}_g\) yields a natural compactification of it and of \(\mathcal{C}_g\), and these spaces have many desirable geometric properties (for example, the singularities of \(\mathcal{M}_g\) are quite mild, and the complement of the set of smooth curves in it is a union of subvarieties of codimension 1, which are themselves naturally smooth and cross nearly transversely).

There are also natural compactifications of the spaces \(\mathcal{H}_d\) in case \(r \geq 3\), namely the (union of the relevant components of the) Hilbert schemes of all space curves of degree \(d\) and genus \(g\) in \(\mathbb{P}^r\), though these are probably not the “right” choices. Even for plane curves, the situation is not entirely clear: it is an important current problem to determine a “good” compactification of the set of plane curves having only ordinary double points as singularities.

This leaves as a crucial problem the discovery of good compactifications of the spaces \(\mathcal{Pic}^d\) and \(\mathcal{G}_d\). To the extent that one succeeds in this one can undertake the very fruitful exploration of the (rational) maps between the compactifications of the spaces involved in the above diagram. The recent solution of the long standing “Severi problem” by Harris [1986], see also Ran [1986]—the proof that the set of plane curves with degree \(d\) having a fixed number of nodes as singularities forms an irreducible variety—is an example of the success of this program in a case where one can piece together at least sufficiently good approximations to solutions of the compactification problems cited.

The problem of linear series on singular curves. To compactify \(\mathcal{Pic}^d\) and \(\mathcal{G}_d\), what is needed is a good theory of line bundles and
linear series on singular curves, or at least on stable curves. The standard definitions of both objects make perfect sense for any curve, even any variety, but do not have the right properties in the singular case. At least for irreducible singular curves one can get a good compactification of $\text{Pic}^d$ by including rank one torsion free sheaves along with line bundles, and the resulting “compactified Jacobian” has been studied for some time (D’Souza [1973], Oda-Seshadri [1979], Altman-Iarrobino-Kleiman [1977], Altman-Kleiman [1979] and [1980], . . . , Gokhale [1986]).

Unfortunately, the necessary extension to reducible curves seems far more troublesome. The rather complicated theory of “Neron models” of elliptic curves (see for example Artin [1986]) may be thought of as a step in the necessary direction. Fairly satisfactory theories are in hand in the case of series of dimension 1 (work of Beauville [1977] and Harris-Mumford [1982]) or series on certain curves called curves of compact type (work of Eisenbud-Harris [1986]), and there is work on the general case (Ran [1985]). The known special cases suffice for many applications, but there is probably a small gold mine awaiting a general insight.

II. The nature of the moduli space. We have now said enough about the foundations of the general theory to be able to discuss, in somewhat more depth, a problem on which a great deal of recent progress has been made. It is a problem which first arose explicitly in the work of Severi in the 1920s. It concerns the “nature,” in a sense to be made precise below, of the moduli space of curves.

The dimension and tangent bundle of the moduli variety. After the existence question for the moduli space $\mathcal{M}_g$ is settled, perhaps the first question one might ask about it is its dimension. The question is answered by two out of the three constructions $\mathcal{M}_g$ referred to above, but it can be answered independently of them—indeed Riemann did this, and one can still read his computation with profit. We will now explain yet another method, this one with a more modern flavor, which actually leads to a computation of the tangent bundle to $\mathcal{M}_g$, a computation that we shall need later.

The Kodaira-Spencer map. Let $C$ be a smooth projective curve of genus $g$, and let $f : X \to B$ be a family of curves containing $C$ as the fiber over a point $p \in B$. Suppose that $X$ and $B$ are smooth, and let $t$ be a tangent vector to $B$ at $p$. If the family $f$ were trivial (that is, $X \cong C \times B$), or even trivial to first order in the direction $t$ in some appropriate sense, we could lift the vector $t$ to a field of tangent vectors to $X$ defined at every point of $C = f^{-1}(p)$. We can therefore use the failure of the existence of such a lifting to “measure” the nontriviality of $f$.

