1

First Ideas: Complex Manifolds, Riemann Surfaces, and Projective Curves

This chapter presents some elementary (and not so elementary) ideas in continual use throughout the book. For more details and further information see, for example, Ahlfors [1979], Bliss [1933], Clemens [1980], Farkas and Kra [1992], Hurwitz and Courant [1964], Kirwan [1992], Reyssat [1989], Springer [1981], and/or Weyl [1955]. These are all perfectly accessible to beginners; further references will be given as we go along.

1.1 The Riemann Sphere

Let $\mathbb{R}^3$ be the 3-dimensional (real) space of points $x = (x_1, x_2, x_3)$ and let $|x| = \sqrt{x_1^2 + x_2^2 + x_3^2}$ be the distance from x to the origin $o = (0, 0, 0)$. $\mathcal{M}$ is the unit sphere $|x| = 1$ and $\mathbb{C}$ is its equatorial plane $x_3 = 0$, identified as the complex numbers via the map $(x_1, x_2, 0) \mapsto x_1 + \sqrt{-1}x_2$. $\mathcal{M}$ is temporarily punctured at the north pole $n = (0, 0, 1)$ and the rest ($x_3 < 1$) is mapped 1:1 onto $\mathbb{C}$ by the projection p depicted in profile in Fig. 1.1. The rule is: Sight from n through the point $x \in \mathcal{M}$, the projection $p(x)$ being the intersection of this line of sight with $\mathbb{C}$. Obviously, $p(x)$ and $x_1 + \sqrt{-1}x_2$ lie on the same ray of $\mathbb{C}$; also the triangles $n, o, p(x)$ and $n, q = (0, 0, x_3), x$ are similar, so

$$|p(x)| = \frac{\text{distance}[o, p(x)]}{\text{distance}[o, n]} = \frac{\text{distance}[q, x]}{\text{distance}[q, n]} = \frac{\sqrt{x_1^2 + x_2^2}}{1 - x_3}.$$

In short,

$$p(x) = \frac{x_1 + \sqrt{-1}x_2}{1 - x_3}.$$

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

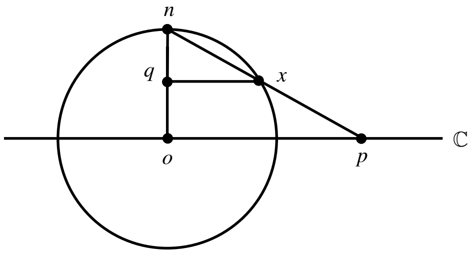

Figure 1.1. The Riemann sphere.

This is the **stereographic projection** of the cartographers. Denote it by p_+ to distinguish it from the analogous projection

$$p_-(x) = \frac{x_1 + \sqrt{-1}x_2}{1 + x_3}$$

of $\mathcal{M} \cap (x_3 > -1)$ produced by sighting from the south pole $(0, 0, -1)$. Now, for $-1 < x_3 < 1$, both maps are available and

$$[p_-(x)]^{-1} = \frac{1 + x_3}{x_1 + \sqrt{-1}x_2} = \frac{1 + x_3}{x_1^2 + x_2^2\,(= 1 - x_3^2)} \times (x_1 - \sqrt{-1}x_2)$$

$$= \frac{x_1 - \sqrt{-1}x_2}{1 - x_3} = [p_+(x)]^*,$$

the star being complex conjugation, so the two images are anticonformally related. Replacing $p_-(x)$ by $[p_-(x)]^*$ produces the following situation: $\mathcal{M}$ is covered by two open patches $U_+ = \mathcal{M} \cap (x_3 < 1)$ and $U_- = \mathcal{M} \cap (x_3 > -1)$, each provided with a **local coordinate**: $z_+ = p_+(x)$ for U_+ and $z_- = [p_-(x)]^*$ for U_-. Most points of $\mathcal{M}$ lie in the overlap $U_- \cap U_+$, and for them the two competing coordinates are conformally related: $z_- = 1/z_+$. This object $[\mathcal{M} +$ patches $+$ projections$]$ is the **Riemann sphere**, alias the extended plane $\mathbb{C} + \infty$, the so-called **point at infinity** being identified with the north pole $n = (0, 0, 1)$.

Exercise 1. Prove that p_+ maps spherical circles into plane circles or lines and vice versa. *Hints*: $x \bullet e = \cos\theta$ marks off a spherical circle for any unit vector e and any angle $0 < \theta \leq \pi/2$. Check that $p_+(x) = a + \sqrt{-1}b$ satisfies $(1 - x_3)(ae_1 + be_2) + x_3 e_3 = \cos\theta$ and $a^2 + b^2 = (1 - x_3)^{-1}(1 + x_3)$. Then eliminate x_3.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Exercise 2. Prove that the map $p_+ : \mathcal{M} \to \mathbb{C}$ is conformal, that is, angle preserving, spherical angles being measured in the natural way.

1.2 Complex Manifolds

A **2-dimensional manifold** or **surface** $\mathcal{M}$ is a geometrical figure that looks in the small like an (open) disk. To be precise, this means three things: (1) $\mathcal{M}$ is a topological space covered by a countable number of open **patches** U. (2) The typical patch U is equipped with a **patch map** $\mathfrak{p} \to x(\mathfrak{p})$ of points $\mathfrak{p} \in U$ to the open unit disk $D = \{(x_1, x_2) : x_1^2 + x_2^2 < 1\} \subset \mathbb{R}^2$. This map is 1:1, continuous, and onto; it provides U with **local coordinates** $x = x(\mathfrak{p})$. (3) An ambiguity arises if the point $\mathfrak{p} \in \mathcal{M}$ lies in the overlap $U_- \cap U_+$ of two patches so that two competing coordinates $x_-(\mathfrak{p})$ and $x_+(\mathfrak{p})$ are available; in this case, the composite map $x_-(\mathfrak{p}) \to \mathfrak{p} \to x_+(\mathfrak{p})$, and likewise its inverse, is required to be continuous; see Fig. 1.2.

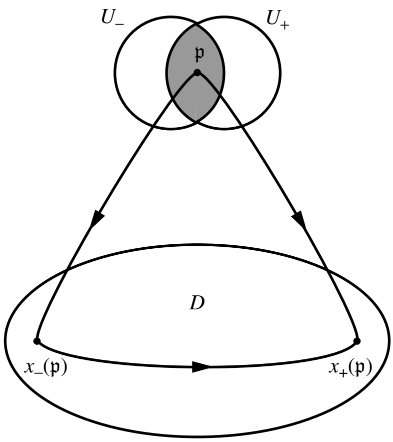

Figure 1.2. Local coordinates on $\mathcal{M}$.

Examples. The plane; the (open) half-plane, disk, or annulus; the sphere or the cylinder; the surface of a doughnut (torus) or a pretzel; the Möbius strip.

- $\mathcal{M}$ is **connected** if it comes in one piece, that is, if any two of its points can be joined by a nice curve; this feature is assumed from now on without further comment.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

- $\mathcal{M}$ is **simply connected** if it has no punctures, holes, or handles, that is, if every closed curve (**loop**) can be shrunk to a point in $\mathcal{M}$. This is so for the disk, half-plane, and sphere, but not for the annulus, cylinder, torus, or pretzel.
- $\mathcal{M}$ is **compact** if every cover of it by open sets admits a finite subcover. In this case it is possible to pick a number $0 < r < 1$ so that the images of the closed disk of radius r under a finite number of the patch maps already cover $\mathcal{M}$. This is so for the sphere, torus, and pretzel, but not for the disk, plane, or cylinder.
- $\mathcal{M}$ is **orientable** if the relation between patch maps preserves the sense of (say, counterclockwise) rotation. This is so for the Riemann sphere of Section 1, but is not possible on a Möbius band.
- $\mathcal{M}$ is **smooth** if competing local coordinates x_- and x_+ on overlaps $U_- \cap U_+$ are smoothly related, that is, if $x_-(\mathfrak{p})$ is an infinitely differentiable function of $x_+(\mathfrak{p})$ and vice versa. Then you may speak of smooth functions $f\colon \mathcal{M} \to \mathbb{R}$, of which you ask that $f(\mathfrak{p})$ be a smooth function of the local coordinate $x(\mathfrak{p})$ on any patch. Plainly, there can be no competition in this regard: On overlaps, $f(\mathfrak{p})$ is a smooth function of both $x_-(\mathfrak{p})$ and $x_+(\mathfrak{p})$ or of neither.
- $\mathcal{M}$ acquires the more subtle structure of a **complex manifold** or **Riemann surface** if the complex local coordinates or **parameters** $z(\mathfrak{p}) = x_1(\mathfrak{p}) + \sqrt{-1}x_2(\mathfrak{p})$ are conformally related on overlaps; orientability is necessary for this. Then it makes sense to speak of the class $\mathbf{K}(\mathcal{M})$ of functions $f\colon \mathcal{M} \to \mathbb{C} + \infty$ of **rational character** defined by the requirement that in the vicinity of any point $\mathfrak{p}_0$, $f(\mathfrak{p})$ have an expansion $w^d[c_0 + c_1 w + c_2 w^2 + \cdots](d > -\infty, c_0 \neq 0)$ in powers of $w = z(\mathfrak{p}) - z(\mathfrak{p}_0)$. Naturally the expansion changes if the local parameter is changed, but the number d does not, so it is permissible to speak of a **root** of multiplicity d if $d > 0$ and of a **pole** of multiplicity $-d$ if $d < 0$; d is the **degree** of f at $\mathfrak{p}_0$.

Exercise 1. $\mathbf{K}(\mathcal{M})$ is a field.

Exercise 2. Check the statement that the degree d is independent of the local parameter.

Now for some easy examples.

Example 1. It is needless to pause over the complex structure of the plane $\mathbb{C}$ except to note that it has a *global parameter* $z(\mathfrak{p}) = x_1 + \sqrt{-1}x_2$. This example is too simple, as is the disk, half-plane, or annulus, or any other open part of $\mathbb{C}$ which obtains a complex structure by mere inheritance.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Example 2. The cylinder is the quotient of $\mathbb{C}$ by its (arithmetic) subgroup $\mathbb{Z}$, so it, too, obtains a complex structure by inheritance, and likewise the (square) torus, which is the quotient of $\mathbb{C}$ by the lattice $\mathbb{Z} \oplus \sqrt{-1}\mathbb{Z}$; see Fig. 1.3.

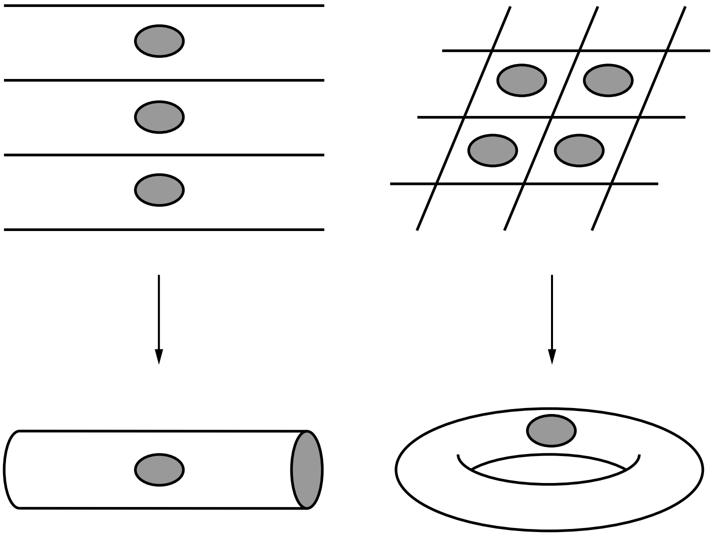

Figure 1.3. Complex structures on the cylinder and the torus.

Example 3. The sphere is more interesting. The stereographic projections of Section 1 provide it with a complex structure after self-evident adjustments; for instance, $z_+(U_+) = \mathbb{C}$ is not the unit disk, but no matter.

Example 4. The **projective line** $\mathbb{P}^1$ is the family of all complex lines in the 2-dimensional complex space $\mathbb{C}^2$: In detail, $\mathbb{C}^2$ is punctured at the origin and two of its points are identified if they lie on the same (complex) line, that is, (a, b) is identified with (a', b') if $a' = ca$ and $b' = cb$ for some nonvanishing complex number c. $\mathbb{P}^1$ is covered by two patches $U_+ = \mathbb{C} \times 1$ and $U_- = 1 \times \mathbb{C}$, provided with self-evident local parameters: $z_+(\mathfrak{p}) = z$ for $\mathfrak{p} = (z, 1) \in U_+$ and $z_-(\mathfrak{p}) = z$ for $\mathfrak{p} = (1, z) \in U_-$; on the overlap $U_- \cap U_+ = \{(a, b) \in \mathbb{C}^2 : ab \neq 0\}$, you have the identifications $(a/b, 1) \equiv (a, b) \equiv (1, b/a)$ and so also the relation of local parameters: $z_+(\mathfrak{p}) = [z_-(\mathfrak{p})]^{-1}$. This is the *same rule* as for the Riemann sphere of Section 1. In short, the projective line and the Riemann sphere are identical (as complex manifolds).

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Exercise 3. $\mathbb{P}^1$ is compact. That is obvious from its identification with the sphere, but do it from scratch, from the original definition of compactness.

A Little Topology. Let $\mathcal{M}$ be any compact surface, with complex structure or not, and let it be **triangulated** by cutting it up into little (topological) triangles having (in sum) c corners, e edges, and f faces (triangles). Reyssat [1989] has a nice proof that this is always possible. Then (remarkable fact!) the Euler number $\chi = c - e + f$ is always the same: 2 for the sphere, 0 for the torus, -2 for the pretzel, and so forth, that is, it depends only upon the surface and not upon the particular triangulation in hand. This number determines the topology of $\mathcal{M}$ completely. In fact, $\mathcal{M}$ is necessarily a **handlebody**, that is, a (topological) sphere with $g = 1 - (1/2)\chi(\mathcal{M})$ handles attached: 0 for the sphere, 1 for the torus, 2 for the pretzel, and so on; this number is the **genus** of $\mathcal{M}$. Hurwitz and Courant [1964: 497–534] present an elementary proof; see also Coxeter [1980] for more information and Euler [1752] who started it all. The next items are illustrative.

Example. The spherical triangulation seen in Fig. 1.4 has 6 corners (the black spots), 12 edges, and 8 faces for an Euler number of $6 - 12 + 8 = 2$.

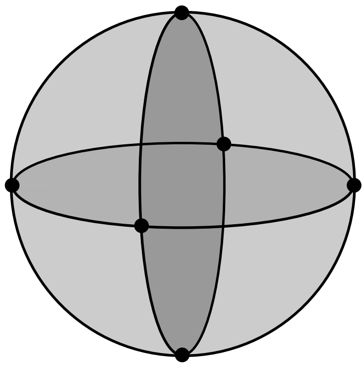

Figure 1.4. Triangulation of the sphere: $c = 6$, $e = 12$, $f = 8$.

Exercise 4. Check by hand that the Euler number and so also the genus of the sphere does not depend upon the triangulation. *Hint*: The sphere can be laid out flat on the plane by cutting all edges that meet at some particular corner. Now count.

Exercise 5. Repeat for higher handlebodies: torus, pretzel, and so on; especially, check that the genus $1 - (1/2)\chi(\mathcal{M})$ really is the handle number.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Preview. It is a fact that any handlebody can be provided with a complex structure, as was already seen for sphere and torus and will appear for the pretzel in Section 12. This can be done in one and only one way for the sphere, but already the torus admits infinitely many (conformally) distinct complex structures; see Section 2.6. It is this unobvious fact that prompted the adjective *subtle* in first speaking of complex manifolds.

1.3 Rational Functions

The function field $\mathbf{K} = \mathbf{K}(\mathbb{P}^1)$ of the projective line is easy to compute: It is just the field $\mathbb{C}(z)$ of rational functions of $z = p_+(x)$.

Proof. $f \in \mathbf{K}$ is a function of rational character of $z_+ = p_+(x)$ on the patch $x_3 < 1$, and likewise of $z_- = [p_-(x)]^*$ on $x_3 > -1$. It follows that f has, in the first patch, a finite number of poles $\mathfrak{p}_i (1 \le i \le m)$, repeated according to their multiplicity, and the possibility of an extra pole of multiplicity n at $\infty =$ the north pole for a total count of $n+m = d$. View $f(\mathfrak{p})$ as a function of $z = p_+(x)$ and let $z_1, \dots, z_m$ be the projections of $\mathfrak{p}_1, \dots, \mathfrak{p}_m$. Then the product $Q(z)$ of $f(\mathfrak{p})$ and $P(z) = (z - z_1) \times \cdots \times (z - z_m)$ is pole-free in $\mathbb{C}$ and of limited growth at ∞: $|Q(z)| \le$ a constant multiple of $|z|^d$ far out. Now use Cauchy's formula for a big circle of radius R:

$$\frac{D^p Q(0)}{p!} = \frac{1}{2\pi\sqrt{-1}} \oint \frac{Q(z)dz}{z^{p+1}}$$

to check that

$$|D^p Q(0)| \le \text{ a constant multiple of } R^d \times R^{-p-1} \times 2\pi R = o(1)$$

for $R \to \infty$ if $p > d$. The upshot is that Q is a polynomial of degree $\le d$ and f is a ratio Q/P, as advertised.

Exercise 1. Check the estimate of Q at ∞.

The **degree** of $f \in \mathbf{K}$ is the total number of its poles, counted according to multiplicity, ∞ included, that is, deg $f = d = n + m$, and it is plain from its representation as rational function that f has the same number of roots, counted likewise, according to multiplicity. But also $f - c \in \mathbf{K}$ has the same number of poles as f for any complex number c, so f takes on every complex value, ∞ included, d times. In short, d is also the **topological degree** of f as a map of $\mathbb{P}^1$ to itself, taking $f(\mathfrak{p}) = \infty$ at poles. This is a general principle for functions of rational character on compact Riemann surfaces $\mathcal{M}$: As maps of $\mathcal{M}$ to $\mathbb{P}^1$, they take on every value the same number of times; see Section 16.

