Chapter 1

From i to z: the basics of complex analysis

1.1. The field of complex numbers

The field $\mathbb{C}$ of complex numbers is obtained by adjoining i to the field $\mathbb{R}$ of reals. The defining property of i is $i^2 + 1 = 0$ and complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are added componentwise and multiplied according to the rule

$$z_1 \cdot z_2 = x_1x_2 - y_1y_2 + i(x_1y_2 + x_2y_1),$$

which follows from $i^2 + 1 = 0$ and the distributional law. The *complex conjugate* of $z = x + iy$ is $\bar{z} = x - iy$ and we have $|z|^2 := z\bar{z} = x^2 + y^2$. Therefore, every $z \neq 0$ has a multiplicative inverse given by $\frac{1}{z} := \bar{z}|z|^{-2}$ and $\mathbb{C}$ becomes a field. Since complex numbers z can be represented as points or vectors in $\mathbb{R}^2$ in the Cartesian way, we can also assign polar coordinates (r, θ) to them. By definition, $r = |z|$ and $z = r(\cos\theta + i\sin\theta)$. The addition theorems for cosine and sine imply that

$$z_1 \cdot z_2 = |z_1||z_2|(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)),$$

which reveals the remarkable fact that complex numbers are multiplied by *multiplying their lengths and adding their angles.* In particular, $|z_1z_2| = |z_1||z_2|$. This shows that power series behave as in the real case with respect to convergence, i.e.,

$$\sum_{n=0}^{\infty} a_n z^n \quad \text{converges on } |z| < R \text{ and diverges for every } |z| > R,$$

$$\text{where } R^{-1} = \limsup_{n\to\infty} |a_n|^{\frac{1}{n}},$$

where the sense of convergence is relative to the length metric $|\cdot|$ on $\mathbb{C}$ which is the same as the Euclidean distance on $\mathbb{R}^2$ (the reader should verify the triangle inequality). The formula for R of course follows from comparison with the geometric series, and $R = 0$ and $R = \infty$ are allowed. Note that the convergence is absolute on the disk $|z| < R$ and uniform on every compact subset of that disk. Moreover, the series diverges for *every* $|z| > R$ as can be seen by the comparison test. We can also write $R = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right|$, provided this limit exists. The first example that comes to mind here is

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n, \qquad |z| < 1.$$

Another example is of course

$$E(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}, \tag{1.1}$$

which converges absolutely and uniformly on every compact subset of $\mathbb{C}$. Expanding $(z_1 + z_2)^n$ via the binomial theorem shows that $E(z_1 + z_2) = E(z_1)E(z_2)$. In particular, we see that $E(z) \neq 0$ for any $z \in \mathbb{C}$. Indeed, if $E(z_0) = 0$, then $1 = E(0) = E(z_0)E(-z_0) = 0$, a contradiction.

Next, recall the definition of the Euler constant e: consider the ordinary differential equation $\dot{y} = y$ with $y(0) = 1$ which has a unique solution $y(t)$ for all $t \in \mathbb{R}$. Then set $e := y(1)$. Let us solve our differential equation iteratively (this is an example of the general Picard method). Indeed, let us assume that there is a C^1 solution up to some time $t > 0$. Applying the fundamental theorem of calculus then yields

$$y(t) = 1 + \int_0^t y(s)\, ds.$$

Now replace $y(s)$ in the integral by the right-hand side and iterate:

$$\begin{aligned} y(t) &= 1 + \int_0^t y(s)\, ds = 1 + t + \int_0^t (t-s) y(s)\, ds = \cdots \\ &= \sum_{j=0}^{n} \frac{t^j}{j!} + \frac{1}{n!} \int_0^t (t-s)^n y(s)\, ds. \end{aligned}$$

We bound

$$\frac{1}{n!} \left| \int_0^t (t-s)^n y(s)\, ds \right| \leq \frac{t^{n+1}}{(n+1)!} \max_{0 \leq s \leq t} |y(s)|,$$

which vanishes as $n \to \infty$. Therefore, our presumed C^1 solution is uniquely represented by the infinite series

$$y(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!},$$

which of course is nothing other than the usual series expansion for e^t. To finish the argument, one now verifies that the series is in fact infinitely differentiable and satisfies the differential equation and initial condition. For the differentiability see Lemma 1.3 below.

Viewing the exponential function as the solution of $\dot{y} = y$ allows for a transparent derivation of the homomorphism property of the exponential. Indeed, it is none other than the group property of flows which implies

$$y(t_2)y(t_1) = y(t_1 + t_2)$$

which proves that $y(t) = e^t$ for every rational t, and then by continuity for every real t, and motivates why we *define*

$$e^t := \sum_{j=0}^{\infty} \frac{t^j}{j!} \qquad \forall\, t \in \mathbb{R}.$$

Hence our series $E(z)$ above is used as the *definition* of e^z for all $z \in \mathbb{C}$. We have the group property $e^{z_1+z_2} = e^{z_1}e^{z_2}$, and by comparison with the power series of cos and sin on $\mathbb{R}$, we arrive at the famous Euler formula

$$e^{i\theta} = \cos\theta + i\sin\theta$$

for all $\theta \in \mathbb{R}$. This, in particular, shows that $z = re^{i\theta}$ where (r, θ) are the polar coordinates of z. This in turn implies that

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$$

for every $n \geq 1$ (de Moivre's formula). Now suppose that $z = re^{i\theta}$ with $r > 0$ (this just means that $z \neq 0$). Then by the preceding,

$$z = e^{\log r + i\theta} \quad \text{or} \quad \log z = \log r + i\theta.$$

Note that the logarithm is not well-defined since θ and $\theta + 2\pi n$ for any $n \in \mathbb{Z}$ both have the property that exponentiating leads to z. Similarly,

$$\left(r^{\frac{1}{n}} e^{i\frac{\theta}{n}} e^{\frac{2\pi i k}{n}}\right)^n = z \quad \forall\, 1 \leq k \leq n,$$

which shows that there are n different possibilities for $\sqrt[n]{z}$. Later on we shall see how these functions become single-valued on their natural Riemann surfaces. Let us merely mention at this point that the complex exponential is most naturally viewed as the covering map

$$\mathbb{C} \to \mathbb{C}^* := \mathbb{C} \setminus \{0\},$$
$$z \mapsto e^z.$$

showing that $\mathbb{C}$ is the universal cover of $\mathbb{C}^*$.

1.2. Holomorphic, analytic, and conformal

But for now, we of course wish to differentiate functions defined on some open set $\Omega \subset \mathbb{C}$. There are two relevant notions of derivative here and we will need to understand how they relate to each other.

The first is the crucial linearization idea from multivariable calculus and the second copies the idea of difference quotients from calculus. In many ways, the former is preferable to the latter. In what follows we shall either use U or Ω to denote planar regions, i.e., open and connected subsets of $\mathbb{R}^2$. Also, we will identify $z = x + iy$ with the real pair (x, y) and will typically write a complex-valued function as $f(z) = u(z) + iv(z) = (u, v)(z)$ where $u, v : \mathbb{C} \to \mathbb{R}$.

Definition 1.1. Let $f : \Omega \to \mathbb{C}$.

(a) We say that $f \in C^1(\Omega)$ if there exists df, a continuous 2×2 matrix-valued function, such that

$$f(z + h) = f(z) + df(z)(h) + r(z, h) \qquad \forall\, z, z + h \in \Omega, \tag{1.2}$$

$$\lim_{h \to 0} \frac{r(z, h)}{|h|} = 0, \tag{1.3}$$

where $df(z)(h)$ means the matrix $df(z)$ acting on the vector h. It is understood that $h \neq 0$ in the limit $h \to 0$.

(b) We say that f is *holomorphic* on Ω if

$$f'(z) := \lim_{w \to z} \frac{f(w) - f(z)}{w - z}$$

exists for all $z \in \Omega$ and is continuous on Ω. We denote this by $f \in \mathcal{H}(\Omega)$. A function $f \in \mathcal{H}(\mathbb{C})$ is called *entire*.

The function $r(z, h)$ is the deviation of the function $f(z + h)$ from its linear approximation $f(z) + df(z)h$. Differentiability means that this deviation vanishes faster than linearly as $h \to 0$. Note that (b) is equivalent to the existence of a function $f' \in C(\Omega)$ such that

$$f(z + h) = f(z) + f'(z)h + o(|h|) \qquad |h| \to 0,$$

where $f'(z)h$ is the product between the complex numbers $f'(z)$ and h. Hence we conclude that the holomorphic functions are precisely those functions in $C^1(\Omega)$ in the sense of (a) for which the differential $df(z)$ acts as a linear map via multiplication by a complex number.

We should remark that the condition $f' \in C(\Omega)$ can be dropped from part (b) of Definition 1.1 and replaced by complex differentiability alone. We shall elucidate this issue in Goursat's theorem below; see Theorem 1.34. But for now we prefer to add the continuity assumption on the derivative.

Obvious examples of holomorphic maps are the powers $f(z) = z^n$ for all $n \in \mathbb{Z}$ (if n is negative, then we exclude $z = 0$). They satisfy $f'(z) = nz^{n-1}$ by the binomial theorem. Also, since we can do algebra in $\mathbb{C}$ the same way we did over $\mathbb{R}$ it follows that the basic differentiation rules like the sum, product, quotient, and chain rules continue to hold for holomorphic functions. Let us demonstrate this for the chain rule: if $f \in \mathcal{H}(\Omega)$, $g \in \mathcal{H}(\Omega')$ and $f(\Omega) \subseteq (\Omega')$, then we know from the C^1-chain rule that

$$(g \circ f)(z+h) = (g \circ f)(z) + Dg(f(z))Df(z)h + o(|h|) \qquad |h| \to 0.$$

From (b) above we infer that $Dg(f(z))$ and $Df(z)$ act as multiplication by the complex numbers $g'(f(z))$ and $f'(z)$, respectively. Thus, we see that $g \circ f \in \mathcal{H}(\Omega)$ and $(g \circ f)' = g'(f)f'$. We leave the product and quotient rules to the reader.

It is clear that all polynomials are holomorphic functions. In fact, we can generalize this to all power series within their disk of convergence. Let us make this more precise.

Definition 1.2. We say that $f : \Omega \to \mathbb{C}$ is *analytic* (or $f \in \mathcal{A}(\Omega)$) if f is represented by a convergent power series expansion on a neighborhood around every point of Ω.

We now establish that analytic functions are holomorphic.

Lemma 1.3. $\mathcal{A}(\Omega) \subset \mathcal{H}(\Omega)$.

Proof. Suppose $z_0 \in \Omega$ and

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \qquad \forall\, z \text{ such that } |z - z_0| < r(z_0),$$

where $r(z_0) > 0$. For simplicity, we may assume that $z_0 = 0$ and set $r_0 = r_0(0)$. Fix $|z| < r_0$ and write

$$f(z+h) = f(z) + g(z)h + r(z,h),$$

where $g(z) := \sum_{n=1}^{\infty} n a_n z^{n-1}$. It is important to note that this series still converges on $|z| < r_0$ as can be seen from the formula for the radius of convergence, for example. We need to prove that $\frac{r(z,h)}{|h|} \to 0$ as $h \to 0$. If $z = 0$, then this is immediate since in that case $g(z) = a_1$. We have

$$r(z,h) = \sum_{n=2}^{\infty} a_n B_n(z,h),$$

$$B_n(z,h) = (z+h)^n - z^n - nz^{n-1}h = \sum_{k=2}^{n} \binom{n}{k} h^k z^{n-k},$$

since $B_0 = B_1 = 0$. Let $|z| + |h| < r_1 < r_2 < r_0$. Then we have $|a_n| \le \frac{M}{r_2^n}$ for all $n \ge 0$ and some $M < \infty$. Also,

$$
\begin{aligned}
|B_n(z,h)| &\le |h|^2 \sum_{k=0}^{n-2} \binom{n}{k+2} |h|^k |z|^{n-2-k} \\
&\le |h|^2 \sum_{k=0}^{n-2} n^2 \binom{n-2}{k} |h|^k |z|^{n-2-k} \qquad (1.4)\\
&= |h|^2 n^2 (|z| + |h|)^{n-2},
\end{aligned}
$$

for any $n \ge 2$. Hence

$$
\begin{aligned}
|r(z,h)| &\le \sum_{n=2}^{\infty} \frac{M}{r_2^n} n^2 |h|^2 (|z| + |h|)^{n-2} \qquad (1.5)\\
&\le \frac{|h|^2 M}{r_2^2} \sum_{n=0}^{\infty} (n+2)^2 \Big(\frac{r_1}{r_2}\Big)^n.
\end{aligned}
$$

We conclude that for some constant C,

$$|r(z,h)| \le C\,|h|^2,$$

which is more than we need. Thus,

$$f'(z) = \sum_{n=0}^{\infty} n a_n (z - z_0)^{n-1} \qquad \forall\, |z - z_0| < r(z_0),$$

as desired. $\square$

In fact, one can differentiate any number of times and

$$f^{(k)}(z) = \sum_{n=0}^{\infty} (n)_k\, a_n (z - z_0)^{n-k} \qquad \forall\, |z - z_0| < r(z_0),$$

where

$$(n)_k = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

This establishes the well-known relation $a_n = \frac{f^{(n)}(z_0)}{n!}$ for all $n \ge 0$.

We note that with e^z defined as above, $(e^z)' = e^z$ from the series representation (1.1). It is a remarkable fact of basic complex analysis that one has equality in Lemma 1.3, i.e., $\mathcal{A}(\Omega) = \mathcal{H}(\Omega)$. In order to establish this equality, we need to be able to integrate; see the section about integration below. Before we delve into the fundamental integration theory, we first develop the geometric aspects of holomorphic functions in some detail.

Recall that $f = u+iv = (u, v)$ belongs to $C^1(\Omega)$ if and only if the partials u_x, u_y, v_x, v_y exist and are continuous on Ω. If $f \in \mathcal{H}(\Omega)$, then by letting w (in Definition 1.1) approach z along the x or y-directions, respectively,

$$f'(z) = u_x + iv_x = -iu_y + v_y$$

so that

$$u_x = v_y, \qquad u_y = -v_x, \tag{1.6}$$

which is the same as

$$f_x + if_y = 0.$$

These relations are known as the *Cauchy-Riemann equations*. They are equivalent to the property that

$$df = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \rho A \quad \text{for some} \quad \rho \geq 0, \quad A \in \mathrm{SO}(2, \mathbb{R}).$$

In other words, at each point where a holomorphic function f has a non-vanishing derivative, its differential df is a conformal matrix: it preserves angles and the orientation between vectors. Conversely, if $f \in C^1(\Omega)$ has the property that df is proportional to a rotation everywhere on Ω, then $f \in \mathcal{H}(\Omega)$. Let us summarize these important observations.

Theorem 1.4. *A complex-valued function $f \in C^1(\Omega)$ is holomorphic if and only if the Cauchy-Riemann equations hold in Ω. This is equivalent to df being the composition of a rotation and a dilation (possibly by zero) at each point in Ω.*

Proof. We already saw that the Cauchy-Riemann equation is necessary. Conversely, since $f \in C^1(\Omega)$ as $(\xi, \eta) \to 0$,

$$\begin{aligned} u(x+\xi, y+\eta) &= u(x,y) + u_x(x,y)\xi + u_y(x,y)\eta + o(|(\xi,\eta)|), \\ v(x+\xi, y+\eta) &= v(x,y) + v_x(x,y)\xi + v_y(x,y)\eta + o(|(\xi,\eta)|). \end{aligned}$$

Using that $u_x = v_y$ and $u_y = -v_x$ we obtain, with $\zeta = \xi + i\eta$,

$$f(z+\zeta) - f(z) = (u_x + iv_x)(z)(\xi + i\eta) + o(|\zeta|),$$

which of course proves that $f'(z) = u_x(z) + iv_x(z) = v_y(z) - iu_y(z)$ as desired. The second part was already discussed above. $\square$

The following notion is of central importance for all of complex analysis:

Definition 1.5. A function $f \in C^1(\Omega)$ is called *conformal* if $df \neq 0$ in Ω and df preserves the angle and orientation at each point.