Now $T$ can certainly be lifted to a tangent vector at any given point of $C$, and thus there is a covering $C = \bigcup_i U_i$ of $C$ by open sets such that $t$ can be lifted to a field of tangent vectors $T_i$ to $X$ along $U_i$. Since $T_i$ and $T_j$ are both defined along $U_i \cap U_j$, and project to the same vector $t$ at $p \in B$, their difference $T_i - T_j$ is a section over $U_i \cap U_j$ of the tangent bundle $\Theta_C$ to $C$. This collection of differences defines a Čech 1-cocycle on
C with coefficients in $\Theta_C$, and thus the construction defines a map, called the Kodaira-Spencer map, from the tangent space to $B$ at $p$ to $H^1(\Theta_C)$.

It turns out that the Kodaira-Spencer map above extends to a sufficiently small analytic neighborhood of $p$, and actually captures all the analytic local data of the family: One can define a “versal” family of curves $f_0: X_0 \to B_0$ with base space $B_0$ a small neighborhood of the origin in $H^1(\Theta_C)$, and the given family $f$, restricted to a small analytic neighborhood of $p$, can be obtained by pulling back this versal family along the extended Kodaira-Spencer map. (The case of complex manifolds described above was first treated by Kodaira-Spencer [1958]. Related constructions of deformation theory extend the theory to higher-dimensional varieties, singular varieties, and other still more complicated objects—see Illusie [1971].)

In particular, we get a map $\varphi$ from $B_0 \subset H^1(\Theta_C)$ to a neighborhood of the point representing the isomorphism class of $C$ in $\mathcal{M}_s$. The mapping property of $B_0$ indicated above suggests that $\varphi$ might be an isomorphism, and indeed it “usually” is. To understand the general picture, consider the case when $C$ has an automorphism $\alpha$ that does not extend to nearby curves. $\alpha$ induces an automorphism $\alpha_*$ of $H^1(\Theta_C)$. Since the construction of the versal family over $H^1(\Theta_C)$ is functorial, $\alpha$ extends to an automorphism of $X_0$ covering $\alpha_*$. On the other hand, since $\alpha$ does not extend to curves near $C$, $\alpha_*$ must act nontrivially, so that several of the fibers of $f_0$ are isomorphic, and the corresponding points of $B_0$ will be identified by $\varphi$.

In case the genus $g$ is $\geq 2$, the general curve has no automorphisms at all, and thus no automorphism that a particular curve happens to have can be extended to nearby curves. The above argument suggests, correctly,
that in this case a neighborhood of the point corresponding to \( C \) in \( \mathcal{M}_g \) will have the form \( B_b / \text{Aut}(C) \). In particular, we see from this that \( \mathcal{M}_g \) is nonsingular away from the set of curves with automorphisms, and also that if \( C \) has no automorphisms, then the tangent space to \( \mathcal{M}_g \) at the point corresponding to \( C \) is \( H^1(\Theta_C) \). By Serre duality, the cotangent space to \( \mathcal{M}_g \) at \( C \) is then \( H^0(\omega_C \otimes \Theta_C^{-1}) = H^0(\omega_C^2) \). As the degree of \( \omega_C^2 \) is \( 4g-4 \), we can compute the dimension of this space, which is also the dimension of \( \mathcal{M}_g \), by using the Riemann-Roch Theorem: we get

\[
\dim \mathcal{M}_g = 3g - 3.
\]

**The structure of the moduli variety.** We now know that \( \mathcal{M}_g \) is an irreducible algebraic variety, and we know its dimension. What question should we ask about it next? The only sensible approach is to study the variety concretely for as many \( g \) as possible, and try to guess what might be its properties in general.

For \( g = 0 \), everything is trivial: \( \mathcal{M}_0 \) is a single point, corresponding to \( \mathbb{P}^1 \).

The study of \( \mathcal{M}_1 \) is one of the most classical and beautiful parts of complex analysis. A Riemann surface of genus one is a complex torus, and as such can be realized as the space of complex numbers, its universal covering space, modulo a lattice, which after rotation and scaling can be taken to have basis \( 1 \) and \( \tau \), with \( \tau \) in the upper half plane \( \mathbb{H} \). This representation is not unique, since the automorphism group \( \text{Sl}(2, \mathbb{Z}) \) (the “modular group”) acts; the moduli space is thus the quotient \( \mathbb{H} / \text{Sl}(2, \mathbb{Z}) \). The famous “modular function” \( j \) maps \( \mathbb{H} \) to the complex line, and is invariant under the action of \( \text{Sl}(2, \mathbb{Z}) \); it provides an isomorphism \( \mathcal{M}_1 \cong \mathbb{C} \).

