Moduli Spaces

David D. Ben-Zvi

Many of the most important problems in mathematics concern CLASSIFICATION. One has a class of mathematical objects and a notion of when two objects should count as equivalent. It may well be that two equivalent objects look superficially very different, so one wishes to describe them in such a way that equivalent objects have the same description and inequivalent objects have different descriptions.

Moduli spaces can be thought of as *geometric* solutions to *geometric* classification problems. In this article we shall illustrate some of the key features of moduli spaces, with an emphasis on the moduli spaces of RIEMANN SURFACES. (Readers unfamiliar with Riemann surfaces may find it helpful to begin by reading about them in Part III.) In broad terms, a *moduli problem* consists of three ingredients.

Objects: which geometric objects would we like to describe, or *parametrize*?
Equivalences: when do we identify two of our objects as being isomorphic, or "the same"?
Families: how do we allow our objects to vary, or modulate?

In this article we will discuss what these ingredients signify, as well as what it means to *solve* a moduli problem, and we will give some indications as to why this might be a good thing to do.

Moduli spaces arise throughout ALGEBRAIC GEOMETRY, DIFFERENTIAL GEOMETRY and ALGEBRAIC TOPOLOGY. (Moduli spaces in topology are often referred to as classifying spaces.) The basic idea is to give a geometric structure to the *totality* of the objects we are trying to classify. If we can understand this geometric structure, then we obtain powerful insights into the geometry of the objects themselves. Furthermore, moduli spaces themselves are rich geometric objects in their own right. They are "meaningful" spaces, in that any statement about their geometry has a "modular" interpretation, in terms of the original classification problem. As a result, when one investigates them one can often reach much further than one can with other

spaces. Moduli spaces such as the moduli of ELLIPTIC CURVES (which we discuss below) play a central role in a variety of areas that have no immediate link to the geometry being classified, in particular in ALGEBRAIC NUMBER THEORY and ALGEBRAIC TOPOLOGY. Moreover, the study of moduli spaces has benefited tremendously in recent years from interactions with physics (in particular with string theory). These interactions have led to a variety of new questions and new techniques.

1 Warmup: The Moduli Space of Lines in the Plane

Let us begin with a problem that looks rather simple, but that nevertheless illustrates many of the important ideas of moduli spaces.

Problem. Describe the collection of all lines in the real plane $\mathbb{R}^2$ that pass through the origin.

To save writing, we are using the word "line" to mean "line that passes through the origin." This classification problem is easily solved by assigning to each line L an essential parameter, or *modulus*, a quantity which we can calculate for each line and which will help us tell different lines apart. All we have to do is take standard Cartesian coordinates x, y on the plane and measure the angle $\theta(L)$ between the line L and the x-axis, taken in counterclockwise fashion. We find that the possible values of θ are those for which $0 \leqslant \theta < \pi$, and that for every such θ there is exactly one line L that makes an angle of θ with the x-axis. So as a *set*, we have a complete solution to our classification problem: the set of lines L, known as $\mathbb{RP}^1$—the real projective line—is in one-to-one correspondence with the half-open interval $[0, \pi)$.

However, we are seeking a *geometric* solution to the classification problem. What does this entail? We have a natural notion of when two lines are near each other, which our solution should capture—in other words, the collection of lines has a natural TOPOLOGY. So far, our solution does not reflect the fact that lines L for which the angle $\theta(L)$ is close to π are almost horizontal: they are therefore close to the x-axis (for which $\theta = 0$) and to the lines L with $\theta(L)$ close to zero. We need to find some way of "wrapping round" the interval $[0, \pi)$ so that π becomes close to 0.

One way to do this is to take not the half-open interval $[0, \pi)$ but the closed interval $[0, \pi]$, and then to "identify" the points 0 and π. (This idea can easily be made formal by defining an appropriate EQUIVALENCE RELATION.) If π and 0 are regarded as the same, then numbers close to π are close to numbers close to 0. This is a way of saying that if you attach the two ends of a line segment together, then, topologically speaking, you obtain a circle.

A more natural way of achieving the same end is suggested by the following geometric construction of $\mathbb{RP}^1$. Consider the unit circle $S^1 \subset \mathbb{R}^2$. To each point $s \in S^1$, there is an obvious way of assigning a line $L(s)$: take the line that passes through s and the origin. Thus, we have a *family of lines parametrized by* S^1, that is, a map (or function) $s \mapsto L(s)$ that takes points in S^1 to lines in our set $\mathbb{RP}^1$. What is important about this is that we already know what it means for two points in S^1 to be close to each other, and the map $s \mapsto L(s)$ is continuous. However, this map is a two-to-one function rather than a bijection, since s and $-s$ always give the same line. To remedy this, we can identify each s in the circle S^1 with its antipodal point $-s$. We then have a one-to-one correspondence between $\mathbb{RP}^1$ and the resulting quotient space (which again is topologically a circle), and this correspondence is continuous in both directions.

The key feature of the space $\mathbb{RP}^1$, considered as the *moduli space* of lines in the plane, is that it captures the ways in which lines can *modulate*, or vary continuously in families. But when do families of lines arise? A good example is provided by the following construction. Whenever we have a continuous curve $C \subset \mathbb{R}^2 \setminus 0$ in the plane, we can assign to each point c in C the line $L(c)$ that passes through 0 and c. This gives us a family of lines parametrized by C. Moreover, the function that takes c to $L(c)$ is a continuous function from C to $\mathbb{RP}^1$—so the parametrization is a continuous one.

Suppose, for example, that C is a copy of $\mathbb{R}$ realized as the set of points $(x, 1)$ at height 1. Then the map from C to $\mathbb{RP}^1$ gives an isomorphism between $\mathbb{R}$ and the set $\{L : \theta(L) \neq 0\}$, which is the subset of $\mathbb{RP}^1$ consisting of all lines apart from the x-axis. Put more abstractly, we have an intuitive notion of what it means for a collection of lines through the origin to depend continuously on some parameters, and this notion is captured precisely by the geometry of $\mathbb{RP}^1$:

for instance, if someone tells me they have a continuous 37-parameter family of lines in $\mathbb{R}^2$, this is the same as saying that they have a continuous map from $\mathbb{R}^{37}$ to $\mathbb{RP}^1$, which sends a point $v \in \mathbb{R}^{37}$ to a line $L(v) \in \mathbb{RP}^1$. (More concretely, we could say that the real function $v \mapsto \theta(L(v))$ on $\mathbb{R}^{37}$ is continuous away from the locus where θ is close to π. Near this locus we could use instead the function ϕ that measures the angle from the y-axis.)