Thus, the holomorphic functions are precisely those C^1 functions which are conformal at all points at which $df \neq 0$. Note that $f(z) = \bar{z}$ belongs to $C^1(\mathbb{C})$ but is not conformal since it reverses orientations. Also note that $f(z) = z^2$ is holomorphic everywhere but doubles angles at $z = 0$ (in the sense that curves crossing at 0 at angle α get mapped onto curves intersecting at 0 at angle 2α), so conformality is lost there.

A particularly convenient—as well as insightful—way of distinguishing holomorphic functions from C^1 functions is given by the $\partial_z, \partial_{\bar{z}}$ calculus. Assume that $f \in C^1(\Omega)$. Then the real-linear map $df(z)$ can be written as the sum of a complex-linear (meaning that $T(\lambda v) = \lambda T(v)$ for all complex λ) and a complex anti-linear transformation (meaning that $T(\lambda v) = \bar{\lambda} T(v)$). We now make this simple linear algebra fact completely explicit.

Lemma 1.6. *If $T : V \to W$ is an $\mathbb{R}$-linear map between complex vector spaces, then there is a unique representation $T = T_1 + T_2$ where T_1 is complex linear and T_2 complex anti-linear. The latter property means that $T_2(\lambda \vec{v}) = \bar{\lambda} T_2(\vec{v})$.*

Proof. Uniqueness follows since a $\mathbb{C}$-linear map which is simultaneously $\mathbb{C}$-anti-linear vanishes identically. For existence, set

$$T_1 = \frac{1}{2}(T - iTi), \qquad T_2 = \frac{1}{2}(T + iTi).$$

Then $T_1 i = i T_1$ and $T_2 i = -i T_2$, $T = T_1 + T_2$, as desired. $\square$

In other words, there exist complex numbers $w_1(z), w_2(z)$ such that

$$df(z) = w_1(z)\, dz + w_2(z)\, d\bar{z},$$

where dz is the identity map and $d\bar{z}$ the reflection about the real axis; in the previous formula, these maps are then followed by multiplication by the complex numbers w_1 and w_2, respectively. We are using here that all complex-linear transformations on $\mathbb{R}^2$ are given by multiplication by a complex number, whereas the complex anti-linear ones become complex linear by composing them with a reflection. To find w_1 and w_2 observe that since $dz = dx + idy$, $d\bar{z} = dx - idy$,

$$\begin{aligned} df &= \partial_x f\, dx + \partial_y f\, dy = \partial_x f\, \frac{1}{2}(dz + d\bar{z}) + \partial_y f\, \frac{1}{2i}(dz - d\bar{z}) \\ &= \frac{1}{2}(\partial_x f - i\partial_y f)\, dz + \frac{1}{2}(\partial_x f + i\partial_y f)\, d\bar{z} \\ &=: \partial_z f\, dz + \partial_{\bar{z}} f\, d\bar{z}. \end{aligned}$$

Thus, $f \in \mathcal{H}(\Omega)$ if and only if $f \in C^1(\Omega)$ and $\partial_{\bar{z}} f = 0$ in Ω. This means that the Cauchy-Riemann system is the same as $\partial_{\bar{z}} f = 0$. To see this explicitly, we write $f = u + iv$ whence $\partial_{\bar z} f = 0$ becomes the familiar form

$$u_x - v_y = 0, \quad u_y + v_x = 0;$$

see (1.6).

As an application of this formalism we record the following crucial fact: for any $f \in \mathcal{H}(\Omega)$,

$$d(f(z)\,dz) = \partial_z f\, dz \wedge dz + \partial_{\bar z} f\, d\bar z \wedge dz = 0,$$

which means that $f(z)\,dz$ is a *closed differential form.* This property is *equivalent* to the homotopy invariance of the Cauchy integral that we will encounter below. As a further example of the $dz, d\bar z$ formalism, we leave it to the reader to verify the chain rules (where we write $f = f(z), g = g(w)$)

$$\begin{aligned} \partial_z(g \circ f) &= \big[(\partial_w g) \circ f\big]\, \partial_z f + \big[(\partial_{\bar w} g) \circ f\big]\, \partial_z \bar f, \\ \partial_{\bar z}(g \circ f) &= \big[(\partial_w g) \circ f\big]\, \partial_{\bar z} f + \big[(\partial_{\bar w} g) \circ f\big]\, \partial_{\bar z} \bar f, \end{aligned} \tag{1.7}$$

as well as the representation of the Laplacian $\Delta = 4\frac{\partial^2}{\partial z \partial \bar z}$.

These ideas will be of particular importance once we discuss differential forms on Riemann surfaces.

1.3. The Riemann sphere

To continue our introductory chapter, we next turn to the simple but important idea of extending the notion of analyticity to functions that take the value ∞. In a similar vein, we can make sense of functions being analytic at $z = \infty$. To start with, we define the one-point compactification of $\mathbb{C}$, which we denote by $\mathbb{C}_\infty$, with the usual basis for the topology: the neighborhoods of ∞ are the complements of all compact sets. Thus, $\mathbb{C}_\infty \simeq S^2$ in the homeomorphic sense. Somewhat deeper as well as much more relevant for complex analysis is the fact that $\mathbb{C} \simeq S^2 \setminus \{p\}$ in the sense of *conformal equivalence* where $p \in S^2$ is arbitrary. Conformality here is expressed in terms of the standard Euclidean metric applied to the respective tangent spaces.

This equivalence is realized by the well-known stereographic projection; see the problems to this chapter, Chapter 4 below, as well as Figure 1.1. If the circle in that figure is the unit circle, $N = (0, 1)$, and $X = (x, y)$, then $Z = \frac{x}{1-y}$ as the reader will easily verify using similarity of triangles. This identifies the stereographic projection as the map

$$\Phi : S^2 \setminus \{(0,0,1)\} \to \mathbb{C},\ X = (x_1, x_2, x_3) \mapsto \frac{x_1 + ix_2}{1 - x_3}.$$

The stereographic projection preserves angles as well as circles; see Problem 1.6. We will see in Chapter 4 that

$$\mathbb{C}_\infty \simeq S^2 \simeq \mathbb{C}P^1$$

in the sense of conformal equivalences, and each of these Riemann surfaces are called the *Riemann sphere*. The space $\mathbb{C}P^1$ is the complex projective line.

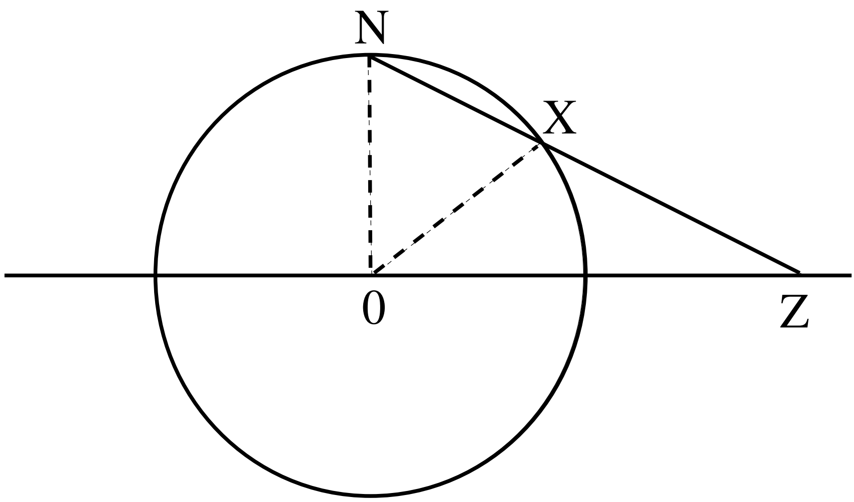

Figure 1.1. Stereographic projection

It is now clear how to extend the domain and range of holomorphic maps to

$$f : \mathbb{C} \to \mathbb{C}_\infty, \quad f : \mathbb{C}_\infty \to \mathbb{C}, \quad f : \mathbb{C}_\infty \to \mathbb{C}_\infty. \tag{1.8}$$

First, we need to require that f is continuous in each case. Second, we require f to be holomorphic both near points where it assumes values in the complex plane, as well as where ∞ is attained. In the latter case, it is natural to reduce to the finite case by considering $1/f$. To be specific, if $f(z_0) = \infty$ for some $z_0 \in \mathbb{C}$, then we say that f is holomorphic close to z_0 if and only if $\frac{1}{f(z)}$ is holomorphic around z_0. To make sense of f being analytic at $z = \infty$ with values in $\mathbb{C}$, we require that $f(\frac{1}{z})$ is holomorphic around $z = 0$. For the final example in (1.8), if $f(\infty) = \infty$, then f is analytic around $z = \infty$ if and only if $1/f(1/z)$ is analytic around $z = 0$. We remark that $z \mapsto \frac{1}{z}$ is conformal as a map from $\mathbb{C}_\infty \to \mathbb{C}_\infty$; this is a tautology in view of our choice of chart at $z = \infty$. On the other hand, if we interpret $\mathbb{C}_\infty$ as the Riemann sphere, then one needs to use here that stereographic projection is a conformal map.

We shall see later in this chapter that the holomorphic maps $f : \mathbb{C}_\infty \to \mathbb{C}$ are constants (indeed, such a map would have to be entire and bounded and therefore constant by Liouville's theorem, see Corollary 1.22 below). On the other hand, the maps $f : \mathbb{C} \to \mathbb{C}_\infty$ are precisely the *meromorphic ones* which we shall encounter in the next chapter. Finally, the holomorphic maps $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ are precisely the rational functions $\frac{P(z)}{Q(z)}$ where P, Q are polynomials. To see this, one shows that any such f is necessarily meromorphic with only finitely many poles in $\mathbb{C}$ and possibly a pole at $z = \infty$; see the following chapter for an explanation of these terms.

1.4. Möbius transformations

If we now accept that the holomorphic, and thus conformal, maps $\mathbb{C}_\infty \to \mathbb{C}_\infty$ are precisely the rational ones, then it is clear how to identify the *conformal automorphisms* (or *automorphisms*) amongst these maps. By this we mean the bijections whose inverse is also conformal. Indeed, in that case the fundamental theorem of algebra implies that P and Q both have degree one or less. This naturally leads to the following definition, which of course can be read ad hoc without the preceding motivation. Based on the argument of the previous paragraph (which the reader can safely accept on first reading), the lemma identifies all automorphisms of $\mathbb{C}_\infty$.

Lemma 1.7. *Every $A \in \mathrm{GL}(2,\mathbb{C})$ defines a transformation*

$$T_A(z) := \frac{az+b}{cz+d}, \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

which is holomorphic as a map from $\mathbb{C}_\infty \to \mathbb{C}_\infty$. It is called a fractional linear *or* Möbius transformation. *The map $A \mapsto T_A$ only depends on the equivalence class of A under the relation $A \sim B$ if and only if $A = \lambda B$, $\lambda \in \mathbb{C}^*$. In other words, the family of all Möbius transformations is the same as*

$$\mathrm{PSL}(2,\mathbb{C}) := \mathrm{SL}(2,\mathbb{C})/\{\pm \mathrm{id}\}. \tag{1.9}$$

We have $T_A \circ T_B = T_{A\circ B}$ and $T_A^{-1} = T_{A^{-1}}$. In particular, every Möbius transformation is an automorphism of $\mathbb{C}_\infty$.

Proof. It is clear that each T_A is a holomorphic map $\mathbb{C}_\infty \to \mathbb{C}_\infty$. The composition law $T_A \circ T_B = T_{A\circ B}$ and $T_A^{-1} = T_{A^{-1}}$ are simple computations that we leave to the reader. In particular, T_A has a conformal inverse and is thus an automorphism of $\mathbb{C}_\infty$. If $T_A = T_{\widetilde{A}}$ where $A, \widetilde{A} \in \mathrm{SL}(2,\mathbb{C})$, then

$$T_A'(z) = \frac{ad-bc}{(cz+d)^2} = T_{\widetilde{A}}'(z) = \frac{\widetilde{a}\widetilde{d} - \widetilde{b}\widetilde{c}}{(\widetilde{c}z + \widetilde{d})^2}$$

and thus $cz + d = \pm(\widetilde{c}z + \widetilde{d})$ under the assumption that

$$ad - bc = \widetilde{a}\widetilde{d} - \widetilde{b}\widetilde{c} = 1.$$

Hence A and $\widetilde{A}$ are the same matrices in $\mathrm{SL}(2,\mathbb{C})$ possibly up to a choice of sign, which establishes (1.9). $\square$

Fractional linear transformations enjoy many important properties which can be checked separately for each of the following four elementary transformations. In particular, Lemma 1.8 below proves that the group $\mathrm{PSL}(2,\mathbb{C})$ has four generators (in the sense of group theory).

Lemma 1.8. *Every Möbius transformation is the composition of four elementary maps:*

- *translations* $z \mapsto z + z_0$*,* $z_0 \in \mathbb{C}$*,*
- *dilations* $z \mapsto \lambda z$*,* $\lambda > 0$*,*
- *rotations* $z \mapsto e^{i\theta} z$*,* $\theta \in \mathbb{R}$*,*
- *inversion* $z \mapsto \frac{1}{z}$ *.*

Proof. If $c = 0$, then $T_A(z) = \frac{a}{d}z + \frac{b}{d}$. This is generated by the first three types of transformations. If $c \neq 0$, then we also require an inversion. In fact, one has

$$T_A(z) = \frac{bc - ad}{c^2} \frac{1}{z + \frac{d}{c}} + \frac{a}{c},$$

and we are done. $\square$

We now consider some examples. Since the imaginary axis is the perpendicular bisector of the segment $(-1, 1)$, it follows that the map $z \mapsto w = \frac{z-1}{z+1}$ takes the imaginary axis onto the circle $\{|w| = 1\}$ (since $|z - 1| = |z + 1|$ for $z \in i\mathbb{R}$). Moreover, $1 \mapsto 0$ so the right half-plane is taken onto the disk $\mathbb{D} := \{|w| < 1\}$. Similarly, $z \mapsto \frac{2z-1}{2-z}$ takes $\mathbb{D}$ onto itself with the boundary going onto the boundary. If we include all lines into the family of circles (they are circles passing through ∞), then these examples can serve to motivate the following lemma.

Lemma 1.9. *Fractional linear transformations take circles; onto circles.*

Proof. In view of the previous lemma, the only case requiring an argument is the inversion. Thus, let $|z - z_0| = r$ be a circle and set $w = \frac{1}{z}$. Then

$$\begin{aligned} 0 &= |z|^2 - 2\,\mathrm{Re}(\bar{z}z_0) + |z_0|^2 - r^2 \\ &= \frac{1}{|w|^2} - 2\frac{\mathrm{Re}(wz_0)}{|w|^2} + |z_0|^2 - r^2. \end{aligned}$$

If $|z_0| = r$, then one obtains the equation of a line in w. Note that this is precisely the case when the circle passes through the origin. Otherwise, we obtain the equation

$$0 = \left| w - \frac{\bar{z}_0}{|z_0|^2 - r^2} \right|^2 - \frac{r^2}{(|z_0|^2 - r^2)^2},$$

which is a circle. A line is given by an equation

$$2\,\mathrm{Re}(z\bar{z}_0) = a,$$

which transforms into $2\,\mathrm{Re}(z_0 w) = a|w|^2$. If $a = 0$, then we obtain another line through the origin. Otherwise, we obtain the equation $|w - z_0/a|^2 = |z_0/a|^2$ which is a circle.

An alternative argument invokes the Riemann sphere and uses the fact that stereographic projection preserves circles; see the problem section. Indeed, note that the inversion $z \mapsto \frac{1}{z}$ corresponds to a rotation of the Riemann sphere about the x_1 axis (the real axis of the plane). Since such a rotation preserves circles, a fractional linear transformation does, too. $\square$

Since $Tz = \frac{az+b}{cz+d} = z$ is a quadratic equation[1] for any Möbius transformation T, we see that T can have at most two fixed points unless it is the identity.