Alternatively, a curve of genus 1 with a chosen base point can be embedded, by means of the linear series of all rational functions with poles of order at most 3 at the base point and no poles elsewhere, as a cubic curve in the projective plane; after a linear transformation of the plane it will have an equation of the form \( y^2 = x(x-1)(x-\lambda) \). The value of \( \lambda \) is then determined up to a finite group; the field of functions invariant under this group is generated by the rational function

\[
j = 256 \frac{\lambda^2(\lambda - 1)^2}{(\lambda^2 - \lambda + 1)^3}
\]

which is the same parameter as before on \( \mathcal{M}_1 \cong \mathbb{C} \).

\( \mathcal{M}_2 \): Any smooth curve of genus 2 can be represented in a unique way as a ramified double covering of \( \mathbb{P}^1 \), branched over 6 distinct points. In fact, this map to \( \mathbb{P}^1 \) corresponds to the linear series obtained by taking all of the sections of the cotangent bundle to the curve. The 6 branch points are easily seen to determine the curve; for example, if they are \( \lambda_1, \ldots, \lambda_6 \in \mathbb{C} \subset \mathbb{P}^1 \), then the curve may be realized through the equation

\[
y^2 = (x - \lambda_1) \cdots (x - \lambda_6).
\]

Writing down this equation, we can say that we have written down the general (indeed, every smooth) curve of genus 2.
Of course, the points $\lambda_i$ are obviously not unique: any automorphism of $\mathbb{P}^1$ will carry six points to another six points giving the same curve. Thus the moduli space will be the quotient of the sixfold symmetric product $(\mathbb{P}^1)^6$ (minus diagonals) by the action of the group $\text{PSL}(2, \mathbb{C})$. But we can do better than this: the automorphism group can be used to carry 3 of the 6 points to 0, 1, and $\infty$, with only a finite ambiguity left over—which 3 points did we choose, and in which order?—so that $\mathcal{M}_3$ may be realized as the quotient of an open subset of the threefold product $(\mathbb{P}^1)^3$ by a finite group. Igusa [1960] has carried through the analysis completely, and shown that the moduli space is in fact $\mathbb{C}^3/\mathbb{Z}_3$ with a certain explicit action of the cyclic group $\mathbb{Z}_3$, so we know things very precisely here.

$\mathcal{M}_3$: If $C$ is a general curve of genus 3, then the 3-dimensional vector space of sections of the cotangent bundle of $C$ yields an embedding of $C$ as a curve of degree 4 in $\mathbb{P}^2$ ("general" here means that $C$ belongs to the complement of the codimension 1 set of curves possessing a $g_2^1$—the "hyperelliptic" curves. If $C$ is hyperelliptic, the map associated to $\omega_C$ maps $C$ 2 to 1 onto a conic). Further, any smooth curve of degree 4 in $\mathbb{P}^2$ will have genus 3. Thus an open subset of the moduli space, at least, can be realized as an open subset in the $\mathbb{P}^1^4$ of smooth quartic curves modulo the automorphism group $\text{PSL}(3, \mathbb{C})$ of $\mathbb{P}^2$.

$\mathcal{M}_4$: To give one more example explicitly, it is a standard (and elementary) fact that a general curve $C$ of genus 4 may be embedded in $\mathbb{P}^1$ by the linear system of sections of its cotangent bundle (the exceptions as before being the hyperelliptic curves). For even more general curves, it is furthermore the case that the image of this embedding will be the curve of intersection of a smooth quadric $Q$ with a cubic surface $S$ (the exception here will be curves with a "vanishing theta-null"—that is, curves possessing a divisor $D$ with 2$D$ linearly equivalent to the canonical divisor and $r(D) \geq 1$—whose canonical images lie on a quadric cone rather than a smooth quadric. To see this, note that if $\sigma$ and $\tau$ are sections of $\mathcal{O}(D)$, then $\alpha = \sigma^2$, $\beta = \sigma \cdot \tau$ and $\gamma = \tau^2$ will be sections of the canonical bundle satisfying $\alpha \gamma = \beta^2$, so that the canonical image of $C$ will lie on a quadric cone). We may thus realize the space $\mathcal{M}_4$ as the quotient of the space of pairs $(Q, S)$ (or, to be marginally more efficient, we could fix the quadric $Q$ and realize $\mathcal{M}_4$ as the quotient of the projective space of cubics modulo those containing $Q$).