1.1 Other Families

The idea of families of lines leads to various other geometric structures on the space $\mathbb{RP}^1$, and not just its topological structure. For example, we have the notion of a *differentiable* family of lines in the plane, which is a family of lines for which the angles vary differentiably. (The same ideas apply if we replace "differentiable" by "measurable," "C^∞," "real analytic," etc.) To parametrize such a family appropriately, we would like $\mathbb{RP}^1$ to be a differentiable manifold, so that we can calculate derivatives of functions on it. Such a structure on $\mathbb{RP}^1$ can be specified by using the angle functions θ and ϕ defined in the previous section. The function θ gives us a coordinate for lines that are not too close to the x-axis, and ϕ gives us a coordinate for lines that are not too close to the y-axis. We can calculate derivatives of functions on $\mathbb{RP}^1$ by writing them in terms of these coordinates. One can justify this differentiable structure on $\mathbb{RP}^1$ by checking that for any differentiable curve $C \subset \mathbb{R}^2 \setminus 0$ the map $c \mapsto L(c)$ comes out as differentiable. This means that if $L(c)$ is not close to the x-axis, then the function $x \mapsto \theta(L(x))$ is differentiable at $x = c$, and similarly for ϕ and the y-axis. The functions $x \mapsto \theta(L(x))$ and $\mapsto \phi(L(x))$ are called *pullbacks*, because they are the result of converting, or "pulling back," θ and ϕ from functions defined on $\mathbb{RP}^1$ to functions defined on C.

We now can state the fundamental property of $\mathbb{RP}^1$ as a differentiable space.

A differentiable family of lines in $\mathbb{R}^2$ parametrized by a differentiable manifold X is the same thing as a function from X to $\mathbb{RP}^1$, taking a point x to a line $L(x)$, such that the pullbacks $x \mapsto \theta(L(x))$ and $x \mapsto \phi(L(x))$ of the functions θ, ϕ are differentiable functions.

We say that $\mathbb{RP}^1$ (with its differentiable structure) is the *moduli space* of (differentiably varying families of) lines in $\mathbb{R}^2$. This means that $\mathbb{RP}^1$ carries the *universal*

differentiable family of lines. From the very definition, we have assigned to each point of $\mathbb{RP}^1$ a line in $\mathbb{R}^2$, and these lines vary differentiably as we vary the point. The above assertion says that *any* differentiable family of lines, parametrized by a space X, is described by giving a map $f : X \to \mathbb{RP}^1$ and assigning to $x \in X$ the line $L(f(x))$.

1.2 Reformulation: Line Bundles

It is interesting to reformulate the notion of a (continuous or differentiable) family of lines as follows. Let X be a space and let $x \mapsto L(x)$ be an assignment of lines to points in X. For each point $x \in X$, we place a copy of $\mathbb{R}^2$ at x; in other words, we consider the Cartesian product $X \times \mathbb{R}^2$. We may now visualize the line $L(x)$ as living in the copy of $\mathbb{R}^2$ that lies over x. This gives us a continuously varying collection of lines $L(x)$ parametrized by $x \in X$, otherwise known as a *line bundle* over X. Moreover, this line bundle is embedded in the "trivial" vector bundle $X \times \mathbb{R}^2$, which is the constant assignment that takes each x to the plane $\mathbb{R}^2$. In the case when X is $\mathbb{RP}^1$ itself, we have a "tautological" line bundle: to each point $s \in \mathbb{RP}^1$, which we can think of as a line L_s in $\mathbb{R}^2$, it assigns that very same line L_s.

Proposition. *For any topological space X there is a natural bijection between the following two sets:*

  (i) *the set of continuous functions $f : X \to \mathbb{RP}^1$; and*
  (ii) *the set of line bundles on X that are contained in the trivial vector bundle $X \times \mathbb{R}^2$.*

This bijection sends a function f to the corresponding pullback of the tautological line bundle on $\mathbb{RP}^1$. That is, the function f is mapped to the line bundle $x \mapsto L_{f(x)}$. (This is a pullback, because it converts L from a function defined on $\mathbb{RP}^1$ to a function defined on X.)

Thus, the space $\mathbb{RP}^1$ carries the *universal* line bundle that sits in the trivial $\mathbb{R}^2$ bundle—any time we have a line bundle sitting in the trivial $\mathbb{R}^2$ bundle, we can obtain it by pulling back the universal (tautological) example on $\mathbb{RP}^1$. (VECTOR BUNDLES are discussed in more detail in Part III.)

1.3 Invariants of Families

Associated with any continuous function f from the circle S^1 to itself is an integer known as its *degree.* Roughly speaking, the degree of f is the number of times $f(x)$ goes round the circle when x goes round once. (If it goes backwards n times, then we say that the degree is $-n$.) Another way to think of the degree is as the number of times a typical point in S^1 is passed by $f(x)$ as x goes round the circle, where we count this as $+1$ if it is passed in the counterclockwise direction and -1 if it is passed in the clockwise direction.

Earlier, we showed that the circle S^1, which we obtained by identifying the end points of the closed interval $[0, \pi]$, could be used to parametrize the moduli space $\mathbb{RP}^1$ of lines. Combining this with the notion of degree, we can draw some interesting conclusions. In particular, we can define its so-called "winding number." Suppose that we are given a continuous function γ from the circle S^1 into the plane $\mathbb{R}^2$ and suppose that it avoids 0. The image of this map will be a closed loop C (which may cross itself). This defines for us a map from S^1 to itself: first do γ to obtain a point c in C, then work out $L(c)$, which belongs to $\mathbb{RP}^1$, and finally use the parametrization of $\mathbb{RP}^1$ to associate with $L(c)$ a point in S^1 again. The degree of the resulting composite map will be *twice* the number of times that γ, and hence C, winds around 0, so half this number is defined to be the winding number of γ.

More generally, given a family of lines in $\mathbb{R}^2$ parametrized by some space X, we would like to measure the "manner in which X winds around the circle." To be precise, given a function ϕ from X to $\mathbb{RP}^1$, which defines the parametrized family of lines, we would like to be able to say, for any map $f : S^1 \to X$, what the winding number is of the composition ϕf, which takes a point x in S^1 to its image $f(x)$ in X and from there to the corresponding line $\phi(f(x))$ in the family. Thus, the map ϕ gives us a way of assigning to each function $f : S^1 \to X$ an integer, the winding number of ϕf. The way this assignment works does not change if ϕ is continuously deformed: that is, it is a topological invariant of ϕ. What it does depend on is the class that ϕ belongs to in the first COHOMOLOGY GROUP of X, $\mathrm{H}^1(X, \mathbb{Z})$. Equivalently, to any line bundle on a space X which is contained in the trivial $\mathbb{R}^2$-bundle, we have associated a cohomology class, known as the *Euler class* of the bundle. This

is the first example of a *characteristic class* for vector bundles. It demonstrates that if we understand the topology of moduli spaces of classes of geometric objects, then we can define topological invariants for families of those objects.

2   The Moduli of Curves and Teichmüller Spaces

We now turn our attention to perhaps the most famous examples of moduli spaces, the moduli spaces of curves, and their first cousins, the *Teichmüller spaces*. These moduli spaces are the geometric solution to the problem of classification of compact Riemann surfaces, and can be thought of as the "higher theory" of Riemann surfaces. The moduli spaces are "meaningful spaces," in that each of their points stands for a Riemann surface. As a result, any statement about their geometry tells us something about the geometry of Riemann surfaces.