It is also clear that every Möbius transformation has at least one fixed point. The map $Tz = z + 1$ has exactly one fixed point, namely $z = \infty$, whereas $Tz = \frac{1}{z}$ has two, $z = \pm 1$.

Lemma 1.10. *A fractional linear transformation is determined completely by its action on three distinct points. Given distinct points $z_1, z_2, z_3 \in \mathbb{C}_\infty$, there exists a unique fractional linear transformation T with $Tz_1 = 0$, $Tz_2 = 1$, $Tz_3 = \infty$.*

Proof. For the first statement, suppose that S, T are Möbius transformations that agree at three distinct points. Then the Möbius transformation $S^{-1} \circ T$ has three fixed points and is thus the identity. For the second statement, let

$$Tz := \frac{z - z_1}{z - z_3} \frac{z_2 - z_3}{z_2 - z_1}$$

in case $z_1, z_2, z_3 \in \mathbb{C}$. If any one of these points is ∞, then we obtain the correct formula by passing to the limit here. $\square$

Lemma 1.10 suggests the following definition.

[1] Strictly speaking, this is a quadratic equation provided $c \neq 0$; if $c = 0$ one, obtains a linear equation with a fixed point in $\mathbb{C}$ and another one at $z = \infty$.

Definition 1.11. The cross ratio of four points $z_0, z_1, z_2, z_3 \in \mathbb{C}_\infty$ is defined as

$$[z_0 : z_1 : z_2 : z_3] := \frac{z_0 - z_1}{z_0 - z_3} \frac{z_2 - z_3}{z_2 - z_1}.$$

The relevance of this quantity lies with its invariance under Möbius transformations.

Lemma 1.12. *The cross ratio of any four distinct points is preserved under Möbius transformations. Moreover, four distinct points lie on a circle if and only if their cross ratio is real.*

Proof. Let z_1, z_2, z_3 be distinct and let $Tz_j = w_j$ for T be a Möbius transformation and $1 \le j \le 3$. Then for all $z \in \mathbb{C}$,

$$[w : w_1 : w_2 : w_3] = [z : z_1 : z_2 : z_3] \text{ provided } w = Tz.$$

This follows from the fact that the cross ratio on the left-hand side defines a Möbius transformation $S_1 w$ with the property that $S_1 w_1 = 0, S_1 w_2 = 1, S_1 w_3 = \infty$, whereas the right-hand side defines a transformation S_0 with $S_0 z_1 = 0, S_0 z_2 = 1, S_0 z_3 = \infty$. Hence $S_1^{-1} \circ S_0 = T$ as claimed. The second statement is an immediate consequence of the first and the fact that for any three distinct points $z_1, z_2, z_3 \in \mathbb{R}$, a fourth point z_0 has a real-valued cross ratio with these three if and only if $z_0 \in \mathbb{R}$. □

It is evident what symmetry of two points relative to a line means: they are reflections of each other relative to the line. While it is less evident what symmetry relative to a circle of finite radius means, the cross ratio allows for a reduction to the case of lines.

Definition 1.13. Let $z_1, z_2, z_3 \in \Gamma$ where $\Gamma \subset \mathbb{C}_\infty$ is a circle. We say that z and z^* are symmetric relative to Γ if

$$\overline{[z : z_1 : z_2 : z_3]} = [z^* : z_1 : z_2 : z_3].$$

Obviously, if $\Gamma = \mathbb{R}$, then $z^* = \bar{z}$. In other words, if Γ is a line, then z^* is the reflection of z across that line. If Γ is a circle of finite radius, then we can reduce matters to this case by an inversion.

Lemma 1.14. *Let $\Gamma = \{|z - z_0| = r\}$. Then for any $z \in \mathbb{C}_\infty$,*

$$z^* = \frac{r^2}{\bar{z} - \bar{z}_0}.$$

Proof. It suffices to consider the unit circle. Then

$$\overline{[z : z_1 : z_2 : z_3]} = [\bar{z} : z_1^{-1} : z_2^{-1} : z_3^{-1}] = [1/\bar{z} : z_1 : z_2 : z_3].$$

In other words, $z^* = \frac{1}{\bar{z}}$. The general case now follows from this via a translation and dilation. □

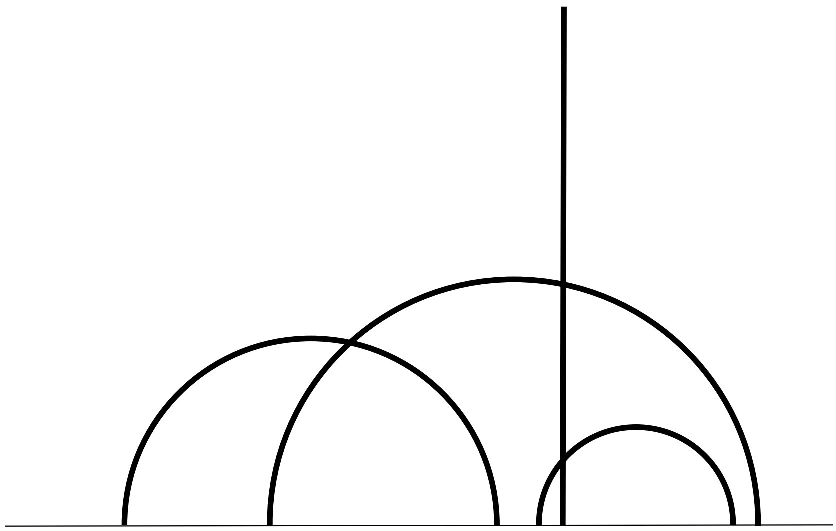

Figure 1.2. Geodesics in the hyperbolic plane

1.5. The hyperbolic plane and the Poincaré disk

Möbius transformations are important for several reasons. We now present a connection to geometry, which can be skipped on first reading. It requires familiarity with basic notions of Riemannian manifolds, such as metrics, isometry group, and geodesics. In the 19th century there was much excitement surrounding non-Euclidean geometry and there is an important connection between Möbius transformations and hyperbolic geometry: the isometries of the hyperbolic plane $\mathbb{H}$ are precisely those Möbius transformations which preserve it. Let us be more precise. Consider the upper half-plane model of the hyperbolic plane given by

$$\mathbb{H} = \{z \in \mathbb{C} \,:\, \operatorname{Im} z > 0\}, \quad ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{d\bar{z}\,dz}{(\operatorname{Im} z)^2}.$$

The subgroup of $\mathrm{PSL}(2,\mathbb{C})$ which preserves the upper half-plane is precisely $\mathrm{PSL}(2,\mathbb{R})$. Indeed, considering the action on three points on the real line, one sees that $z \mapsto \frac{az+b}{cz+d}$ preserves $\mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$ if and only if $a, b, c, d \in \lambda\mathbb{R}$ for some $\lambda \in \mathbb{C}^*$. In other words, the stabilizer of $\mathbb{R}$ (as a set) is $\mathrm{PGL}(2,\mathbb{R})$ which contains $\mathrm{PSL}(2,\mathbb{R})$ as an index two subgroup. The latter preserves the upper half-plane, whereas those matrices with negative determinant interchange the upper with the lower half-plane. It is easy to check (see the problems at the end of this chapter) that $\mathrm{PSL}(2,\mathbb{R})$ operates transitively on $\mathbb{H}$ and preserves the metric: for the latter, one computes

$$w = \frac{az+b}{cz+d} \Longrightarrow \frac{d\bar{w}\,dw}{(\operatorname{Im} w)^2} = \frac{d\bar{z}\,dz}{(\operatorname{Im} z)^2}.$$

In particular, the geodesics are preserved under $\mathrm{PSL}(2,\mathbb{R})$. Since the metric does not depend on x it follows that all vertical lines are geodesics. Now consider $\mathrm{Stab}(i)$, which are all Möbius transformations that fix i. Thus, $\frac{ai+b}{ci+d} = i$ whence $a = d, b = -c$. Since we can assume that $a^2 + b^2 = 1$, it follows that we can set $a + bi = e^{i\theta}$. But then

$$T'(i) = \frac{1}{(ci+d)^2} = e^{2i\theta}$$

acts as a rotation in the tangent space of $\mathbb{H}$ at $z_0 = i$. This property carries over to other $z_0 \in \mathbb{H}$.

Since isometries preserve geodesics, the latter are precisely all circles which intersect the real line at a right angle (with the vertical lines being counted as circles of infinite radius). From this it is clear that the hyperbolic plane satisfies all axioms of Euclidean geometry with the exception of the parallel axiom: there are many "lines" (i.e., geodesics) passing through a point which is not on a fixed geodesic that do not intersect that geodesic. Let us now prove the famous Gauss-Bonnet theorem which describes the hyperbolic area of a triangle whose three sides are geodesics (those are called geodesic triangles). We remark in passing that the following theorem is a special case of a much more general statement about integrating the Gaussian curvature over a geodesic triangle on a general two-dimensional Riemannian manifold. From this perspective, Theorem 1.15 expresses that $\mathbb{H}$ has constant sectional curvature equal to -1. But we shall make no use of this fact here.

Theorem 1.15. *Let T be a geodesic triangle with angles $\alpha_1, \alpha_2, \alpha_3$. Then* $\mathrm{Area}(T) = \pi - (\alpha_1 + \alpha_2 + \alpha_3)$.

Proof. There are four essentially distinct types of geodesic triangles, depending on how many of its vertices lie on $\mathbb{R}_\infty$. Up to equivalences via transformations in $\mathrm{PSL}(2,\mathbb{R})$ (which are isometries and therefore also preserve the area) we see that it suffices to consider precisely those cases described in Figure 1.3. Let us start with the case in which exactly two vertices belong to $\mathbb{R}_\infty$ as shown in that figure (the second triangle from the right). Without loss of generality one vertex coincides with 1, the other with ∞, and the circular arc lies on the unit circle with the projection of the second

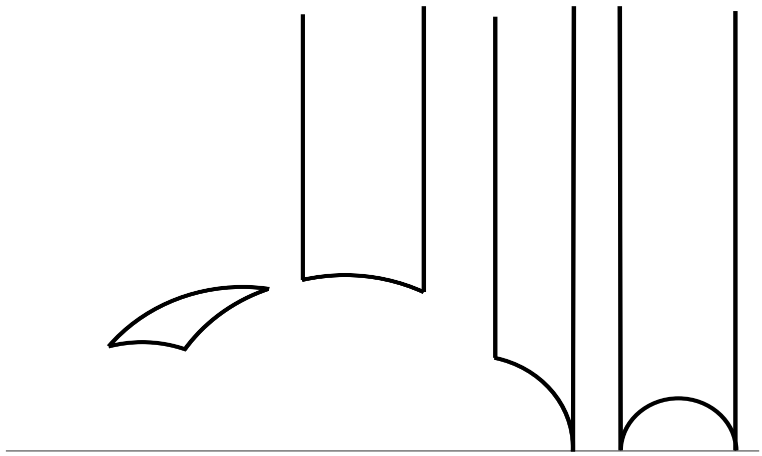

Figure 1.3. Geodesic triangles

finite vertex onto the real axis being x_0. Then

$$\begin{aligned}\operatorname{Area}(T) &= \int_{x_0}^{1}\int_{y(x)}^{\infty}\frac{dx\,dy}{y^2} = \int_{x_0}^{1}\frac{dx}{\sqrt{1-x^2}}\\ &= \int_{\alpha_0}^{0}\frac{d\cos\phi}{\sqrt{1-\cos^2(\phi)}} = \alpha_0 = \pi-\alpha_1,\end{aligned}$$

as desired since the other two angles are zero. By additivity of the area we can deal with the other two cases in which at least one vertex is real. We leave the case where no vertex lies on the (extended) real axis to the reader; the idea is to use Figure 1.4. $\square$

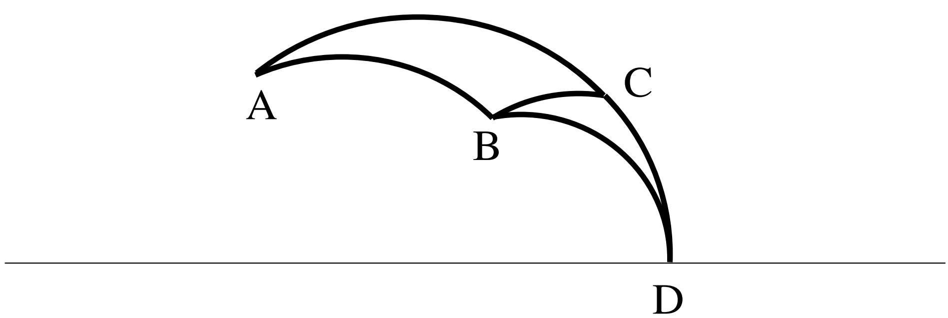

Figure 1.4. The case of no real vertex

We leave it to the reader to generalize the Gauss-Bonnet theorem to geodesic polygons. Many interesting questions about Möbius transformations remain, for example, how to characterize those that correspond to rotations of the sphere, or how to determine all finite subgroups of $\mathrm{PSL}(2,\mathbb{C})$.

The upper half-plane is mapped onto the disk $\mathbb{D}$ by the Möbius transformation $z \mapsto \frac{z-i}{z+i}$. This allows us to map the non-Euclidean geometry that we established on the upper half-plane onto the disk. For example, since Möbius transformations take circles onto circles, we conclude that the geodesics in this geometry on the disk are segments of circular arcs that intersect $\partial\mathbb{D}$ at right angles.

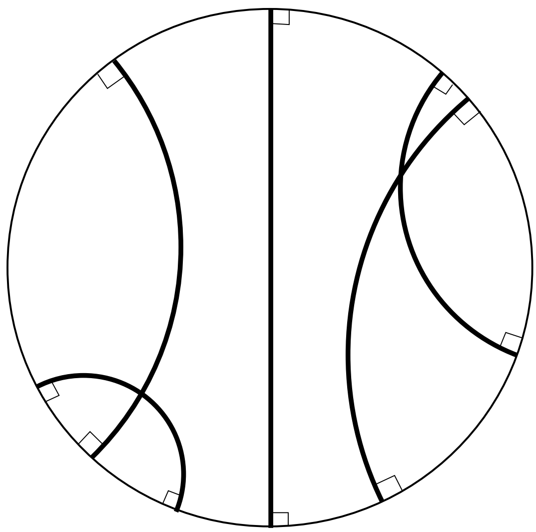

Figure 1.5. The Poincaré disk

Problem 1.13 introduces the natural metrics in these non-Euclidean geometries. Relative to these metrics, the geodesics we observed above are the shortest paths between any two points on them.

1.6. Complex integration, Cauchy theorems

We now develop our complex calculus further. The following definition introduces the complex integral. Indeed, it is the only definition which preserves the fundamental theorem of calculus for holomorphic functions.

Definition 1.16. For any C^1 curve $\gamma : [0,1] \to \Omega$ and any complex-valued $f \in C(\Omega)$ we define

$$\int_\gamma f(z)\,dz = \int_0^1 f(\gamma(t))\gamma'(t)\,dt. \tag{1.10}$$

If γ is a closed curve ($\gamma(0) = \gamma(1)$), then we also write $\oint_\gamma f(z)\,dz$ for this integral (note that an orientation needs to be specified). By taking sums, this definition extends to continuous curves γ which are piecewise C^1 on finitely many closed intervals.

The integral is $\mathbb{C}$-valued, and $f(\gamma(t))\gamma'(t)$ is understood as multiplication of complex numbers. From the chain rule, we deduce the fundamental fact that the line integrals of this definition do not depend on any particular C^1 parametrization of the curve as long as the orientation is preserved (hence, there is no loss in assuming that γ is parametrized by $0 \le t \le 1$).