Alternately, if we project the canonical image of the curve $C$ from a point $p$ of $C$ to the plane $\mathbb{P}^2$, we see that $C$ may be realized as a plane quintic with exactly two double points, corresponding to the two lines of $Q$ passing through $p$. The space of quintics with two double points may in turn be rationally parametrized, by first freely specifying the two double points $P$ and $Q$ in $\mathbb{P}^2$, and then letting the coefficients of the quintic polynomial vary in the vector space of quintics double at $P$ and $Q$.

Further examples. We could continue this list, with slowly increasing difficulty, for some time; this was in fact done by Severi up to $g = 10$. It seems to us salient, as it apparently seemed also to Severi, that in each of these cases, there is at least a large open set of $\mathcal{M}_g$, which is the image of a
large open subset of \( \mathbb{C}^n \) for some \( n \). A variety with this property is said to be unirational.

To show systematically that \( \mathcal{M}_g \) is unirational for \( g \leq 10 \), one can, with Severi, proceed as follows. By the Brill-Noether theorem stated above, a general curve of genus \( g \) will have a linear system \( g^d_2 \)—and so will be realized as a plane curve of degree \( d \)—if and only if \( \rho = g - 3(g - d + 2) \geq 0 \), that is, \( d \geq 2g/3 + 2 \) (the cases of genus 3 and 4 discussed above are examples of this). Moreover, a general such representation will have only nodes as singularities; so that an open subset of \( \mathcal{M}_g \) will be a quotient of an open subset of the space of plane curves of degree \( d \) with \( \delta = (d - 1)(d - 2)/2 - g \) double points. In case the number \( \delta \) of double points is relatively small—precisely, if \( \delta < (d + 1)(d + 2)/6 \), so that there will exist a nonzero vector space of polynomials of degree \( d \) vanishing to order two at \( \delta \) arbitrarily assigned points of the plane—then as in the case of genus 4 above we see that this space, and hence \( \mathcal{M}_g \), may be rationally parametrized. The two conditions \( d \geq 2g/3 + 2 \) and \( (d - 1)(d - 2)/2 - g < (d + 1)(d_2)/6 \) can be satisfied for some \( d \) precisely when \( g \leq 10 \).

Severi's argument unfortunately ignores the subtle fact that the nodes of a general plane model of a general curve need not \textit{a priori} be general points, and will not be such, in fact, except under special circumstances. In recent times Arbarello and Cornalba [1983] have, however, given a correct version of the argument, so that we now know in this way that \( \mathcal{M}_g \) is unirational for \( g \leq 10 \).

The approach via plane curves is not the only one. Sernesi in [1981] used a representation of a general curve of genus 12 as a space curve in \( \mathbb{P}^3 \) to rationally parametrize \( \mathcal{M}_{12} \); and Chang and Ran [1984] were able to carry out similar analyses in genera 11, 12 and 13. Beyond 13, however, no moduli space \( \mathcal{M}_g \) is known to be unirational; new methods seem to be required.

The rest of this paper will be devoted to the question of the unirationality of \( \mathcal{M}_g \), a question which has motivated a great deal of work in this field. Before plunging into this we prepare with a digression.

**The Meaning of Rationality, Unirationality, and Uniruledness.**

One basic question one can ask about any variety \( X \) is whether or not it is \textit{rational}; that is, whether or not the field of meromorphic functions on \( X \) is purely transcendental over \( \mathbb{C} \), or, equivalently, whether or not some open subset of \( X \) is isomorphic to an open subset of \( \mathbb{C}^n \). In the case of curves, this is clearly a fundamental question: a smooth curve \( X \) is rational if and only if it is the Riemann sphere. In higher dimensions, however, rationality becomes much harder to ascertain—for example, it is not clear whether, in a family \( \pi: \mathcal{X} \to B \) of smooth varieties, the locus \( \{ b \in B : X_b \text{ is rational} \} \) is open, or closed, or neither. Indeed, the notion appears less fundamental in the higher-dimensional case.