Exercise 2. Clarify the statement: $f \in \mathbf{K}(\mathbb{P}^1)$ is an analytic map of $\mathbb{P}^1$ to itself.

Exercise 3. Check that the roots and poles of $f \in \mathbf{K}(\mathbb{P}^1)$ can be placed any way you like, provided only that they are the same in number. This is not the case for any other compact Riemann surface: It is already false for the torus; see Section 2.7, item 4.

Exercise 4. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ be any collection of points on $\mathbb{P}^1$, repetition permitted, and let $\mathfrak{L}$ be the space of functions $f \in \mathbf{K}(\mathbb{P}^1)$ having these poles or softer; for example, f is permitted a pole at $\mathfrak{p}_1$, of degree no more than the number of its repetitions. $\mathfrak{L}$ is a vector space over $\mathbb{C}$. Prove that its (complex) dimension is $n + 1$.

Besides its functions of rational character, $\mathbb{P}^1$ also carries **differentials of rational character**. These are the objects ω expressible patchwise as $c(z)dz$ with coefficients c of rational character in the local parameter $z = z(\mathfrak{p})$. If the parameter is changed from $z_+ = z$ to $z_- = w$ on the overlap $U_- \cap U_+$, then the coefficient changes in the natural way, from c to $c \times (dz/dw)$. The differential ω has a root or pole of degree d at the point $\mathfrak{p}_0$ if its coefficient does so. The residue of ω at $\mathfrak{p}_0$ is the integral $(2\pi\sqrt{-1})^{-1} \oint \omega$ taken about a small circle enclosing $\mathfrak{p}_0$. ω is a **differential of the first kind** if it is pole-free, of the **second kind** if it has poles but only vanishing residues, and of the **third kind** otherwise.

Exercise 5. dz is a differential of the second kind on $\mathbb{P}^1$: It has 2 poles at ∞. $df = f'(z)dz$ is likewise of the second kind for any $f \in \mathbf{K}(\mathbb{P}^1)$. $z^{-1}dz$ is different, being of the third kind, in agreement with the fact that the logarithm is not single-valued. Check all that.

Exercise 6. $\mathbb{P}^1$ has no differentials of the first kind besides $\omega = 0$.

Exercise 7. Check that the total degree (roots $-$ poles) of a differential of rational character on $\mathbb{P}^1$ is necessarily -2.

1.4 Luroth's Theorem*

The star means that you may skip this section, but *do* note the following fact which will be useful later: *Any subfield of* $\mathbf{K} = \mathbf{K}(\mathbb{P}^1)$ *containing more than the constant field* $\mathbb{C}$ *is isomorphic to* $\mathbf{K}$ *itself.* This is Luroth's theorem [1876].

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Example. The subfield $\mathbf{K}_0$ of functions $f \in \mathbf{K}$ invariant under the involution $z \mapsto 1/z$ is the field $\mathbb{C}(w)$ of rational functions of $w = z + 1/z$. The latter is viewed as a map from one projective line (the cover) to a second projective line (the base); it is of degree 2. The field $\mathbf{K}$ of the cover is likewise of degree 2 over the field $\mathbf{K}_0$ of the base in view of $z = [w \pm \sqrt{w^2 - 4}]/2$. It is this type of counting that is the key to the present proof. It is not the usual proof in that it mixes standard field theory with nonstandard geometric considerations. It is precisely this type of mixture that we want to emphasize in this book. Van der Waerden [1970] presents the standard proof; see also Hartshorne [1977].

A Little Algebra. Not much is needed. The letter $\mathbf{K}$ denotes a field over the rational numbers $\mathbb{Q}$. The degree of a big field $\mathbf{K}$ (the **extension**) over a smaller field $\mathbf{K}_0$ (the **ground field**) is the dimension of $\mathbf{K}$ as a vector space over $\mathbf{K}_0$, denoted by $[\mathbf{K}\colon \mathbf{K}_0] \leq \infty$. If the degree is not infinite, then the powers y^n, $n \geq 0$, of an element $y \in \mathbf{K}$ cannot be independent over $\mathbf{K}_0$, so y is a root of some polynomial $P(x) = x^n + c_1 x^{n-1} + \cdots + c_n$ with coefficients from $\mathbf{K}_0$, and y is **algebraic** over $\mathbf{K}_0$. $\mathbf{K}_0[x]$ is the ring of such polynomials. The **field polynomial** of y over $\mathbf{K}_0$ is the irreducible polynomial $P(x) = x^d + c_1^{d-1} + \cdots$ of class $\mathbf{K}_0[x]$ that it satisfies. The extended field $\mathbf{K}_1 = \mathbf{K}_0(y)$ of rational functions of y with coefficients from $\mathbf{K}_0$, obtained by adjunction of y to $\mathbf{K}_0$, is spanned by the d powers $1, y, \ldots, y^{d-1}$; in particular, $[\mathbf{K}_1\colon \mathbf{K}_0] = d$. The roots $x_1 = y, x_2, \ldots, x_d$ of $P(x) = 0$ are necessarily simple. They are adjoined to the ground field $\mathbf{K}_0$ to produce the **splitting field** $\mathbf{K}_2 = \mathbf{K}_0(x_1, \ldots, x_d)$ of $P(x)$. This is the smallest extension of $\mathbf{K}_0$ in which $P(x)$ splits into factors of degree 1: $P(x) = (x - x_1) \cdots (x - x_d)$; it can be realized as the quotient field $\mathbf{K}_0[x]$ modulo $P(x)$. The simplicity of the roots implies that the **discriminant** $\Delta = \prod_{i<j}(x_i - x_j)^2$ does not vanish. This quantity, together with any other symmetric polynomial in the roots, belongs to the ground field $\mathbf{K}_0$. The only other fact that will be needed is that if the extended field $\mathbf{K}$ is obtained from the ground field by the adjunction of n such algebraic elements $y_i (1 \leq i \leq n)$, then there is a single **primitive element** y_0 that does the job at one stroke: $\mathbf{K} = \mathbf{K}_0(y_0)$. Artin [1953], Lang [1984], Pollard [1950], and/or Stillwell [1994] are recommended as refreshers and for more information.

Exercise 1. Prove directly that the discriminant Δ is, itself, a polynomial in the so-called **elementary symmetric functions**

$$\sigma_1 = \sum x_i, \sigma_2 = \sum_{i<j} x_i x_j, \ldots, \sigma_d = x_1 \cdots x_d.$$

Exercise 2. Deduce $\Delta \in \mathbf{K}_0$.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Exercise 3. What is the discriminant of the general cubic $x^3 + ax^2 + bx + c$?

Aside. Luroth's theorem illustrates, in the simplest circumstances, an important theme of complex geometry: The *complex structure* of a compact Riemann surface $\mathcal{M}$ is determined by the *algebraic structure* of its function field $\mathbf{K}(\mathcal{M})$; see Section 15 under **rational curves** for more information and also Section 2.13 for the case of the torus. The characteristic feature of the rational function field $\mathbf{K} = \mathbb{C}(z)$ is that it is of infinite degree over the ground field $\mathbb{C}$ and isomorphic to any proper intermediate field. As to the geometry of $\mathbb{P}^1$, if $\mathcal{M}$ is a compact complex manifold and if $\mathbf{K} = \mathbf{K}(\mathcal{M})$ is a copy of $\mathbb{C}(z)$ then $\mathbf{K} = \mathbb{C}(f)$ for some distinguished $f \in \mathbf{K}$. Now view f as a map of $\mathcal{M}$ to $\mathbb{P}^1$: It has a degree d just like an ordinary rational function as expounded in Section 3. Besides, it is a fact that $\mathbf{K}$ separates points of $\mathcal{M}$, so $d = 1$ and f maps $\mathcal{M}$ 1:1 onto $\mathbb{P}^1$. In short, as a complex manifold, $\mathcal{M}$ *is* $\mathbb{P}^1$.

Proof of Luroth's theorem. The first item of business is to check that if f_0 is any nonconstant rational function, then the *algebraic degree* of $\mathbf{K} = \mathbb{C}(z)$ over $\mathbf{K}_0 = \mathbb{C}(f_0)$ is the same as the *topological degree* d_0 of f_0. The first degree is finite because $f_0 = a_0/b_0$ with coprime $a_0, b_0 \in \mathbb{C}[z]$ and $P(x) = a_0(x) - f_0 b_0(x) \in \mathbf{K}_0[x]$ has $x = z$ as a root; moreover,

$$\begin{aligned} d = [\mathbf{K}\colon \mathbf{K}_0] \le \deg P = &\text{ the larger of the degrees of } a_0 \text{ and } b_0 \\ = &\ \deg f_0 = d_0. \end{aligned}$$

Now let $P_0 = x^d + s_1 x^{d-1} + \cdots + s_d \in \mathbf{K}_0[x]$ be the field polynomial of z over $\mathbf{K}_0$ and observe that *for most values c of f_0* three things happen: (1) c is not a pole of any coefficient $s_n (n \le d)$; (2) $f_0(z) = c$ has d_0 simple roots in $\mathbb{C}$; (3) $z^d + s_1(c) z^{d-1} + \cdots + s_d(c)$ vanishes at each of these. But that makes $d_0 \le d$ and equality prevails: $d_0 = d$.

Now comes the proof of Luroth's theorem itself; compare Fig. 1.5. Let the intermediate field $\mathbf{K}_1$ lie properly above the constant field $\mathbb{C}$ so that $n = [\mathbf{K}\colon \mathbf{K}_1] \le \min\{\deg f_0 \colon f_0 \in \mathbf{K}_1\} < \infty$ and let $P_1(x) = x^n + r_1 x^{n-1} + \cdots + r_n \in \mathbf{K}_1[x]$ be the field polynomial of z over $\mathbf{K}_1$. Then r_1 (or some other of its coefficients) is not constant, z being of infinite degree over $\mathbb{C}$. It is to be proved that $\mathbf{K}_1 = \mathbb{C}(r_1)$, producing the whole of that field by its adjunction to $\mathbb{C}$. Now write $r_1 = a_1/b_1$ with $a_1, b_1 \in \mathbb{C}[z]$, and so on, and clear denominators in $P_1(x)$ to produce $P_2(x) = c_0 x^n + c_1 x^{n-1} + \cdots + c_n \in \mathbb{C}[z][x]$. This divides $P_3(x) = a_1(x) b_1(z) - a_1(z) b_1(x) \in \mathbb{C}[z][x]$ and comparison of degrees with

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

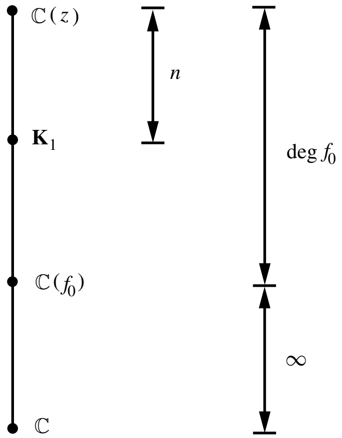

Figure 1.5. The extensions.

regard to z produces

$$\begin{aligned}\deg P_3 &= \text{the larger of the degrees of } a_1 \text{ and } b_1\\ &= \deg r_1\\ &\leq \text{the larger of the degrees of } c_0 \text{ and } c_1\\ &\leq \deg P_2,\end{aligned}$$

whereupon the divisibility of P_3 by P_2 implies $P_3 = P_2 \times P_4$ with $P_4 \in \mathbb{C}[x]$ independent of z. But $P_4(x)$ divides P_3 only if it divides both $a_1(x)$ and $b_1(x)$, and as these are coprime, so P_4 must be constant, whereupon comparison of degrees with regard to x produces

$$\begin{aligned}[\mathbf{K}\colon \mathbb{C}(r_1)] &= \deg r_1 \quad \text{(by the first item of business)}\\ &= \deg P_3\\ &= \deg P_2\\ &= [\mathbf{K}\colon \mathbf{K}_1].\end{aligned}$$

The proof is finished: The degrees of $\mathbf{K}$ over $\mathbf{K}_1$ and over $\mathbb{C}(r_1) \subset \mathbf{K}_1$ match only if $\mathbb{C}(r_1) = \mathbf{K}_1$.

Exercise 4. Check that r_1 is of minimal positive degree in $\mathbf{K}_1$. *Hint*: Any other such primitive element of $\mathbf{K}_1$ must be of the same degree n.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

1.5 Automorphisms of $\mathbb{P}^1$

An automorphism of a complex manifold $\mathcal{M}$ is a 1:1 analytic map of $\mathcal{M}$ onto itself. The inverse of such a map is likewise analytic, so the automorphisms form a group $\Gamma(\mathcal{M})$. $\Gamma(\mathcal{M})$ is easily identified for $\mathcal{M} = \mathbb{P}^1$: It comprises all the rational functions of degree 1, alias the **fractional linear substitutions** or **Möbius transformations**, of the form

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

Exercise 1. Explain the proviso $ad - bc \neq 0$.

The numbers a, b, c, d can be scaled to make $ad - bc = 1$ without changing the map, so you may associate to each automorphism the 2×2 complex matrix $[ab/cd]$ of determinant 1. This effects an isomorphism between $\Gamma(\mathbb{P}^1)$ and the **special linear group** $SL(2, \mathbb{C})$ of all such matrices. Actually, that is not quite right: The map is unaffected by the substitution $a, b, c, d \mapsto -a, -b, -c, -d$, so what you really have is an isomorphism with the **projective special linear group** $PSL(2, \mathbb{C})$ which is the quotient of $SL(2, \mathbb{C})$ by its center $(\pm 1)\times$ the identity. The symbol $[ab/cd]$ signifies either the 2×2 matrix or the associated map, as the context requires, $[ab/cd]$ and its negative being identified until further notice.

Exercise 2. Check all that. Chiefly it is required to prove that the association of $\Gamma(\mathbb{P}^1)$ to $PSL(2, \mathbb{C})$ respects the group operations.

Exercise 3. What is $\Gamma(\mathbb{C})$?

Exercise 4. $\Gamma(\mathbb{P}^1)$ is generated by (1) **translations** $z \mapsto z + a$, (2) **magnifications** (including rotations) $z \mapsto bz$, and (3) the **inversion** $z \mapsto -1/z$. Check it.

Exercise 5. Deduce that $\Gamma(\mathbb{P}^1)$ preserves the class of circles and lines.

Exercise 6. $SO(3)$ is the group of proper (i.e., orientation-preserving) rotations of $\mathbb{R}^3$, realized as (real) 3×3 orthogonal matrices of determinant $+1$. Prove that every such rotation is an automorphism of $\mathbb{P}^1$.

$\Gamma(\mathbb{P}^1)$ moves any three distinct points to any other three points, as you will check: For example, $z \mapsto [(b-a)/(c-b)] \times [(c-z)/(z-a)]$ moves a, b, c to $\infty, 1, 0$ with the natural interpretation of the map if one of the points is at infinity.

Exercise 7. The action of an automorphism on three distinct points specifies it completely; especially, it is the identity if it fixes three points. Why?

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

The (one and only) automorphism $z \mapsto w$ that moves z_1, z_2, z_3 to w_1, w_2, w_3 can be expressed as

$$\frac{z_1 - z_2}{z_2 - z_3}\frac{z_3 - z}{z - z_1} = \frac{w_1 - w_2}{w_2 - w_3}\frac{w_3 - w}{w - w_1}.$$

It moves the additional point z_4 to w_4 if and only if the zs and ws have the same **cross ratio**:

$$\frac{z_1 - z_2}{z_2 - z_3}\frac{z_3 - z_4}{z_4 - z_1} = \frac{w_1 - w_2}{w_2 - w_3}\frac{w_3 - w_4}{w_4 - w_1},$$

with the natural interpretation if any point is at infinity.

Exercise 8. Check that $\Gamma(\mathbb{P}^1)$ preserves cross ratios.

The cross ratio is changed under the action of the **symmetric group** S_4 of permutations of the four "letters" 1, 2, 3, 4. It is invariant under the subgroup[1]

$$K(1234) = \text{id},\ (2143),\ \ (3412),\ \ (4321),$$

so to understand the action, it is permissible to place z_4 at infinity and to study the action of $S_4/K \sim S_3$(= the symmetric group on three letters) on the reduced ratio $x = (z_1 - z_2)/(z_3 - z_2)$. This is seen in Table 1.5.1.

Table 1.5.1. *Action of S_3 on the reduced cross ratio*

(123)	(321)	(132)	(231)	(312)	(213)
x	$1/x$	$1 - x$	$1/(1 - x)$	$(x - 1)/x$	$x/(x - 1)$

Exercise 9. $x \neq 0, 1, \infty$. Why?

Exercise 10. Check the table.

The substitutions $x \mapsto x, 1/x, 1-x, 1/(1-x), (x-1)/x, x/(x-1)$ comprise the **group of anharmonic ratios** $\mathfrak{H}$, which is isomorphic to S_3. **Harmonic ratios** arise when two anharmonic ratios of x coincide; the nomenclature goes back to Chasles (1852). The group will play a small but important role later; see Section 7 under **Platonic solids** and also Chapter 4.

Exercise 11. The harmonic ratios are of two kinds: $x = -1, 1/2, 2$ is the *harmonic proportion* of nineteenth-century projective geometry; it corresponds,

[1] (2143) stands for the permutation $1234 \to 2143$, not the cycle $2 \to 1 \to 4 \to 3 \to 2$.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

for example, to the four points $0, 2/3, 1, 2$. Otherwise, $x^2 - x + 1 = 0$, x being a primitive root of -1 such as $\omega = e^{\pi\sqrt{-1}/3}$; the points may now be taken to form an equilateral triangle $1, \omega, 0$ (and ∞). Check that no other harmonic ratios exist.

Exercise 12. Check that the **Schwarzian derivative** $(f')^{-2}[f'f''' - (3/2)(f'')^2]$ commutes with the action of $\Gamma(\mathbb{P}^1) = PSL(2, \mathbb{C})$ on the independent variable; see Ford [1972: 98–101] for more information and for applications of this intriguing object.