We turn first to the objects. Recall that a *Riemann surface* is a topological surface X (connected and oriented) to which a *complex structure* has been given. Complex structures can be described in many ways, and they enable us to do complex analysis, geometry, and algebra on the surface X. In particular, they enable us to define HOLOMORPHIC (complex-analytic) and MEROMORPHIC FUNCTIONS on open subsets of X. To be precise, X is a two-dimensional MANIFOLD, but the charts are thought of as open subsets of $\mathbb{C}$ rather than of $\mathbb{R}$, and the maps that glue them together are required to be holomorphic. An equivalent notion is that of a *conformal structure* on X, which is the structure needed to make it possible to define angles between curves in X. Yet another important equivalent notion is that of *algebraic structure* on X, making X into a *complex-algebraic curve* (leading to the persistent confusion in terminology: a Riemann surface is two dimensional, and therefore a surface, from the point of view of topology or the real numbers, but one dimensional, and therefore a curve, from the point of view of complex analysis and algebra). An algebraic structure is what allows us to speak of polynomial, rational, or algebraic functions on X, and is usually specified by realizing X as the set of solutions to polynomial equations in complex projective space $\mathbb{CP}^2$ (or $\mathbb{CP}^n$).

In order to speak of a classification problem, let alone a moduli space, for Riemann surfaces we must next specify when we regard two Riemann surfaces as equivalent. (We postpone the discussion of the final ingredient, the notion of families of Riemann surfaces, to Section 2.2.) To do this, we must give a notion of *isomorphism* between Riemann surfaces: when should two Riemann surfaces X and Y be identified, or thought of as giving two equivalent realizations of the same underlying object of our classification? This issue was hidden in our toy example of classifying lines in the plane: there we simply identified two lines if and only if they were *equal* as lines in the plane. This naive option is not available to us with the more abstractly defined Riemann surfaces. If we considered Riemann surfaces realized concretely as subsets of some larger space—for example, as solution sets to algebraic equations in complex projective space—we could similarly choose to identify surfaces only if they were equal as subsets. However, this is too fine a classification for most applications—what we care about is the *intrinsic geometry* of Riemann surfaces, and not incidental features that result from the particular way we choose to realize them.

At the other extreme, we might choose to ignore the extra geometric structure that makes a surface into a Riemann surface. That is, we could identify two Riemann surfaces X and Y if they are topologically equivalent, or homeomorphic (the "coffee mug is a doughnut" perspective). The classification of compact Riemann surfaces up to topological equivalence is captured by a single positive integer, the genus g ("number of holes") of the surface. Any surface of genus zero is homeomorphic to the Riemann sphere $\mathbb{CP}^1 \simeq S^2$, any surface of genus one is homeomorphic to a torus $S^1 \times S^1$, and so on. Thus, in this case there is no issue of "modulation"—the classification is solved by giving a list of possible values of a single discrete invariant.

However, if we are interested in Riemann surfaces *as Riemann surfaces* rather than simply as topological manifolds, then this classification is too crude: it completely ignores the complex structure. We would now like to refine our classification to remedy this defect. To this end, we say that two Riemann surfaces X and Y are (conformally, or holomorphically) *equivalent* if there is a topological equivalence between them that preserves the geometry, i.e., a homeomorphism that

preserves the angles between curves, or takes holomorphic functions to holomorphic functions, or takes rational functions to rational functions. (These conditions are all equivalent.) Note that we still have at our disposal our discrete invariant—the genus of a surface. However, as we shall see, this invariant is not fine enough to distinguish between all inequivalent Riemann surfaces. In fact, it is possible to have families of inequivalent Riemann surfaces that are parametrized by *continuous* parameters (but we cannot make proper sense of this idea until we have said precisely what is meant by a family of Riemann surfaces). Thus, the next step is to fix our discrete invariant and to try to classify all the different isomorphism classes or Riemann surfaces with the same genus by assembling them in a natural geometric fashion.

An important step towards this classification is the *uniformization theorem*. This states that any SIMPLY CONNECTED Riemann surface is holomorphically isomorphic to one of the following three: the Riemann sphere $\mathbb{CP}^1$, the complex plane $\mathbb{C}$, or the upper half-plane $\mathbb{H}$ (equivalently, the unit disc D). Since the UNIVERSAL COVERING SPACE of any Riemann surface is a simply connected Riemann surface, the uniformization theorem provides an approach to classifying arbitrary Riemann surfaces. For instance, any COMPACT Riemann surface of genus zero is simply connected, and in fact homeomorphic to the Riemann sphere, so the uniformization theorem already solves our classification problem in genus zero: up to equivalence, $\mathbb{CP}^1$ is the *only* Riemann surface of genus zero, and so in this case the topological and conformal classifications agree.

2.1 Moduli of Elliptic Curves

Next, we consider Riemann surfaces whose universal cover is $\mathbb{C}$, which is the same as saying that they are quotients of $\mathbb{C}$. For example, we can look at a quotient of $\mathbb{C}$ by $\mathbb{Z}$, which means that we regard two complex numbers z and w as equivalent if $z - w$ is an integer. This has the effect of "wrapping $\mathbb{C}$ round" into a cylinder. Cylinders are not compact, but to get a compact surface we could take a quotient by $\mathbb{Z}^2$ instead: that is, we could regard z and w as equivalent if their difference is of the form $a + b$i, where a and b are both integers. Now $\mathbb{C}$ is wrapped round in two directions and the result is a torus with a complex (or, equivalently, conformal or algebraic) structure. This is a

compact Riemann surface of genus one. More generally, we can replace $\mathbb{Z}^2$ by any lattice L, regarding z and w as equivalent if $z - w$ belongs to L. (A *lattice* L in $\mathbb{C}$ is an additive subgroup of $\mathbb{C}$ with two properties. First, it is not contained in any line. Second, it is *discrete*, which means that there is a constant $d > 0$ such that the distance between any two points in L is at least d. Lattices are also discussed in Section **??** of THE GENERAL GOALS OF MATHEMATICAL RESEARCH in Part I. A *basis* for a lattice L is a pair of complex numbers u and v belonging to L such that every z in L can be written in the form $au + bv$ with a and b integers. Such a basis will not be unique: for example, if $L = \mathbb{Z} \oplus \mathbb{Z}$, then the obvious basis is $u = 1$ and $v = $ i, but $u = 1$ and $v = 1 + $ i would do just as well.) If we take a quotient of $\mathbb{C}$ by a lattice, then we again obtain a torus with complex structure. It turns out that any compact Riemann surface of genus one can be produced in this way.

From a topological point of view, any two tori are the same, but once we consider the complex structure we start to find that different choices of lattice may lead to different Riemann surfaces. Certain changes to L do *not* have an effect: for example, if we multiply a lattice L by some nonzero complex number λ, then the quotient surface $\mathbb{C}/L$ will not be affected. That is, $\mathbb{C}/L$ is naturally isomorphic to $\mathbb{C}/\lambda L$. Therefore, we need only worry about the difference between lattices when one is not a multiple of the other. Geometrically, this says that one cannot be obtained from the other by a combination of rotation and dilation.