There are, however, related conditions, equivalent to rationality for one-dimensional varieties, that are easier to test for in practice and that are, conjecturally at least, related to the most important numerical invariants of a variety. One is \textit{unirationality}: we say that a variety \( X \) is unirational
if the field of meromorphic functions on $X$ is a subfield of a purely transcendental extension of $\mathbb{C}$; or, equivalently, if there exists a surjective algebraic map from an open subset of $\mathbb{C}^n$ to an open subset of $X$. Both conditions amount simply to saying that most of the elements of $X$ may be rationally parametrised: if we regard $X$ as a subvariety of some fixed projective space $\mathbb{P}^N$, then the condition says that there exist rational functions $F_0(t_0, \ldots, t_n), \ldots, F_N(t_0, \ldots, t_n)$ such that the points

$$(F_0(t_0, \ldots, t_n) : \cdots : F_N(t_0, \ldots, t_n)) \in \mathbb{P}^N,$$

for freely varying $(t_0, \ldots, t_n)$, fill up most of $X$.

Perhaps still more important is the weaker condition of uniruledness: we say that $X$ is uniruled if, though a general point of $X$, there passes at least one rational curve; equivalently, $X$ is uniruled if there exists an algebraic map $\varphi$ from an open subset of a product variety $Y \times \mathbb{P}^1$ onto an open subset of $X$ such that $\varphi$ is not constant on the fibers $\{\text{point}\} \times \mathbb{P}^1 \subset Y \times \mathbb{P}^1$. We have, of course,

$$\text{rational } \Rightarrow \text{ unirational } \Rightarrow \text{ uniruled};$$

all three conditions are equivalent for curves, and the first two for surfaces, but in higher dimensions the three are distinct.

Part of the interest in the condition of uniruledness is the (partly conjectural) criterion for it in terms of more geometric objects, the “pluricanonical divisors.” A “pluricanonical form” on a smooth compact variety $X$ is by definition a nonzero global section of a tensor power of the line bundle $\omega_X$, itself defined as the highest exterior power of the cotangent bundle of $X$. The significance of these comes partially from the fact that, although defined in terms of the whole variety, the existence of, and indeed the dimension of the space of, pluricanonical forms of a given type depends only on dense open subsets of $X$; that is, if a dense open subset of $X$ is isomorphic to a dense open subset of a smooth compact variety $X'$, then the sections of a given power of $\omega_X$ are in natural correspondence with the sections of the same power of $\omega_{X'}$. Further, if $X$ has the form $Y \times \mathbb{P}^1$ for some $Y$, then essentially because differential forms on $\mathbb{P}^1$ (of which there are none) would have to be involved in any pluricanonical form on $X$, $X$ has no pluricanonical forms at all. This proves that a uniruled variety has no pluricanonical forms. Conversely, it is conjectured as part of the “classification theory for higher dimensional varieties” that a variety with no pluricanonical forms must be uniruled; this is known for curves (easy), for surfaces (one of the triumphs of Castelnuovo in the “golden age” about 100 years ago) and for 3-folds (by a very recent theorem of Miyaoka-Mori [1986]).

In many important cases it is useful to interpret the existence of pluricanonical forms in terms of the existence of certain subvarieties of codimension 1. The idea is that the locus $Y$ along which a section of a given line bundle on a variety $X$ vanishes determines the bundle, and the determination can sometimes be made simply in terms of the number of points in which $Y$ meets various curves lying in $X$. Thus to check whether $X$ is uniruled one might try to characterize the zero loci of pluricanonical
forms—called pluricanonical divisors—in terms of their intersection numbers with certain known curves, and then to ask whether subvarieties with the desired intersection properties actually exist.

The interpretation of unirationality and uniruledness of moduli. To say that $\mathcal{M}_g$ is unirational amounts to saying that there is a procedure with “free parameters” that describes nearly all curves of genus $g$, i.e., that we can write down a curve $C \subset \mathbf{P}^r$ of genus $g$ in such a way that the general curve of genus $g$ is obtained by varying the coefficients of the defining equations of $C$ as rational functions of parameters $t_1, \ldots, t_n$. Informally, this amounts to saying just that “we can write down the general curve of genus $g$.”

On the other hand, the statement that $\mathcal{M}_g$ is not uniruled would imply, for example, that there is a dense open set of curves $C$ of genus $g$ such that if a surface in a projective space has one of these curves $C$ as a hyperplane section, then the surface is ruled by lines over $C$, that is, it is the union of the family of lines corresponding to a map of $C$ to the Grassmannian of lines in projective space. Thus unirationality and uniruledness or their lack have rather dramatic consequences.

