Surfaces and their Geometry

The geometric perspective on (non-abelian) groups usually stems from how they act on spaces. It will be convenient to focus on the modular group, and this example will lead us to discussing both conformal and algebraic geometry.

1 Introduction

The hyperbolic plane H, having both a rich geometry and many connections to number theory, will be a good choice for the space being acted on, and it will be convenient to choose the upper half-plane model consisting of the complex numbers τ with $\operatorname{Im}\tau > 0$. The metric imposed on H can be expressed as

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

meaning that arc lengths are computed with this choice of ds. An interesting case to consider is the modular group $\mathrm{PSL}_2(\mathbb{Z})$ with action given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tau = \frac{a\tau + b}{c\tau + d} \qquad \operatorname{Im}\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{(ad - bc)\operatorname{Im}\tau}{|c\tau + d|^2}$$

Such transformations are called fractional linear.

Definition. A lattice in $\mathbb{C}$ is an additive group $\Lambda = \omega_1\,\mathbb{Z} + \omega_2\,\mathbb{Z}$ where the periods ω_1 and ω_2 are linearly independent over the reals.

Lemma. Two such pairs $\alpha, \beta \in \mathbb{C}^2$ generate the same lattice iff

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

for some matrix in $\mathrm{GL}_2(\mathbb{Z})$.

Proof. The reverse implication follows from the matrix being invertible, so it remains to show the forward implication. Assuming they generate the same lattice, we can express the α_i as an integer linear combination of β_i and vice versa. This corresponds to an invertible matrix (over the integers). □

A common theme for groups of geometric interest is that of being finitely generated, and $\mathrm{SL}_2(\mathbb{Z})$ is no exception.

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Theorem. The matrices S and T generate $\mathrm{SL}_2(\mathbb{Z})$.

Proof. Let G be the subgroup generated by S and T.

$$S \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix} \qquad T^n \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+nc & b+nd \\ c & d \end{pmatrix}$$

We will show that for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ there is some $g \in G$ so that $g\gamma = \pm T^n$. This is sufficient because $S^2 = -I$. Thanks to Euclid's division algorithm, we know the lower left entry can zeroed by some $g \in G$.

$$g\gamma = \begin{pmatrix} \pm 1 & - \\ 0 & \pm 1 \end{pmatrix}$$

Obviously $\det = 1$ forces such a diagonal, and this is of the desired form. □

There is a great wealth of examples found as *discrete* subgroups Γ of the much larger group $\mathrm{PSL}_2(\mathbb{R})$. These groups are discrete in the sense of having the discrete topology when equipped with the operator norm.

Remark. When a group G acts on a space X, it is usually insightful to consider the space of orbits X/G equipped with the quotient topology[1].

Constructing a torus as a quotient is something covered in every topology course. However this construction omits the geometry the torus inherits, which turns out to depend on the lattice used. Because the torus is constructed from translations of a plane, it is naturally equipped with a *flat* metric. Surfaces with many holes pose as more interesting examples.

1.1 Elliptic Functions

A recurring theme throughout mathematics is to study the maps between objects, rather than the objects themselves. To understand *complex tori*, we study the meromorphic functions on them. These are doubly periodic functions on the complex plane, possibly with some poles.

These so-called elliptic functions have a *fundamental* domain from which all other values are determined. This domain is contained in a closed disk, which is compact. Here and onwards we follow the relevant content of [1] and [2].

[1]For topological prerequisites see the appendix.

Proposition. Every holomorphic elliptic function is constant.

Proof. A continuous function on a compact set is bounded, and so, being bounded and entire, this follows immediately from Liouville's theorem. □

So all interesting examples of elliptic functions have some (finite) number of poles in the fundamental domain. The properties of meromorphic functions are usually determined by the poles.

In an attempt to construct such an elliptic function, we consider summing terms $(w-\lambda)^{-n}$ for each lattice point $\lambda \in \Lambda$. Unless n is taken sufficiently large, extra terms may need to be inserted to handle convergence. The Weierstrass elliptic functions are constructed in this way, having a pole at each lattice point

$$\wp(w) = \frac{1}{w^2} + \sum_{\lambda \neq 0} \left(\frac{1}{(w-\lambda)^2} - \frac{1}{\lambda^2} \right)$$

and are known to convergence absolutely and (locally) uniformly away from the poles, that is, on compact subsets of the domain. We can check absolute convergence by examining the tail, which for large λ is of order λ^{-3}. Assuming uniform convergence, we are allowed to differentiate $\wp(w)$ term by term to get

$$\wp'(w) = -2 \sum_{\lambda} \frac{1}{(w-\lambda)^3}$$

and, as before, any shift merely changes the order of the terms, so this is also elliptic. The derivative also being elliptic is obvious however, as it follows immediately from $\wp(w)$ being determined *entirely* by the fundamental domain.