We shall repeatedly encounter the question of estimating the absolute value of the integral in (1.10). Passing the absolute values inside of the integral yields

$$\begin{aligned} \Big|\int_\gamma f(z)\,dz\Big| &\le \int_0^1 |f(\gamma(t))|\,|\gamma'(t)|\,dt \\ &\le \sup_{0\le t\le 1} |f(\gamma(t))| \int_0^1 |\gamma'(t)|\,dt = ML(\gamma), \end{aligned} \tag{1.11}$$

where $L(\gamma)$ is the length of γ, and M is an upper bound for $|f|$ on γ. This is our standard estimate on the size of complex integrals.

Let us now investigate the fundamental theorem of calculus in this context.

Proposition 1.17. *From the chain rule, we immediately obtain the following: if $f \in \mathcal{H}(\Omega)$, then*

$$\int_\gamma f'(z)\,dz = f(\gamma(1)) - f(\gamma(0))$$

for any γ as in the definition. In particular,

$$\oint_\gamma f'(z)\,dz = 0 \qquad \forall \text{ closed curves } \gamma \text{ in } \Omega.$$

Proof. It follows from the chain rule that

$$\int_\gamma f'(z)\,dz = \int_0^1 f'(\gamma(t))\gamma'(t)\,dt = \int_0^1 \frac{d}{dt} f(\gamma(t))\,dt = f(\gamma(1)) - f(\gamma(0))$$

for any γ as in the definition. □

Perhaps the most fundamental complex line integral is the one in the following lemma. It shows that not every integral over a closed loop vanishes.

Lemma 1.18. *Let γ_r be the circle $\{|z| = r\}$, $r > 0$ fixed, with the counter-clockwise orientation. Then*

$$\oint_{\gamma_r} z^n \, dz = \begin{cases} 0 & \text{if } n \neq -1, \\ 2\pi i & \text{if } n = -1, \end{cases} \tag{1.12}$$

where n is an arbitrary integer.

Proof. By direct computation, with $\gamma_r(t) := re^{it}$, $r > 0$,

$$\oint_{\gamma_r} z^n \, dz = \int_0^{2\pi} r^n e^{int} \, rie^{it} \, dt = \begin{cases} 0 & \text{if } n \neq -1, \\ 2\pi i & \text{if } n = -1. \end{cases}$$

In $\Omega = \mathbb{C}^*$, the function $f(z) = z^n$ has the primitive $F_n(z) = \frac{z^{n+1}}{n+1}$ provided $n \neq -1$. This explains why we obtain 0 for all $n \neq -1$. $\square$

From the $n = -1$ case of the previous lemma, we realize that $\frac{1}{z}$ *does not* have a (holomorphic) primitive on $\mathbb{C}^*$. This issue merits further investigation (for example, we need to answer the question whether $\frac{1}{z}$ has a *local primitive* on $\mathbb{C}^*$—this is indeed the case and this primitive is a branch of $\log z$).

In order to answer such questions, we need to develop some general tools. The most fundamental of those, the Cauchy theorem, gives a sufficient criterion for the path independence of complex line integrals.

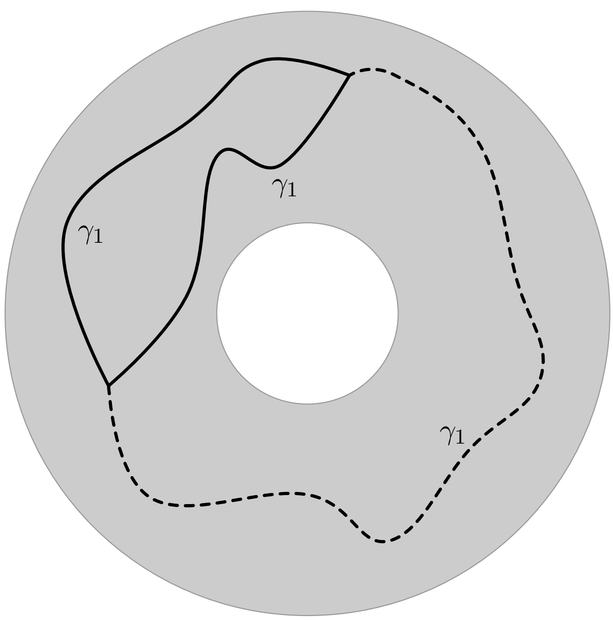

Figure 1.6. Two homotopic curves γ_1 and γ_2. The curve γ_3 is not homotopic to γ_1 or γ_2

Figure 1.6 shows two curves, namely γ_1 and γ_2, which are homotopic within the annular region they lie in. The dashed curve γ_3 is not homotopic to either of them within the annulus.

Theorem 1.19. *Let $\gamma_0, \gamma_1 : [0,1] \to \Omega$ be C^1 curves*[2] *with $\gamma_0(0) = \gamma_1(0)$ and $\gamma_0(1) = \gamma_1(1)$ (the fixed endpoint case) or $\gamma_0(0) = \gamma_0(1)$, $\gamma_1(0) = \gamma_1(1)$ (the closed case). Assume that they are C^1-homotopic*[3] *in the following sense: there exists a continuous map $H : [0,1]^2 \to \Omega$ with $H(t,0) = \gamma_0(t)$, $H(t,1) = \gamma_1(t)$ and such that $H(\cdot, s)$ is a C^1 curve for each $0 \le s \le 1$. Moreover, in the fixed endpoint case we assume that H freezes the endpoints, whereas in the closed case we assume that each curve from the homotopy is closed. Then*

$$\int_{\gamma_0} f(z)\,dz = \int_{\gamma_1} f(z)\,dz$$

for all $f \in \mathcal{H}(\Omega)$. In particular, if γ is a closed curve in Ω which is homotopic to a point, then

$$\oint_{\gamma} f(z)\,dz = 0. \tag{1.13}$$

In particular, this is the case if γ is the boundary of a subregion $\Omega_1 \subset \Omega$ such that $\overline{\Omega_1}$ is diffeomorphic to a closed disk.

Proof. We first note the important fact that $f(z)\,dz$ is a closed form. Indeed,

$$d(f(z)\,dz) = \partial_z f(z)\,dz \wedge dz + \partial_{\bar z} f(z)\,d\bar z \wedge dz = 0,$$

by the Cauchy-Riemann equation $\partial_{\bar z} f = 0$. Thus, Cauchy's theorem is a special case of the homotopy invariance of the integral over closed forms which in turn follows from Stokes's theorem. Let us briefly recall the details: since a closed form is locally exact, we first note that

$$\oint_{\eta} f(z)\,dz = 0$$

for all closed curves η which fall into sufficiently small disks, say. But then we can triangulate the homotopy so that

$$\int_{\gamma_0} f(z)\,dz - \int_{\gamma_1} f(z)\,dz = \sum_j \oint_{\eta_j} f(z)\,dz = 0,$$

where the sum is over a finite collection of small loops which constitute the triangulation of the homotopy H. The more classically minded reader might

[2] This can of course be relaxed to piecewise C^1, which means that we can write the curve as a finite concatenation of C^1 curves. The same comment applies to the homotopy.

[3] In light of commonly used terminology it is probably best to refer to this as *homotopic through C^1 curves*, but for simplicity, we shall continue to abuse terminology and use C^1-homotopic.

prefer to use Green's formula, which we now recall: if R is a rectangle, say, and if a, b are C^1 on R, up to the boundary, then

$$\oint_{\partial R} a(x,y)\,dx + b(x,y)\,dy = \iint_R (-a_y(x,y) + b_x(x,y))\,dx\,dy.$$

This formula extends to far more general regions such as those diffeomorphic to a disk and bounded by finitely many C^1 curves. Suppose therefore that $U \subset \Omega$ is such a region. Then returning to our function f as above we obtain

$$\begin{aligned}\oint_{\partial U} f(z)\,dz &= \oint_{\partial U} u\,dx - v\,dy + i(u\,dy + v\,dx)\\ &= \iint_U (-u_y - v_x)\,dx\,dy + i\iint_U (-v_y + u_x)\,dx\,dy = 0,\end{aligned}$$

where the final equality sign follows from the Cauchy-Riemann equations. □

For the Cauchy theorem it is of course absolutely essential that curves (or "contours" as they are commonly referred to) are deformed *within* the region of holomorphy. Indeed, Lemma 1.18 with $n = -1$ shows that an integral around a closed loop need not vanish. The issue here is, of course, that the circle γ_r cannot be contracted to a point *without hitting the origin.* In a similar vein, (1.13) might fail if γ cannot be "filled in" to a region Ω_1 which lies entirely within Ω; see Lemma 1.18.

The Cauchy theorem is typically applied to simple configurations, such as two circles which are homotopic to each other in the region of holomorphy of some function f. As an example, we now derive the Cauchy formula, a tool of fundamental importance to mathematics.

Proposition 1.20. *Let $\overline{D(z_0,r)} \subset \Omega$ and $f \in \mathcal{H}(\Omega)$. Then*

$$f(z) = \frac{1}{2\pi i}\oint_\gamma \frac{f(\zeta)}{\zeta - z}\,d\zeta \quad \text{where } \gamma(t) = z_0 + re^{it} \tag{1.14}$$

for all $z \in D(z_0,r)$.

Proof. Fix any $z \in D(z_0,r)$ and apply Theorem 1.19 to the region $U_\varepsilon := D(z_0,r) \setminus D(z,\varepsilon)$ where $\varepsilon > 0$ is small. The importance of U_ε of course lies with the fact that $\zeta \mapsto \frac{f(\zeta)}{\zeta - z}$ is holomorphic in this region for any $\varepsilon > 0$. Moreover, the two boundary circles of U_ε are homotopic to each other relative to the region $\Omega \setminus \{z\}$. Therefore, by Theorem 1.19,

$$
\begin{aligned}
0 &= \frac{1}{2\pi i}\int_{\partial U_\varepsilon} \frac{f(\zeta)}{z-\zeta}\,d\zeta = \frac{1}{2\pi i}\int_{\partial D(z_0,r)} \frac{f(\zeta)}{z-\zeta}\,d\zeta \\
&\quad - \frac{1}{2\pi i}\int_{\partial D(z,\varepsilon)} \frac{f(\zeta)-f(z)}{z-\zeta}\,d\zeta - \frac{f(z)}{2\pi i}\int_{\partial D(z,\varepsilon)} \frac{1}{z-\zeta}\,d\zeta \\
&= \frac{1}{2\pi i}\int_{\partial D(z_0,r)} \frac{f(\zeta)}{z-\zeta}\,d\zeta + O(\varepsilon) + f(z) \quad \text{as } \varepsilon \to 0.
\end{aligned}
\tag{1.15}
$$

To pass to the last line we bounded the first integral in (1.15), which involves the ratio $\frac{f(\zeta)-f(z)}{z-\zeta}$, via (1.11), noting that differentiability implies that

$$\Big|\frac{f(\zeta)-f(z)}{z-\zeta}\Big| \le M \quad \forall\, 0<|z-\zeta|\le \varepsilon_0,$$

where $\varepsilon_0 > 0$ is small and fixed. This gives the $O(\varepsilon)$ in the final equality. Furthermore, we used the $n=-1$ case of (1.12) to pass to the third term of the last line:

$$-\frac{f(z)}{2\pi i}\int_{\partial D(z,\varepsilon)} \frac{1}{z-\zeta}\,d\zeta = f(z)$$

and we are done. □

It is clear both from the statement and the proof that Cauchy's formula is intimately tied up with the $n=-1$ case of (1.12). The Cauchy formula is remarkable for many reasons. Indeed, it implies that a holomorphic function in a disk is determined by its values on the boundary of the disk.

1.7. Applications of Cauchy's theorems

It also implies the astonishing fact that holomorphic functions are in fact analytic. This is done by noting that the integrand in (1.14) is analytic relative to z. In other words, we reduce ourselves to the geometric series.

Corollary 1.21. $\mathcal{A}(\Omega) = \mathcal{H}(\Omega)$. *In fact, every $f \in \mathcal{H}(\Omega)$ is represented by a convergent power series on $D(z_0, r)$ where $r = \mathrm{dist}(z_0, \partial\Omega)$.*

Proof. We proved in Lemma 1.3 that analytic functions are holomorphic. For the converse, we use the previous proposition to conclude that with γ

given by a circle $\{|\zeta - z_0| = r_0\} \subset \Omega$ and z lying inside of this circle,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{\zeta - z_0 - (z - z_0)}\, d\zeta \\ &= \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{\zeta - z_0} \sum_{n=0}^{\infty} \left(\frac{z - z_0}{\zeta - z_0}\right)^n d\zeta \\ &= \sum_{n=0}^{\infty} \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta\ (z - z_0)^n. \end{aligned}$$

The interchange of summation and integration is justified due to uniform and absolute convergence of the series which follows from $\left|\frac{z-z_0}{\zeta-z_0}\right| = \frac{|z-z_0|}{r_0} < 1$. Thus, we obtain that f is analytic and, moreover,

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n$$

converges on $|z - z_0| < \operatorname{dist}(z_0, \partial\Omega)$ with

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_\gamma \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta \tag{1.16}$$

for any $n \geq 0$. □

Consider the function $f(x) = \frac{1}{1+x^2}$ on the real line. Around $x = 0$ this function has a convergent Taylor series

$$f(x) = \sum_{n=0}^{\infty} (-1)^n x^{2n}, \quad |x| < 1.$$

But there is no apparent reason why the radius of convergence should be $R = 1$ in this case. Indeed, $f(x)$ remains smooth on the whole real line. However, when viewed over the complex numbers, $f(z) = \frac{1}{1+z^2}$ immediately reveals the reason for this barrier: f ceases to be holomorphic at $z = \pm i$ whence $R = 1$.

In summary, in contrast to power series over $\mathbb{R}$, over $\mathbb{C}$ there is an explanation for the radius of convergence: $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ has finite and positive radius of convergence R if and only if $f \notin \mathcal{H}(\Omega)$ for every Ω which compactly contains $D(z_0, R)$. We immediately obtain a number of corollaries from this. Recall that an entire function is holomorphic on the whole complex plane.

Corollary 1.22. *(a) Cauchy's estimates: Let $f \in \mathcal{H}(\Omega)$ with $|f(z)| \leq M$ on Ω. Then*

$$|f^{(n)}(z)| \leq \frac{M n!}{\operatorname{dist}(z, \partial\Omega)^n}$$

for every $n \geq 0$ and all $z \in \Omega$.

(b) Liouville's theorem: If $f \in \mathcal{H}(\mathbb{C}) \cap L^\infty(\mathbb{C})$, then f is constant. In other words: bounded entire functions are constant. More generally, if $|f(z)| \le C(1+|z|^N)$ for all $z \in \mathbb{C}$, for some fixed integer $N \ge 0$ and a finite constant C, then f is a polynomial of degree at most N.

Proof. (a) follows by putting absolute values inside (1.16), in other words we use (1.11). For

(b) apply (a) to $\Omega = D(0, R)$ and let $R \to \infty$. This shows that $f^{(k)} \equiv 0$ for all $k > N$. □

Part (b) has a famous consequence, namely the *fundamental theorem of algebra.*

Proposition 1.23. *Every polynomial $P \in \mathbb{C}[z]$ of positive degree has a complex zero; in fact, it has exactly as many zeros over $\mathbb{C}$ (counted with multiplicity) as its degree.*

Proof. Suppose $P(z) \in \mathbb{C}[z]$ is a polynomial of positive degree and without zero in $\mathbb{C}$. Then $f(z) := \frac{1}{P(z)}$ is an entire function. We claim that this function vanishes at infinity. Indeed,

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$$

with $n \ge 1$ and $a_n \ne 0$. Thus,

$$|P(z)| \ge |a_n||z|^n - \sum_{k=0}^{n-1} |a_k||z|^k \ge \frac{1}{2}|a_n| R^n$$

for all $|z| = R$, with R large. Hence $|P(z)| \to \infty$ as $|z| \to \infty$, and f is bounded. By Liouville, f is constant and so P is constant contrary to the assumption of positive degree. So $P(z_0) = 0$ for some $z_0 \in \mathbb{C}$. Factoring out $z - z_0$ we conclude inductively that P has exactly $\deg(P)$ many complex zeros, as desired. □

We now return to the problem of finding holomorphic primitives to complex-valued functions f on a region Ω. Since holomorphic functions are analytic, and therefore have holomorphic derivatives of any order, it follows that f needs to be itself holomorphic. However, as the example $f(z) = z^{-1}$ on $\Omega = \mathbb{C} \setminus \{0\}$ shows, not every region Ω is admissible (compare Lemma 1.18 and Proposition 1.17). Of key importance here is the notion of simple connectivity.