The question of the unirationality of $\mathcal{M}_g$. In view of the difficulty geometers experienced for a half a century in trying to extend Severi’s result beyond genus 10, the suspicion naturally arose that there might not be a rational parametrization of the moduli of curves of genus $g$ for all $g$ (to paraphrase Barry Mazur, “While working on a proof by day, spend your nights looking for counterexamples”).

How one might prove that $\mathcal{M}_g$ is not unirational is not so clear, unless one is lucky and $\mathcal{M}_g$ is not even uniruled. In that case, as indicated above, the “natural” thing to do is to search for pluricanonical forms on $\mathcal{M}_g$. Paradoxically, this method of proving that a variety is not unirational requires that the variety be compact—even though unirationality is a property of every algebraic open (that is, Zariski open) subset. It also requires that the variety be smooth, something the moduli space is also not.

Luckily, we have a good compactification of the moduli space $\mathcal{M}_g$, namely the moduli space $\overline{\mathcal{M}}_g$ of stable curves. The problem then becomes one of determining, if we can, whether or not there are pluricanonical forms on $\overline{\mathcal{M}}_g$, or, equivalently, if the divisor class on $\mathcal{M}_g$ represented by this line bundle contains any effective divisors.

As for the singularities, a close analysis shows that the method still works in the presence of some very mild singularities (Tai [1982]). As we have seen, the singularities of $\overline{\mathcal{M}}_g$ look like the singularities of spaces of the form $H^0(\omega_C^{\otimes 2})^*/\text{Aut}(C)$. A detailed analysis by Harris and Mumford of the possibilities for the action of $\text{Aut}(C)$ on $H^0(\omega_C^{\otimes 2})^*$ revealed that the singularities appearing could be safely ignored.

Line bundles on the moduli space. We now consider the question of describing line bundles on the moduli space of curves, the goal being to be able to characterize the pluricanonical divisors. Since $\mathcal{M}_g$ is obtained from $\overline{\mathcal{M}}_g$ by adding divisors on the boundary, and the classes of those divisors are easily seen to be independent, knowing the group of line bundles on
\( \mathcal{M}_g \) is equivalent to knowing it on \( \mathcal{M}_g \), and we will focus on the latter problem. It turns out that the situation is reasonably simple. There are two approaches, one through cohomology classes and the other through homology classes, which together will give us sufficient information.

**Cohomology—the nontriviality of families of curves.** First we look at the cohomology classes (Chern classes) associated to line bundles; by work of MacLean and Harer (see Harer [1988]) the first Betti number of \( \mathcal{M}_g \) is zero, so by the argument using the exponential sequence to determine the Picard group as explained above, any line bundle is determined by its first Chern class, an element of \( H^2(\mathcal{M}_g, \mathbb{Z}) \).

Now (ignoring torsion for the moment) a cohomology class can be thought of as a linear function on cycles, measuring their nontriviality. Since a cycle in \( \mathcal{M}_g \) is roughly a family of curves, a cohomology class in \( H^2(\mathcal{M}_g, \mathbb{Z}) \) may thus be thought of as a device that measures the nontriviality of a family \( f: X \to B \) of smooth curves. Associated with any such family there is a vector bundle \( E_f \) on \( B \), called the Hodge bundle, whose fiber at a point \( b \in B \) with preimage \( f^{-1}b = C \) is \( H^0(\omega_C) \) (more precisely, \( E_f \) is the push-forward of the relative dualizing bundle on \( X \)). If \( f \) were a trivial family, the Hodge bundle would be a trivial bundle, so we may measure the nontriviality of \( f \) by the nontriviality of the Hodge bundle. Explicitly, if we apply this construction to the universal curve \( \pi: \mathcal{C}^0_\mathcal{M}_g \to \mathcal{M}_g^0 \), and use the fact that the complement of \( \mathcal{M}_g^0 \) has codimension \( \geq 2 \) in \( \mathcal{M}_g \) (except along a certain locus in the boundary, which requires special care) we get a fundamental class, associated to the line bundle \( \lambda = \wedge^g E_\pi \) usually also denoted \( \lambda \in H^2(\mathcal{M}_g, \mathbb{Z}) \).