1.2 Cubic Equations

We see from the definition of the Weierstrass elliptic functions that they are even functions, and so must have an expansion around $w = 0$ of the form

$$\wp(w) = \frac{1}{w^2} + 0 + aw^2 + bw^4 + \cdots$$

where the constant term is the tail evaluated at $w = 0$. But now consider

$$\wp''(w) - 6\wp(w)^2 = -10a + \cdots$$

and note that this is holomorphic and elliptic (this should be obvious). So this is constant, and we have found that the Weierstrass elliptic function satisfies

$$\frac{d}{dw}(\wp'(w)^2 - 4\wp(w)^3 + 20a\wp(w)) = 2\wp'(w)(\wp''(w) - 6\wp(w)^2 + 10a) = 0$$

or

$$\wp'(w)^2 = 4\wp(w)^3 - g_2\wp(w) - g_3$$

for suitable coefficients. So we can use elliptic functions to parameterise solutions to cubic equations of the form

$$y^2 = 4x^3 - g_2 x - g_3$$

known as an elliptic curve when there are no cusps or self-intersections.

2 Surface Theory

When working with spaces defined by gluing together small patches, it is natural to ask that the gluing maps, also known as transition functions, are *nice*. In our case this will mean holomorphic. The basic idea of a surface is a Hausdorff space with open sets U_α covering X and homeomorphisms $\phi_\alpha : U_\alpha \to V_\alpha$ to open sets of $\mathbb{C}$. We say X is a Riemann surface when $\phi_\beta \phi_\alpha^{-1}$ is always holomorphic.

Each pair (U_α, ϕ_α) is called a chart on X, and the collection of all charts is called an atlas. The compatibility of our atlas allows us to make global definitions merely in terms of charts. This is important because each chart corresponds to a choice of local coordinates on the surface. We say, for example, that a map $F : X \to Y$ between Riemann surfaces is holomorphic when $\psi_\beta F \phi_\alpha^{-1}$ is always holomorphic for suitable charts ϕ_α and ψ_β.

Example (Riemann sphere). The space $\mathbb{CP}^1$ has standard coordinates $(x : y)$ giving affine patches for $y \neq 0$ and $x \neq 0$. Both patches only exclude one point, and give charts with affine coordinates x/y and y/x respectively. Away from zero and infinity, we change coordinates by $w \mapsto w^{-1}$ and this is holomorphic.

Example (complex tori). The quotient group $\mathbb{C}/\Lambda$ is a set of equivalence classes and is naturally equipped with the quotient topology. A small enough disk in the plane would contain unique representatives. So we use the inclusions of these disks into the plane to construct the charts. Any two choices of local coordinates would be related by $w \mapsto w + \lambda$ which is clearly holomorphic.

2.1 Holomorphic Maps

The implicit and inverse function theorems also hold for complex functions. We can use these to deduce that every holomorphic map between Riemann surfaces looks like $w \mapsto w^k$ near each point.

Lemma. Let f be a holomorphic function on an open neighbourhood U of $0 \in \mathbb{C}$ with $f(0) = 0$, but with f not identically zero. Then there is a unique integer $k \geq 1$ such that on some smaller neighbourhood V we can find a holomorphic function g with $g'(0) \neq 0$ and $f(w) = g(w)^k$ on V.

Proof. Suppose $f(w) = a_k w^k (1 + b_1 w + b_2 w^2 + \cdots)$ is the expansion near $w = 0$, where $a_k \neq 0$. If w is sufficiently small, we can write

$$g(w) = {a_k}^{1/k} w (1 + b_1 w + b_2 w^2 + \cdots)^{1/k}$$

by using the implicit function theorem to make a choice of root. We clearly have $g'(0) = {a_k}^{1/k} \neq 0$ and so existence is established. Note that $k = 1$ iff

$$f'(0) = k g(0)^{k-1} g'(0) \neq 0$$

and otherwise $k - 1$ is the multiplicity of the zero of f' at $w = 0$. So the uniqueness of k is clear. $\square$

Theorem. Let X and Y be connected Riemann surfaces and $F : X \to Y$ a non-constant holomorphic map. For each point $x \in X$, there is a unique integer $k = k_x \geq 1$ such that we can find charts around x and $F(x)$ in which F is represented by the map $w \mapsto w^k$ with $w = 0$ corresponding to x.