1.6 Spherical Geometry

The sphere inherits from the ambient space $\mathbb{R}^3$ its customary (round) geometry, and it is easy to see that the shortest path **(geodesic)** joining two points is an arc of a great circle. Introduce spherical polar coordinates: $x_1 = \sin\varphi\cos\theta$, $x_2 = \sin\varphi\sin\theta$, $x_3 = \cos\varphi$ in which $0 \le \varphi \le \pi$ is the colatitude measured from the north pole and $0 \le \theta < 2\pi$ is longitude. The line element (of arc length) is $ds = \sqrt{(d\varphi)^2 + \sin^2\varphi(d\theta)^2}$. The rotation group $SO(3)$ acts in a distance-preserving way, so you may as well take the first point to be the north pole $n = (1, 0, 0)$. The second point lies on a great circle passing through n, as seen in Fig. 1.6. Plainly, any longitudinal deviation from the great-circle path makes the journey longer.

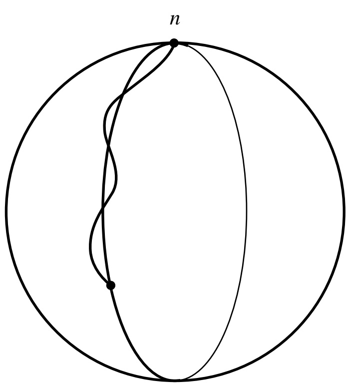

Figure 1.6. Geodesic path on $\mathbb{P}^1$.

Exercise 1. Write this up carefully and find a nice expression for the (shortest) distance.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Exercise 2. Express the spherical distance between two points in the language of the stereographic projection. *Hint*: The line element is $2(1+r^2)^{-1}|dz|$ with $r = |z|$.

Exercise 3. A pair of great circles cuts the spherical surface into two pairs of congruent **lunes**. Check that a lune of angle θ has area 2θ.

Exercise 4. A spherical triangle Δ is formed by joining three points by pieces of great circles. Gauss [1827] found that the area of Δ is the sum of its three interior angles diminished by π. Check this. *Hint*: Each interior angle is marked off by two great circles, determining a lune; see Fig. 1.7.

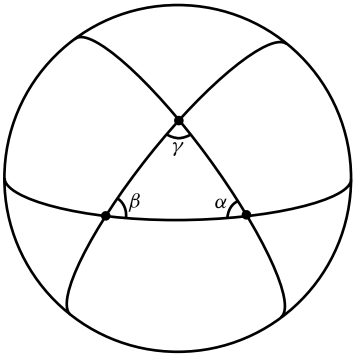

Figure 1.7. Geodesic triangles on $\mathbb{P}^1$.

Exercise 5. Find all the coverings or **tessellations** of the spherical surface by f congruent equilateral triangles, e triangles meeting at each corner. *Hint*: $4/f = 6/e - 1$, the only solutions being $(e, f) = (3, 4), (4, 8), (5, 20)$. To what Platonic solids do these correspond? Compare Section 7.

Exercise 6. Prove that the area of a spherical disk of (geodesic) radius r satisfies

$$A(r) = \pi r^2 \times [1 - r^2/12 + O(r^4)].$$

The number

$$k = \lim_{r \downarrow 0} \frac{12}{\pi} r^{-4}[\pi r^2 - A(r)] = +1$$

measures the deviation of the (round) geometry of the sphere from the (flat) geometry of the plane. It is the **Gaussian curvature** of the sphere; obviously, the plane has curvature $k = 0$; compare Section 9 for the (hyperbolic) geometry of the half-plane with curvature $k = -1$.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

1.7 Finite Subgroups and the Platonic Solids

The finite subgroups of $\Gamma = \Gamma(\mathbb{P}^1)$ were already known to Kepler [1596] through their connection with the five **Platonic solids** of antiquity. This will be explained in a moment. The present section serves also as a prototype for the considerations of Chapter 4 about arithmetic subgroups of $PSL(2, \mathbb{R})$; compare Ford [1972: 127–36] and Coxeter [1963].

Let Γ_0 be such a subgroup of order $d < \infty$ and suppose (what is no loss) that no element of Γ_0 fixes ∞, the identity excepted, of course. This can always be achieved by preliminary conjugation of Γ_0 in the ambient group: In fact, the nontrivial elements of Γ_0 have (in sum) at most $2d - 2$ fixed points and almost any conjugation will move them all away from ∞. Let $c \in \mathbb{P}^1 - \infty$ be distinct from any of these fixed points. Then the **orbit** $\Gamma_0 c = \{gc\colon g \in \Gamma_0\}$ is simple, that is, it is comprised of d distinct points. Pick two such finite points a and b with different orbits. The product $j(z)$ of $(gz - a)(gz - b)^{-1}$, taken over the substitutions $g \in \Gamma_0$, is a rational function of z, of degree d, with three properties. (1) It is invariant under the action of Γ_0. (2) It separates orbits; especially, it has d simple roots and/or poles at the orbit of a/b. (3) Any other (rational) invariant function of Γ_0 is a rational function of j. In short, the field of invariant functions of Γ_0 is $\mathbb{C}(j)$; for this reason, j is called the **absolute invariant** of Γ_0.

Proof of (1). This is self-evident.

Proof of (2). The rational function j is of degree d as it has (simple) poles at the orbit of b and no others. The rest follows from the fact that j takes each value d times and so takes distinct values at distinct orbits of Γ_0.

Proof of (3). Let $\mathbf{K}_0 = \mathbf{K}(\Gamma_0)$ be the field of invariant functions. It is a subfield of $\mathbf{K} = \mathbb{C}(z)$, properly including the constant field $\mathbb{C}$ because j is in it. By Luroth's theorem, $\mathbf{K}_0 = \mathbb{C}(j_0)$ for some $j_0 \in \mathbf{K}_0$ of minimal degree d_0. But this function takes the same value d times on any simple orbit, so $d_0 \geq d = \deg j$. In short, j is of minimal degree and j_0 is a rational function of it, of degree 1.

Exercise 1. The derivative j' of j is of degree $2d$. Why? Note that its roots come from exceptional (nonsimple) orbits.

The next step in the classification of the subgroups of Γ_0 is to describe the family of points fixed by some nontrivial element of Γ_0. It is divided into $h < \infty$ **classes k** according to the action of Γ_0; $n(\mathbf{k}) < \infty$ is the number of points in the class. The function j is invariant, so it takes the same value $j(\mathbf{k})$ at

Table 1.7.1. *Multiplicities*

$m(\mathbf{k}_1)$	$m(\mathbf{k}_2)$	$m(\mathbf{k}_3)$	d
2	2	m	$2m$
2	3	3	12
2	3	4	24
2	3	5	60

each point of $\mathbf{k}$, the latter being a full orbit of Γ_0; moreover, this value is taken each time with the same multiplicity $m(\mathbf{k})$.

Exercise 2. Check that.

The aim of the next few lines is to recompute the degree $2d$ of j' with the aid of these numbers $n(\mathbf{k})$ and $m(\mathbf{k})$ so as to obtain a relation between them. $d = n(\mathbf{k}) \times m(\mathbf{k})$ is plain, j being of degree d. The situation at ∞ has now to be clarified: ∞ is not a pole of j by choice of b, so you have an expansion in powers of the local parameter $1/z$: $j(z) = c_0 + c_1 z^{-1} + \cdots$ in which $c_1 \neq 0$ because the orbit of ∞ is simple. This means that j' has a double root at ∞ for a new count of its degree as per its roots:

$$2d = \deg j' = 2 + \sum_{\mathbf{k}} n(\mathbf{k})[m(\mathbf{k}) - 1],$$

which is to say

$$h - 2 + \frac{2}{d} = \sum_{\mathbf{k}} \frac{1}{m(\mathbf{k})},$$

as you will check using $d = n \times m$. This relation gives rise to a complete list of the possible values of d, h, n, and m. The fact is that *either $d \geq 2$ is arbitrary, $h = 2$, and $m(\mathbf{k}) = d$, or else $h = 3$ and the numbers m fall into one of the four patterns of Table 1.7.1.*

Proof. $m(\mathbf{k}) \geq 2$ (why?) so $h - 2 + 2/d \leq h/2$, and this is contradictory unless $h \leq 3$, that is, $h = 2$ or 3, $h = 1$ being the case $d = 1$, which is trivial. (Why is that?)

Case 1. $h = 2$. There are two classes and $[m(\mathbf{k}_1)]^{-1} + [m(\mathbf{k}_2)]^{-1} = 2/d$. Now $m \leq d$ divides d, and if $m < d$, then it is $\leq d/2$ already and the identity cannot balance. In short, $m = d$ for both classes and each comprises just one point.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Table 1.7.2. *Platonic solids*

Solid	Faces	Edges	Corners	m
tetrahedron	4 triangles	6	4	3
cube	6 squares	12	8	3
dodecahedron	12 pentagons	30	20	5
octahedron	8 triangles	12	6	4
icosahedron	20 triangles	30	12	3

Case 2. $h = 3$. Now $[m(\mathbf{k}_1)]^{-1} + [m(\mathbf{k}_2)]^{-1} + [m(\mathbf{k}_3)]^{-1} = 1 + 2/d$ and if all three multiplicities were three or more, you would have $1 + 2/d \leq 1$, so that is out, and *some* class, say the first, has multiplicity two: $m(\mathbf{k}_1) = 2$. $m(\mathbf{k}_2) \leq m(\mathbf{k}_3)$ can also be assumed, and you have $[m(\mathbf{k}_2)]^{-1} + [m(\mathbf{k}_3)]^{-1} = 1/2 + 2/d$.

Case 2−. $m(\mathbf{k}_2) = 2$. Then $d = 2m(\mathbf{k}_3)$, which is line 1 of Table 1.7.1.

Case 2+. $m(\mathbf{k}_2) \geq 3$. $m(\mathbf{k}_2) = 4$ or more is contradictory in view of $[m(\mathbf{k}_2)]^{-1} + [m(\mathbf{k}_3)]^{-1} \leq 1/2 < 1/2 + 2/d$, so you have $m(\mathbf{k}_2) = 3$ and $[m(\mathbf{k}_3)]^{-1} = 1/6 + 2/d$. It follows that $m(\mathbf{k}_3) = 3, 4,$ or 5 producing lines 2, 3, 4 of Table 1.7.1 with $d = 12, 24, 60$. The proof is finished.

Table 1.7.1 displays the possibilities. Now they must be realized concretely. The role of absolute invariant may be played by $(aj+b)(cj+d)^{-1}$ for any substitution $[ab/cd]$ of $\Gamma = PSL(2, \mathbb{C})$, so you may assign the values $j(\mathbf{k}) = 0, 1, \infty$ (or $0, \infty$) to the $h = 3$ (or 2) classes; also, Γ_0 can be conjugated in Γ, permitting you to distribute all (or two) of the points $0, 1, \infty$ among the distinct classes as you will. This freedom permits the group and its absolute invariant to be brought to standard form, with the final result that *each pattern of multiplicities is realized by just one subgroup* $\Gamma_0 \subset \Gamma$, *up to conjugation; in particular, lines* 2–4 *of Table 1.7.1 are realized by the proper* (= *orientation-preserving*) *symmetries of the five Platonic solids of antiquity listed in Table 1.7.2*; compare Fig. 1.8. The idea will now be illustrated in the two simplest cases.

Example 1. $h = 2, m(\mathbf{k}) = d$. The two classes $\mathbf{k}_1$ and $\mathbf{k}_2$ are single points at which j takes its value d-fold; they may be placed at 0 and ∞ and assigned the values $j(\mathbf{k}_1) = 0$ and $j(\mathbf{k}_2) = \infty$. Then you may take $j(z) = z^d$ (how come?), and the substitutions of Γ_0 fixing, as they do, both 0 and ∞, must be of the form $g\colon z \mapsto \omega z$, ω being a dth root of unity because of the invariance of j, and every one of these roots must be employed, Γ_0 having d elements. In short, Γ_0 *is the cyclic group of rotations about the north pole by the dth roots of unity.*

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

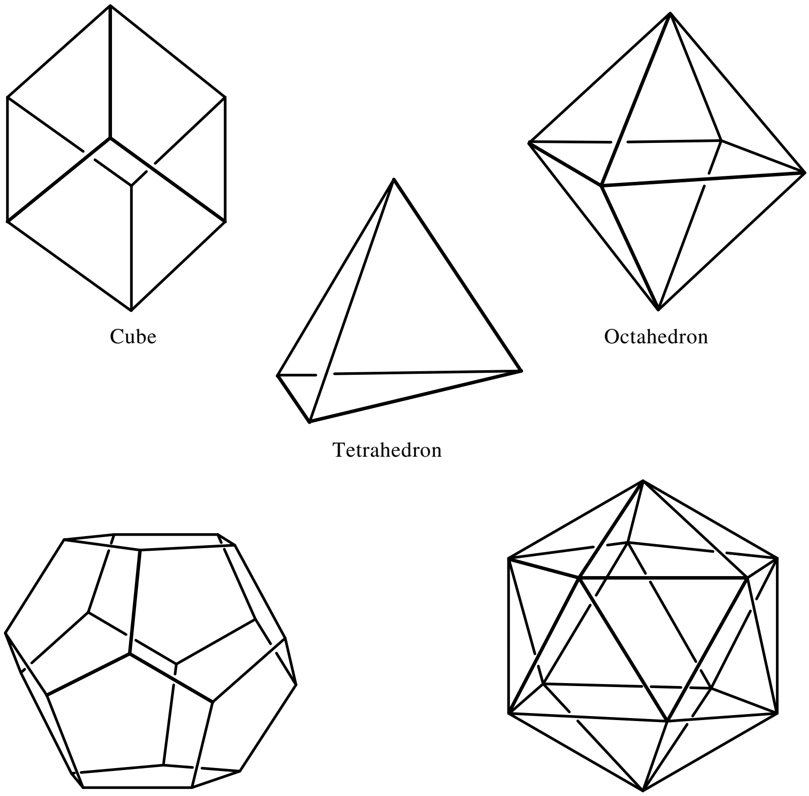

Figure 1.8. The five Platonic solids.

Example 2. $h = 3, m(\mathbf{k}_1) = 2, m(\mathbf{k}_2) = 2, m(\mathbf{k}_3) = m, d = 2m$. This is line 1 of Table 1.7.1. Let $j(\mathbf{k}_1) = 0$, $j(\mathbf{k}_2) = 1$, and $j(\mathbf{k}_3) = \infty$ and let $\mathbf{k}_3$, which has just $2 = d/m$ points, be the pair $0, \infty$, so that j has m poles at each of these points. Γ_0 contains a substitution $g \neq$ the identity fixing ∞, and this must also fix 0, $\mathbf{k}_3$ being an orbit of the group. Then g is of the form $z \mapsto \omega z$, and ω can only be an mth root of unity, the invariant function j having m poles at ∞. Any substitution of Γ_0 not of this type still preserves $\mathbf{k}_3$ and so must exchange 0 and ∞, that is, it is of the form $z \mapsto \omega/z$. Then $j(z) = z^{-m} \prod (z - a)^2$, up to a constant multiplier, the product being taken over $a \in \mathbf{k}_1$, and ω can only be an mth root of $\prod a^2$, by the invariance of j. But now $d = 2m$ forces every mth root of unity to appear in the first round, and placing 1 in the class $\mathbf{k}_1$, as you may, requires that $\mathbf{k}_1$, itself, is just the mth roots of unity; that $\prod a^2 = 1$; that every ω in the second round is an mth root of unity, too; and that each of these must also appear. In short, Γ_0 is the *so-called dihedral group comprised of the north-pole*

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

rotations by m*th roots of unity, with the involution* $z \mapsto 1/z$ *adjoined*; it is the symmetry group of a degenerate polyhedron formed by pasting together two copies of the regular m-sided polygon, each divided into m slices of the pie.

Exercise 3. Check that the absolute invariant is $j(z) = z^m + z^{-m}$, up to trivialities.

Exercise 4. The group $\mathfrak{H}$ of anharmonic ratios $x, 1/x, 1 - x, 1/(1 - x)$, $(x - 1)/x, x/(x - 1)$ of Section 5 illustrates the case $d = 6$. Use Luroth's theorem to check that

$$j_6(x) = \frac{4}{27} \frac{(x^2 - x + 1)^3}{x^2(1 - x)^2}$$

is an absolute invariant and reduce it to $(x^3 + x^{-3} + 2)/4$ by conjugation. The function j_6, harmless as it may look, plays an important role in several subjects discussed later; see especially, ex. 2.12.3 on the classification of complex tori and Section 4.6 on automorphic functions of $PSL(2, \mathbb{Z})$. Quite a big role for such a little function; compare Ford [1972: 127–36] and also Klein [1884].

Example 3: The Platonic solids. A regular polyhedron in 3-dimensional space has f faces, say, and if each face has n sides and if m faces meet at each corner, then it has $e = nf/2$ edges and $c = nf/m$ corners for an Euler number of $2 = f - e + c = f \times (1 - n/2 + n/m)$.

Exercise 5. Check that $n \leq 5$; after all, for $n = 6$, you have the hexagonal tessellation of the plane and do not get a 3-dimensional object at all.

Case 1. $n = 3$. $4m = f(6 - m)$ so $3 \leq m \leq 5$, and these choices are all permitted: $m = 3$ with $f = 4$ (for the tetrahedron), $m = 4$ with $f = 8$ (octahedron), and $m = 5$ with $f = 20$ (icosahedron).

Case 2. $n = 4$. $2m = f(4 - m)$, the only possibility being $m = 3$ with $f = 6$, that is, a cube.

Case 3. $n = 5$. $4m = f(10 - 3m)$ and only $m = 3$ with $f = 12$ will do (dodecahedron).