Notice that by taking the quotient $\mathbb{C}/L$ we obtain not just a "naked" Riemann surface, but one equipped with an "origin"—a distinguished point $e \in E$, which is the image of the origin $0 \in \mathbb{C}$—in other words, an *elliptic curve*:

Definition. An elliptic curve (over $\mathbb{C}$) is a Riemann surface E of genus one, equipped with a marked point $e \in E$. Elliptic curves, up to isomorphism, are in bijection with lattices $L \subset \mathbb{C}$ up to rotation.

Remark. In fact, since $L \subset \mathbb{C}$ is a *subgroup* of the Abelian group $\mathbb{C}$, the elliptic curve $E = \mathbb{C}/L$ is naturally an Abelian group, with e as its identity element. This is an important motivation for keeping e as part of the data that defines an elliptic curve. A more subtle reason for remembering the location of e when we speak of E is that it helps us to define E more

uniquely. This is useful, because any surface E of genus one has lots of symmetries, or *automorphisms*: there is always a holomorphic automorphism of E taking any point x to any other given point y. (If we think of E as a group, these are achieved by translations.) Thus, if someone hands us another genus-one surface E', there may be no way to identify E and E', or there may be infinitely many ways: we can always compose a given isomorphism between them with a self-symmetry of E. As we will discuss later, automorphisms haunt almost every moduli problem, and are crucial when we consider the behavior of families. It is usually convenient to "rigidify" the situation somewhat, so that the possible isomorphisms between different objects are less "floppy" and more uniquely determined. In the case of elliptic curves, distinguishing the point e achieves this by reducing the symmetry of E. Once we do that, there is usually at most one way to identify two elliptic curves (one way, that is, that takes origin to origin).

We see that Riemann surfaces of genus one (with the choice of a marked point) can be described by concrete "linear algebra data": a lattice $L \subset \mathbb{C}$, or rather the equivalence class consisting of all nonzero scalar multiples λL of L. This is the ideal setting to study a classification, or moduli, problem. The next step is to find an explicit parametrization of the collection of all lattices, up to multiplication, and to decide in what sense we have obtained a geometric solution to the classification problem.

In order to parametrize the collection of lattices, we follow a procedure used for all moduli problems: first parametrize lattices together with the choice of some additional structure, and then see what happens when we forget this choice. For every lattice L we choose a basis $\omega_1, \omega_2 \in L$: that is, we represent L as the set of all integer combinations $a\omega_1 + b\omega_2$. We do this in an *oriented* fashion: we require that the *fundamental parallelogram* spanned by ω_1 and ω_2 is positively oriented. (That is, the numbers 0, ω_1, $\omega_1 + \omega_2$, and ω_2 list the vertices of the parallelogram in an counterclockwise order. From the geometric point of view of the elliptic curve E, L is the FUNDAMENTAL GROUP of E, and the orientation condition says that we generate L by two loops, or "meridians," $A = \omega_1$, $B = \omega_2$, which are oriented, in that their oriented intersection number $A \cap B$ is equal to $+1$ rather than -1.) Since we are interested in lattices only up to multiplication, we can

multiply L by a complex number so as to turn ω_1 into 1 and hence ω_2 into $\omega = \omega_2/\omega_1$. The orientation condition now says that ω is in the upper half-plane $\mathbb{H}$, i.e., its imaginary part is positive, $\operatorname{Im} \omega > 0$. Conversely, any complex number $\omega \in \mathbb{H}$ in the upper half-plane determines a unique oriented lattice $L = \mathbb{Z}1 \oplus \mathbb{Z}\omega$—that is, the set of all integer combinations $a + b\omega$ of 1 and ω—and no two of these lattices are related by a rotation.

What does this tell us about elliptic curves? We saw earlier that an elliptic curve is defined by a lattice L and an identity e. Now we have seen that if we give L some extra structure, namely an oriented basis, then we can parametrize it by a complex number $\omega \in \mathbb{H}$. This makes precise for us the "additional structure" that we want to place on elliptic curves. We say that a *marked* elliptic curve is an elliptic curve E, e together with the choice of an oriented basis ω_1, ω_2 for the associated lattice (fundamental group) L of E. The point is that any lattice has infinitely many different bases, which lead to many automorphisms of E. By "marking" one of these bases, we stop them being automorphisms.

2.2 Families and Teichmüller Space

With our new definition, we can summarize the earlier discussion by saying that marked elliptic curves are in bijection with points $\omega \in \mathbb{H}$ of the upper half-plane. The upper half-plane is, however, much more than just a *set* of points: it carries a host of geometric structures, in particular a topology and a complex structure. In what sense do these structures reflect geometric properties of marked elliptic curves? In other words, in what sense is the complex manifold $\mathbb{H}$, known in this context as the *Teichmüller space* $\mathcal{T}_{1,1}$ of genus-one Riemann surfaces with one marked point, a geometric solution to the problem of classifying marked elliptic curves?

In order to answer this question, we need the notion of a continuous family of Riemann surfaces, and also the notion of a complex-analytic family. A *continuous family of Riemann surfaces* parametrized by a topological space S—for example, by the circle S^1—is the assignment of a Riemann surface X_s to every point s of S, which "varies continuously." In our example of the moduli of lines in the plane, a continuous family of lines was characterized by the property that the angles

between the lines and the x- or y-axes defined continuous functions of the parameters. Geometrically defined collections of lines, such as those produced by a curve C in the plane, then gave rise to continuous families. More abstractly, a continuous family of lines defined a line *bundle* over the parameter space. A good criterion for a family of Riemann surfaces is likewise that any "reasonably defined" geometric quantity that we can calculate for every Riemann surface should vary continuously in the family. For example, a classical construction of Riemann surfaces of genus g comes from taking $4g$-gons and gluing opposite sides together. The resulting Riemann surface is fully determined by the edge-lengths and angles of the polygon. Therefore, a continuous family of Riemann surfaces described in this fashion should be precisely a family such that the edge-lengths and angles give *continuous* functions of the parameter set.

In more abstract topological terms, if we have a collection $\{X_s, \ s \in S\}$ of Riemann surfaces depending on points in a space S and we wish to make it into a continuous family, then we should give the union $\bigcup_{s \in S} X_s$ itself the structure of a topological space $\mathcal{X}$, which should simultaneously extend the topology on each individual X_s. The result is called a *Riemann surface bundle*. Associated with $\mathcal{X}$ is the map that takes each point x to the particular s for which x belongs to X_s. We should demand that this map is continuous, and perhaps more (it could be a fibration, or fiber bundle). This definition has the advantage of great flexibility. For example, if S is a complex manifold, then in just the same way we can speak of a *complex-analytic family of Riemann surfaces* $\{X_s, \ s \in S\}$ parametrized by S: now we ask for the union of the X_s to carry not just a topology but a complex structure (i.e., it should form a complex manifold), extending the complex structure on the fibers and mapping holomorphically to the parameter set. The same holds with "complex-analytic" replaced by "algebraic." These abstract definitions have the property that if our Riemann surfaces are described in a concrete way—cut out by equations, glued from coordinate patches, etc.—then the coefficients of our equations or gluing data will vary as complex-analytic functions in our family precisely when the family is complex analytic (and likewise for continuous or algebraic families).