Proof. Initially choose arbitrary charts so that F is represented by $f = \psi F \phi^{-1}$ and the previous lemma gives us $f = g^k$ with $g'(0) \neq 0$. The inverse function theorem says g is a homeomorphism in some restricted (co)domain, so we now change the chart about x by composing with g. The uniqueness of k_x is clear. □

The points where $k_x > 1$ are known as ramification points and are particularly important. They are the zeros of $f'(w)$ for any choice of charts, and as such are also known as the critical points. A loop winding once around any point x is mapped to a loop winding k_x times around the image point.

Lemma. The zeros of a non-constant holomorphic map form a discrete set.

Proof. Suppose f is holomorphic on some connected open set U, so that in particular $N = f^{-1}(\{0\})$ is closed in U, and let $c \in N$. Note that f is not identically zero, and so there is a least m with $f^{(m)}(c) \neq 0$ by analyticity. Now we can write $f(w) = (w - c)^m g(w)$ for some g non-zero near c. □

By using charts, we can apply this lemma to show that the ramification points of $F : X \to Y$ form a discrete set. It is common to require that F is proper, so that the image of this set is also discrete. The image points are known as branch points, because the map looks locally like the branched covering $w \mapsto w^k$ with branch point $w = 0$.

2.2 Complex Tori

When a holomorphic map has no ramification points, it must preserve the angle between any two curves and is therefore *conformal*. To see this, choose suitable charts so that the map is represented by $w \mapsto w$. Then clearly the angle does not change (assuming that such a notion of angle is well-defined). It is not hard to find curves through ramification points for which this fails, so the inverse function theorem tells us that a map is conformal iff it is locally an isomorphism onto its image. Obviously bijective conformal maps are exactly isomorphisms.

Example. The automorphisms of the complex plane are the affine maps. [1, p. 41]

This fact, together with the idea of covering spaces, helps us to classify complex tori as follows [3]. We work with lattices upto *homothety* so that $s\Lambda$ is considered the same as Λ when $s \neq 0$. Note that this is an equivalence relation.

Lemma. Every automorphism of $\mathbb{C}$ gives an isomorphism of complex tori.

Proof. Suppose $T(w) = aw + b$ with $a \neq 0$. Then $w + \Lambda$ is mapped to $T(w) + a\Lambda$ and we have a map $\mathbb{C}/\Lambda \to \mathbb{C}/a\Lambda$. This has an inverse given by T^{-1} and is thus the desired isomorphism. A linear T would be a homothety of the lattices. □

Theorem. Complex tori are isomorphic iff their lattices are equivalent.

Proof. Suppose $F : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ is an isomorphism and note that this gives a doubly periodic function on the plane. Now lift F to $T : \mathbb{C} \to \mathbb{C}$ which, being an automorphism, must be affine linear. By a translation, we can assume T is in fact linear, and therefore T maps Λ_1 to Λ_2. Otherwise F would not be one-to-one. But this means T gives a homothety of lattices. In the other direction, we already have our homothety T and we apply the previous lemma. □

The automorphisms that fix a marked point $p \in \mathbb{C}/\Lambda$ are then just those homotheties $s\Lambda = \Lambda$ since the translations are now excluded. Every marked torus is isomorphic to one with lattice $\mathbb{Z} + \tau\,\mathbb{Z}$ where $\operatorname{Im}\tau > 0$ and we wish to understand when these lattices are equivalent.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} a\tau + b \\ c\tau + d \end{pmatrix}$$

Any non-trivial homothety of such lattices must involve a rotation, as the lattice points along the real axis are exactly the integers. Moreover, there must be a lattice point $c\tau + d$ which is mapped to 1 by the homothety. Hence the scaling constant is $s = (c\tau + d)^{-1}$ and we change basis as above, choosing

$$\operatorname{Im}\left(\frac{a\tau + b}{c\tau + d}\right) > 0$$

so that the associated invertible, integer matrix has positive determinant, and therefore $ad - bc = 1$. The fundamental domains of these two bases must have equal area, both being equal to the surface area of the torus. Conversely, the modular group action on τ produces equivalent lattices, as shown by the change of basis above.

We have just shown that the points of $\mathrm{H}/\mathrm{PSL}_2(\mathbb{Z})$ are in bijection with the isomorphism classes of marked tori. For this to be of any importance, we would also need to show that this correspondence is *natural*. Naturality may mean that the geometry of this space reflects the geometry of tori.