Instead of using the global sections of the cotangent bundle of \( C \) in the above construction, we could rather have taken the global sections of a tensor power, say the square, \( H^0(\omega_C^{\otimes 2}) \), obtaining a bundle \( E \) on \( \mathcal{M}_g^0 \) whose fiber at the point corresponding to a curve \( C \) is \( H^0(\omega_C^{\otimes 2}) \). We have already shown that this is the same as the fiber of the cotangent bundle to \( \mathcal{M}_g \), and indeed \( E \) turns out to be the cotangent bundle.

On the other hand, it is again a theorem of Harer's (using some earlier work of Mumford) that \( H^2(\mathcal{M}_g, \mathbb{Z}) = \mathbb{Z} \). A theorem of Arbarello and Cornalba [1987] shows that \( H^2(\mathcal{M}_g, \mathbb{Z}) \) is actually generated by \( \lambda \), so that the first Chern class of each of the bundle \( E \) is an integral multiple of \( \lambda \). To see this directly for the canonical bundle \( \omega \), and to compute the multiple in this case, we can use the Grothendieck-Riemann-Roch formula (Borel-Serre [1958]). Essentially, whenever we have a family of line bundles on curves, parametrized by a variety \( B \), we can define an object on \( B \) that will be, in nice cases, a bundle whose fiber at each point \( b \in B \) is the space of global sections of the original line bundle on the fiber over \( b \). The Grothendieck-Riemann-Roch formula then measures the nontriviality of this bundle, by determining its Chern character. (The ordinary Riemann-Roch-formula is just the 0th graded piece of the Grothendieck-Riemann-Roch, inasmuch as it tells us the rank of this bundle.) When we apply the Grothendieck-Riemann-Roch formula to the family of line bundles consisting of the
square of the canonical bundle on each curve, we find that
\[ \omega_{\tilde{\mathcal{M}}_g} \cong c_1(E) = 13 \cdot \lambda. \]

We can extend the computation to all of \( \tilde{\mathcal{M}}_g \) to obtain the formula
\[ (*) \quad \omega_{\tilde{\mathcal{M}}_g} \cong 13 \cdot \lambda - 2 \cdot \delta \]

where \( \delta \) denotes the fundamental class of the boundary \( \mathcal{M}_g - \mathcal{M}_g \). (This formula is actually accurate only in the “functorial Picard group”—the moduli space is singular along one component of the boundary, and this causes difficulties which we are ignoring.)

It is worth remarking—though irrelevant to our story—that the “13” occurring here is, according to string theorists, the “reason” why the dimension of the universe in which we live (in one popular model) is 26! See for example Nelson [1987].

**Homology—loci characterized by properties of curves.** The approach above gave us a complete description of the line bundles on \( \tilde{\mathcal{M}}_g \). But we still need to know whether the pluricanonical line bundles have nonzero sections (it is interesting to ask more generally which bundles have nonzero sections). In this second approach we try to describe the codimension 1 subvarieties of moduli, called divisors, directly. Here we are lucky: while our knowledge of the geometry of \( \mathcal{M}_g \) is far from sufficient to allow us to describe divisors on \( \mathcal{M}_g \) by writing down equations for them, there is another way. Since the points of \( \mathcal{M}_g \) correspond to curves, any time we write down a geometric property enjoyed by some, but not all curves, we arrive at a subset of \( \mathcal{M}_g \) that is usually a (locally) closed subvariety. Thus, for example, the locus of hyperelliptic curves is a subvariety of, and hence describes a cycle in \( \mathcal{M}_g \), as are the loci of trigonal (or more generally \( k \)-gonal) curves, smooth plane curves of degree \( d \) (or more generally curves realizable as curves of degree \( d \) in \( \mathbb{P}^r \); see the next paragraph), curves with a given automorphism group, curves with a given type of Weierstrass point, and so on. To find divisors, we just have to look for those conditions that define subvarieties of codimension one; for example, the locus on \( \mathcal{M}_g \) of curves with a nonnormal Weierstrass point is a union of two divisors whose classes have been determined by Diaz [1985] and Cukierman [1987]; the locus of curves with a vanishing theta-null (equivalently, a semicanonical pencil) has a divisorial component whose class has been determined by Mumford and Teixidor [1986].