Definition 1.24. A region Ω is called simply-connected if every closed curve can be contracted to a point. In other words, if there is a homotopy between any closed curve and a constant curve.

In the definition it makes no difference if we use continuous curves or (piecewise) C^1 curves.

Also, by the assumed connectivity of Ω we conclude that the point to which we contract can be chosen arbitrarily.

Proposition 1.25. *Let Ω be simply-connected. Then for every $f \in \mathcal{H}(\Omega)$ such that $f \neq 0$ everywhere on Ω there exists $g \in \mathcal{H}(\Omega)$ with $e^{g(z)} = f(z)$. The function g is unique up to additive constants $2\pi i n$, $n \in \mathbb{Z}$ and any such choice of g is called a branch of the logarithm of f. Furthermore, for any $n \geq 1$ there exists $f_n \in \mathcal{H}(\Omega)$ with $(f_n(z))^n = f(z)$ for all $z \in \Omega$.*

In particular, if $\Omega \subset \mathbb{C}^*$ is simply-connected, then there exists $g \in \mathcal{H}(\Omega)$ with $e^{g(z)} = z$ everywhere on Ω. Such a g is called a branch of $\log z$. *Similarly, there exist holomorphic branches of any $\sqrt[n]{z}$ on Ω, $n \geq 1$.*

Proof. If $e^g = f$, then $g' = \frac{f'}{f}$ in Ω. So fix any $z_0 \in \Omega$ and define

$$g(z) := \int_{z_0}^{z} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta,$$

where the integration path joins z_0 to z and consists of a finite number of line segments (say). We claim that $g(z)$ does not depend on the choice of path. First note that $\frac{f'}{f} \in \mathcal{H}(\Omega)$ due to analyticity and nonvanishing of f. Second, since Ω is simply-connected, any two curves with coinciding initial and terminal points are homotopic to each other via a piecewise C^1 homotopy. Thus, Theorem 1.19 yields the desired equality of the integrals. It is now an easy matter to check that $g'(z) = \frac{f'(z)}{f(z)}$. Indeed,

$$\frac{g(z+h) - g(z)}{h} = \int_0^1 \frac{f'(z+th)}{f(z+th)}\, dt \to \frac{f'(z)}{f(z)} \quad \text{as } h \to 0.$$

So $g \in \mathcal{H}(\Omega)$ and $(fe^{-g})' \equiv 0$ shows that $e^g = cf$ where c is some constant different from zero and therefore $c = e^k$ for some $k \in \mathbb{C}$. Hence $e^{g(z)-k} = f(z)$ for all $z \in \Omega$ and we are done with the logarithm.

For the roots, set $f_n(z) := e^{g(z)/n}$. Then $(f_n(z))^n = e^{g(z)} = f(z)$ as desired. $\square$

The equivalence between holomorphic and analytic functions clearly has far-reaching consequences. We now present some results in this direction which heavily rely on properties of power series.

We begin with the *uniqueness theorem.* The name derives from the fact that two functions $f, g \in \mathcal{H}(\Omega)$ are identical if the set $\{z \in \Omega : f(z) = g(z)\}$ has an accumulation point inside of Ω.

Throughout, for any disk D, the punctured disk D^* denotes D with its center removed.

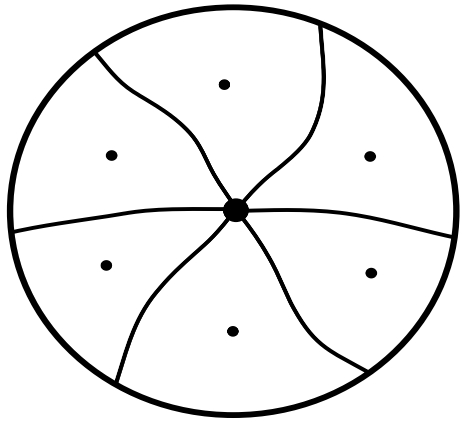

Figure 1.7. A branch point

Proposition 1.26. *Let $f \in \mathcal{H}(\Omega)$. Then the following are equivalent:*

- $f \equiv 0$,
- *for some* $z_0 \in \Omega$, $f^{(n)}(z_0) = 0$ *for all* $n \geq 0$,
- *the set* $\{z \in \Omega : f(z) = 0\}$ *has an accumulation point in* Ω.

Proof. We only need to show that third property implies the identical vanishing. Let $z_n \to z_0 \in \Omega$ as $n \to \infty$, where $f(z_n) = 0$ for all $n \geq 1$. Suppose $f^{(m)}(z_0) \neq 0$ for some $m \geq 0$. Then

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k = a_N (z - z_0)^N (1 + O(z - z_0)) \quad \text{as } z \to z_0$$

locally around z_0 where $N \geq 0$ is minimal with the property that $a_N \neq 0$. But then it is clear that f does not vanish on some punctured disk $D(z_0, r)^*$, contrary to assumption. Thus, $f^{(n)}(z_0) = 0$ for all $n \geq 0$ and thus $f \equiv 0$ locally around z_0. Since Ω is connected, it then follows that $f \equiv 0$ on Ω. Indeed, define $S := \{z \in \Omega \mid f \text{ vanishes locally around } z\}$. Then S is evidently open, but it is also closed by what we have just shown. Since $S \neq \emptyset$ it follows that $S = \Omega$ (recall Ω connected). This settles the equivalences. $\square$

Next, we describe the rigid mapping properties of holomorphic functions via a *normal form* representation (1.17). Amongst other things, this yields the *open mapping theorem* for nonconstant analytic functions.

Proposition 1.27. *Assume that f is holomorphic and is not constant. Then at every point $z_0 \in \Omega$ there exist a positive integer n and a holomorphic function h locally at z_0 such that*

$$f(z) = f(z_0) + [(z - z_0)h(z)]^n, \qquad h(z_0) \neq 0. \tag{1.17}$$

In particular, there are disks $D(z_0, \rho), D(f(z_0), r)$ *with the property that every* $w \in D(f(z_0), r)^*$ *has precisely* n *pre-images under* f *in* $D(z_0, \rho)^*$*. If* $f'(z_0) \neq 0$*, then* f *is a local* C^∞ *diffeomorphism. Finally, every nonconstant holomorphic map is an open map (i.e., it takes open sets to open sets).*

Proof. If f' does not vanish identically, let us first assume that $f'(z_0) \neq 0$. We claim that *locally* around z_0, the map $f(z)$ is a C^∞ diffeomorphism from a neighborhood of z_0 onto a neighborhood of $f(z_0)$ and, moreover, that the inverse map to f is also holomorphic. Indeed, in view of Theorem 1.4, the differential *df* is invertible at z_0. Hence by the usual inverse function theorem from multivariable calculus[4] we obtain the statement about diffeomorphisms. Furthermore, since *df* is conformal locally around z_0, its inverse is, too, and so f^{-1} is conformal and thus holomorphic. If $f'(z_0) = 0$, then there exists some positive integer n with $f^{(n)}(z_0) \neq 0$. But then from the power series representation

$$f(z) = f(z_0) + (z - z_0)^n g(z)$$

with $g \in \mathcal{H}(\Omega)$ satisfying $g(z_0) \neq 0$. By Proposition 1.25 we can write $g(z) = (h(z))^n$ for some $h \in \mathcal{H}(U)$ where U is a neighborhood of z_0 and $h(z_0) \neq 0$, whence (1.17). Now consider the function

$$\varphi(z) := (z - z_0)h(z).$$

By construction, $f(z) = f(z_0) + \varphi(z)^n$. We note that

$$\varphi'(z_0) = h(z_0) \neq 0.$$

By the preceding analysis of the $n = 1$ case we conclude that $\varphi(z)$ is a local diffeomorphism. But this implies that any $w \neq f(z_0)$ which lies in a small neighborhood of $f(z_0)$ admits exactly n solutions to the equation $f(z) = w$. Indeed, this equation is equivalent with the equation $w - f(z_0) = \varphi(z)^n$. Indeed, each of the n^{th} roots of $w - f(z_0)$ (which are all distinct) has a unique pre-image under $\varphi(z)$. In summary, f has the stated n-to-one mapping property. The openness is now also evident. □

Figure 1.7 shows the case of $n = 6$. The dots symbolize the six pre-images of some point. We remark that any point $z_0 \in \Omega$ for which $n \geq 2$ is called a *branch point*. The branch points are precisely the zeros of f' in Ω and therefore form a discrete subset of Ω (this means that every point of this subset is isolated from the other ones).

Corollary 1.28. *Suppose* f *is analytic on* Ω *and a bijection between* Ω *and* Ω'*. Then* $f' \neq 0$ *on* Ω *and* $f^{-1} : \Omega' \to \Omega$ *(the inverse map of* f*) is analytic.*

[4] In the following chapter we will use Rouché's theorem of complex analysis to circumvent this recourse through real calculus.

Proof. Proposition 1.27 implies that $f'(z) \neq 0$ for every $z \in \Omega$ so f is conformal. By real-variable calculus, the differential of f^{-1} at $w = f(z)$ is the inverse of the differential of f at z, in symbols $Df^{-1}(w) = (Df(z))^{-1}$. This matrix is again a composition of a rotation with a dilation and so f^{-1} is conformal and thus analytic. By the open mapping theorem, Ω' is open and is also connected. □

The maps described by the previous corollary play a central role in complex analysis and geometry.

Definition 1.29. A mapping as in Corollary 1.28 is called a conformal isomorphism between Ω and Ω', or simply an isomorphism. This establishes an equivalence relation between regions. If $\Omega = \Omega'$, then we call the map an automorphism. These maps form a group, denoted by $\mathrm{Aut}(\Omega)$.

It is of course a natural problem to determine when two regions are isomorphic. In other words, we wish to classify regions up to conformal equivalence. We shall discuss this problem in the wider and more suitable framework of Riemann surfaces, where the solution goes by the name of *uniformization theorem*.

Proposition 1.27 has an important implication known as the *maximum principle*.

Corollary 1.30. *Let $f \in \mathcal{H}(\Omega)$. If there exists $z_0 \in \Omega$ with $|f(z)| \leq |f(z_0)|$ for all $z \in \Omega$, then f is constant. If Ω is bounded and f is continuous on $\overline{\Omega}$, then $|f(z)| \leq \max_{\zeta \in \partial\Omega} |f(\zeta)|$ for all $z \in \Omega$. Equality can occur here only if f is constant.*

Proof. If f is not constant, then $f(\Omega)$ is open contradicting that $f(z_0) \in \partial f(\Omega)$, which is required by $|f(z)| \leq |f(z_0)|$ on Ω. The second part is a consequence of the first one and the fact that continuous functions on compact domains attain their supremum. □

In the second part of the maximum principle it is essential to assume that Ω is bounded. Indeed, consider $f(z) = e^{e^z}$ on the strip $-\frac{\pi}{2} < \mathrm{Im}(z) < \frac{\pi}{2}$. Then on the boundary lines f is bounded, but it is clearly unbounded on the strip. A positive statement can still be made, provided we exclude rapid growth of the type exhibited by this function. This goes by the name of Phragmen-Lindelöf theorems; see for example Problem 3.7 below.

The following Schwarz lemma serves both as a useful tool as well as a nice illustration of the maximum principle. $\mathbb{D}$ denotes the open unit disk centered at the origin.

Lemma 1.31. *Suppose $f : \mathbb{D} \to \mathbb{D}$ is analytic and $f(0) = 0$. Then $|f(z)| \leq |z|$ for all $z \in \mathbb{D}^*$ with equality if and only if f is a rotation. If f is bijective, from $\mathbb{D}$ onto itself, then f is a rotation.*

Proof. The power series of f around $z = 0$ converges on $\mathbb{D}$ and does not contain a constant term. Consider $g(z) := \frac{f(z)}{z}$, which is therefore analytic on $\mathbb{D}$. Let $0 < r < 1$ and set

$$\max_{|z|=r} |g(z)| =: M_r.$$

Then $\limsup_{r \to 1^-} M_r \leq 1$ whence $|f(z)| \leq |z|$ by the maximum principle. If $|g(z)| = 1$ for some $z \neq 0$, then g is a constant. In other words, $f(z) = e^{i\theta} z$ for some real θ. Now suppose that f is invertible. Then we can apply the same reasoning to f^{-1} which implies that $|f^{-1}(w)| \leq |w|$ which is the same as $|f(z)| \geq |z|$ for all z. So the first part implies that f is a rotation. $\square$

Natural examples for Lemma 1.31 are $f(z) = e^{i\theta} z^k$ where $k \geq 1$ is an integer and $\theta \in \mathbb{R}$. Note that the only case of a bijection is $k = 1$.

In Problem 1.12 the reader will find some extensions of the Schwarz lemma. The main application of the Schwarz lemma is in characterizing all conformal automorphisms of the disk.

Proposition 1.32. *All conformal automorphisms of the disk are given by*

$$\Big\{ \varphi_{a,\theta}(z) := e^{i\theta} \frac{z - a}{1 - z\bar{a}} \;\Big|\; |a| < 1, \theta \in \mathbb{R} \Big\}.$$

Proof. It is elementary to check that $|\varphi_{a,\theta}(z)| = 1$ for all $|z| = 1$ (see the first problem of this chapter). Moreover, $\varphi_{a,\theta}(a) = 0$ so the Möbius transform $\varphi_{a,\theta}$ takes $\mathbb{D}$ to itself. Furthermore, we leave it to the reader to check that all Möbius transformations that preserve the disk are of this form.

Now suppose $\psi \in \mathrm{Aut}(\mathbb{D})$. Then $\psi(a) = 0$ and $\varphi_{a,\theta} \circ \psi^{-1} \in \mathrm{Aut}(\mathbb{D})$ and it preserves the origin. So by the Schwarz lemma, this map is a rotation whence the result follows. $\square$

To conclude our presentation of integration theory, we present Morera's theorem (a kind of converse to Cauchy's theorem).

Theorem 1.33. *Let $f \in C(\Omega)$ and suppose $\mathcal{T}$ is a collection of triangles in Ω which contains all sufficiently small triangles[5] in Ω. If*

$$\oint_{\partial T} f(z)\, dz = 0 \qquad \forall\, T \in \mathcal{T}, \tag{1.18}$$

then $f \in \mathcal{H}(\Omega)$.

[5] This means that every point in Ω has a neighborhood in Ω such that all triangles which lie inside that neighborhood belong to $\mathcal{T}$.

Proof. The idea is to find a local holomorphic primitive of f. Thus, assume $D(0,r) \subset \Omega$ is a small disk and set

$$F(z) := \int_0^z f(\zeta)\, d\zeta = z\int_0^1 f(tz)\, dt$$

for all $|z| < r$. Writing out our assumption (1.18) over each of the three line segments bounding the triangle with vertices $\{0, z, z+h\}$, we conclude for $|z| < r$ and h small that

$$F(z+h) - F(z) = \int_z^{z+h} f(\zeta)\, d\zeta$$

where the integration is along a straight line segment connecting z to $z+h$. Parametrizing this line segment as $z + th$ with $0 \le t \le 1$ we obtain

$$\frac{F(z+h) - F(z)}{h} = \int_0^1 f(z+th)\, dt \to f(z)$$

as $h \to 0$. This shows that $F \in \mathcal{H}(D(0,r))$ and therefore also $F' = f \in \mathcal{H}(D(0,r))$. Hence $f \in \mathcal{H}(\Omega)$ as desired. □

Now suppose $\{f_n\}_{n\ge 1}$ is a sequence of analytic functions on Ω which converges uniformly on compact subsets to some function f. Then f is clearly continuous. Furthermore, using the notation of Theorem 1.33 we see that (1.18) holds since it does for each f_n by Cauchy's theorem; taking limits then yields the same property for f. So f is again analytic. We shall develop this powerful principle in more detail in the next chapter.