As a reality check, note that a (continuous, analytic, or other) family of Riemann surfaces parametrized by a single *point* $s = S$ is indeed just a single Riemann surface X_s. Just as in this simple case we wish to consider Riemann surfaces only up to equivalence, so there is a notion of equivalence or isomorphism of two analytic families $\{X_s\}$ and $\{X'_s\}$ parametrized by the same space S. We simply regard the families as equivalent if the surfaces X_s and X'_s are isomorphic for every s, and if the isomorphism depends analytically on s.

Armed with the notion of family, we can now formulate the characteristic property that the upper half-plane possesses when we think of it as the moduli space of marked elliptic curves. We define a continuous or analytic family of marked elliptic curves to be a family where the underlying genus-one surfaces vary continuously or analytically, while the choice of basepoint $e_s \in E_s$ and the basis of the lattice L_s vary continuously.

The upper half-plane $\mathbb{H}$ plays a role for marked elliptic curves that is similar to the role played by $\mathbb{RP}^1$ for lines in the plane. The following theorem makes this statement precise.

Theorem. *For any topological space S, there is a one-to-one correspondence between continuous maps from S to $\mathbb{H}$ and isomorphism classes of continuous families of marked elliptic curves parametrized by S. Similarly, there is a one-to-one correspondence between analytic maps from any complex manifold S to $\mathbb{H}$ and isomorphism classes of analytic families of marked elliptic curves parametrized by S.*

If we apply the theorem in the case where S is a single point, it simply tells us that the points of $\mathbb{H}$ are in bijection with the isomorphism classes of marked elliptic curves, as we already knew. However, it contains more information: it says that $\mathbb{H}$, with its topology and complex structure, *embodies the structure* of marked elliptic curves and the ways in which they can modulate. At the other extreme, we could take $S = \mathbb{H}$ itself, mapping S to $\mathbb{H}$ by the identity map. This expresses the fact that $\mathbb{H}$ itself carries a family of marked elliptic curves, i.e., the collection of Riemann surfaces defined by $\omega \in \mathbb{H}$ fit together into a complex manifold fibering over $\mathbb{H}$ with elliptic curve fibers. This family is called the *universal family*, since by the theorem any family is "deduced" (or pulled back) from this one universal example.

2.3   From Teichmüller Space to Moduli Space

We have arrived at a complete and satisfying picture for the classification of elliptic curves when we choose in addition a marking (that is, an oriented basis of the associated lattice $L = \pi_1(E)$). What can we say about elliptic curves themselves, without the choice of marking? We somehow need to "forget" the marking, by regarding two points of $\mathbb{H}$ as equivalent if they correspond to two different markings of the same elliptic curve.

Now, given any two bases of the group (or lattice) $\mathbb{Z} \oplus \mathbb{Z}$, there is an invertible 2×2 matrix with integer entries that takes one basis to the other. If the two bases are *oriented*, then this matrix will have determinant 1, which means that it is an element

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2\,\mathbb{Z}$$

of the group of invertible unimodular matrices over $\mathbb{Z}$. Similarly, given any two oriented bases (ω_1, ω_2) and (ω_1', ω_2') of a lattice L, which can be thought of as oriented identifications of L with $\mathbb{Z} \oplus \mathbb{Z}$, there is a matrix $A \in \mathrm{SL}_2\,\mathbb{Z}$ such that $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$. If we now consider the normalized bases $(1, \omega)$ and $(1, \omega')$, where $\omega = \omega_1/\omega_2$ and $\omega' = \omega_1'/\omega_2'$, then we obtain a transformation of the upper half-plane. It is given by the formula

$$\omega' = \frac{a\omega + b}{c\omega + d}.$$

That is, the group $\mathrm{SL}_2\,\mathbb{Z}$ is acting on the upper half-plane by linear fractional (or Möbius) transformations with integer coefficients, and two points in the upper half-plane correspond to the same elliptic curve if one can be turned into the other by means of such a transformation. If this is the case, then we should regard the two points as equivalent: that is how we formalize the idea of "forgetting" the marking. Note also that the scalar matrix $-\,\mathrm{Id}$ in $\mathrm{SL}_2\,\mathbb{Z}$, which negates both ω_1 and ω_2, acts trivially on the upper half-plane, so that we in fact get an action of $\mathrm{PSL}_2\,\mathbb{Z} = \mathrm{SL}_2\,\mathbb{Z}/\{\pm\,\mathrm{Id}\}$ on $\mathbb{H}$.

So we come to the conclusion that *elliptic curves (up to isomorphism) are in bijection with orbits of* $\mathrm{PSL}_2\,\mathbb{Z}$ *on the upper half-plane, or equivalently with points of the quotient space* $\mathbb{H}/\,\mathrm{PSL}_2\,\mathbb{Z}$. This quotient space has a natural quotient topology, and in fact can be given a complex-analytic structure, which, it turns out, identifies it with the complex plane $\mathbb{C}$ itself. To see this one uses the classical MODULAR FUNCTION $j(z)$, a complex-analytic function on $\mathbb{H}$ which is invariant under the modular group $\mathrm{PSL}_2\,\mathbb{Z}$ and which therefore defines a natural coordinate $\mathbb{H}/\,\mathrm{PSL}_2\,\mathbb{Z} \to \mathbb{C}$.

It appears that we have solved the moduli problem for elliptic curves: we have a topological, and even complex-analytic, space $\mathfrak{M}_{1,1} = \mathbb{H}/\,\mathrm{PSL}_2\,\mathbb{Z}$ whose points are in one-to-one correspondence with isomorphism classes of elliptic curves. This already qualifies $\mathfrak{M}_{1,1}$ as the *coarse moduli space* for elliptic curves, which means it is as good a moduli space as we can hope for. However, $\mathfrak{M}_{1,1}$ fails an important test for a moduli space that $\mathcal{T}_{1,1}$ passed (as we saw in Section 2.2): it is *not* true, even for the circle $S = S^1$, that every continuous family of elliptic curves over S corresponds to a map from S to $\mathfrak{M}_{1,1}$.

The reason for this failure is the problem of automorphisms. These are equivalences from E to itself, that is, complex-analytic maps from E to E that preserve the basepoint e. Equivalently, they are given by complex-analytic self-maps of $\mathbb{C}$ that preserve 0 and the lattice L. Such a map must be a rotation: that is, multiplication by some complex number λ of modulus 1. It is easy to check that for most lattices L in the plane, the only rotation that sends L to itself is multiplication by $\lambda = -1$. Note that this is the same -1 that we quotiented out by to pass from $\mathrm{SL}_2\,\mathbb{Z}$ to $\mathrm{PSL}_2\,\mathbb{Z}$. However, there are two special lattices that have greater symmetry. These are the *square lattice* $L = \mathbb{Z} \cdot 1 \oplus \mathbb{Z} \cdot \mathrm{i}$, corresponding to the fourth root of unity i, and the *hexagonal lattice* $L = \mathbb{Z} \cdot 1 \oplus \mathbb{Z} \cdot \mathrm{e}^{2\pi\mathrm{i}/6}$, corresponding to a sixth root of unity. (Note that the hexagonal lattice is also represented by the point $\omega = \mathrm{e}^{2\pi\mathrm{i}/3}$.) The square lattice, which corresponds to the elliptic curve formed by gluing the opposite sides of a square, has as its symmetries the group $\mathbb{Z}/4\mathbb{Z}$ of rotational symmetries of the square. The hexagonal lattice, which corresponds to the elliptic curve formed by gluing the opposite sides of a regular hexagon, has as its symmetries the group $\mathbb{Z}/6\mathbb{Z}$ of rotational symmetries of a hexagon.