An important class of examples of such divisors in \( \mathcal{M}_g \) is provided by the Brill-Noether theorem. This theorem says that a general curve \( C \) of genus \( g \) will possess a linear system \( g_d^r \) if and only if the Brill-Noether number \( \rho = g - (r + 1)(g - d + r) \) is nonnegative; thus, whenever \( \rho < 0 \), the locus \( \mathcal{M}_{g,r,d} \) of curves possessing such a series is a proper subvariety of \( \mathcal{M}_g \). More generally, it seems sometimes to be the case, especially when \( \rho \) is negative with small absolute value, that \( \mathcal{M}_{g,r,d} \) will have codimension \( -\rho \); in particular, it turns out that whenever \( \rho = -1 \), \( \mathcal{M}_{g,r,d} \) will have a unique component of codimension one (and is conjecturally irreducible). The class of this component—or rather of its closure in \( \tilde{\mathcal{M}}_g \)—was first
determined in case $r = 1$ by Harris and Mumford [1982], and for arbitrary $r$ by Eisenbud and Harris [1986]; the dual cohomology class is given by the formula

$$\mathcal{M}_{g,r,d} \sim c_{g,r,d} \left[ \left( 6 + \frac{12}{g+1} \right) \cdot \lambda - \delta_0 - b_1 \cdot \delta_1 - \cdots \right]$$

where all the $b_i$ are greater than one, and $c_{g,r,d}$ is an explicitly known rational number.

**The result.** The significance of this computation is that, because $\mathcal{M}_{g,r,d}$ is by definition an effective divisor on $\mathcal{M}_g$, it asserts the existence of certain meromorphic functions on $\mathcal{M}_g$. Specifically, it says that there exist sections of the line bundle $\mathcal{L}^m$ on $\mathcal{M}_g$ having zeroes of order at least $m \cdot (6 + 12/(g+1))^{-1}$ along every boundary component $\Delta_\alpha$ of $\mathcal{M}_g$; in particular, it says that whenever the number $6 + 12/(g+1)$ is less than the ratio 13/2 appearing in the formula (*) for the canonical class of $\mathcal{M}_g$, then there will exist holomorphic pluricanonical differentials on $\mathcal{M}_g$. Since, as observed, a uniruled variety can never have such forms, we deduce the

**Theorem (Harris-Mumford [1982], Harris [1982], Wolfert [1983], Eisenbud-Harris [1986]).** For any $g \geq 23$, the moduli space $\mathcal{M}_g$ of stable curves possesses pluricanonical forms, and hence is not unirational, or even uniruled.

The argument given above takes care of the case when $g+1$ is composite, so that there are integers $r$ and $d$ such that $\rho = -1$, and thus divisors of the form $\mathcal{M}_{g,r,d}$; other divisors can be used when $g+1$ is prime, as in Harris [1982]. Also, it should be pointed out that the conclusion given here is substantially weaker than may actually be deduced. In fact, the argument outlined above shows that for $g > 23$ the moduli space $\mathcal{M}_g$ is in fact a variety of general type, meaning that the dimension $P_m$ of the space of sections of $\omega^{\otimes m}$ grows as a polynomial in $m$ of degree dim$(\mathcal{M}_g) = 3g - 3$.

**A speculation.** The question of whether the moduli space $\mathcal{M}_g$ is unirational for values of $g$ between 14 and 22 remains open. As for the related question of whether $\mathcal{M}_g$ admits pluricanonical forms, this is equivalent to the question of what linear combinations of the divisor classes $\Delta$ and $\delta = \sum \delta_\alpha$ are represented by effective divisors, which is in turn related to the question of what inequalities are satisfied by the degrees of $\lambda$ and $\delta$ for a one-parameter family of curves containing the general curve of genus $g$. Chang and Ran [1984], [1986] have constructed families showing that in fact $\mathcal{M}_g$ has no pluricanonical forms when $g \leq 13$ or $g = 15$, and natural conjectures of Harris and Morrison [1987] suggest that this is so for the remaining open cases, that is for all $g < 23$. Putting this together with the conjectures from the classification theory of higher dimensional varieties, we get

**Conjecture.** $\mathcal{M}_g$ is uniruled iff $g < 23$. 

REFERENCES


O. Teichmüller, Extremale quasikonforme Abbildungen und quadratische Differentialen, Preuss. Akad. 22 (1940).


M. Teixidor, Irreducibility of the variety of curves with even semicanonical pencil, preprint (1986).


DEPARTMENT OF MATHEMATICS, BRANDEIS UNIVERSITY, WALTHAM, MASSACHUSETTS 02254

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RHODE ISLAND 02912