A well-known consequence of Morera's theorem is the fact that complex differentiability of a function everywhere in a region implies holomorphy. In other words, the continuity of the derivative need not be imposed in the definition of holomorphic functions. This fact is known as Goursat's theorem.

Theorem 1.34. *Let f be complex differentiable everywhere in a region Ω. Then $f \in \mathcal{H}(\Omega)$.*

Proof. Given a triangle Δ which lies entirely within Ω (with its interior), we define the basic subdivision process as follows: mark the three midpoints of the sides of the triangle and draw a new triangle with these points as vertices. This divides Δ into four congruent subtriangles. Denote them by $\Delta^{(j)}$, $1 \le j \le 4$.

Suppose $\left|\oint_\Delta f(z)\, dz\right| = \delta > 0$. Then for some $1 \le j \le 4$,

$$\left|\oint_{\Delta^{(j)}} f(z)\, dz\right| \ge \frac{\delta}{4}.$$

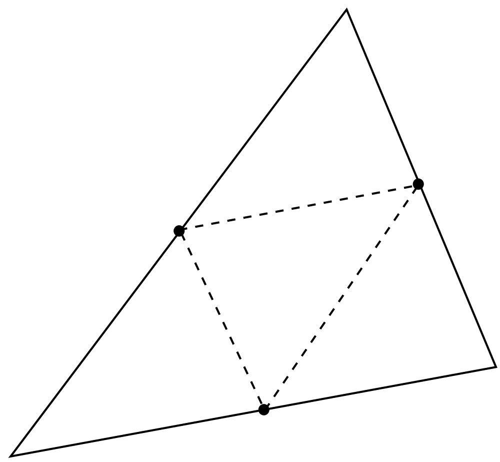

Figure 1.8. The triangle division in Goursat's theorem

Applying the basic subdivision process repeatedly, we see that there exists a sequence Δ_ℓ of nested triangles with $\Delta_0 := \Delta$ such that $\Delta_{\ell+1} \subset \Delta_\ell$ is obtained from Δ_ℓ by a Euclidean motion and a dilation by a factor $\frac{1}{2}$. Moreover,

$$\Big| \oint_{\Delta_\ell} f(z)\, dz \Big| \geq \frac{\delta}{4^\ell} \qquad \forall\, \ell \geq 0. \tag{1.19}$$

To lead this to a contradiction, we denote z_0 the unique point in the intersection:

$$\{z_0\} = \bigcap_\ell \Delta_\ell.$$

By the definition of differentiability,

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + r(z)$$

where for any $\varepsilon > 0$ we have $|r(z)| < \varepsilon |z - z_0|$ provided $|z - z_0|$ is small enough. But since

$$\oint_{\Delta_\ell} (f(z_0) + f'(z_0)(z - z_0))\, dz = 0,$$

we see that

$$\Big| \oint_{\Delta_\ell} f(z)\, dz \Big| \leq L(\Delta_\ell) \max_{z \in \partial\Delta_\ell} |r(z)| \leq \varepsilon 4^{-\ell}$$

for large ℓ. This contradicts (1.19). By Morera's theorem we conclude that f is analytic. $\square$

1.8. Harmonic functions

This section is merely an introduction to this rich area which is explored in more detail in a later chapter.

Definition 1.35. A function $u : \Omega \to \mathbb{C}$ is called harmonic if $u \in C^2(\Omega)$ and $\Delta u = u_{xx} + u_{yy} = 0$.

Typically, harmonic functions are taken to be real-valued but there is no need to make this restriction in general. The following result explains the ubiquity of harmonic functions in complex analysis.

Proposition 1.36. *If $f \in \mathcal{H}(\Omega)$, then* $\mathrm{Re}(u), \mathrm{Im}(v)$ *are harmonic in Ω.*

Proof. First, $u := \mathrm{Re}(f), v := \mathrm{Im}(f) \in C^\infty(\Omega)$ by analyticity of f. Second, by the Cauchy-Riemann equations,

$$u_{xx} + u_{yy} = v_{yx} - v_{xy} = 0, \qquad v_{xx} + v_{yy} = -u_{yx} + u_{xy} = 0,$$

as claimed. □

This motivates the following definition.

Definition 1.37. Let u be harmonic on Ω and real-valued. We say that v is the harmonic conjugate of u if v is harmonic and real-valued on Ω and $u + iv \in \mathcal{H}(\Omega)$.

Let us first note that a harmonic conjugate, if it exists, is unique up to constants; indeed, if not, then we would have a real-valued harmonic function v on Ω such that $iv \in \mathcal{H}(\Omega)$. But from the Cauchy-Riemann equations we would then conclude that $\nabla v = 0$ or v is constant by connectedness of Ω.

This definition presents us with the question of whether every harmonic function on a region of $\mathbb{R}^2$ has a harmonic conjugate. The classical example for the failure of this is $u(z) = \log|z|$ on $\mathbb{C}^*$; in view of the complex logarithm, the unique harmonic conjugate v with $v(1) = 0$ would have to be the polar angle which is not defined on $\mathbb{C}^*$. However, in view of Proposition 1.25 it is defined and harmonic on every simply-connected region in $\mathbb{C}^*$. As the following proposition explains, this is a general fact.

Proposition 1.38. *Let Ω be simply-connected and u real-valued and harmonic on Ω. Then $u = \mathrm{Re}(f)$ for some $f \in \mathcal{H}(\Omega)$ and f is unique up to an additive imaginary constant.*

Proof. We already established the uniqueness property. To obtain existence, we need to solve the Cauchy-Riemann equations. In other words, we need to find a potential v to the vector field $(-u_y, u_x)$ on Ω, i.e., $\nabla v = (-u_y, u_x)$. If v exists, then it is $C^2(\Omega)$ and

$$\Delta v = -u_{yx} + u_{xy} = 0,$$

hence v is harmonic. Define

$$v(z) := \int_{z_0}^{z} -u_y\,dx + u_x\,dy$$

where the line integral is along a curve connecting z_0 to z which consists of finitely many line segments, say. If γ is a closed curve of this type in Ω, then by Green's theorem,

$$\oint_\gamma -u_y\,dx + u_x\,dy = \iint_U (u_{yy} + u_{xx})\,dx\,dy = 0$$

where $\partial U = \gamma$ (this requires Ω to be simply-connected). So the line integral defining v does not depend on the choice of curve and v is therefore well-defined on Ω. Furthermore, as usual one can check that $\nabla v = (-u_y, u_x)$ as desired. A quick but less self-contained proof is as follows: the differential form

$$\omega := -u_y\,dx + u_x\,dy$$

is closed since $d\omega = \Delta u\,dx \wedge dy = 0$. Hence it is locally exact and by simple connectivity of Ω, exact on all of Ω. In other words, $\omega = dv$ for some smooth real-valued function v on Ω as desired. □

From this, we can easily draw several conclusions about harmonic functions. We begin with the important observation that a conformal change of coordinates preserves harmonic functions.

Corollary 1.39. *Let u be harmonic in Ω and $f : \Omega_0 \to \Omega$ holomorphic. Then $u \circ f$ is harmonic in Ω_0.*

Proof. Locally around every point of Ω, there is a v such that $u + iv$ is holomorphic. Since the composition of holomorphic functions is again holomorphic, the statement follows. There is of course a direct way of checking this: since $\Delta = 4\partial_z\partial_{\bar z}$ one has from the chain rule (1.7)

$$\partial_z(u \circ f) = (\partial_w u) \circ f\,\partial_z f + (\partial_{\bar w} u) \circ f\,\overline{\partial_{\bar z} f} = (\partial_w u) \circ f\,f'$$

and thus, furthermore,

$$\Delta(u \circ f) = 4\partial_{\bar z}\partial_z(u \circ f) = 4(\partial_{\bar w}\partial_w u) \circ f |f'|^2 = |f'|^2\,(\Delta u) \circ f$$

whence the result. □

Next, we describe the well-known mean-value and maximum properties of harmonic functions.

Corollary 1.40. *Let u be harmonic on Ω. Then $u \in C^\infty(\Omega)$, u satisfies the mean-value property*

$$u(z_0) = \int_0^1 u(z_0 + re^{2\pi i t})\,dt \qquad \forall\, r < \operatorname{dist}(z_0, \partial\Omega), \tag{1.20}$$

and u obeys the maximum principle: if u attains a local maximum or minimum in Ω, then u is constant. In particular, if Ω is bounded and $u \in C(\overline{\Omega})$, then

$$\min_{\partial\Omega} u \le u(z) \le \max_{\partial\Omega} u \qquad \forall\, z \in \Omega,$$

where equality can be attained only if u is constant.

Proof. Let $U \subset \Omega$ be simply-connected, say a disk. By Proposition 1.38, $u = \mathrm{Re}(f)$ where $f \in \mathcal{H}(U)$. Since $f \in C^\infty(U)$, so is u. Moreover,

$$f(z_0) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z - z_0}\, dz = \int_0^1 f(z_0 + re^{2\pi i t})\, dt.$$

Passing to the real part proves (1.20). For the maximum principle, suppose that u attains a local extremum on some disk in Ω. Then it follows from (1.20) that u has to be constant on that disk. Since any two points in Ω are contained in a simply-connected subregion of Ω, we conclude from the existence of conjugate harmonic functions on simply-connected regions as well as the uniqueness theorem for analytic functions that u is globally constant. $\square$

The mean-value property already characterizes harmonic functions. For this, see the chapter on harmonic functions. It is important to note that harmonic functions are not tied to dimension two. The defining equation $\Delta u = 0$ applies to any dimension. The concept of a conjugate harmonic function of course does not apply in the same form as we saw it here. However, the mean-value property and maximum principles do apply. See the final problem of this chapter.

Notes

All of this material is standard, but the presentation and organization of all the basic results in this chapter, as well as the following one is perhaps unusual. The section on non-Euclidean geometry is meant to illustrate the concept of complex differentiability, which should be seen as a geometric one. For the same reason, we discuss Möbius transformations before complex integration. The Poincaré disk will appear again numerous times in this text, for example in Chapter 4 in the construction of the covering space of the twice punctured plane, but of course also in Chapter 8 as part of the Uniformization Theorem.

For much more on Möbius transformations and how they relate to geometry, group theory, etc., we urge the reader to consult the nice book by Jones and Singerman [**47**]. Later in the text we will study groups of Möbius transforms by the name of Fuchsian groups, which play a crucial role in the classification of non-simply-connected Riemann surfaces. See Chapters 4 and 8, as well as the book by Katok [**50**].

Some texts, such as Lang [**55**], establish the Cauchy theorem without any reference to Green (or Stokes) and the Cauchy-Riemann equations. Rather, one argues directly as in Goursat's theorem; see Theorem 1.34. The author believes, however, that the approach through Cauchy-Riemann is much more fundamental, as it shows that the 1-form $f'(z)\,dz$ being closed is the reason why Cauchy's theorem is true. The extra C^1 condition required in the definition of holomorphic functions is a small price to pay in comparison to the loss of transparency. The $dz, d\bar{z}$ formalism is known as Wirtinger calculus.

1.9. Problems

Problem 1.1. Let $\{z_j\}_{j=1}^n \subset \mathbb{C}$ be distinct points and $m_j > 0$ for $1 \le j \le n$. Assume $\sum_{j=1}^n m_j = 1$ and define $z = \sum_{j=1}^n m_j z_j$. Prove that every line ℓ through z *separates* the points $\{z_j\}_{j=1}^n$ unless all of them are collinear. Here "separates" means that there are points from $\{z_j\}_{j=1}^n$ on both sides of the line ℓ (without being on ℓ).

Problem 1.2. Suppose $p_0 > p_1 > p_2 > \cdots > p_n > 0$. Prove that all zeros of the polynomial $P(z) = \sum_{j=0}^n p_j\, z^j$ lie in $\{|z| > 1\}$.

Problem 1.3. Let $P(z) = \sum_{j=0}^n a_j\, z^j$ be a polynomial of degree $n \ge 1$ with all roots inside the unit circle $|z| < 1$. Define $P^*(z) = z^n \bar{P}(z^{-1})$ where $\bar{P}(z) = \sum_{j=0}^n \bar{a}_j\, z^j$. Show that all roots of

$$P(z) + P^*(z) = 0$$

lie on the unit circle $|z| = 1$. Do the same for $P(z) + e^{i\theta} P^*(z) = 0$, with $\theta \in \mathbb{R}$ arbitrary.

Problem 1.4. In the problem we express simple geometric shapes via relations between complex numbers.

(a) Let $a, b \in \mathbb{C}$ and $k > 0$. Describe the set of points $z \in \mathbb{C}$ which satisfy
$$|z - a| + |z - b| \le k.$$
(b) Let $|a| < 1$, $a \in \mathbb{C}$. The plane $\mathbb{C}$ is divided into three subsets according to whether
$$w = \frac{z - a}{1 - \bar{a}z}$$
satisfies $|w| < 1$, $|w| = 1$, or $|w| > 1$. Describe these sets (in terms of z).

Problem 1.5. Find a Möbius transformation that takes $\{|z - i| < 1\}$ onto $\{|z - 2| < 3\}$. Do the same for $\{|z + i| < 2\}$ onto $\{x + y \ge 2\}$. Is there a Möbius transformation that takes

$$\{|z - i| < 1\} \cap \{|z - 1| < 1\}$$

onto the first quadrant? How about $\{|z-2i|<2\}\cap\{|z-1|<1\}$ and $\{|z-\sqrt{3}|<2\}\cap\{|z+\sqrt{3}|<2\}$ onto the first quadrant?

Problem 1.6. Let $\Phi : S^2 \to \mathbb{C}_\infty$ be the stereographic projection

$$(x_1, x_2, x_3) \mapsto \frac{x_1 + ix_2}{1 - x_3}.$$

(a) Give a detailed proof that Φ is conformal.

(b) Define a metric $d(z,w)$ on $\mathbb{C}_\infty$ as the Euclidean distance of $\Phi^{-1}(z)$ and $\Phi^{-1}(w)$ in $\mathbb{R}^3$. Find a formula for $d(z,w)$. In particular, find $d(z,\infty)$.

(c) Show that circles on S^2 go to circles or lines in $\mathbb{C}$ under Φ.

Problem 1.7. Prove (1.20) without using complex analysis. In other words, use only real-variable methods. Show that your proof carries over to all dimensions, and thus obtain the maximum principle for harmonic functions in all dimensions.

Problem 1.8. Find the holomorphic function $f(z) = f(x+iy)$ with real part

$$\frac{x(1+x^2+y^2)}{1+2x^2-2y^2+(x^2+y^2)^2}$$

and such that $f(0) = 0$.

Problem 1.9. This exercise highlights properties of infinite series of complex numbers, and how they differ from real series:

(a) Suppose $\{z_j\}_{j=1}^\infty \subset \{z \in \mathbb{C} : \operatorname{Re} z \geq 0\}$ is a given sequence. True or false: if both $\sum_{j=1}^\infty z_j$ and $\sum_{j=1}^\infty z_j^2$ converge, then $\sum_{j=1}^\infty |z_j|^2$ also converges.

(b) True or false: there are sequences of complex numbers $\{z_j\}_{j=1}^\infty$ such that for *each* integer $k \geq 1$ the infinite series $\sum_{j=1}^\infty z_j^k$ converges, but fails to converge absolutely.

Problem 1.10. Discuss the mapping properties of $z \mapsto w = \frac{1}{2}(z + z^{-1})$ on $|z| < 1$. Is it one-to-one there? What is the image of $|z|<1$ in the w-plane? What happens on $|z| = 1$ and $|z| > 1$? What is the image of the circles $|z| = r < 1$, and of the ray $\operatorname{Arg} z = \theta$ emanating from zero?