Coxeter [1963] presents a beautiful discussion of these issues. Now come the symmetry groups of the solids. Let S_n be the symmetric group of permutations of n letters $\{1, 2, 3, \ldots, n\}$ and A_n the alternating group of permutations of parity $+1$, preserving the sign of the square root of the discriminant $\sqrt{\Delta} = \prod_{i<j}(x_i - x_j)$. Line 2 of Table 1.7.2 is realized by the tetrahedral group A_4, line 3 by the group S_4 of the cube or of its dual the octahedron, and line 4 by the icosahedral group A_5 which is also the group of the (dual) dodecahedron. Here *duality* is effected by placing a corner at the center of each face of the original solid.

Exercise 6. Make yourself a cardboard icosahedron and check its group A_5 by hand.

The orders of the groups are easy to compute. If the solid has f faces and each face has e edges, then you can map face 1 to any other and rotate the displaced face in e different ways for a total count of $e \cdot f = 3 \cdot 4 = 12$ for the tetrahedron, $6 \cdot 4 = 24$ for the cube, and $5 \cdot 12 = 60$ for the icosahedron. The tetrahedron has four corners which any symmetry permutes; odd permutations such as (2134) are improper, so the group can only be A_4. The cube has 24 symmetries. These can be enumerated as follows: Corner 1 can be sent to its own (front) face or to the back, and to the bottom or to the top, for a count of four; there are three further possibilities for corner 2 and two for corner 3, for a total count of $4 \cdot 3 \cdot 2 = 24$. This mode of counting identifies the group as S_4. The dodecahedron is a bit more complicated. There are 20 corners which may be divided into four families of five, each family belonging to a single face and labeled 12345. A symmetry maps corner 1 of family 1 to one of the other four families and ascribes to it a new label; this can be done in $4 \cdot 5 = 20$ ways. The two corners of family 1 adjacent to corner 1 can then be placed in three ways for a total count of 60. It follows that the group is part of S_5; it can only be A_5.

Exercise 7. Why?

Now imagine such a Platonic solid inscribed in a sphere and project its edges outward, from the center to the surface, to obtain a tessellation of the sphere. Its symmetries appear as elements of $\Gamma = PSL(2, \mathbb{C})$, so these Platonic groups appear among the subgroups Γ_0; indeed, you now see that these exhaust the possibilities.

Proof for line 2 of Table 1.7.1. $n(\mathbf{k}_3) = 4$ and Γ_0 permutes these four points, its action being determined by what it does to any three of them. $\Gamma_0 \subset S_4$ follows and $d = 12$ identifies it as A_4, S_4 having no other subgroup of index 2.

Proof for line 3. This is the same, only now $d = 24$ is the order of S_4.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Proof for line 4. This is the same as for line 2, only now with $A_5 \subset S_5$.

Klein [1884: 110–43] computed the absolute invariants for the most symmetrical disposition of the solids:

$$j_2(z) = \frac{(z^4 - 2\sqrt{-3}z^2 + 1)^3}{(z^4 + 2\sqrt{-3}z^2 + 1)^3} \qquad \text{for the tetrahedron,}$$

$$j_3(z) = \frac{(z^8 + 14\,z^4 + 1)^3}{108\,z^4(z^4 - 1)^4} \qquad \text{for the cube,}$$

$$j_4(z) = \frac{(-z^{20} + 228\,z^{15} - 494\,z^{10} - 228\,z^5 - 1)^3}{1728\,z^5(z^{10} + 11\,z^5 - 1)^5} \qquad \text{for the dodecahedron.}$$

The phrase *symmetrical disposition* is explained for the dodecahedron: j_4 takes the value 1 with multiplicity 2 at the 30 midpoints of the edges, 0 with multiplicity 3 at the 20 corners, and ∞ with multiplicity 5 at the midpoints of the 12 faces, in accord with line 4 of Table 1.7.1. The claim is that line 4 can be realized only in this way, up to conjugation.

Proof for the tetrahedron. The class $\mathbf{k}_1$ contains six points, of multiplicity 2 each, with assigned value $j(\mathbf{k}_1) = 1$. Let $0, 1, \infty$ be placed in $\mathbf{k}_1$. Then Γ_0 contains a nontrivial substitution $z \mapsto a+bz$ fixing ∞, and you must have $b^2 = 1$ to fix the form of $j(z) = 1+c/z^2+\cdots$. The possibility $b = +1$ is excluded by the fact that the substitution is of order at most 12, so $b = -1$, and conjugation of Γ_0 by the substitution $z \mapsto z - a/z$ reduces a to 0, with the result that j is invariant under the substitution $z \mapsto -z$ and so a function of z^2. Now Γ_0 also contains a substitution $z \mapsto a + b/z$ that sends $0 \in \mathbf{k}_1$ to ∞. This means that, besides 0 and ∞, $\mathbf{k}_1$ also contains the six points $\pm a$, $\pm(a+b/a)$, and $\pm(a-b/a)$ for a total count of eight, which would be too many if they were distinct. Take $a \neq 0$. Then $0 = a \pm b/a$ is one possible conjunction, in which case $b = \pm a^2$, and the class-preserving substitution $z \mapsto a + b/(\pm z) = a + a^2/z$ sends a to $a \times 2, 3/2, 5/3, 8/5, 3/8, 21/13, \ldots$, which is too many points. Otherwise, $a = -(a \pm b/a)$, in which case $b = \pm 2a^2$, and $z \mapsto a + b/(\pm z) = a + 2a^2/z$ sends a to $a \times 3, 5/3, 11/5, 21/4, 43/21, 85/43, \ldots$, which is also wrong, so $a = 0$ and a further conjugation of Γ_0 reduces b to $+1$, with the result that $j(z) = j(1/z)$. Now assign the values $j(\mathbf{k}_2) = 0$ and $j(\mathbf{k}_3) = \infty$ and reflect that each of these classes contains four points of multiplicity 3, whence $j(z)$ is of the form $(z^4 + 2\alpha z^2 + 1)^3/(z^4 + 2\beta z^2 + 1)^3$. It remains to pin down the numbers α and β. Let $z \mapsto (z + a)(z + b)^{-1}$ be a substitution of Γ_0 mapping

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

∞ to $1 \in \mathbf{k}_1$. Then $a/b \in \mathbf{k}_1$ and of the eight points $0, \infty, \pm 1, \pm a/b, \pm b/a$ not more than six can be distinct.

Case 0. $ab = 0$. Then Γ_0 contains a substitution of the form $z \mapsto 1 + cz$, the number c being a primitive nth root of unity, and this leads to a contradiction in every case $n = 2, 3, 4, 6$ (n divides 12).

- $n = 2$: $c = -1$ and that is impossible since, together with $z \mapsto 1 - z$, $z \mapsto -z \mapsto 1 + z$ also belongs to Γ_0.
- $n = 3$ produces, from ± 1, four additional points $\pm(1 + c), \pm(1 - c)$, and this is too many.
- $n = 4$: $c = \pm\sqrt{-1}$ and $z \mapsto \pm z \mapsto 1 + \sqrt{-1}z$ produces, from ± 1, the distinct points $\pm 1 \pm \sqrt{-1}$.
- $n = 6$ produces, from 1, five additional distinct points $1+c+\cdots+c^m$ $(m \leq 5)$.

Case 1. $1 = a/b$ is not possible, but $1 = -a/b$ *is*, in which case the substitutions known to date produce new points $\pm(1+a)(1-a)^{-1}$ and $\pm(1-a)(1+a)^{-1}$ of $\mathbf{k}_1$ which cannot be distinct. This forces $a^2 = -1$ so, besides $0, \infty$, and ± 1, $\mathbf{k}_1$ also contains $\pm(1 + \sqrt{-1})(1 - \sqrt{-1})^{-1} = \pm\sqrt{-1}$.

Case 2. $1 \neq -a/b$. Then $a/b = \pm b/a$ implies $a = \pm\sqrt{-1}b$, and the same result is obtained: $\mathbf{k}_1$ contains $a/b = \pm\sqrt{-1}$.

It follows that

$$1 = j(\pm 1) = \left(\frac{1+\alpha}{1+\beta}\right)^3 \quad \text{and} \quad 1 = j(\pm\sqrt{-1}) = \left(\frac{1-\alpha}{1-\beta}\right)^3$$

from which follow $\alpha^2 = \beta^2$ and $3\alpha + \alpha^3 = 3\beta + \beta^3$. But $\alpha \neq \beta$ since $j \not\equiv 1$, so $\alpha = -\beta$ and $\alpha^2 = -3$, that is, $\alpha = \pm\sqrt{-3}$ and $\beta = \mp\sqrt{-3}$, as per the formula for $j = j_2$ previously displayed.

Proof for the dodecahedron. The values $j(\mathbf{k}_1) = 1$, $j(\mathbf{k}_2) = 0$, and $j(\mathbf{k}_3) = \infty$ are assigned and the point ∞ is placed in the class $\mathbf{k}_3$. Γ_0 contains a nontrivial substitution A fixing ∞. This must be of the form $z \mapsto a + bz$ with $b^5 = 1$ to fix the pole of $j(z) = cz^5 + \cdots$. Now A is of finite order n and $A^n z = a(1 + b + \cdots + b^{n-1}) + b^n z$ so $b \neq 1$ can only be a primitive fifth root of unity, $b = \omega = e^{2\pi\sqrt{-1}/5}$, say, and you can conjugate the group by $z \mapsto z - a(1 - \omega)^{-1}$ to bring A to its simplest form: $z \mapsto \omega z$. Then j is a function of z^5. Now place the point 0 in the class $\mathbf{k}_3$. The group Γ_0 contains a substitution B: $z \mapsto a + b/z$ mapping 0 to ∞. If $a \neq 0$, B will map ∞ to a new point a of $\mathbf{k}_3$. Then $a, \omega a, \ldots, \omega^4 a$ belong to $\mathbf{k}_3$, and new applications of the substitutions A and B produce further points of $\mathbf{k}_3$: $\omega^i a + \omega^j b/a (0 \le i, j < 5)$. But $\mathbf{k}_3$ contains just 12 points, including ∞, so at most 11 of these new points are distinct, and a picture will convince you that this is not possible. The upshot is that $a = 0$ and a further conjugation reduces b to 1 and j to the form

$$\frac{(z^{10} - az^5 + b + a/z^5 + 1/z^{10})^3}{f(z^5 + c - 1/z^5)^5} = \frac{(z^{20} - az^{15} + bz^{10} + az^5 + 1)^3}{fz^5(z^{10} + cz^5 - 1)^5}$$

with undetermined constants a, b, c, and f, in which the top accounts for the 20 points of the class $\mathbf{k}_2$ together with the pole at ∞, and the bottom for $\mathbf{k}_3 - \infty$. The final step is to require that j take the value 1 with multiplicity 2 at each of the 30 points of the class $\mathbf{k}_1$. Now the derived function j' is of degree 70, having four poles at ∞ and six more at each of the other points of $\mathbf{k}_3$; it vanishes 40-fold on class $\mathbf{k}_2$, so the condition to be imposed is that the square of j', with its poles and its roots of the class $\mathbf{k}_2$ removed, should be a constant multiple of

$$(z^{20} - az^{15} + bz^{10} + az^5 + 1)^3 - fz^5(z^{10} + cz^5 - 1)^5.$$

The rest of the computation elicits the values $a = 228$, $b = 494$, $c = 11$, $f = -1728$. This is omitted, as there seems to be no slick way to do it.

1.8 Automorphisms of the Half-Plane

The group $\Gamma(\mathbb{H})$ of automorphisms of the open upper half-plane $\mathbb{H} = \{z = x_1 + \sqrt{-1}x_2 : x_2 > 0\}$ may be identified with the real special linear group $PSL(2, \mathbb{R})$.

Proof. Let g be an automorphism of $\mathbb{H}$ mapping $\sqrt{-1}$ to $a + \sqrt{-1}b \, (b > 0)$ and observe that $g_1 = b^{-1}(g - a)$ fixes $\sqrt{-1}$. Let h be the standard map $z \mapsto (z - \sqrt{-1})(z + \sqrt{-1})^{-1}$ of $\mathbb{H}$ to the disk $|h| < 1$. Then $g_2 = h \circ g_1 \circ h^{-1}$ is an automorphism of the disk fixing the origin and $|g_2(z)/z| \le 1$ by application of the maximum modulus principle on the perimeter. The same idea applies to the inverse map g_2^{-1}, so the reciprocal modulus $|z/g_2(z)| = |g_2^{-1}(z')/z'|$ with $z' = g_2(z)$ is likewise ≤ 1. The upshot is that $g_2(z)/z$ is of modulus $\equiv 1$ and so must be constant. This pretty trick is due to H. A. Schwartz. The rest is computation if you like. A better way is to note that $g = [ba/01] \circ h^{-1} \circ g_2 \circ h$ is a fractional linear substitution preserving the completed line $\mathbb{R} + \infty$ bordering $\mathbb{H}$ and to deduce that it belongs to $PSL(2, \mathbb{R})$.

Exercise 1. Do it.

Exercise 2. Check that a nontrivial element $[ab/cd]$ of $\Gamma(\mathbb{H}) = PSL(2, \mathbb{R})$ is conjugate to (1) a magnification ($d = 1/a, b = c = 0, a > 0$), (2) a translation ($a = d = 1, c = 0, b \in \mathbb{R}$), or (3) a rotation ($a = d = \cos\theta, -b = c = \sin\theta, \theta \neq 0, \pi$) according as the absolute value of its trace is $> 2, = 2$, or < 2.

Exercise 3. Use ex. 2 to prove that any subgroup of $PSL(2, \mathbb{R})$ isomorphic to $\mathbb{Z}^2$ comes as close to the identity as you like. *Hint*: What commutes with a translation *is* a translation.

Exercise 4. Show that $\Gamma = PSL(2, \mathbb{R})$ will move one point of $\mathbb{H}$ to $\sqrt{-1}$ and any *second* point to some position on the imaginary half-line.

1.9 Hyperbolic Geometry

This is a model of the non-Euclidean geometry discovered by Bolyai, Lobatchevsky, and Gauss about 1820; it is connected in a beautiful way to $PSL(2, \mathbb{R})$ and $\mathbb{H}$. The latter can be equipped with the **hyperbolic line element** $ds = x_2^{-1}\sqrt{dx_1^2 + dx_2^2}$ of Liouville and Beltrami (1868) and Klein (1870), which was rediscovered by Poincaré [1882] and usually called by his name, relative to which the vicinity of each point looks like a mountain pass or saddle point: two ridges rising, one on either hand, two valleys falling away. The geodesics of this geometry are semicircles with centers on the bordering line $x_2 = 0$ and/or vertical lines, and $\Gamma(\mathbb{H}) = PSL(2, \mathbb{R})$ is its group of proper (= orientation-preserving) rigid motions, playing the same role as the rotation group $SO(3)$ for the (round) spherical geometry of Section 6 or the Euclidean motion group (translations and rotations) for the (flat) geometry of $\mathbb{R}^2$. See Milnor [1982] for historical remarks, Pogorelov [1967] and Beardon [1983] for further information, and Section 4.9 for a deep connection to complex function theory.

Exercise 1. Check that the hyperbolic line element is invariant under the action of $\Gamma(\mathbb{H})$.

Now let $\sqrt{-1}a$ and $\sqrt{-1}b$ ($0 < a < b$) be two points of the imaginary half-line. Inspection of the line element shows that any deviation from the vertical path joining them makes the journey longer, so *that* is the geodesic, and a self-evident application of ex. 8.4 confirms the statement made before that the general geodesic is either a semicircular arc meeting $\mathbb{R}$ at 90° or else a vertical line. The identification of $\Gamma(\mathbb{H})$ as the rigid motions is a by-product.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Exercise 2. The hyperbolic distance from $\sqrt{-1}a$ to $\sqrt{-1}b$ is $\int_a^b x_2^{-1} dx_2 = \log(b/a)$. Confirm the general formula for the distance between any two points of $\mathbb{H}$: $\cosh(d(p,q)) = 1 + (1/2)\times$ the flat distance between p and q, squared, divided by the product of their heights. *Hint*: The right-hand side had better be invariant under the action of $\Gamma(\mathbb{H})$.

Exercise 3. Prove that, for fixed θ, the point

$$x = x_1 + \sqrt{-1}x_2 = \frac{\sqrt{-1}e^r \cos\theta - \sin\theta}{\sqrt{-1}e^r \sin\theta + \cos\theta}$$

traverses a geodesic issuing from $\sqrt{-1}$ as r runs from $-\infty$ to ∞, and that, for fixed $r > 0$, it passes (twice) about the geodesic circle with center at $\sqrt{-1}$ and (hyperbolic) radius r as θ runs from 0 to 2π. The latter is also a *flat* circle of radius $\sinh r$ centered at $\sqrt{-1}\cosh r$, as you will check.

Exercise 4. Deduce that the line element $(x_2)^{-1}\sqrt{(dx_1)^2 + (dx_2)^2}$ can be expressed as $\sqrt{(dr)^2 + (\sinh r)^2(d\theta)^2}$ and use it to evaluate a) the area $A(r) = 2\pi(\cosh r - 1)$ of the hyperbolic circle of radius r and b) the curvature $k = \lim_{r\downarrow 0} \frac{12}{\pi} r^{-4}[\pi r^2 - A(r)] = -1$; compare ex. 6.6.

A down-to-earth picture of curvature $k = -1$ is obtained from the simple surface of revolution $y = h(x)$ seen in Fig. 1.9. The so-called principal curvatures are

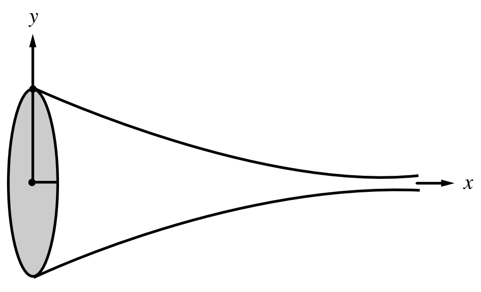

Figure 1.9. A surface with curvature -1.