We see that the number of automorphisms of an elliptic curve jumps discontinuously at the special points $\omega = \mathrm{i}$ and $\omega = \mathrm{e}^{2\pi\mathrm{i}/6}$. This already suggests that something might be wrong with $\mathfrak{M}_{1,1}$ as a moduli space. Note that we avoided this problem with

the moduli $\mathcal{T}_{1,1}$ of *marked* elliptic curves, since there are no automorphisms of an elliptic curve that also preserve the marking. Another place we might have observed this problem with $\mathfrak{M}_{1,1}$ is when we passed to the quotient $\mathbb{H}/\operatorname{PSL}_2 \mathbb{Z}$. We avoided the automorphism $\lambda = -1$ by quotienting by $\operatorname{PSL}_2 \mathbb{Z}$ rather than $\operatorname{SL}_2 \mathbb{Z}$. However, the two special points i and $\mathrm{e}^{2\pi\mathrm{i}/6}$ are preserved by integer Möbius transformations of $\mathbb{H}$ other than the identity, and they are the only points with that property. This means that the quotient $\mathbb{H}/\operatorname{PSL}_2 \mathbb{Z}$ naturally comes with conical singularities at the points corresponding to these two orbits: one looks like a cone with angle π, and the other like a cone with angle $\frac{2}{3}\pi$. (To see why this is plausible, imagine the following simpler instance of the same phenomenon. If for every complex number z you identify z with $-z$, then the result is to wrap the complex plane round into a cone with a singularity at 0. The reason 0 is singled out is that it is preserved by the transformation $z \mapsto -z$. Here the angle would be π because the identification of points is two-to-one away from the singularity and π is half of 2π.) It is possible to massage these singularities away using the j-function, but they are indicating a basic difficulty.

So why do automorphisms form an obstacle to the existence of "good" moduli spaces? We can demonstrate the difficulty by considering an interesting continuous family of marked elliptic curves parametrized by the circle $S = S^1$. Let $E(\mathrm{i})$ be the "square" elliptic curve that we considered earlier, based on the lattice of integer combinations of 1 and i. Next, for every t between 0 and 1, let E_t be a copy of $E(\mathrm{i})$. Thus, we have taken the constant, or "trivial," family of elliptic curves over the closed unit interval $[0, 1]$, where every curve in the family is $E(\mathrm{i})$. Now we identify the elliptic curves at the two ends of this family, not in the obvious way, but by using the automorphism given by a 90° rotation, or multiplication by i. This means that we are looking at the family of elliptic curves over the circle where each member of the family is a copy of the elliptic curve $E(\mathrm{i})$, but these copies twist by 90° as we go around the circle.

It is easy to see that there is no way to capture this family of elliptic curves by means of a map from S^1 to the space $\mathfrak{M}_{1,1}$. Since all of the members of the family are isomorphic, each point of the circle should map to the same point in $\mathfrak{M}_{1,1}$ (the equivalence class of i in $\mathbb{H}$). But the constant map $S^1 \to \{\mathrm{i}\} \in \mathfrak{M}_{1,1}$ classifies

the *trivial* family $S^1 \times E_i$ of elliptic curves over S^1, that is, the family where every curve is equal to $E(\mathrm{i})$ but the curves *do not* twist as we go around! Thus, there are more families of elliptic curves than there are maps to $\mathfrak{M}_{1,1}$—the quotient space $\mathbb{H}/\operatorname{PSL}_2 \mathbb{Z}$ cannot handle the complications caused by automorphisms. A variant of this construction applies to complex-analytic families with S^1 replaced by $\mathbb{C}^\times$. This is a very general phenomenon in moduli problems: whenever objects have nontrivial automorphisms, we can imitate the construction above to get nontrivial families over an interesting parameter set, all of whose members are the same. As a result, they cannot be classified by a map to the set of all isomorphism classes.

What do we do about this problem? One approach is to resign ourselves to having coarse moduli spaces, which have the right points and right geometry but do not quite classify arbitrary families. Another approach is the one that leads to $\mathcal{T}_{1,1}$: we can fix markings of one kind or another, which "kill" all automorphisms. In other words, we choose enough extra structure on our objects so that there do not remain any (nontrivial) automorphisms that preserve all this decoration. In fact, one can be far more economical than picking a basis of the lattice L and obtaining the infinite covering $\mathcal{T}_{1,1}$ of $\mathfrak{M}_{1,1}$: one can fix a basis of L only up to some congruence (for example, of $L/2L$). Finally, we can simply learn to come to terms with the automorphisms, keeping them as part of the data, resulting in "spaces" where points have internal symmetries. This is the notion of an ORBIFOLD, or STACK, which is flexible enough to deal with essentially all moduli problems.

3 Higher-Genus Moduli Spaces and Teichmüller Spaces

We would now like to generalize as much as possible of the picture of elliptic curves and their moduli to higher-genus Riemann surfaces. For each g we would like to define a space $\mathfrak{M}_g$, called the *moduli space of curves of genus g*, that classifies compact Riemann surfaces of genus g and tells us how they modulate. Thus, the points of $\mathfrak{M}_g$ should correspond to our objects, compact Riemann surfaces of genus g, or, to be more accurate, equivalence classes of such surfaces, where two surfaces are considered

to be equivalent if there is a complex-analytic iso-
morphism between them. In addition, we would like
$\mathfrak{M}_g$ to do the best it can to embody the structure
of continuous families of genus-g surfaces. Likewise,
there are spaces $\mathfrak{M}_{g,n}$ parametrizing "n-punctured"
Riemann surfaces of genus g. This means we consider
not "bare" Riemann surfaces, but Riemann surfaces
together with a "decoration" or "marking" by n dis-
tinct labeled points (punctures). Two of these are con-
sidered to be equivalent if there is a complex-analytic
isomorphism between them that takes punctures to
punctures and preserves labels. Since there are Rie-
mann surfaces with automorphisms, we do not expect
$\mathfrak{M}_g$ to be able to classify all families of Riemann sur-
faces: that is, we will expect examples similar to the
twisted square-lattice construction discussed earlier.
However, if we consider Riemann surfaces with enough
extra markings, then we will be able to obtain a mod-
uli space in the strongest sense. One way to choose
such markings is to consider $\mathfrak{M}_{g,n}$ with n large enough
(for fixed g). Another approach will be to mark gen-
erators of the FUNDAMENTAL GROUP, leading to the
Teichmüller spaces $\mathcal{T}_g$ and $\mathcal{T}_{g,n}$. We now outline this
process.