2.3 Hyperbolic Surfaces

Since the Riemann sphere is also a projective space, it is natural to discuss the projective transformations. Being linear transformations of $\mathbb{C}^2$ they look like

$$(x : y) \mapsto (ax + by : cx + dy) \qquad ad - bc \neq 0$$

which is an action of $\mathrm{PGL}_2(\mathbb{C})$ defined by linear fractional transformation, as discussed in the introduction. These *Möbius transformations* are exactly the automorphisms of the Riemann sphere [1, p. 41]. In the standard chart,

$$w \mapsto \frac{aw + b}{cw + d} \qquad \infty \mapsto \frac{a}{c}$$

so that ∞ is fixed exactly when $c = 0$ and the map is affine linear.

Lemma. Any Möbius transformation with 3 fixed points must be the identity.

Proof. The fixed points are exactly the solutions to $cw^2 + (d-a)w - b = 0$ and so there are at most two, unless $c = d - a = b = 0$ and the transformation is just the identity. □

Theorem. Every Möbius transformation is uniquely determined by three points.

Proof. Suppose f and g be Möbius transformations agreeing at three distinct points. Then $g^{-1}f$ has three fixed points and must be the identity. □

Remark. Now mark four points $w_i \neq \infty$ and consider the Möbius transformation

$$w \mapsto \frac{(w - w_1)(w_2 - w_3)}{(w - w_3)(w_2 - w_1)}$$

so that in particular $w_1 \mapsto 0$, $w_2 \mapsto 1$ and $w_3 \mapsto \infty$. Any four distinct points can be transformed to those three, along with some $\mu \neq \infty$. By composing transformations, we see this μ to be invariant under any transformation of the original four *ordered* points. This allows us to define the j*-invariant*.

Every surface to which we have assigned a notion of distance has also a notion of curvature, and at each point this curvature is either positive, negative or zero. These also go by the names elliptic, hyperbolic and flat. Since the Riemann sphere is elliptic, and the complex tori are flat, so we now construct the upper half-plane H as an example of a hyperbolic Riemann surface.

Any open set in $\mathbb{C}$ is clearly a Riemann surface, this includes the unit disk and the upper half-plane. Both of these are models of the hyperbolic plane, and are conformally equivalent, that is, isomorphic via the Möbius transformation

$$\tau \mapsto \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \tau$$

sending the real axis to the unit circle by a rational parametrization. We could easily check this [1, p. 162] but the result is not important for us.

The motivation for studying the modular group comes from $\mathrm{PSL}_2(\mathbb{R})$ being the group of orientation preserving isometries of H [1, p. 163]. As Möbius transformations, we also see this as the group of automorphisms. This shows that the metric we imposed on H is indeed consistent with the conformal structure inherited from $\mathbb{C}$. Moreover, this metric is known to be hyperbolic.

3 Elliptic Curves

Algebraic curves such as the familiar $x^2+y^2=1$ have a notion of degree, found by counting the number of intersection points with any line. To make sure this is well-defined, i.e. independent of the choice of line, we work over an algebraically closed field, include so-called points at infinity, and count with multiplicity.

Suppose F is a homogeneous polynomial over the complex numbers, that is, each term has the same algebraic degree. Then we can talk about the *projective* curve $F=0$ in $\mathbb{C}\mathrm{P}^n$ consisting of those points where F vanishes. Note that homogeneity is necessary for this to make sense.

We now homogenise the elliptic curve from the introduction to get

$$E : y^2 z = 4x^3 - g_2 x z^2 - g_3 z^3$$

whose set of complex projective points is denoted by $E(\mathbb{C}) \subset \mathbb{C}\mathrm{P}^2$. By exploiting properties of the Weierstrass elliptic function for Λ we can write

$$w \mapsto (\wp(w) : \wp'(w) : 1)$$

$$\lambda \mapsto (0 : 1 : 0)$$

to define a map $\mathbb{C}/\Lambda \to E(\mathbb{C})$. Note that the chosen point at infinity for the elliptic curve follows from setting $z=0$ in the equation above. Assuming this is an isomorphism of Riemann surfaces, we can carry the group structure of the complex torus over to the elliptic curve.

The *j-invariant* is an important tool for understanding elliptic curves and assigns to each complex torus Σ a complex number $\langle\Sigma\rangle$ in such a way that two tori are isomorphic iff they have the same invariant. We can use the Weierstrass elliptic functions to compute this invariant.

The Weierstrass elliptic functions are meromorphic functions on marked tori, and thus give holomorphic maps $\mathbb{C}/\Lambda \to \mathbb{C}\mathrm{P}^1$. A little covering theory [4, p. 61] shows that there exists no *unramified* covering of the sphere by a torus, so these maps must have ramification points. The marked point 0 maps to ∞ and in suitable charts the function looks like $w^2(1+O(w^4))$. Since we only care about the function near $w=0$ we see that $k=2$ at the marked point. Explicitly, we are computing the expansion of $1/\wp(w)$ around the lattice points.