$k_1 = -h''[1 + (h')^2]^{-3/2}$ in the plane of the paper and $k_2 = 1/h\sqrt{1 + (h')^2}$ in the plane perpendicular to the paper and the surface. Their product is the Gaussian curvature k, so the condition $k = -1$ becomes $h'' = h[1 + (h')^2]^2$, producing a convex function starting at $h(0) = 1$ and vanishing at ∞, expressed

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

explicitly as

$$x = \sqrt{1-h^2} + \log \frac{1-\sqrt{1-h^2}}{h}. \tag{1}$$

Exercise 5. Check this.

Exercise 6. Compute the hyperbolic area of a geodesic triangle with interior angles α, β, γ.

Pogorelov [1967: 127–67] is recommended as an introduction to this circle of ideas; see also Stillwell [1992] which is nice and elementary.

1.10 Projective Curves

Clemens [1980] presents a splendid account of this subject; Kirwan [1992] is fine, too. The projective line $\mathbb{P}^1$ you know. The projective plane $\mathbb{P}^2$ is similarly defined: It is the space of complex lines in $\mathbb{C}^3$ obtained by identifying two points of $\mathbb{C}^3 - 0$ that differ by a nonvanishing complex multiplier. It is covered by three patches $U_1 = \mathbb{C} \times \mathbb{C} \times 1$, $U_2 = \mathbb{C} \times 1 \times \mathbb{C}$, and $U_3 = 1 \times \mathbb{C} \times \mathbb{C}$, and, with a self-evident extension of the terminology of Section 2, it is a compact manifold of (complex) dimension 2. The higher projective spaces $\mathbb{P}^d$ ($d \geq 3$) follow the same pattern.

Exercise 1. A *line* in $\mathbb{P}^2$ is determined by the vanishing of a linear form $ax + by + cz = 0$ with fixed $(a, b, c) \in \mathbb{C}^3 - 0$. Check that it is nothing but a copy of the projective line $\mathbb{P}^1$.

To explain projective curves, think first of the *real* quadratic curves (conic sections) of the schoolroom. They come in three varieties: circle, hyperbola, and parabola, typified by $x^2 + y^2 = 1$, $xy = 1$, and $x^2 = y$; see Fig. 1.10. A different and, in many respects, a superior picture is obtained by complexifying x and y; it is even better to compactify everything, by taking a projective standpoint in $\mathbb{P}^2$, so as to take into account what is happening at *infinity*. The idea is due to Bezout [1779]. The case of the circle will convey the idea. The old curve $x^2 + y^2 = 1$ in $\mathbb{R}^2$ is replaced by the locus $x^2 + y^2 + z^2 = 0$ in $\mathbb{C}^3 - 0$. This makes projective sense, each term being of the same degree, and so defines a **projective curve** in $\mathbb{P}^2$ of 1 complex (= 2 real) dimension. The old curve (and more) is seen in the patch $\mathbb{C}^2 \times \sqrt{-1} = \mathbb{C}^2 \times 1$, projectively, but now you pick up two new (compactifying) points "at infinity": $(1, \pm\sqrt{-1}, 0)$ in $(1 \times \mathbb{C}^2) \cap (\mathbb{C} \times 1 \times \mathbb{C})$. The same recipe applies to the projective hyperbola

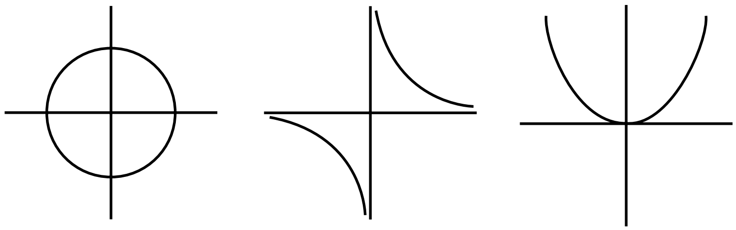

Figure 1.10. Quadratic curves.

$xy - z^2 = 0$ and to the parabola $x^2 - yz = 0$, which are evidently the same in $\mathbb{P}^2$. In fact, *all three types of curves fall together into a single projective class*. The (1:1 projective) substitution

$$x \mapsto x + \sqrt{-1}y, \;\; y \mapsto x - \sqrt{-1}y, \;\; z \mapsto \sqrt{-1}z$$

converts $xy = z^2$ into $x^2 + y^2 + z^2 = 0$. In short, up to such substitutions, the conic sections are indistinguishable in $\mathbb{P}^2$. But what does the circle $\mathbf{C}$: $x^2 + y^2 + z^2 = 0$ really look like in $\mathbb{P}^2$? The pretty answer is: *the projective line*. The proof is easy. $\mathbb{P}^1$ is viewed as $\mathbb{C} + \infty$ and provided with the parameter w. Then $x = (1/2)(w + 1/w)$ and $y = (1/2\sqrt{-1})(w - 1/w)$ solve $x^2 + y^2 = 1$ provided $w \neq 0, \infty$. This presents, in a 1:1 manner, the finite part of $\mathbf{C}$ that lies in the patch $\mathbb{C}^2 \times \sqrt{-1}$: in fact, $w = x + \sqrt{-1}y$. To cope with the points at ∞, the correspondence $[x, y, z] = [(w + 1/w)/2, \sqrt{-1}(w - 1/w)/2, \sqrt{-1}]$ is expressed in the projectively equivalent forms $[(1+1/w^2)/2, (1-1/w^2)/2\sqrt{-1}, \sqrt{-1}/w]$ and $[(w^2 + 1)/2, (w^2 - 1)/2\sqrt{-1}, \sqrt{-1}w]$: The first places the north pole $w = \infty$ in correspondence with $[1/2, 1/2\sqrt{-1}, 0] = [1, -\sqrt{-1}, 0]$ projectively; the second places the south pole $w = 0$ in correspondence with the second point at infinity $[1, \sqrt{-1}, 0]$.

Moral. The recipe places the whole projective circle in faithful (rational) correspondence with the projective line; in particular, *the totality of real solutions of* $x^2 + y^2 = 1$ *is obtained by numerical specialization of the rational functions* $x = \frac{1}{2}(w + 1/w)$ *and* $y = \frac{1}{2\sqrt{-1}}(w - 1/w)$. Do you recognize them?

More generally, any irreducible polynomial $P \in \mathbb{C}[x, y]$ defines, by its vanishing, a projective curve $\mathbf{X}$ in $\mathbb{P}^2$: The individual terms $x^m y^n$ of P are brought up to a common (minimal) degree d by powers of a new variable z. Then the vanishing of P makes projective sense, and the rest is as before; in particular, the original curve $\mathbf{X}_0 = \mathbb{C}^2 \cap \{P(x, y) = 0\}$ is seen in the patch $\mathbb{C}^2 \times 1$. $\mathbb{P}^2$ is compact and $\mathbf{X}$ inherits that; it is also connected, automatically, but that lies deeper.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Amplification 1. **X** need not be a projective line; that is the exception, not the rule. For example, if e_1, e_2, e_3 are distinct complex numbers, then the cubic $y^2 = (x - e_1)(x - e_2)(x - e_3)$ is a torus in $\mathbb{P}^2$ and cannot be identified with $\mathbb{P}^1$ for topological reasons; see Section 1.12 for pictures and Section 2.11 for a full explanation of this striking geometric fact.

Amplification 2. **X** inherits from $\mathbb{P}^2$ the structure of a complex manifold, with exceptions at a few places. Let $P \in \mathbb{C}[x, y, z]$ be the homogeneous polynomial that defines **X** by its vanishing. Then, with the notation $P_1 = \partial P/\partial x$, and so forth, the form $P_1 dx + P_2 dy + P_3 dz$ vanishes on **X**, so if, for example, you place yourself in the patch $\mathbb{C}^2 \times 1$ and if (P_1, P_2) does not vanish, then you can get rid of dz $(z = 1)$ and use the implicit function theorem in the small to solve for y in terms of x (or vice versa), placing a whole patch of neighboring points $\mathfrak{p} = (x, y) \in \mathbf{X}$ in faithful correspondence with a little disk by means of the *local parameter* $x = x(\mathfrak{p})$; see Section 13 for more details from another point of view. The exceptions alluded to are the **singular points** of **X** at which the gradient (P_1, P_2, P_3) vanishes. If the gradient never vanishes the curve is termed **nonsingular**.

Exercise 2. Check that the circle $x^2 + y^2 + z^2 = 0$ is nonsingular.

Exercise 3. The cubic **X**: $y^2 = x^2 + x^3$ has just one singular point: (0, 0, 1), as you will check.

This example is more typical. At $x = 0$, the two analytic branches $y = \pm x\sqrt{1 + x}$ of **X** cross, accidentally so to speak, so, in the small, the curve looks like two complex lines touching at one point and not like a disk; see Fig. 1.11. This can be cured. The branches are distinguished at $x = 0$ by their slopes $y' = \pm 1$, which suggests a new attitude toward **X**, described afresh as the common roots (in $\mathbb{P}^3$) of the *two* relations $xz = y$ and $z^2 = 1 + x$; in fact, $z = y/x$ takes distinct values ± 1 at $x = 0$, depending on the branch, so the old point (0, 0) splits into two separate points $(0, 0 \pm 1)$. The present discussion is just to whet the appetite; it is continued in Section 11, and also in Section 13. For further information, Bliss [1933], Walker [1978], Kirwan [1992], Shafarevich [1977], and Mumford [1976] are recommended, in order of sophistication.

1.11 Covering Surfaces

We review a few topological facts needed for the further study of projective curves and/or Riemann surfaces. The ideas go back to Poincaré. Ahlfors [1973], Forster [1981], and Massey [1991] are recommended for more information.

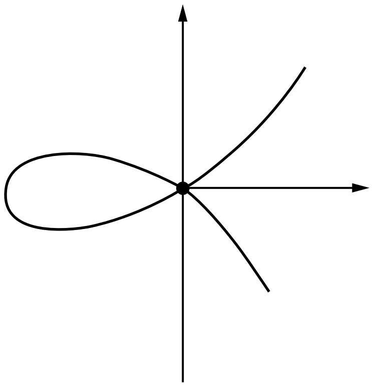

Figure 1.11. Crossing of branches.

Covering Spaces. Let $\mathcal{M}$ and $\mathcal{K}$ be surfaces, that is, topological manifolds of real dimension 2. $\mathcal{K}$ is an (unramified) **cover** of $\mathcal{M}$ if it admits a **projection** p onto $\mathcal{M}$ with the characteristic feature that every point $\mathfrak{p} \in \mathcal{M}$ has an open neighborhood U such that $p^{-1}(U)$ breaks up into disjoint open pieces V of $\mathcal{K}$, finite or countable in number, the restrictions $p\colon V \to U$ being homeomorphisms; see Fig. 1.12. The cardinality of the **fiber** $p^{-1}(\mathfrak{p})$ is independent of the point $\mathfrak{p} \in \mathcal{M}$; it is the **degree** or **sheet number** of the cover.

Exercise 1. Check that $p(z) = z^k (k \in \mathbb{N})$ is a projection from $\mathbb{C} - 0$ to itself. What happens if 0 is included?

Exercise 2. Check that the exponential $\exp\colon \mathbb{C} \to \mathbb{C} - 0$ is a projection.

Exercise 3. Let ω be a complex number of positive imaginary part and let $\mathbb{L}$ be the lattice $\mathbb{Z} \oplus \omega\mathbb{Z}$. The map $p\colon \mathbb{C} \to \mathbf{X} = \mathbb{C}/\mathbb{L}$, reducing the plane modulo the lattice $\mathbb{L}$, induces a topology on the torus $\mathbf{X}$. Check that p is a projection.

Universal Cover. Among all the surfaces that cover $\mathcal{M}$, there is a largest one. This is its **universal cover** $\mathcal{K}$, distinguished by the property that if $\mathcal{K}_0$ is any other cover of $\mathcal{M}$, then $\mathcal{K}$ covers $\mathcal{K}_0$. $\mathcal{K}$ is unique up to homeomorphisms and **simply connected**, that is, any closed loop in $\mathcal{K}$ can be shrunk to a point. This feature, too, distinguishes the universal cover among its competitors. $\mathcal{K}$ is formed by a) fixing a base point $\mathfrak{o} \in \mathcal{M}$; b) stacking up, over the general point $\mathfrak{p} \in \mathcal{M}$, covering points $\mathfrak{q}$, one to each deformation class of curves leading from $\mathfrak{o}$ to $\mathfrak{p}$ in $\mathcal{M}$; and c) lifting up to $\mathfrak{q}$ the topology in the vicinity of $\mathfrak{p}$, as indicated in Fig. 1.12. The projection is $p : \mathfrak{q} \to \mathfrak{p}$.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

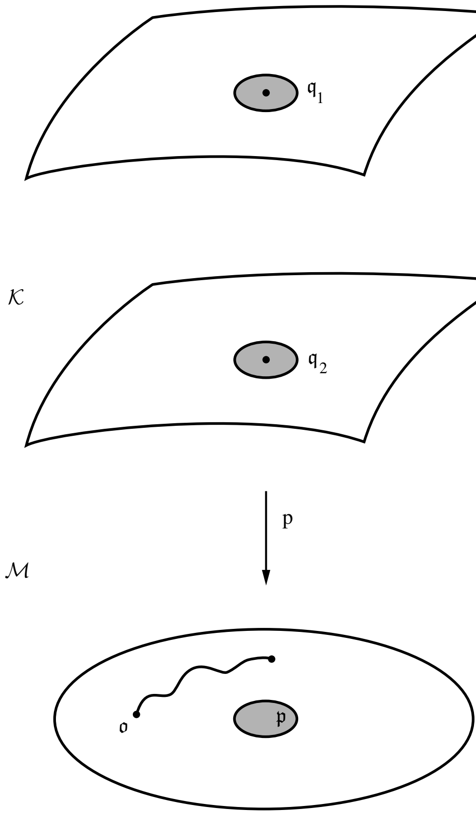

Figure 1.12. The universal cover.

Exercise 4. The universal cover of the torus $\mathbf{X} = \mathbb{C}/\mathbb{L}$ is the plane $\mathbb{C}$, by ex. 3. What is the universal cover of the disk, annulus, sphere, cylinder, or the once-punctured plane?

Lifting and Covering Maps. The universal cover $\mathcal{K}$ is provided with a group $\Gamma(\mathcal{K})$ of **covering maps**. These are homeomorphisms of $\mathcal{K}$ that commute with the projection $p: \mathcal{K} \to \mathcal{M}$; in particular, covering maps preserve fibers. Let $\mathfrak{o}$ be a point of $\mathcal{M}$, fix a point $\mathfrak{o}_1$ of $\mathcal{K}$ covering it, and select a loop in the base starting and ending at $\mathfrak{o}$, as in Fig. 1.13. The beginning of the loop lifts unambiguously to a patch about $\mathfrak{o}_1$ via the inverse projection, and this lifting can be continued without obstruction until the moving base point returns to $\mathfrak{o}$. The lifted loop ends at a point $\mathfrak{o}_2$ covering $\mathfrak{o}$ and the map $\mathfrak{o}_1 \mapsto \mathfrak{o}_2$ may be extended to a covering map of the whole of $\mathcal{K}$ by a self-evident continuation.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

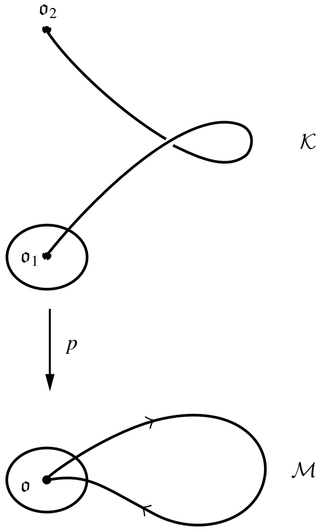

Figure 1.13. Lifting of loops.

Exercise 5. Check all that by means of pictures. Think of an example in which the lifted curve is not closed.

Exercise 6. Prove that every covering map arises in this way.

Exercise 7. Prove that covering maps have no fixed points, the identity excepted.

Fundamental Group. The loops of $\mathcal{M}$, starting and ending at $\mathfrak{o}$, fall into deformation classes; and it is easy to see that these classes form a group: The formation of classes respects the composition of loops effected by passing first about loop 1 and then about loop 2; the identity is the class of the trivial loop = the point $\mathfrak{o}$ itself; the inverse is the class of the loop run backward; and so on. This is the **fundamental group** $\pi_1(\mathcal{M})$ of the surface.

Exercise 8. What is the fundamental group of the annulus, sphere, torus, once-punctured plane, or twice-punctured plane? *Answer*: $\mathbb{Z}$, id, $\mathbb{Z}^2$, $\mathbb{Z}$, the free group on two generators.

Exercise 9. The covering map attached to a loop depends only upon the class of the latter. Why? Deduce that the group $\Gamma(\mathcal{K})$ of the cover is isomorphic to the fundamental group of the base.

More Structure. $\mathcal{K}$ may be equipped with any extra structure, smooth or complex, that $\mathcal{M}$ enjoys: Just lift it up via the inverse projection. Then the covering maps appear as diffeomorphisms of $\mathcal{K}$ if $\mathcal{M}$ is smooth, and as conformal automorphisms if $\mathcal{M}$ is a complex manifold.

Exercise 10. Give all the necessary details in the complex case.

Monodromy. The universal cover leads to a simple proof of the **monodromy theorem** of classical function theory. Let $\mathfrak{f}_0$ be a function element, that is, a convergent power series in the local parameter of a small patch of a complex manifold $\mathbf{X}$, and suppose its continuation along paths of $\mathbf{X}$ is unobstructed. This leads to a cover $\mathcal{K}_0$ of $\mathbf{X}$ whose points are pairs comprising a point $\mathfrak{p}$ of $\mathbf{X}$ and a function element $\mathfrak{f}$ at $\mathfrak{p}$ obtained by continuation of $\mathfrak{f}_0$. Now let $\mathbf{X}$ be simply connected, so that it is its own universal cover. Then $\mathbf{X} = \mathcal{K}_0$, which is to say that the process of continuation of $\mathfrak{f}_0$ leads to a single-valued function on $\mathbf{X}$. This is the monodromy theorem. The more conventional proof is to note that a function element produced by continuation is insensitive to small (and so also to large) deformations of the path as long as the endpoints of the latter are fixed.