To construct the space $\mathfrak{M}_g$, we return to the uni-
formization theorem. Any compact surface X of genus
$g > 1$ has as its universal cover the upper half-plane $\mathbb{H}$,
so it is represented as a quotient $X = \mathbb{H}/\Gamma$, where Γ
is a representation of the fundamental group of X as a
subgroup of conformal self-maps of $\mathbb{H}$. The group of all
conformal automorphisms of $\mathbb{H}$ is $\mathrm{PSL}_2\,\mathbb{R}$, the group
of linear fractional transformations with real coeffi-
cients. The fundamental groups of all compact genus-
g Riemann surfaces are isomorphic to a fixed abstract
group Γ_g, with $2g$ generators A_i, B_i $(i = 1, \ldots, g)$
and one relation: that the product of all commutators
$A_i B_i A_i^{-1} B_i^{-1}$ is the identity. A subgroup $\Gamma \subset \mathrm{PSL}_2\,\mathbb{R}$
that acts on $\mathbb{H}$ in such a way that the quotient $\mathbb{H}/\Gamma$
is a Riemann surface (technically, the action should
have no fixed points and should be properly discon-
tinuous) is known as a FUCHSIAN GROUP. Thus, the
analog of the representation of elliptic curves by lat-
tices $L \simeq \mathbb{Z} \oplus \mathbb{Z}$ in the plane is the representation of
higher-genus Riemann surfaces as $\mathbb{H}/\Gamma$, where Γ is a
Fuchsian group.

The Teichmüller space $\mathcal{T}_g$ of genus-g Riemann sur-
faces is the space that solves the moduli problem
for genus-g surfaces, when they come with a mark-

ing of their fundamental group. This means that our
objects are genus-g surfaces X plus a set of generators
A_i, B_i of $\pi_1(X)$, which give an isomorphism between
$\pi_1(X)$ and Γ_g, up to conjugation.[1] Our equivalences
are complex-analytic maps that preserve the mark-
ings. Finally, our continuous (respectively, complex-
analytic) families are continuous (complex-analytic)
families of Riemann surfaces with continuously vary-
ing markings of the fundamental group. In other
words, we are asserting the existence of a topological
space/complex manifold $\mathcal{T}_g$, with a complex-analytic
family of marked Riemann surfaces over it, and the
following strong property.

The characteristic property of $\mathcal{T}_g$. *For any topo-
logical space (respectively, complex manifold) S, there
is a bijection between continuous maps (respectively,
holomorphic maps) $S \to \mathcal{T}_g$ and isomorphism classes
of continuous (respectively, complex-analytic) families
of marked genus-g surfaces parametrized by S.*

3.1 Digression: "Abstract Nonsense"

It is interesting to note that, while we have yet to see
why such a space exists, it follows from general, non-
geometric principles—CATEGORY THEORY or "abstract
nonsense"—that it is completely and uniquely deter-
mined, both as a topological space and as a com-
plex manifold, by this characteristic property. In a
very abstract way, every topological space M can be
uniquely reconstructed from its set of points, the set
of paths between these points, the set of surfaces span-
ning these paths, and so on. To put it differently, we
can think of M as a "machine" that assigns to any
topological space S the set of continuous maps from S
to M. This machine is known as the "functor of points
of M." Similarly, a complex manifold M provides a
machine that assigns to any other complex manifold S
the set of complex-analytic maps from S to M. A curi-
ous discovery of category theory (the *Yoneda lemma*)
is that for very general reasons (having nothing to do
with geometry), these machines (or functors) uniquely
determine M as a space, or a complex manifold.

---

1. Note that while the fundamental group of X depends on the
choice of a basepoint, $\pi_1(X, x)$ and $\pi_1(X, y)$ may be identified by
choosing a path from x to y, and the different choices are related
by conjugation by a loop. Thus, if we are willing to identify sets of
generators A_i, B_i when they differ only by a conjugation, then we
can ignore the choice of a basepoint.

Any moduli problem in the sense we have described (giving objects, equivalences, and families) also gives such a machine, where to S we assign the set of all families over S, up to isomorphism. So *just by setting up the moduli problem* we have already uniquely determined the topology and complex structure on Teichmüller space. The interesting part then is to know whether or not there *actually exists* a space giving rise to the same machine we have constructed, whether we can construct it explicitly, and whether we can use its geometry to learn interesting facts about Riemann surfaces.

3.2 Moduli Spaces and Representations

Coming back to earth, we discover that we have a fairly concrete model of Teichmüller space at our disposal. Once we have fixed the marking $\pi_1(X) \simeq \Gamma_g$, we are simply looking at all ways to represent Γ_g as a Fuchsian subgroup of $\mathrm{PSL}_2\,\mathbb{R}$. Ignoring the Fuchsian condition for a moment, this means finding $2g$ real matrices (up to $\pm\,\mathrm{Id}$) $A_i, B_i \in \mathrm{PSL}_2\,\mathbb{R}$ satisfying the commutator relation of Γ_g. This gives an explicit set of (algebraic!) equations for the entries of the $2g$ matrices, which determine the space of all representations $\Gamma_g \to \mathrm{PSL}_2\,\mathbb{R}$. We must now quotient out by the action of $\mathrm{PSL}_2\,\mathbb{R}$ that simultaneously conjugates all $2g$ matrices to obtain the *representation variety* $\mathrm{Rep}(\Gamma_g, \mathrm{PSL}_2\,\mathbb{R})$. This is analogous to considering lattices in $\mathbb{C}$ up to rotation, and is motivated by the fact that the quotients of $\mathbb{H}$ by two conjugate subgroups of $\mathrm{PSL}_2\,\mathbb{R}$ will be isomorphic.

Once we have described the space of all representations of Γ_g into $\mathrm{PSL}_2\,\mathbb{R}$, we can then single out Teichmüller space as the subset of the representation variety that consists of Fuchsian representations of Γ_g into $\mathrm{PSL}_2\,\mathbb{R}$. Luckily this subset is *open* in the representation variety, which gives a nice realization of $\mathcal{T}_g$ as a topological space—in fact, $\mathcal{T}_g$ is homeomorphic to $\mathbb{R}^{6g-6}$. (This can be seen very explicitly in terms of the *Fenchel–Nielsen* coordinates, which parametrize a surface in $\mathcal{T}_g$ via a cut-and-paste procedure involving $3g - 3$ lengths and $3g - 3$ angles.) We may now try to "forget" the marking $\pi_1(X) \cong \Gamma_g$, to obtain the moduli space $\mathfrak{M}_g$ of unmarked Riemann surfaces. In other words, we would like to take $\mathcal{T}_g$ and identify any two points that represent the same underlying Riemann surface with different markings. This identification is achieved by the action of a group, the genus-g *mapping class group* MCG_g or *Teichmüller modular group*, on $\mathcal{T}_g$, which generalizes the modular group $\mathrm{PSL}_2\,\mathbb{Z}$ that acts on $\mathbb{H} = \mathcal{T}_{1,1}$. (The mapping class group is defined as the group of all self-diffeomorphisms of a genus-g surface—remember that all such surfaces are topologically the same—modulo those diffeomorphisms that act trivially on the fundamental group.) As in the case of elliptic curves, Riemann surfaces with automorphisms correspond to points in $\mathcal{T}_g$ fixed by some subgroup of MCG_g, and give rise to singular points in the quotient $\mathfrak{M}_g = \mathcal{T}_g/\mathrm{MCG}_g$.