Lemma. Every complex torus is branched cover of the Riemann sphere with exactly four branch points, and each ramification point with index 2.

We now define the j-invariant as discussed in [1, pp. 91–94]. Consider the equivalence classes of complex tori with ordered branch points. These are just quadruples of distinct points on the sphere modulo the Möbius action. But we can always choose these branch points to be 0, 1, ∞ and some $\mu \in \mathbb{C}\setminus\{0,1\}$ being uniquely determined by the equivalence class. We would like to know how μ changes as we permute the branch points. By considering the *cross-ratio* formula shown in the previous section, we see that the kernal of this S_4 action is isomorphic to the Klein four-group. If we quotient out by this subgroup, we get a faithful action of S_3. Careful consideration of this action shows that

$$\langle \Sigma \rangle = \frac{(\mu^2 - \mu + 1)^3}{\mu^2(\mu - 1)^2}$$

is an invariant quantity. The specifics of this formula[2] guarantee that no complex number is missed, and two cross-ratios $\mu \in \mathbb{C} \setminus \{0, 1\}$ result in the same value iff they belong to the same orbit of the S_3 action permuting the first three branch points. Having taken care of the arbitrary choice, the choice of quadruple, we now have unique complex number for each isomorphism class of torus.

[2] The factor of 256 usually seen in the literature is omitted.

A Topological Prerequisites

Suppose we have a collection of objects and morphisms, equipped with a notion of composition. An isomorphism is any morphism $f : X \to Y$ for which there exists some morphism $g : Y \to X$ behaving as an inverse, that is, satisfying

$$gf = \mathrm{id}_X \text{ and } fg = \mathrm{id}_Y$$

where id denotes the identity. When the objects are topological spaces, and the morphisms continuous maps, we obtain the definition of a homeomorphism. Throughout the chapter on surface theory, the objects are Riemann surfaces and the morphisms are holomorphic maps.

Category theorists tell us to forget the objects and only work with the morphisms. On the other hand, the purpose of topological spaces is that continuity is defined wholly in terms of the open sets. The latter favours the descriptive language of *near* and *small*.

A.1 Analytic Properties

That any two points can be separated by disjoint open sets, is the most basic assumption we will make about our spaces, and is known as the Hausdorff separation axiom. This allows us to construct *sufficiently small* neighbourhoods around points of interest, excluding any bad points.

Another property of note is that of compactness, with the intuition being that the space is finite in extent. Formally, we say that every cover has a finite subcover, where a cover is any collection of open sets covering the space. As usual, there is also a weaker, local version saying that every point has a compact neighbourhood. This definition is complemented by the notion of a proper map, one where the preimages of compact sets are compact.

In a similar vain, it is common to assume that a space is second-countable, that is, admits a countable base of open sets.

A.2 Algebraic Constructions

A quotient map $q : Y \to Y/_\sim$ identifies points of Y by mapping each point to its equivalence class. The quotient topology is the finest topology for which q is continuous. Explicitly, we define $U \subseteq Y/_\sim$ to be open iff $q^{-1}(U)$ is open.

Following Hatcher [4, p. 56] we say that continuous $p : Y \to X$ is a covering map if every point in X has an open neighbourhood U which is evenly covered in the sense that $p^{-1}(U)$ is a union of disjoint open sets, called sheets, each of which is mapped homeomorphically onto U by p. A continuous $f : Z \to X$ is said to lift to some continuous $F : Z \to Y$ if we have $pF = f$.

It is easy to see that every covering map is also a quotient map. Let V be an open subset of Y. Then $p(V)$ is also open because V is a union of sheets and the image of any sheet is open. Hence $U \subseteq X$ is open iff $p^{-1}(U)$ is open, which we recognise as the quotient topology where fibers $p^{-1}(x)$ are identified.

A particularly important example of a covering map is $p : S^1 \to S^1$ given by $w \mapsto w^k$ where we view S^1 as the complex unit circle. It is common in low-dimensional topology to use the looser definition of a *branched cover* and this definition allows us to extend p to to $\mathbb{C} \to \mathbb{C}$ with branch point $w = 0$.

References

[1] S.K. Donaldson. *Riemann Surfaces*. Oxford University Press, 2011.

[2] Frances Kirwan. *Complex Algebraic Curves*. Cambridge University Press, 1992.

[3] Richard Hain. *Lectures on Moduli Spaces of Elliptic Curves*. 2008. DOI: 10.48550/ARXIV.0812.1803. URL: https://arxiv.org/abs/0812.1803.

[4] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.