Exercise 11. Why?

1.12 Scissors and Paste

The most general compact orientable surface is a sphere with handles, as noted in Section 2. The number of handles is the genus g: 0 for the sphere, 1 for the torus, 2 for the pretzel, and so on. It is a fact that each of these can be equipped with a complex structure, and that in many distinct ways if $g \geq 1$. This will now be made plausible by the familiar informal method of *scissors and paste*.

Take the simple two-valued function $\sqrt{z}$. It is desired to make a 2-sheeted Riemann surface S on which it can live comfortably as a single-valued function. Let $\mathbb{P}^1$ be the projective line with parameter z. The radical branches at $z = 0$ and at $z = \infty$, but may be made single-valued by cutting $\mathbb{P}^1$ from 0 to ∞ in view of the fact that a circuit enclosing both 0 and ∞ produces two changes of sign and so no change at all. It has, however, different signs at the two banks of the cut, the endpoints 0 and ∞ excepted, and now it is clear what to do. Take two copies of the severed $\mathbb{P}^1$ with the cuts opened up to make holes and paste them together as in Fig. 1.14, matching the banks of the cuts according to the sign of the radical to produce a copy S of $\mathbb{P}^1$. The two original copies of $\mathbb{P}^1$ are the **sheets** of S; naturally, upon erasing the cuts, it will not be clear where one sheet ends and the other begins. The discussion is sloppy: For example, careless pasting could leave a crease in S; presumably, it can be ironed out, but

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

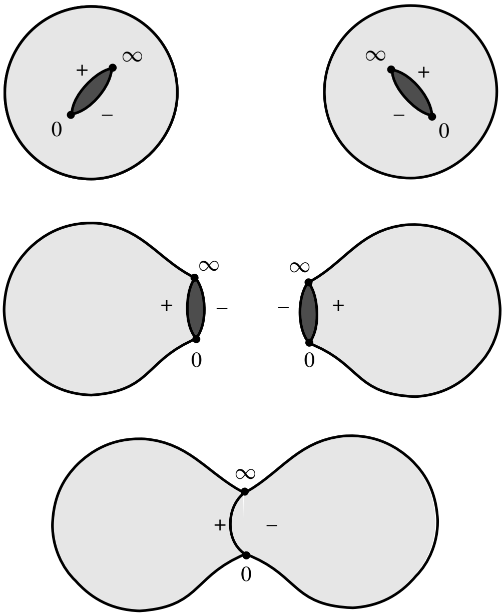

Figure 1.14. The Riemann surface for $\sqrt{z}$.

the more subtle (because less visible) complex structure of S is best treated from another standpoint. Now look at the function z^2 as a 2:1 map of one copy of $\mathbb{P}^1$ to another, with exceptions at 0 and ∞ where it is 1:1 (but still of degree 2 in that 0 is a double root and ∞ a double pole). The map is seen in Fig. 1.15. The great circle seen in the upper sphere represents the inverse stereographic projection of $\mathbb{R} \subset \mathbb{C}$, cutting $\mathbb{P}^1$ into two hemispheres, labeled $+$ and $-$, and z^2 opens up each of these into a full copy of $\mathbb{P}^1$. The inverse map is $\sqrt{z}$ whose Riemann surface S, seen in Fig. 1.14, may now be identified with the covering projective line of Fig. 1.15. The complex structure of S is clarified thereby: Plainly, it is compatible with that of the base except over 0 and ∞ where the branching of the radical takes place. There the cover is **ramified** over the base, its two sheets touching as in Fig. 1.16, or, more realistically, as in Fig. 1.17, in which you see that one revolution about 0 carries you from sheet 1 to sheet 2, and a second revolution (not shown) brings you back to sheet 1. Figures 1.16 and 1.17 hint that the complex structure of the cover goes bad at the ramifications, but *this is not so*: Both cover and base are complex manifolds in themselves; it is just that their complex structures are not the same: At 0, z is local parameter downstairs and $\sqrt{z}$ is local parameter upstairs; at ∞, you must use $1/z$ downstairs and $1/\sqrt{z}$ upstairs. Forster [1981] and Springer [1981] provide more details.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

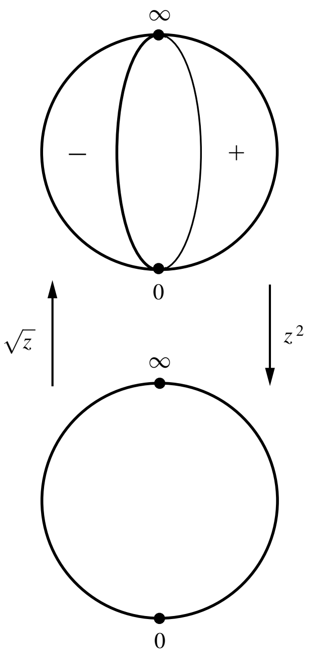

Figure 1.15. The maps $z \mapsto z^2$ and $z \mapsto \sqrt{z}$.

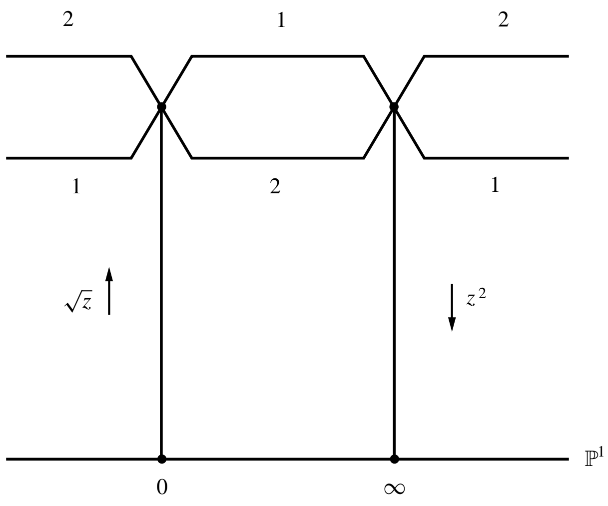

Figure 1.16. The two sheets touch at certain points.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

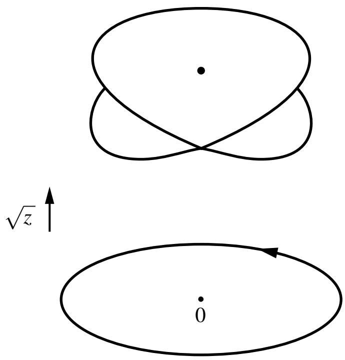

Figure 1.17. The two sheets.

Exercise 1. Give a parallel discussion of the Riemann surface for $\sqrt[d]{z}$ as a d-fold ramified cover of $\mathbb{P}^1$ for $d \geq 3$.

Handlebodies. The same idea applies to the radical $\sqrt{(z-e_1)\cdots(z-e_n)}$ with distinct branch points $e_i (i \leq n)$. For $n = 2$, nothing changes; it is only that e_1 and e_2 play the roles of 0 and ∞. For $n = 3$, the radical branches at e_1, e_2, e_3, and ∞ since a loop enclosing e_1, e_2, and e_3 produces $3(= 1)$ changes of sign. To make the radical single-valued on $\mathbb{P}^1$ requires two cuts, one from e_1 to e_2, say, and one from e_3 to ∞, as in Fig. 1.18. Two copies of this cut $\mathbb{P}^1$ are now pasted together in the manner of Fig. 1.14, but with the very different outcome seen in Fig. 1.18. The complex structure of the torus so produced may be clarified as in Fig. 1.16; for example, $\sqrt{z-e_1}$ is local parameter over e_1 and so on. For $n = 4$, you get the same figure with e_4 in place of ∞, but for $n = 5$ or 6 a pretzel appears, and for general $n = 2g + 1$ or $2g + 2$, a handlebody of genus (= handle number) g. The moral is that every handlebody appears as a Riemann surface; in particular, they all admit a complex structure. More complicated examples abound. The projective curve $y^3 = x - 1/x$ is pretty typical: y branches triply over $x = 0, \infty$, and over the roots ± 1 of $x^2 = 1$, as in Fig. 1.19; in detail

$$\begin{aligned} y &= x^{1/3}\left[1 + \text{ powers of } 1/x\right] && \text{at} \quad x = \infty \\ &= x^{-1/3}\left[-1 + \text{ powers of } x\right] && \text{at} \quad x = 0 \\ &= \sqrt[3]{x-e}\left[\sqrt[3]{2} + \text{ powers of } x - e\right] && \text{at} \quad x = e = \pm 1, \end{aligned}$$

so a little counterclockwise circuit about $x = \infty, 0, -1$, or $+1$ multiplies y by $e^{-2\pi\sqrt{-1}/3}$ for $x = \infty$ or 0 and by $e^{2\pi\sqrt{-1}/3}$ for $x = \pm 1$. Otherwise, y has

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

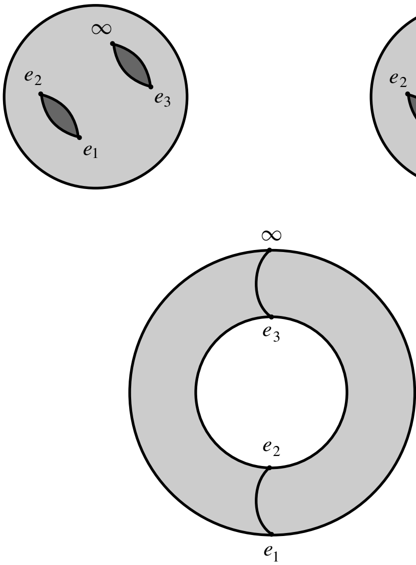

Figure 1.18. The Riemann surface for $\sqrt{(z - e_1)(z - e_2)(z - e_3)}$.

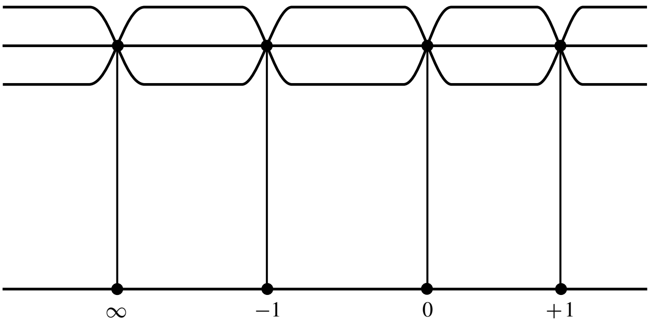

Figure 1.19. The projective curve $y^3 = x - 1/x$.

three distinct values over every value of $x \in \mathbb{P}^1$. Now take three copies of the projective line cut from ∞ to -1 and from 0 to 1 so as to account for the three different determinations of $y = \sqrt[3]{x - 1/x}$. These are seen in Fig. 1.20I. They must be pasted by the numbers. A preliminary pasting of 3 to 3 and 4 to 4 produces Fig. 1.20II; the final handlebody (III) is of genus 2.

Exercise 2. Check the numbering of Fig. 1.20. *Hint*: As you pass just below the real line from $-\infty$ to 1 and just above it to $-\infty$, y changes by the factor

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

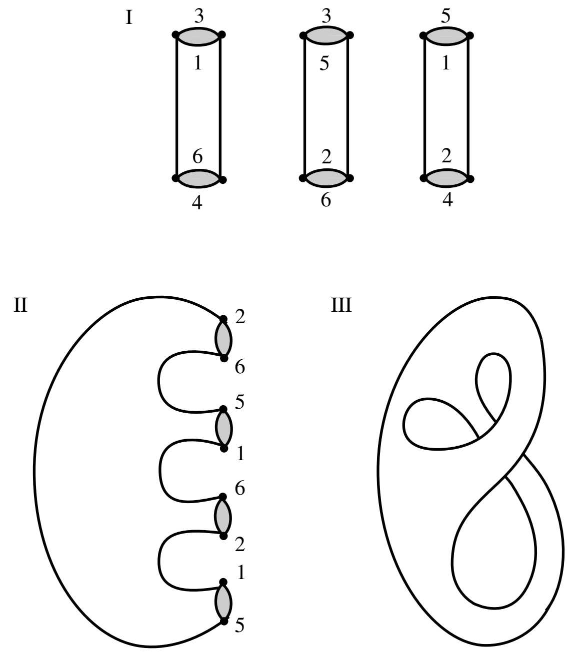

Figure 1.20. Cutting and pasting for $y^3 = x - 1/x$.

$\omega = e^{\pi\sqrt{-1}/3}$ at -1, ω^2 at 0, and ω^2 at 1, and again, on the return trip, by ω^2 at 0 and ω at -1, for a net change of $\omega^8 = \omega^2$.

Exercise 3. Check that the Riemann surface of $y^3 = x^2 - x^{-2}$ has genus 4.

The Helix. The Riemann surface of the logarithm should also be mentioned. The function $w = e^z$ maps each horizontal strip $\mathbb{R} + 2\pi\sqrt{-1} \times [n, n+1)$ of $\mathbb{C}$ faithfully onto the punctured plane $\mathbb{C} - 0$, as in Fig. 1.21, so the inverse function $z = \log w$, defined on the base $\mathbb{C} - 0$, has the covering plane $\mathbb{C}$ as its Riemann surface, each strip constituting a sheet of the latter. A counterclockwise circuit about the puncture downstairs raises the covering point to the next sheet up, by addition of $2\pi\sqrt{-1}$, suggesting that the cover is better viewed as the infinite helix, with $\log z$ itself filling the office of global parameter, opening up the punctured disk $0 < r < 1$ downstairs into the half-plane $(-\infty, 0) \times \sqrt{-1}\mathbb{R}$ upstairs.

Ramified Covers. The Riemann surface of $z^{1/3}$ provides an example of a **ramified cover**: It is a copy of $\mathbb{P}^1$ covering the projective line three-fold, except over 0 and ∞ which are covered once; compare Fig. 1.22. The anomalous covering

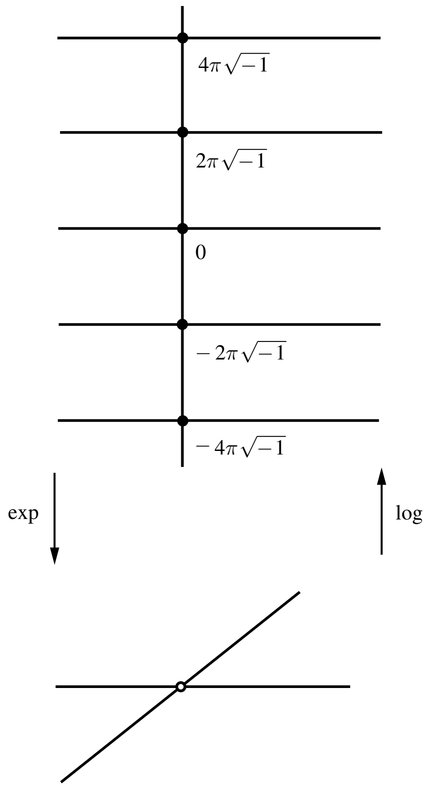

Figure 1.21. The exponential and logarithm.

points are the **ramifications**. More generally, a cover is ramified if $d \geq 2$ sheets meet at a point in the manner of the Riemann surface $z^{1/d}$, $|z| < 1$. The idea of a cover is now widened to admit such ramifications; it is even convenient to admit infinite ramifications as in the Riemann surface of the logarithm. The **index** of a ramification is 1 less than the degree, that is, 1 less than the number of adjacent sheets.

Riemann–Hurwitz Formula. Let $\mathcal{K}$ be a compact orientable manifold (= a handlebody) covering the projective line $\mathbb{P}^1$ with total ramification index r, d sheets, and g handles. Then $r = 2(d + g - 1)$. This is the **Riemann–Hurwitz formula**.

Example 1. The formula precludes unramified covers unless $d = 1$ and $g = 0$ ($\mathcal{K} = \mathbb{P}^1$).

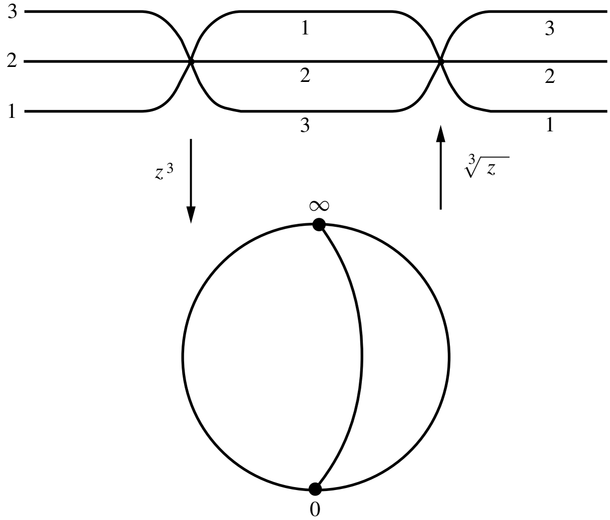

Figure 1.22. Ramified covers.

Example 2. The Riemann surface of $y^3 = x - x^{-1}$ seen in Fig. 1.19, viewed as a ramified cover of the projective line with parameter x, has 4 ramifications of degree 3 apiece for a total index of $4(3-1) = 8$, 3 sheets, and 2 handles: $8 = 2(3+2-1)$.

Proof of the Riemann–Hurwitz formula. The compactness of $\mathcal{K}$ implies that both the ramification index and the sheet number d are finite. Let the base be provided with a fine triangulation in which the projections of the ramified points of the cover appear as corners, and lift it up by the inverse projection to obtain a triangulation of $\mathcal{K}$. The alternating sum: corners $-$ edges $+$ faces is the Euler number: Upstairs its value is $2-2g$; downstairs its value is $2(g=0)$. Now the Euler number of the cover should be $2-2g = 2d$ because everything downstairs appears d-fold upstairs, *but not quite*: A base point that lifts to one or more ramified points of $\mathcal{K}$, of total index m, is covered only $d-m$ times, so there is an imbalance of $-\sum m = -r$ to the right, the true relation being $2-2g = 2d - r$. This is the formula.