Problem 1.11. Let $T(z) = \frac{az+b}{cz+d}$ be a Möbius transformation.

(a) Show that $T(\mathbb{R}_\infty) = \mathbb{R}_\infty$ if and only if we can choose $a, b, c, d \in \mathbb{R}$.

(b) Find all T such that $T(\mathbb{T}) = \mathbb{T}$, where $\mathbb{T} = \{|z| = 1\}$ is the unit circle.

(c) Find all T for which $T(\mathbb{D}) = \mathbb{D}$, where $\mathbb{D} = \{|z| < 1\}$ is the unit disk.

Problem 1.12. Let $f \in \mathcal{H}(\mathbb{D})$ with $|f(z)| < 1$ for all $z \in \mathbb{D}$.

(a) If $f(0) = 0$, show that $|f(z)| \le |z|$ on $\mathbb{D}$ and $|f'(0)| \le 1$. If $|f(z)| = |z|$ for some $z \neq 0$, or if $|f'(0)| = 1$, then f is a rotation.

(b) Without any assumption on $f(0)$, prove that

$$\left|\frac{f(z_1) - f(z_2)}{1 - \overline{f(z_1)}f(z_2)}\right| \le \frac{|z_1 - z_2|}{|1 - \bar{z}_1 z_2|} \quad \forall\, z_1, z_2 \in \mathbb{D} \tag{1.21}$$

and

$$\frac{|f'(z)|}{1 - |f(z)|^2} \le \frac{1}{1 - |z|^2} \quad \forall\, z \in \mathbb{D}. \tag{1.22}$$

Show that equality in (1.21) for some pair $z_1 \neq z_2$ or in (1.22) for some $z \in \mathbb{D}$ implies that $f(z)$ is a fractional linear transformation.

Problem 1.13. This problem discusses the metric properties of the non-Euclidean geometries on the upper half-plane and the disk as introduced in this chapter. Endow $\mathbb{H}$ with the Riemannian metric

$$ds^2 = \frac{1}{y^2}(dx^2 + dy^2) = \frac{1}{(\operatorname{Im} z)^2} dz\, d\bar{z}.$$

The power in the dominator is no accident: it is unique with the property that the resulting metric is scaling invariant. Similarly, equip $\mathbb{D}$ with the metric

$$ds^2 = \frac{4}{(1 - |z|^2)^2} dz\, d\bar{z}.$$

With this metric, $\mathbb{D}$ is known as Poincaré disk. These Riemannian manifolds, which turn out to be isometric, are models of *hyperbolic space*. By definition, for any two Riemannian manifolds M, N a map $f : M \to N$ is called an *isometry* if it is one-to-one, onto, and preserves the metric.

(a) The distance between any two points z_1, z_2 in hyperbolic space (on either $\mathbb{D}$ or $\mathbb{H}$) is defined as

$$d(z_1, z_2) = \inf_\gamma \int_0^1 \|\gamma'(t)\|\, dt$$

where the infimum is taken over all curves joining z_1 and z_2 and the length of $\dot{\gamma}$ is determined by the hyperbolic metric ds. Show that any holomorphic $f : \mathbb{D} \to \mathbb{D}$ or holomorphic $f : \mathbb{H} \to \mathbb{H}$ satisfies

$$d(f(z_1), f(z_2)) \le d(z_1, z_2)$$

for all z_1, z_2 in hyperbolic space.

(b) Determine all orientation preserving isometries of $\mathbb{H}$ to itself, $\mathbb{D}$ to itself, as well as from $\mathbb{H}$ to $\mathbb{D}$.

(c) Determine all geodesics of hyperbolic space as well as its scalar curvature (we are using the terminology of Riemannian geometry).

Problem 1.14. Let $f \in \mathcal{H}(\mathbb{D})$ be such that $\operatorname{Re} f(z) > 0$ for all $z \in \mathbb{D}$, and $f(0) = a > 0$. Prove that $|f'(0)| \leq 2a$. Is this inequality sharp? If so, which functions attain it?

Problem 1.15. Give another — more elementary — proof of the fundamental theorem of algebra; see (Proposition 1.23) following these lines: Let $p(z)$ be a nonconstant polynomial. Show that $|p(z)|$ attains a minimum in the complex plane, say at z_0. If the polynomial $q(z) = p(z + z_0)$ starts with a nonzero constant term, obtain a contradiction by showing that we may find a small z such that $q(z)$ is nearer to the origin than $q(0)$.

Chapter 4

Riemann surfaces: definitions, examples, basic properties

In this chapter we introduce rigorously the concept of a Riemann surface. Historically, this idea arose naturally in the 19th century in an attempt to understand "multi-valued" analytic functions. The prime examples here are the logarithm, as well as algebraic functions. By the latter we mean the analytic continuations of the roots of a polynomial equation $P(w, z) = 0$ relative to w. We will defer the details of this algebraic construction to the following chapter. This chapter mostly presents various basic examples (such as elliptic functions) and basic general geometric-topological properties which Riemann surfaces possess.

4.1. The basic definitions

We begin with the abstract Riemann surface. Needless to say, the historical development was the exact opposite: examples first and then—after more than sixty years of "naive" but no less successful usage of the concept of a Riemann surface—H. Weyl gave the following definition.

Definition 4.1. A *Riemann surface* is a two-dimensional, connected, Hausdorff topological manifold M with a countable base for the topology and with conformal transition maps between charts. That is, there exists a family of open sets $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ covering M and homeomorphisms $\phi_\alpha : U_\alpha \to V_\alpha$ where

$V_\alpha \subset \mathbb{C}$ is some open set so that

$$\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$$

is biholomorphic (in other words, a conformal homeomorphism). We refer to each (U_α, ϕ_α) as a *chart* and to the collection of all charts as an *atlas* of M.

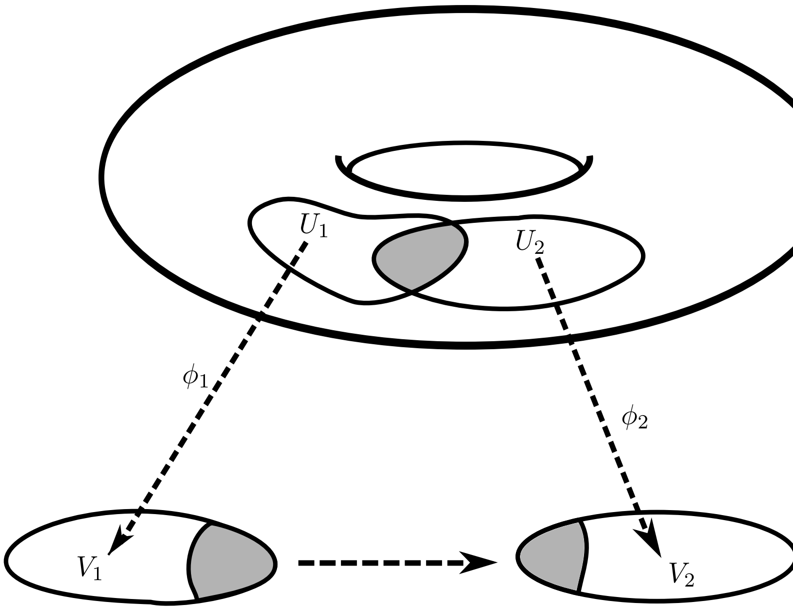

Figure 4.1. Charts and analytic transition maps

Connectivity is a convenient assumption to include. For example, one obtains the uniqueness theorem this way. There is also nothing lost, since we may consider the connected components.

The countability axiom can be dispensed with as it can be shown to follow from the other axioms (this is called Rado's theorem, preceded by the important Poincaré-Volterra theorem), but in all applications it is easy to check directly. Two atlases $\mathcal{A}_1, \mathcal{A}_2$ of M are called equivalent if $\mathcal{A}_1 \cup \mathcal{A}_2$ is an atlas of M. An equivalence class of atlases of M is called a *conformal structure* and a *maximal atlas* of M is the union of all atlases in a conformal structure. We shall often write (U, z) for a chart indicating the fact that $p \mapsto z(p)$ takes U into the complex plane. Moreover, a *parametric disk* is a set $D \subset U$ where (U, z) is a chart with $z(D)$ a Euclidean disk in $\mathbb{C}$. We shall always assume that $\overline{z(D)} \subset z(U)$. By a parametric disk D *centered at* $p \in M$ we mean that (U, z) is a chart with $p \in U$, $z(p) = 0$, and $D = z^{-1}(D(0, r))$ for some $r > 0$.

Throughout this text, we say that the Riemann surface M is an *extension* of the Riemann surface N if $N \subset M$ as an open subset and if the conformal structure of M restricted to N is exactly the conformal structure that N carried to begin with. Note that it follows immediately from the definition that any homeomorphic image of one Riemann surface is another Riemann surface in a natural way: the conformal structure is pushed forward to the target by the homeomorphism.

Definition 4.2. A continuous map $f : M \to N$ between Riemann surfaces is said to be *analytic* if it is analytic in charts, i.e., if $p \in M$ is arbitrary and $p \in U_\alpha$, $f(p) \in V_\beta$ where (U_α, z_α) is a chart of M and (V_β, w_β) is a chart of N, respectively, then $w_\beta \circ f \circ z_\alpha^{-1}$ is analytic where it is defined. We say that f is a *conformal isomorphism* if f is an analytic homeomorphism. If $N = \mathbb{C}$, then one says that f is holomorphic; if $N = \mathbb{C}_\infty$, f is called meromorphic.

We remark that if f is an analytic homeomorphism then the inverse mapping f^{-1} is also analytic. This follows from the fact that this same property holds for maps between open sets in the plane $\mathbb{C}$. It is clear that the meromorphic functions on a Riemann surface form a field. One refers to this field as the *function field*[1] of a surface M.

4.2. Examples and constructions of Riemann surfaces

In this section we discuss a number of examples of Riemann surfaces, some of which will play a formative role in the development of the theory. The first three examples are very basic, and serve to illustrate the main ideas in as simple a context as possible. Examples 4)–7), on the other hand, are more involved and should perhaps only be read after more theory and experience have been acquired. This being said, Examples 4)–7) played a formative role in the development of the theory of Riemann surface. In fact, they each can be seen as an entry point to either the entire theory or a large part of it.

The be specific, Example 4) shows how we may carry out complex analysis on orientable smooth two-dimensional manifolds, such as imbedded compact surfaces in $\mathbb{R}^3$. This insight played a key role during the early stages of Riemann surfaces as it provided a tangible connection with intuitive geometry. The main computational device in this context, which refers to a coordinate-based approach, is the introduction of *isothermal coordinates*.

Example 5) shows how the topological notion of a covering space (see the appendix) arises naturally in Riemann surface theory. In fact, *every non-simply-connected* Riemann surface is obtained from a simply-connected

[1] At least when M is compact, this terminology is commonly used in algebraic geometry.

one by taking the quotient relative to the group of deck transformations. We shall prove this statement in the final chapter of this book which covers the *uniformization theorem*. This is one of the more beautiful and central aspects of the theory; it reveals that up to conformal isomorphisms there are only three simply-connected surfaces: the disk, the plane, and the Riemann sphere.

Example 6) discusses *smooth projective algebraic curves*. These are given by algebraic equations $P(w, z) = 0$ where P is a polynomial. The nonsingular curves are the ones for which the implicit function theorem permits to solve for either z or w. One of the key classical theorems which we establish in this book is that *every compact Riemann surface is obtained from an algebraic equation*. This result has clear connections with Galois theory, in the sense that the polynomial $P(w, z)$ is obtained by thinking in terms of the symmetric polynomials involving roots.

Finally, Example 7) discusses how we may “grow” a Riemann surface from an analytic “germ” (which is nothing other than a convergent power series around a point) by means of all possible analytic continuations along chains of disks, see Figure 2.6. This is precisely where Riemann’s surfaces meet Weierstrass’ *analytisches Gebilde*. We recall in Example 7) what this means. The classical Poincaré-Volterra theorem states that we may obtain at most countably many distinct power series above a fixed base point $z_0 \in \mathbb{C}$ by means of all possibly analytic continuations from a fixed germ at z_0.

In the notes the reader will find more historical comments as well as contemporary references which explain the course of events in detail.

1) *Any open region* $\Omega \subset \mathbb{C}$: Here, a single chart suffices, namely (Ω, z) with z being the identity map on Ω. The associated conformal structure consists of all (U, ϕ) with $U \subset \Omega$ open and $\phi : U \to \mathbb{C}$ biholomorphic. Notice that an alternative, nonequivalent conformal structure is $(\Omega, \bar{z})$ where $\bar{z}$ is the complex conjugation map.

2) *The Riemann sphere* $S^2 \subset \mathbb{R}^3$*, which can be described in three, conformally equivalent, ways*: S^2 (the standard 2-sphere), $\mathbb{C}_\infty$ (the complex plane compactified by a point at infinity), and $\mathbb{C}P^1$ (the complex projective line).

2a) We define a conformal structure on S^2 via two charts

$$(S^2 \setminus (0,0,1), \phi_+), \quad (S^2 \setminus (0,0,-1), \phi_-)$$

where $\phi_\pm$ are the stereographic projections

$$\phi_+(x_1, x_2, x_3) = \frac{x_1 + ix_2}{1 - x_3}, \quad \phi_-(x_1, x_2, x_3) = \frac{x_1 - ix_2}{1 + x_3}$$

from the north, and south pole, respectively; see Figure 1.1 of Chapter 1. If

$$p = (x_1, x_2, x_3) \in S^2$$

with $x_3 \neq \pm 1$, then

$$\phi_+(p)\phi_-(p) = 1.$$

This shows that the transition map between the two charts is $z \mapsto \frac{1}{z}$ from $\mathbb{C}^* \to \mathbb{C}^*$.

2b) The one-point compactification of $\mathbb{C}$ denoted by $\mathbb{C}_\infty := \mathbb{C} \cup \{\infty\}$. The neighborhood base of ∞ in $\mathbb{C}_\infty$ is given by the complements of all compact sets of $\mathbb{C}$. Again there are two charts, namely

$$(\mathbb{C}, z), \quad \bigl(\mathbb{C}_\infty \setminus \{0\}, \frac{1}{z}\bigr)$$

in the obvious sense. The transition map is again given by $z \mapsto \frac{1}{z}$.

2c) The one-dimensional complex projective space

$$\mathbb{C}P^1 := \bigl\{(z, w) \mid (z, w) \in \mathbb{C}^2 \setminus \{(0, 0)\}\bigr\} / \sim$$

where the equivalence relation is $(z_1, w_1) \sim (z_2, w_2)$ if and only if

$$z_2 = \lambda z_1, \ w_2 = \lambda w_1$$

for some $\lambda \in \mathbb{C}^*$. The equivalence class of (z, w) is denoted by $[z : w]$. Our charts are (U_1, ϕ_1) and (U_2, ϕ_2) where

$$\begin{aligned} U_1 &:= \{[z : w] \in \mathbb{C}P^1 \mid w \neq 0\}, \quad \phi_1([z : w]) = \frac{z}{w}, \\ U_2 &:= \{[z : w] \in \mathbb{C}P^1 \mid z \neq 0\}, \quad \phi_2([z : w]) = \frac{w}{z}. \end{aligned}$$

Here, too, the transition map is $z \mapsto \frac{1}{z}$.

To go between $\mathbb{C}P^1$ and $\mathbb{C}_\infty$ we take $[z_1 : z_2] \mapsto \frac{z_1}{z_2}$ where the point at infinity is given by $z_2 = 0$. The stereographic projection takes S^2 onto $\mathbb{C}_\infty$, the north pole being sent to the point at infinity. See also Theorem 4.4.

In the previous chapters, we classified all analytic homeomorphisms of the Riemann sphere: they are exactly the Möbius transforms, in other words $\mathrm{PSL}(2, \mathbb{C})$. We also saw that any such analytic homeomorphism is conformal with a conformal inverse.