Representation varieties, or moduli spaces of representations, are an important and concrete class of moduli spaces that arise throughout geometry, topology, and number theory. Given any (discrete) group Γ, we ask (for example) for a space that parametrizes homomorphisms of Γ into the group of $n \times n$ matrices. The notion of equivalence is given by conjugation by GL_n, and that of families by continuous (or analytic, or algebraic, etc.) families of matrices. This problem is interesting even when the group Γ is $\mathbb{Z}$. Then we are simply considering invertible $n \times n$ matrices (the image of $1 \in \mathbb{Z}$) up to conjugacy. It turns out that there is no moduli space for this problem, even in the coarse sense, unless we consider only "nice enough" matrices: for example, matrices that consist of only a single Jordan block. This is a good example of a ubiquitous phenomenon in moduli problems: one is often forced to throw out some "bad" (unstable) objects in order to have any chance of obtaining a moduli space. (See the paper by Mumford and Suominen (1972) for a detailed discussion.)

3.3 Moduli Spaces and Jacobians

The upper half-plane $\mathbb{H} = \mathcal{T}_{1,1}$, together with the action of $\mathrm{PSL}_2\,\mathbb{Z}$, gives an appealingly complete picture of the moduli problem for elliptic curves and its geometry. The same cannot be said, unfortunately, for the picture of $\mathcal{T}_g$ as an open subset of the representation variety. In particular, the representation variety does not even carry a natural complex structure, so we cannot see from this description the geometry of $\mathcal{T}_g$ as a complex manifold. This failure reflects some of the ways in which the study of moduli spaces is more complicated for genus greater than one. In particular, the moduli spaces of higher-genus surfaces are

not described purely by linear algebra plus data about orientation, as is the case in genus one.

Part of the blame for this complexity lies with the fact that the fundamental group $\Gamma_g \simeq \pi_1(X)$ $(g > 1)$ is no longer Abelian, and in particular it is no longer equal to the first homology group $\mathrm{H}_1(X, \mathbb{Z})$. A related problem is that X is no longer a group. A beautiful solution to this problem is given by the construction of the Jacobian $\mathrm{Jac}(X)$, which shares with elliptic curves the properties of being a torus (homeomorphic to $(S^1)^{2g}$), an Abelian group, and a complex (in fact complex-algebraic) manifold. (The Jacobian of an elliptic curve is the elliptic curve itself.) The Jacobian captures the "Abelian" or "linear" aspects of the geometry of X. There is a moduli space $\mathcal{A}_g$ for such complex-algebraic tori (known as *Abelian varieties*), which does share all of the nice properties and linear algebraic description of the moduli of elliptic curves $\mathfrak{M}_{1,1} = \mathcal{A}_1$. The good news—the Torelli theorem—is that by assigning to each Riemann surface X its Jacobian we embed $\mathfrak{M}_g$ as a closed, complex-analytic subset of $\mathcal{A}_g$. The *interesting* news—the Schottky problem—is that the image is quite complicated to characterize intrinsically. In fact, solutions to this problem have come from as far afield as the study of nonlinear partial differential equations!

3.4   Further Directions

In this section we give hints at some interesting questions about, and applications of, moduli spaces.

Deformations and degenerations.   Two of the main topics in moduli spaces ask which objects are very near to a given one, and what lies very far away. Deformation theory is the calculus of moduli spaces: it describes their infinitesimal structure. In other words, given an object, deformation theory is concerned with describing all its small perturbations (see Mazur (2004) for a beautiful discussion of this). At the other extreme, we can ask what happens when our objects degenerate? Most moduli spaces, for example the moduli of curves, are not compact, so there are families "going off to infinity." It is important to find "meaningful" compactifications of moduli spaces, which classify the possible degenerations of our objects. Another advantage of compactifying moduli spaces is that we can then calculate integrals over the completed space. This is crucial for the next item.

Invariants from moduli spaces.   An important application of moduli spaces in geometry and topology is inspired by quantum field theory, where a particle, rather than follow the "best" classical path between two points, follows all paths with varying probabilities. Classically, one calculates many topological invariants by picking a geometric structure (such as a metric) on a space, calculating some quantity using this structure, and finally proving that the result of the calculation did not depend on the structure we chose. The new alternative is to look at *all* such geometric structures, and integrate some quantity over the space of all choices. The result, if we can show convergence, will manifestly not depend on any choices. String theory has given rise to many important applications of this idea, in particular by giving a rich structure to the collection of integrals obtained in this way. Donaldson and Seiberg–Witten theories use this philosophy to give topological invariants of four-manifolds. Gromov–Witten theory applies it to the topology of symplectic manifolds, and to counting problems in algebraic geometry, such as, How many rational plane curves of degree 5 pass through 14 points in general position? (Answer: 87 304.)

Modular forms.   One of the most profound ideas in mathematics, the LANGLANDS PROGRAM, relates number theory to function theory (harmonic analysis) on very special moduli spaces, generalizing the moduli space of elliptic curves. These moduli spaces (Shimura varieties) are expressible as quotients of symmetric spaces (such as $\mathbb{H}$) by arithmetic groups (such as $\mathrm{PSL}_2\,\mathbb{Z}$). MODULAR and AUTOMORPHIC FORMS are special functions on these moduli spaces, described by their interaction with the large symmetry groups of the spaces. This is an extremely exciting and active area of mathematics, which counts among its recent triumphs the proof of Fermat's last theorem and the Shimura–Taniyama–Weil conjecture (Wiles, Taylor–Wiles, Breuil–Conrad–Diamond–Taylor).

Further Reading

For further reading, historical accounts and bibliographies on moduli spaces, the following articles are highly recommended.

A beautiful and accessible overview of moduli spaces, with an emphasis on the notion of deforma-

tions, is given by Mazur (2004). The articles by Hain (2000) and Looijenga (2000) give excellent introductions to the study of the moduli spaces of curves, perhaps the oldest and most important of all moduli problems. The article by Mumford and Suominen (1972) introduces the key ideas underlying the study of moduli spaces in algebraic geometry.

Hain, R. 2000. Moduli of Riemann surfaces, transcendental aspects. In *School on Algebraic Geometry, Trieste, 1999*, pp. 293–353. ICTP Lecture Notes Series, No. 1. Trieste: The Abdus Salam International Centre for Theoretical Physics.

Looijenga, E. 2000. A minicourse on moduli of curves. In *School on Algebraic Geometry, Trieste, 1999*, pp. 267–291. ICTP Lecture Notes Series, No. 1. Trieste: The Abdus Salam International Centre for Theoretical Physics.

Mazur, B. 2004. Perturbations, deformations and variations (and "near–misses") in geometry. Physics and number theory. *Bulletin of the American Mathematical Society* 41(3):307–336.

Mumford, D. and K. Suominen. 1972. Introduction to the theory of moduli. In *Algebraic Geometry, Oslo, 1970: Proceedings of the Fifth Nordic Summer School in Mathematics* (ed. F. Oort), pp. 171–222. Groningen: Wolters-Noordhoff.