Exercise 4. Extend the Riemann–Hurwitz formula to the case of a ramified cover of a general compact complex manifold.

1.13 Algebraic Functions

The idea of a Riemann surface applies to the projective curves of Section 10. Let $P(x, y)$ be an irreducible polynomial in y with coefficients from $\mathbb{C}[x]$. The roots y of $P(x, y) = 0$ are viewed temporarily as elements of the splitting field of P over the ground field $\mathbb{C}(x)$. They are simple, so their discriminant is a nonvanishing element of $\mathbb{C}(x)$; see Section 4. The totality of these roots is an **algebraic function y** of the **indeterminate x**. Riemann's idea provides a geometrical picture of this, as will now be explained in stages.

The Covering. Fix a point x_0 of the projective line $\mathbb{P}^1$ punctured at ∞ and at the points where either the discriminant or the top coefficient of $P(x, y) = c_0(x)y^d + c_1(x)y^{d-1} + \cdots$ vanishes. Near x_0, you have d distinct numerical roots $y_1, \ldots, y_d$, each of which is capable of being expanded in powers of $z = x - x_0$: $\mathbf{y} = k_0 + k_1 z + \cdots$. The resulting pairs $\mathfrak{p} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = x_0$ and $\mathbf{y} = y_1, \ldots, y_d$ are displayed in Fig. 1.23 stacked up over x_0. The latter is the

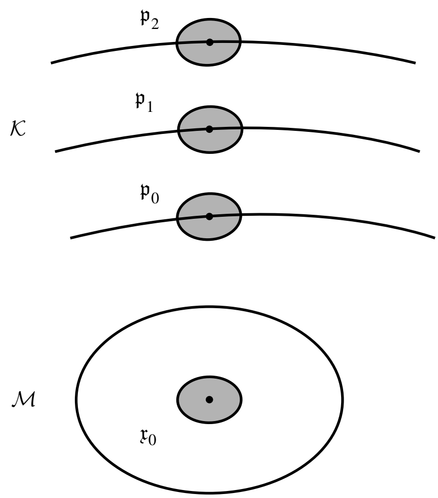

Figure 1.23. Function elements.

base point; **y** is a **function element**; the map $\mathfrak{p}_0 \mapsto x_0$ is the **projection**. Now any such function element $\mathbf{y}_0$ can be reexpanded about any center **x** close to x_0. This produces a patch of points $\mathfrak{p} = (\mathbf{x}, \mathbf{y})$ about $\mathfrak{p}_0 = (\mathbf{x}_0, \mathbf{y}_0)$, and as this patch is faithfully mirrored by the projection $x = \mathbf{x}(\mathfrak{p})$, so it may be equipped with a complex structure by use of the **local parameter** $z(\mathfrak{p}) = \mathbf{x}(\mathfrak{p}) - x_0$. This

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

makes the totality of points $\mathfrak{p}$ into a complex manifold $\mathcal{K}$ covering the punctured base $\mathbb{P}^1$.

Connectivity. The possible disconnectivity of the cover is easily disproved. Any function element $\mathbf{y}_0$ can be continued without obstruction over the punctured base, producing a subfamily $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of the function elements over each base point $\mathbf{x}$, of fixed cardinality $n \le d$. The coefficients of the polynomial $P_1(\mathbf{y}) = (\mathbf{y} - \mathbf{y}_1)\cdots(\mathbf{y} - \mathbf{y}_n)$ are single-valued functions of $\mathbf{x}$. These must be *rational* since any function element obeys an estimate $|\mathbf{y}| \le C_1 \times |\mathbf{x} - \mathbf{x}_0|^{-m}$ at a finite puncture or $|\mathbf{y}| \le C_2 \times |\mathbf{x}|^m$ at ∞. It follows that $P_1(\mathbf{y})$ divides $P(\mathbf{x}, \mathbf{y})$ over $\mathbb{C}(\mathbf{x})$, violating the irreducibility of the latter unless $n = d$; in short, continuation of $\mathbf{y}_0$ produces the full family of function elements over every base point, which is to say that $\mathcal{K}$ is connected.

The Punctures Filled In. This is a pretty application of the monodromy theorem of Section 11. Fix a puncture $\mathbf{x} = 0$, say, and let $0 < |\mathbf{x}| < r_0$ be puncture-free. The left half-plane $\mathcal{H}$ of the Riemann surface of $\log \mathbf{x}$ covers this punctured disk, and any function element $\mathbf{y}$ over a point of the latter can be lifted to the former and continued there without obstruction. This produces a single-valued function of the logarithm because $\mathcal{H}$ is *simply connected*. Now each of the points $\log \mathbf{x} + 2\pi\sqrt{-1}\mathbb{Z}$ covers x and each of the associated function elements is a root of $P(x, y) = 0$. It follows that only a finite number of different function elements appear at $\log \mathbf{x} + 2\pi\sqrt{-1}\mathbb{Z}$ and that the original branch repeats itself after a continuation upward by $2\pi\sqrt{-1}n$ units with minimal $n \le d$ independent of $\mathbf{x}$, the intervening branches being distinct. This means that the continued function element $\mathbf{y}$ may be viewed as a single-valued function of $\mathbf{x}^{1/n}$; it is even of rational character in this parameter at $\mathbf{x} = 0$ in view of the estimate $|\mathbf{y}| \le C_1 \times |\mathbf{x}|^{-m}$ used before. The upshot is that the totality of function elements so produced can be obtained from a single fractional expansion $\mathbf{y}_0 = c_{-k}\mathbf{x}^{-k/n} + \cdots + c_0 + c_1\mathbf{x}^{1/n} + \cdots$ by reexpansion about centers $0 < |\mathbf{x}| < r_0$. $\mathcal{K}$ is now completed at punctures by the insertion of points of this new type $\mathfrak{p}_0 = (0, \mathbf{y}_0)$, and the complex structure is extended by attributing the local parameter $z(\mathfrak{p}) = [\mathbf{x}(\mathfrak{p})]^{1/n}$ to an ambient patch. The new points have the (ramified) aspect of Fig. 1.24, and the completed surface $\mathcal{K}$ is a compact *ramified* cover of the unpunctured base $\mathbb{P}^1$ with a full complex structure: It is *the Riemann surface of the algebraic function* $\mathbf{y}$. Bliss [1933] and Springer [1981] present more details and additional information. Weyl [1955] and Narasimhan [1992] are recommended for a more sophisticated view. $\mathcal{K}$ is a **nonsingular** model of the projective curve defined by $P(x, y) = 0$; compare amplification 2 in Section 10. The present desingularization is an overkill as it

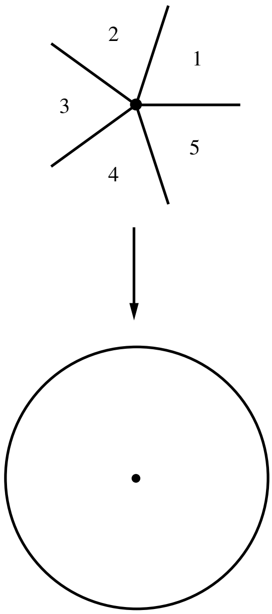

Figure 1.24. The ramified cover.

is realized only in the infinite-dimensional space of points $\mathfrak{p} = (\mathbf{x}, \mathbf{y})$. Actually, a nonsingular model of any projective curve can always be accommodated in $\mathbb{P}^3$; see Shafarevich [1977].

Uniformization. The universal cover $\mathcal{K}$ of a complex manifold can be equipped with the natural complex structure lifted up from the base; it is also simply connected. What are the possibilities? The answer is contained in the celebrated theorem of Klein [1882], Poincaré [1907], and Koebe [1909–14]: Up to conformal equivalence, $\mathcal{K}$ *is either the sphere* $\mathbb{P}^1$, *the plane* $\mathbb{C}$, *or else the disk* $\mathbb{D}$: $|x| < 1$. Note that $\mathbb{P}^1$ is inequivalent to $\mathbb{C}$ or $\mathbb{D}$ on topological grounds already and that $\mathbb{C}$ is inequivalent to $\mathbb{D}$ because an analytic map of $\mathbb{C}$ to $\mathbb{D}$ has a one-point image, by Liouville's theorem: In short, sphere, plane, and disk are genuinely different. The statement includes the earlier Riemann mapping theorem [1851]: *A simply connected region of the sphere omitting two or more points is conformally equivalent to a disk.* It is important to understand what is involved in the general statement, though the proof is not so simple and we refer you to Ahlfors [1973], Springer [1981], or Weyl [1955] for full details. Poincaré himself did not find a fully successful proof. The following informal discussion may be helpful.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Some Hydrodynamics. A pretty hydrodynamical picture was introduced by Klein in his Cambridge lectures [1893]. Let U be a patch of $\mathcal{K}$ equipped with a local parameter $z(\mathfrak{p}) = x_1(\mathfrak{p}) + \sqrt{-1}x_2(\mathfrak{p})$. A steady irrotational flow of an incompressible fluid is specified by a velocity field $\mathbf{v}(\mathfrak{p}) = (v_1, v_2)$ subject to (1) $\mathrm{div}(\mathbf{v}) = \partial v_1/\partial x_1 + \partial v_2/\partial x_2 = 0$ for the incompressibility and (2) $\mathrm{curl}(\mathbf{v}) = \partial v_2/\partial x_1 - \partial v_1/\partial x_2 = 0$ for the irrotational character. (2) implies the existence of a **potential function** p, producing the velocity field $\mathbf{v} = \mathrm{grad} p$, subject to (3) $\Delta p = 0$ in place of (1), and conversely, if $\Delta p = 0$ in U, then $\mathbf{v} = \mathrm{grad}\ p$ is the velocity field of a steady irrotational flow of an incompressible fluid: $\dot{\mathbf{x}}(\mathfrak{p}) = \mathbf{v}(\mathfrak{p})$ with $\mathbf{x}(\mathfrak{p}) = (x_1(\mathfrak{p}), x_2(\mathfrak{p}))$. Let $z'(\mathfrak{p}) = x_1'(\mathfrak{p}) + \sqrt{-1}x_2'(\mathfrak{p})$ be a new local parameter. The new velocity field is also incompressible and irrotational, and the new flow appears in the old coordinates as $\dot{\mathbf{x}} = JJ^\dagger \partial p/\partial \mathbf{x} = c\mathbf{v}$ with Jacobian $J = \partial \mathbf{x}/\partial \mathbf{x}'$, its transpose $J^\dagger$, and the positive factor $c = |\det J|^2 = |dz/dz'|^2$, as you will check by means of the Cauchy–Riemann equations $\partial x_1/\partial x_1' = \partial x_2/\partial x_2'$, $\partial x_1/\partial x_2' = -\partial x_2/\partial x_1'$. The new streamlines are the same as the old; it is only the speed that is changed. This ambiguity of speeds is unavoidable: $\mathcal{K}$ has only a conformal structure, so there is no *preferred* local parameter on any patch.

Exercise 1. Check the new flow.

A Global Flow. $\mathcal{K}$ is equipped with the streamlines of a steady incompressible irrotational flow produced by a single *source* at a point $\mathfrak{o}$, with the understanding that if $\mathcal{K}$ is compact, it will be necessary to destroy fluid at a complementary *sink* $\mathfrak{o}'$. The source is modeled by the potential function $p = \log|z(\mathfrak{p})|$ with local parameter $z(\mathfrak{p})$ vanishing at $\mathfrak{p} = \mathfrak{o}$, but care is needed: If the local parameter is poorly chosen, it may not be possible to extend the flow patchwise over the whole of $\mathcal{K}$ without unpleasant singularities besides the necessary sink in the compact case. What is needed, and this is the hard part of the proof, is the existence of a global function p with $\Delta p = 0$ at ordinary points of $\mathcal{K}$, the singularity $\log|z(\mathfrak{p})|$ at the source, and in the compact case, a second singularity $-\log|z'(\mathfrak{p})|$ at the sink, $z'(p)$ being a local parameter there. This defines a flow over the whole of $\mathcal{K}$ with velocities $\mathbf{v} = (v_1, v_2) = \mathrm{grad}\ p$. The conjugate flow with velocities $(v_2, -v_1)$ is at right angles; it has a (local) potential function q determined up to an additive constant. The associated **circulation** $= \oint dq$ taken about any of the (closed) level lines of p is independent of the particular level line, by Stokes's theorem, and may be evaluated as 2π by shrinking the level line to $\mathfrak{o}$. The new parameter $z(\mathfrak{p}) = \exp\left[p(\mathfrak{p}) + \sqrt{-1}q(\mathfrak{p})\right]$ is now seen to be a single-valued function of rational character on $\mathcal{K}$ with a simple root at the source and, in the compact case, a simple pole at the sink. This function

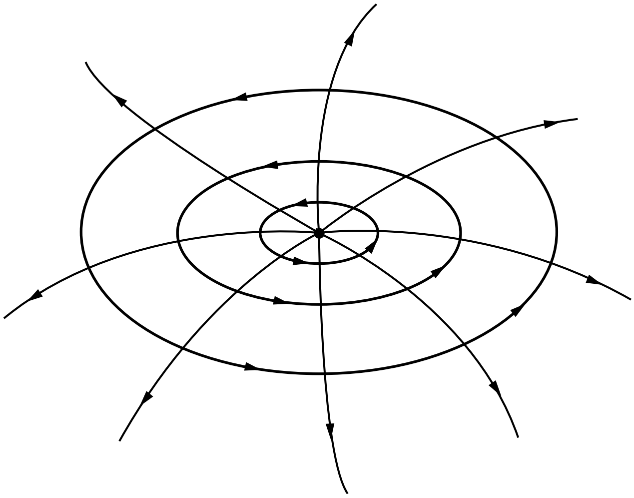

Figure 1.25. Global parameters.

is 1:1 as is plain from the picture (Fig. 1.25) but not so easy to prove: The streamlines issuing from $\mathfrak{o}$ cover the punctured surface simply, except that they come together again at the sink in the compact case; p serves as coordinate *along* the streamline and q tells *which* streamline it is. Three cases are now distinguished according as $\mathcal{K}$ is compact or not and, in the noncompact case, according as p is bounded or not.

Case 1. $\mathcal{K}$ is compact. Then it is a topological sphere and $z{:}\,\mathcal{K} \to \mathbb{P}^1$ is a conformal equivalence between $\mathcal{K}$ and the projective line.

Case 2. $\mathcal{K}$ is noncompact and p is unbounded. Then $z{:}\,\mathcal{K} \to \mathbb{C}$ is a conformal equivalence between $\mathcal{K}$ and the whole complex plane.

Case 3. $\mathcal{K}$ is noncompact and p tends to the finite number $p(\infty)$. Then z maps $\mathcal{K}$ 1:1 onto a disk of radius $r = \exp[p(\infty)]$. This is Riemann's case.

Idea of the proof. Springer [1981] explains the actual construction of p: Let $\wedge^1$ be the class of smooth 1-forms $\omega = \omega_1 dx_1 + \omega_2 dx_2$ on $\mathcal{K}$ and let $*\omega = -\omega_2 dx_1 + \omega_1 dx_2$. Weyl [1955] proved that $\wedge^1$ splits into three pieces, mutually perpendicular relative to the natural quadratic form $\int \omega \wedge^* \bar{\omega}$,[2] of which the first is differentials of smooth functions ($\omega = df = (\partial f/\partial x_1)dx_1 + (\partial f/\partial x_2)dx_2$),

[2] $\bar{\omega}$ is the complex conjugate of ω.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

the second is codifferentials ($\omega = *df$), and the third is the **harmonic differentials** which are simultaneously **closed** ($d\omega = (\partial\omega_2/\partial x_1 - \partial\omega_1/\partial x_2)dx_1 \wedge dx_2 = 0$), like differentials of functions, and also **coclosed** ($d * \omega$), like codifferentials. Now place yourself in the noncompact case and cook up a smooth exact differential ω that imitates $d\log z(\mathfrak{p})$ near $\mathfrak{p} = \mathfrak{o}$ and vanishes at a little distance from it. Then $\omega - \sqrt{-1} * \bar{\omega}$ vanishes near 0 and ∞ and, being smooth, may be split into its three parts: $df_1 + *df_2 + \omega_3$. The new differential $\omega - df_1 = \sqrt{-1} * \omega + *df_2 + \omega_3$ is plainly smooth, exact, and coclosed away from $\mathfrak{o}$, and its real part is the differential of a single-valued harmonic function p on $\mathcal{K} - \mathfrak{o}$ having the required singularity $\log|z(\mathfrak{p})|$ at $\mathfrak{p} = \mathfrak{o}$. The simple connectivity of $\mathcal{K}$ is now used to confirm that $dq = *dp$ is also exact modulo 2π so $z(\mathfrak{p}) = \exp\left[p(\mathfrak{p}) + \sqrt{-1}q(\mathfrak{p})\right]$ is single-valued too. The final point is that $z: \mathcal{K} \to \mathbb{C}$ is 1:1. This is more subtle, so we stop here.

Exercise 2. The map of $\mathcal{K}$ to a sphere, plane, or disk (= half-plane) may be standardized by fixing its values at any three points as you will. Why?

1.14 Examples

The deep content of the Koebe–Poincaré theorem is plain from elementary examples.

Spheres. This is already instructive. The statement is that a complex structure on a topological sphere can be described by a *global parameter* $z: \mathcal{K} \to \mathbb{P}^1$. This is what is meant by saying that *the sphere has just one complex structure*.

Annuli. The sphere is peculiar in this respect; for instance, two annuli equipped with the natural complex structure they inherit from $\mathbb{C}$ are conformally equivalent if and only if they have the same ratio r of inner to outer radii. In short, the inequivalent complex structures of a topological annulus are in faithful correspondence with the numbers $0 < r < \infty$.

Exercise 1. Why? *Hint*: A map of annuli extends by circular reflection to a map of punctured planes.

Punctured Spheres. The once-punctured sphere is the plane, so the twice-punctured sphere is a punctured plane, and its universal cover $\mathcal{K}$ may be viewed as the Riemann surface of the logarithm; in short, $\mathcal{K}$ is a plane. The thrice-punctured sphere (= the doubly punctured plane) is different: *Its universal cover is the disk.*

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Proof. The cover is not compact (why?), so only $\mathcal{K} = \mathbb{C}$ needs to be ruled out. But if $\mathcal{K} = \mathbb{C}$, then the covering group is populated by substitutions of $PSL(2, \mathbb{C})$ fixing ∞; see Section 5. These are of the form $z \mapsto az + b$ and have fixed points if $a \neq 1$, whereas covering maps do not; see ex. 11.7. It follows that the covering group is commutative, contradicting the fact that the fundamental group of the twice-punctured plane is not. A very different proof will be found in Section 4.9.

Picard's Little Theorem [1879]. A very pretty bonus is a proof of the fact that a *nonconstant integral function takes on every complex value with at most one exception*. The exception is real: *the exponential does not vanish.*

Proof. Let the integral function f omit the values a and b. Then $(b-a)^{-1}(f-a)$ omits 0 and 1, so it is permissible to take $a = 0$ and $b = 1$. Now use f to map a small disk into the punctured plane $\mathbb{C} - 0 - 1$, lift the map to the (universal) covering half-plane via any branch of the inverse projection, and map the lift into the unit disk. The composite map $\mathbb{C} \to \mathbb{C} - 0 - 1 \to \mathbb{H} \to \mathbb{D}$ may be continued without obstruction along paths of the plane. This produces a single-valued function in that plane by the monodromy theorem of Section 11, and as its values are confined to the disk, so it must be constant. But the projection is not constant nor is the map from the half-plane to the disk. The only way out is for f to be constant.

Ahlfors [1973: 19–21] presents a beautiful elementary proof, not employing such transcendental aids; see also Nevanlinna [1970: 248–9] for a thorough geometric discussion and Section 4.9 for a reprise.

Hyperbolic Geometry. Another amusing consequence is that the thrice-punctured sphere admits a geometry of constant curvature -1. The cusps of this **horned sphere** are modeled in ex. 9.5; see Fig. 1.26. The point is that the covering group of the punctured sphere is realized by substitutions of $PSL(2, \mathbb{R})$ acting upon the covering half-plane, and as these are rigid motions of the cover, so this geometry drops down to the punctured sphere; see Section 8 and Pogorelov [1967: 166–7] for such matters. It is not possible to do this for the unpunctured sphere. The obstacle is expressed by the Gauss–Bonnet formula which states that if a handlebody $\mathbf{X}$ of genus g is equipped with a geometry with (possibly variable) curvature κ and surface element $d\sigma$, then the value of the **curvatura integra** $\int_{\mathbf{X}} \kappa \, d\sigma$ is 2π times the Euler number $2 - 2g$ of $\mathbf{X}$; for example, it is $+4\pi$ for the sphere. Pogorelov [1967: 164–6] explains this well.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

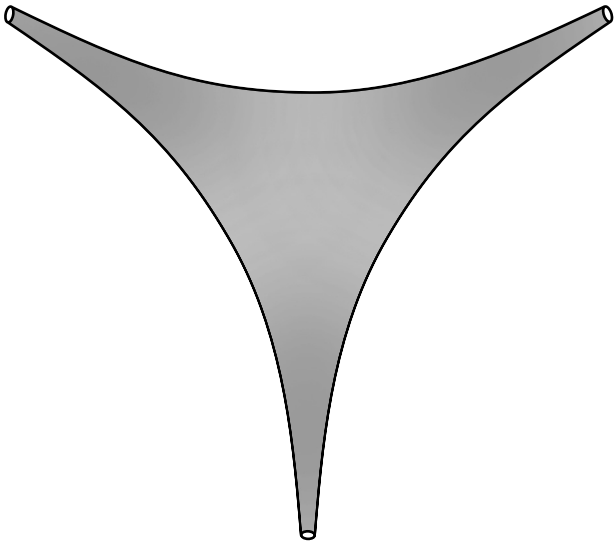

Figure 1.26. The horned sphere.

Higher Handlebodies. The idea behind the horned sphere applies, as well, to higher handlebodies of genus 2 or more: *Their universal covers are always half-planes* so the covering maps appear as substitutions of $PSL(2, \mathbb{R})$ and the geometry of curvature -1 drops down. Note, also, that the covering group must be noncommutative.

Exercise 2. Check this. *Hint*: Two handles produce a nontrivial commutator $aba^{-1}b^{-1}$ in the fundamental group; see Fig. 1.27.

Poincaré [1898] had the attractive idea of producing the universal cover of a higher handlebody by endowing the body with a geometry of curvature -1, lifting this geometry to the cover, and identifying the latter with the hyperbolic half-plane; see Kazdan [1985] for such matters.

Tori. These are different; for example, the curvatura integra vanishes, so a geometry of constant curvature ± 1 is not possible. Let ω be a complex number of positive imaginary part and let $\mathbb{L}$ be the lattice $\mathbb{Z} \oplus \omega\mathbb{Z}$. The torus $\mathbf{X} = \mathbb{C}/\mathbb{L}$ inherits the complex structure of $\mathbb{C}$ and has this plane as its universal cover. The latter divides naturally into cells as in Fig. 1.28. The **fundamental cell** is shaded. The projection $\mathbb{C} \to \mathbf{X}$ is *identification modulo* $\mathbb{L}$ and the covering

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

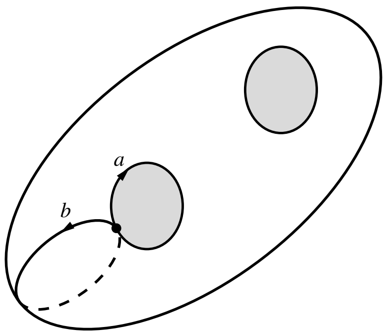

Figure 1.27. Handles produce noncommutativity.

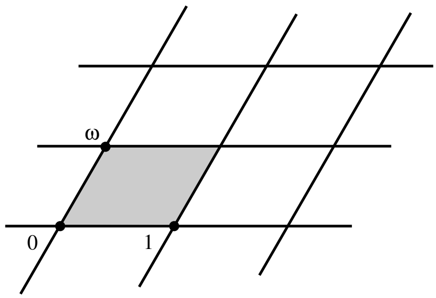

Figure 1.28. The lattice.

group is a copy of the fundamental group $\mathbb{Z}^2 = \mathbb{Z} \oplus \mathbb{Z}$, as it had to be. The fact is that, up to conformal equivalence, every complex torus arises in this way.

Proof. The universal cover of the torus is the plane; intuitively, it is just the torus rolled out vertically and horizontally. Now the covering maps are fixed-point-free conformal self-maps of $\mathbb{C}$, that is, translations $z \mapsto z + c$, as noted under the heading "punctured spheres." The numbers c form a sublattice $\mathbb{L}$ of $\mathbb{C}$, isomorphic to $\mathbb{Z}^2$ and so of the form $\omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ with noncollinear ω_1 and ω_2, as will be verified in Section 2.6. $\mathbf{X}$ is now identified with the quotient $\mathbb{C}/\mathbb{L}$. A trivial map $z \mapsto z/\omega_1$ and, if need be, a change of sign of the ratio $\omega = \omega_2/\omega_1$ brings $\mathbb{L}$ to the standard form $\mathbb{Z} \oplus \omega\mathbb{Z}$ with ω of positive imaginary part.

The identification $\mathbf{X} = \mathbb{C}/\mathbb{L}$ leads to a simple proof of the fact that, unlike the unpunctured sphere, but like the annulus of ex. 1, *the topological torus has many inequivalent conformal structures*. The same is true of the higher handlebodies.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

Proof. Let $\mathbf{X}_1 = \mathbb{C}/\mathbb{L}_1$ and $\mathbf{X}_2 = \mathbb{C}/\mathbb{L}_2$ be complex tori with a conformal map between them and lift this map up to a self-map of their common universal cover $\mathbb{C}$; see Fig. 1.29. The lifted map is single-valued by the monodromy theorem, the continuation being unobstructed; moreover, it is of linear growth at ∞ since a closed path that goes around or through the hole of $\mathbf{X}_1$ leads from a point in one cell of the cover to a new point in an *adjacent* cell. But then the lifted map is a linear function $az + b$ of the parameter z of $\mathcal{K}_1 = \mathbb{C}$, and as it commutes with projections, so you must have $b \in \mathbb{L}_2$ and $a\mathbb{L}_1 \subset \mathbb{L}_2$; indeed,

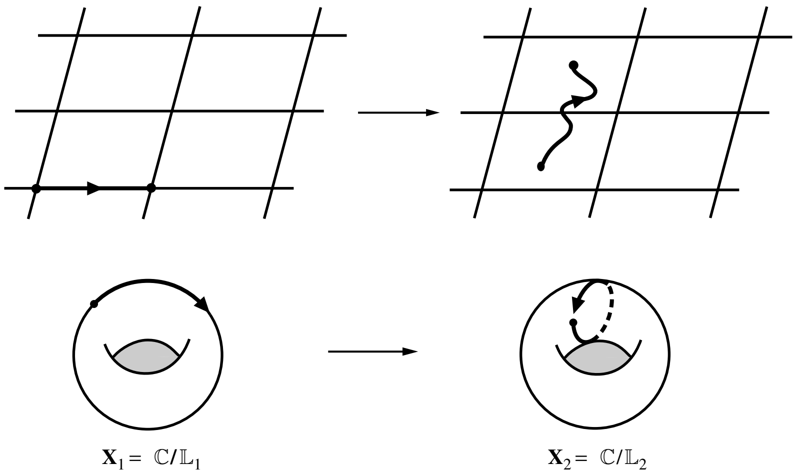

Figure 1.29. Conformal maps between tori.

$a\mathbb{L}_1 = \mathbb{L}_2$ by consideration of the inverse map. The moral is that the tori cannot be conformally equivalent otherwise. For example, the period ratios $\omega = \sqrt{-1}$ and $\omega = (1 + \sqrt{-3})/2$ produce inequivalent tori.

Geometric Explanation. The fact that tori admit many different complex structures is the subject of Section 2.6 and of the first part of Chapter 4. A geometric explanation can be given right now. Let the fundamental cell be

$$\mathfrak{F} = \{x = x_1 + sx_2 + \sqrt{-1}x_2 : 0 \leq x_1 < 1, 0 \leq x_2 < h\}$$

with fixed base [0, 1). This standardization does not affect the number of complex structures. There are two degrees of freedom left: the **slant** s and the **height** $h > 0$. The corresponding torus is made in two steps from the rectangular cell of height $h(s = 0)$: First paste together the vertical sides to make a right cylinder; then twist the upper circle by sh and paste it to the bottom.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

This produces a shear in the amount sx_2 at level $0 \leq x_2 < h$ and changes the complex structure; a change of height usually changes the structure, too.

Example. Let $h = 1$ and let the two slants s_1 and s_2 produce equivalent tori. The corresponding lattices $\mathbb{L} = \oplus(s + \sqrt{-1})\mathbb{Z}$ are related by a multiplication $a\mathbb{L}_1 = \mathbb{L}_2$, as you know, and $|a| = 1$ by comparison of areas of fundamental cells. But then $a \times 1 = i + j(s_2 + \sqrt{-1})$ with $i, j \in \mathbb{Z}$ and $1 = (i + js_2)^2 + j^2$, so either $j = 0$, $a = i = \pm 1$, and $\mathbb{L}_{\nvdash} = \mathbb{L}_{\neq}$, or else $j = \pm 1$, $s_2 = \mp i$, and $a = \pm\sqrt{-1}$. Now note that s_2 may be reduced modulo the period 1 so as to lie in [0,1), the moral being that, with this adjustment, the conformal structure determines the slant.

Exercise 3. Does the conformal structure determines the height if the slant is 0? If not, how far does it go?

1.15 More on Uniformization

The statement of Koebe–Poincaré is called the **uniformization theorem** in the older literature. The name refers to the nineteenth-century usage uniform = one-valued. To clarify matters let $\mathbf{X}$ be a projective curve defined by the vanishing of an irreducible polynomial: $P(x, y) = 0$. Its points are triples $(x, y, z) \in \mathbb{C}^3 - 0$ with projective identifications, so $\mathbf{x} = x/z$ and $\mathbf{y} = y/z$ make projective sense; in fact, they are functions of rational character on $\mathbf{X}$, and the vanishing of P expresses a relation between them, specifying $\mathbf{y}$ as a many-valued function of $\mathbf{x}$ and vice versa. These functions may be promoted to the universal cover $\mathcal{K}$ of $\mathbf{X}$ where they appear as functions $\mathbf{x}, \mathbf{y}$ of rational character, invariant under the action of the covering group. In this way, $\mathbf{X}$ is uniformized by $\mathcal{K}$ in that the totality of points $\mathbf{X}$ is displayed by means of single-valued functions on $\mathcal{K}$(= sphere, plane, or half-plane).

Genus 0. If $\mathbf{X}$ has no handles, then it is a projective line $\mathbb{P}^1$ and is its own universal cover, so $\mathbf{x}$ and $\mathbf{y}$ appear as rational functions of the parameter w of $\mathbb{P}^1$ and the map $w \mapsto (\mathbf{x}, \mathbf{y})$ is a conformal equivalence between $\mathbb{P}^1$ and $\mathbf{X}$. The simplest nontrivial example is provided by the projective circle $\mathbf{X}$: $\mathbf{x}^2 + \mathbf{y}^2 = 1$ of Section 10 and its uniformization by

$$\mathbf{x} = \frac{1}{2}(w + w^{-1}), \quad \mathbf{y} = \frac{1}{2\sqrt{-1}}(w - w^{-1}).$$

Rational Curves. Here is a deeper fact: If $\mathbf{X}$ is a **rational curve** in the sense that there exists a nonconstant map of rational character of $\mathbb{P}^1$ into $\mathbf{X}$, then

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press

X *is of genus* 0 *already and so itself a projective line*, as before; compare Section 14. The idea will be plain from a picture: $\mathbb{P}^1$ is simply connected, so a map of rational character of $\mathbb{P}^1$ into **X** lifts to a map of $\mathbb{P}^1$ into $\mathcal{K}$, by a self-evident application of the monodromy theorem. This is impossible if $\mathcal{K}$ is not a projective line since nonconstant rational functions take all complex values, ∞ included.

Tori. The complex torus **X** cannot be uniformized by **rational functions** since its universal cover $\mathcal{K}$ is not a projective line; see ex. 2.11.3. Indeed, $\mathcal{K} = \mathbb{C}$, **X** is its quotient by the lattice $\mathbb{L}$ as in Section 14, and $\mathbf{x} = x/z$ and $\mathbf{y} = y/z$ may be promoted to functions of rational character on $\mathbb{C}$ having every complex number $\omega \in \mathbb{L}$ as a period. These functions uniformize **X** in the former style, but are not rational. The whole of Chapter 2 is devoted to such **elliptic functions**.

Higher Handlebodies. $\mathcal{K}$ is a half-plane by Section 14. Now the uniformizing functions **x** and **y**, promoted to $\mathcal{K}$, are invariant under a subgroup of $PSL(2, \mathbb{R})$: the so-called **automorphic functions**. The subject lies mostly outside the scope of this book, but see Ford [1972] and Terras [1985] for more information, and Chapters 4 and 5 for a number of special instances and some general information.

1.16 Compact Manifolds as Curves: Finale

The methods sketched in Section 14 can be used to prove the existence of nonconstant functions of rational character on any compact complex manifold $\mathcal{M}$; see, for example, Hurwitz and Courant [1964], Springer [1981], or Weyl [1955]. Let **x** be such a function: It has a finite number of poles and an equal number of roots, as may be seen by integrating $(2\pi\sqrt{-1})^{-1} d \log[\mathbf{x}(\mathfrak{p})]$ about the edges of a triangulation of $\mathcal{M}$, every edge being traversed twice, in opposite directions. This number is its degree d. It follows that **x** takes on every complex value d times, so $\mathcal{M}$ appears as a d-fold ramified covering of $\mathbb{P}^1$ with projection $\mathfrak{p} \mapsto \mathbf{x}(\mathfrak{p})$. Now two cases arise according as $d = 1$ or $d \geq 2$.

Case 1. If $d = 1$, then the projection is a conformal equivalence of $\mathcal{M}$ to $\mathbb{P}^1$; indeed, **x** fills the office of global parameter on $\mathcal{M}$.

Case 2. If $d \geq 2$, a further construction is necessary to produce a function **y** of rational character taking d distinct numerical values $\mathbf{y} = \mathbf{y}_1, \ldots, \mathbf{y}_d$ at the d points of the fiber $x = \mathbf{x}(\mathfrak{p})$, for most values of the base point $x \in \mathbb{P}^1$. Then it is easy to see that $(\mathbf{y} - \mathbf{y}_0) \cdots (\mathbf{y} - \mathbf{y}_d)$ is of rational character in **x** and so represents a polynomial $P \in \mathbb{C}(x)[y]$ satisfied by $\mathbf{x}$ and $\mathbf{y}$; moreover, P is necessarily irreducible over the ground field $\mathbb{C}(x)$, as you will see by continuation of the identity $P(\mathbf{x}, \mathbf{y}) = 0$ over the necessarily connected manifold $\mathcal{M}$. Now comes the punch line: The vanishing of $P(x, y)$ defines a nonsingular projective curve $\mathbf{X}$, as in Section 10, and the map $\mathfrak{p} \mapsto (\mathbf{x}, \mathbf{y})$ of $\mathcal{M}$ to $\mathbf{X}$ is a conformal equivalence; in short, *every compact complex manifold is a projective curve*. The discussion has come full circle.

https://doi.org/10.1017/9781009602129.003 Published online by Cambridge University Press