3) *Any polyhedral surface $S \subset \mathbb{R}^3$ carries the structure of a Riemann surface*: Such a surface S is defined to be a compact topological manifold which can be written as the finite union of *faces* $\{f_i\}$, *edges* $\{e_j\}$, and *vertices* $\{v_k\}$. Any f_i is assumed to be an open subset of a plane in $\mathbb{R}^3$ with line segments as edges (in other words, a planar open polygon), an edge is an open line segment and a vertex a point in $\mathbb{R}^3$, with the obvious relations between them (the boundaries of faces in $\mathbb{R}^3$ are finite unions of edges and

vertices and the endpoints of the edges are vertices; an edge is where two faces meet etc.).

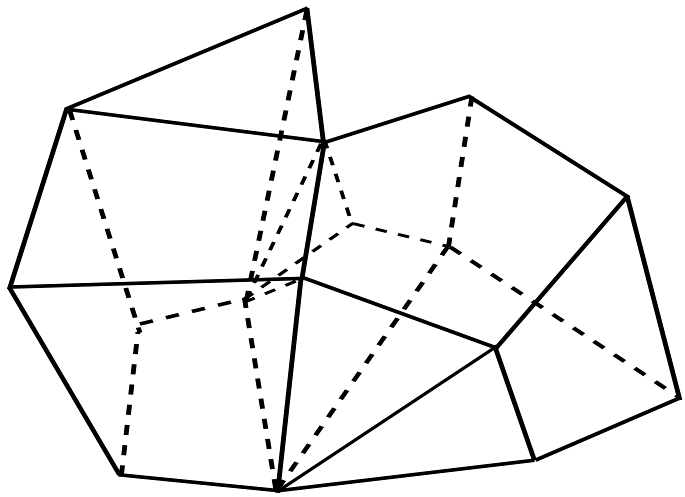

Figure 4.2. Polyhedra are Riemann surfaces

To define a conformal structure on such a polyhedral surface, proceed as follows: each f_i defines a chart (f_i, ϕ_i) where ϕ_i is a Euclidean motion (affine isometry) that takes f_i into $\mathbb{C} = \mathbb{R}^2 \subset \mathbb{R}^3$ where we identify $\mathbb{R}^2$ with the (x_1, x_2)-plane of $\mathbb{R}^3$, say. Each edge e_j defines a chart as follows: let f_{i_1} and f_{i_2} be the two unique faces that meet in e_j. Then $(f_{i_1} \cup f_{i_2} \cup e_j, \phi_j)$ is a chart where ϕ_j is a map that folds $f_{i_1} \cup f_{i_2} \cup e_j$ at the edge so that it becomes piece of a plane, and then maps that plane isometrically into $\mathbb{R}^2$. Finally, at a vertex v_k we define a chart as follows: for the sake of illustration, suppose three faces meet at v_k, say $f_{i_1}, f_{i_2}, f_{i_3}$ with respective angles α_1, α_2, and α_3. Let $\gamma > 0$ be defined so that

$$\gamma \sum \alpha_i = 2\pi, \tag{4.1}$$

and let the chart map these faces together with their edges meeting at v into the plane in such a way that angles get dilated by γ. It is easy to see that this defines a conformal structure (for example, at a vertex, the transition maps are z^γ where γ is as in (4.1)). Note that z^γ is analytic in any sector of opening angle less than 2π and with vertex at the origin.

Note that the complex structure which turns the polyhedral surface into a Riemannian one is completely different from the Euclidean structure inherited from the ambient space.

4) *Any smooth, compact, orientable two-dimensional submanifold of* $\mathbb{R}^3$ *carries the structure of a Riemann surface. In other words, any smooth*

orientable compact surface in $\mathbb{R}^3$ *admits a conformal structure.* There are a number of different ways of thinking about this. One approach is to note that any such surface is homeomorphic to a polyhedral one as in the previous example. In fact, topologically it is equivalent to a sphere with a finite number of handles attached; see the appendix. A conformal structure on the original surface is then obtained by pulling the one on the polygonal surface back.

The condition of orientability is essential: the tangent spaces of any Riemann surface are canonically complex vector spaces of dimension one (follows from conformal change of charts) and the multiplication by i in the tangent space defines the positive orientation. In particular, the Möbius strip does not admit the structure of a Riemann surface. The following lemma makes this precise.

Lemma 4.3. *Let* M *be a Riemann surface. Every tangent space* T_pM *is in a natural way a complex vector space. In particular,* M *is orientable and thus carries a volume form. Moreover, if* $f : M \to N$ *is a* C^1 *map between Riemann surfaces, then* f *is analytic if and only if* $Df(p)$ *is complex linear as a map* $T_pM \to T_{f(p)}N$ *for each* $p \in M$.

Proof. First note that $\sphericalangle(\vec{v}, \vec{w})$ is well-defined in T_pM. We can measure this angle in any chart—because of conformality of the transition maps this does not depend on the choice of chart. The sign of the angle is also well-defined because of the orientation on M. Now let R be a rotation in T_pM by $\frac{\pi}{2}$ in the positive sense. Then we define

$$i\vec{v} := R\vec{v}.$$

It is clear that this turns each T_pM into a complex one-dimensional vector space. To fix an orientation on M, define (v, iv) with $v \in T_pM$, $v \neq 0$, as the positive orientation. Since $f : U \to \mathbb{R}^2$ with $f \in C^1(U)$, $U \subset \mathbb{C}$ open is holomorphic if and only if Df is complex linear, we see via charts that the same property lifts to the Riemann surface case. $\square$

Classically, the observation is simply that a conformal change of coordinates always preserves orientation: As we say in Chapter 1 a conformal map has the property that its (real) differential is a rotation followed by a dilation and thus preserves orientation. Another way of seeing this is that if f is conformal, then $\det(df) = |f'|^2 > 0$ by the Cauchy-Riemann equations whence the (real) differential df preserves orientation.

Historically speaking, the realization that we may carry out complex analysis on orientable two-dimensional surfaces in $\mathbb{R}^3$ (which is another way of saying that they are Riemann surfaces) was the beginning of the entire story. The classical approach involves *isothermal coordinates* to which we now turn. Start from a local surface given by a parametrization $\vec{x} = \vec{x}(u, v)$, where the function $\vec{x} = (x_1, x_2, x_3)$ is smooth with maximal rank, and $(u, v) \subset \Omega \subset \mathbb{R}^2$, is some region in the plane. Then the metric in the tangent space is given by

$$\begin{aligned} ds^2 &= E\,du^2 + 2F\,dudv + G\,dv^2, \\ E(u,v) &= \langle \vec{x}_u(u,v), \vec{x}_u(u,v)\rangle,\ F = \langle \vec{x}_u, \vec{x}_v\rangle,\ G = \langle \vec{x}_v, \vec{x}_v\rangle, \end{aligned} \tag{4.2}$$

with $\langle\cdot,\cdot\rangle$ being the standard Euclidean inner product in $\mathbb{R}^3$. For example, the length of a curve $\vec{\gamma}(t) = \vec{x}(u(t), v(t))$, $a \le t \le b$, is given by

$$L(\vec{\gamma}) = \int_a^b \sqrt{E(u(t),v(t))\,\dot{u}^2(t) + 2F(u(t),v(t))\,\dot{u}(t)\,\dot{v}(t) + G(u(t),v(t))\,\dot{v}(t)^2}\,dt.$$

In a similar fashion we compute angles and areas on the surface in these local coordinates. By positive definiteness of the metric, $W^2 := EG - F^2 > 0$.

Motivated by conformal maps in the plane, we ask if we may change coordinates $u = u(x, y), v = v(x, y)$ such that the metric becomes

$$ds^2 = \mu^2(dx^2 + dy^2)$$

with $\mu = \mu(x, y) > 0$ smooth. Such coordinates are called **isothermal**. To find the differential equations for x, y as functions of u, v, we factor the metric into linear terms:

$$ds^2 = (\omega_1\,du + \omega_2\,dv)(\bar{\omega}_1\,du + \bar{\omega}_2\,dv) = \mu(dx + idy)\,\mu(dx - idy),$$

which implies that we need to solve for

$$\begin{aligned} \omega_1\,du + \omega_2\,dv &= e^{i\alpha}\mu(dx + idy), \\ \bar{\omega}_1\,du + \bar{\omega}_2\,dv &= e^{-i\alpha}\mu(dx - idy). \end{aligned} \tag{4.3}$$

where $\alpha \in \mathbb{R}$. The conditions for ω_1, ω_2 are

$$E = |\omega_1|^2, \quad F = \mathrm{Re}(\omega_1\bar{\omega}_2), \quad G = |\omega_2|^2$$

with solutions

$$\omega_1 = \zeta\sqrt{E}, \quad \omega_2 = \zeta\,\frac{F \pm iW}{\sqrt{E}}, \quad |\zeta| = 1.$$

The choice of ζ then needs to be such that, with $\sigma = \frac{\zeta}{\mu}$,

$$\sigma\,(\omega_1\,du + \omega_2\,dv) = dx + idy = (x_u\,du + x_v\,dv) + i(y_u\,du + y_v\,dv)$$

is a total differential. This means that

$$\sigma\,\omega_1 = x_u + iy_u, \quad \sigma\,\omega_2 = x_v + iy_v,$$

or, eliminating the integrating factor σ,

$$\omega_2(x_u + iy_u) = \omega_1(x_v + iy_v) \quad \text{or} \quad E(x_v + iy_v) = (F + iW)(x_u + iy_u).$$

Solving for y_u, y_v yields

$$y_u = \frac{Fx_u - Ex_v}{W}, \quad y_v = \frac{Gx_u - Fx_v}{W} \tag{4.4}$$

which, upon equating mixed partials, finally gives the following partial differential equation for $x = x(u, v)$:

$$\frac{\partial}{\partial v}\frac{Fx_u - Ex_v}{\sqrt{EG - F^2}} + \frac{\partial}{\partial u}\frac{Fx_v - Gx_u}{\sqrt{EG - F^2}} = 0. \tag{4.5}$$

Similarly, we obtain for $y = y(u, v)$,

$$\frac{\partial}{\partial v}\frac{Fy_u - Ey_v}{\sqrt{EG - F^2}} + \frac{\partial}{\partial u}\frac{Fy_v - Gy_u}{\sqrt{EG - F^2}} = 0. \tag{4.6}$$

Any reader familiar with the Laplace-Beltrami operator on a manifold will recognize (4.5), (4.6) to be precisely of this form (up to a sign); in fact, on a d-dimensional manifold with coordinates $(\xi^1, \xi^2, \dots, \xi^d)$ the Laplacian equation takes the form

$$\Delta f = \frac{\partial}{\partial \xi^i}\Big(\sqrt{g}\, g^{ij}\,\frac{\partial f}{\partial \xi^j}\Big) = 0, \tag{4.7}$$

where $\{g^{ij}\}_{i,j=1}^d = \{g_{ij}\}^{-1}$ is the dual metric and $g = \det(g_{ij})$ the determinant of the metric (but we shall make no use of this fact).

Let us now *assume* that $x = x(u, v)$ is a smooth nondegenerate solution of (4.5), i.e., $dx \neq 0$, on some open set. Then we compute $y = y(u, v)$ from (4.4). This is possible since (4.5) is precisely the integrability condition which guarantees the existence of y. We leave it to reader to check that y is then a solution of (4.6), as well as that the map $(u, v) \mapsto (x, y)$ has maximal rank. In other words, it is a local diffeomorphism.

Now suppose that we had already started with a conformal metric, i.e., $F = 0$ and $E = G$. Then we see that (4.5), (4.6) turn into

$$x_{uu} + x_{vv} = 0, \quad y_{uu} + y_{vv} = 0.$$

In other words, x, y are harmonic in the (u, v)-coordinates. Moreover, (4.4) becomes the standard Cauchy-Riemann system

$$y_u = -x_v, \quad y_v = x_u.$$

This shows us the following: if we were to use the isothermal coordinates (x, y) constructed above to express *any other isothermal coordinates* (ξ, η) as functions of (x, y), then necessarily $\xi + i\eta$ or $\xi - i\eta$ would be **conformal mappings in the plane** relative to $z = x + iy$. In particular, if we choose the isothermal coordinates in agreement with a global orientation, this implies that the changes in coordinates are always analytic.

Or to say this differently, on **any surface** in $\mathbb{R}^3$, at least locally, we may give meaning to analytic and conformal maps. This is a remarkable insight since the entire **local** theory, as developed, for example, in the first three chapters of this book carries over to the surfaces in $\mathbb{R}^3$. "Local" here refers to any description of analytic functions in a small neighborhood of any point. Moreover, orthogonal families of level curves such as those in Section 3.8—which should be interpreted as flow lines of incompressible fluids and their associated equipotential curves—now equally make sense on such surfaces.

Note that we really did not make any use of the ambient space and our analysis carries over to abstract manifolds. So if we accept that we can always find a nondegenerate solution of (4.5), then it is clear that all orientable two-dimensional manifolds admit a Riemann surface structure. Indeed, orientability means precisely that we may choose our isothermal coordinates in agreement with the global sense of orientation. Therefore, they form an atlas on the surface.

However, we have yet to settle the existence question relating to the PDE (4.5). In other words, how can we be sure that there always exists a nondegenerate harmonic function near any point on a surface? If the metric is analytic, i.e., E, F, G are analytic functions of (u, v), then we may for example appeal to the Cauchy-Kowalewskaya theorem which gives many analytic nontrivial solutions. If the metric is not analytic, then we may still solve the elliptic equation (4.5) by other means such as through a variational principle or through the standard weak-solutions approach to elliptic equations (Hilbert space methods). Later we shall rigorously treat a related existence problem in the context of the Hodge theorem; see Chapter 6. For this reason we do not dwell on the existence problem any further.

5) *Covers and universal covers, and their quotients.* This refers to a large class of fundamental objects. We will treat this topic in more detail both, later in this chapter, as well as in Chapters 5 and 8. For the purposes of this list of examples, we limit ourselves to an introduction. We refer the reader to the appendix for the definition of a covering space and its relation to the fundamental group π_1.

$\mathbb{C}\setminus\{0\}$ is covered by the plane $\mathbb{C}$ with covering map $z \mapsto e^z$. On each open horizontal strip of width 2π the map is a bijection. Denote by Γ the subgroup of all Möbius transforms generated by the translation $\gamma : z \mapsto z + 2\pi i$. Then the quotient $\mathbb{C}/\Gamma$ defined by identifying two points of $\mathbb{C}$ if and only if they are mapped onto each other by an element of Γ, is conformally equivalent to $\mathbb{C} \setminus \{0\}$. Hence $\mathbb{C}$ covers $\mathbb{C} \setminus \{0\}$.

Moreover, Γ is precisely the group of deck transformations. This group is isomorphic to $\mathbb{Z}$, which happens to be the fundamental group $\pi_1(\mathbb{C}\setminus\{0\})$. This is no accident, but a *general fact.*

These observations can and should be interpreted in the context of the inverse of the covering map, i.e., the complex logarithm. Thus, we can view $\mathbb{C}$ as the Riemann surface of $\log z$, visualized, however, more easily by means of the usual "infinite helix" sitting above $\mathbb{C}\setminus\{0\}$.

This example suggests something much more general: any subgroup $\Gamma \subset \mathrm{Aut}(\mathbb{C})$ which is "discrete" in a suitable sense leads to a Riemann surface $\mathbb{C}/\Gamma$ covered by $\mathbb{C}$. The *discreteness* here of course refers to the property that we cannot allow arbitrarily close points to be identified under Γ (the technical term is *properly discontinuous group action*). Recall from Chapter 2 that $\mathrm{Aut}(\mathbb{C}) = \{az+b \;:\; a \neq 0\}$; see Proposition 2.10. We leave it to the reader to explore which subgroups have this property. It turns out that the only Riemann surfaces covered by $\mathbb{C}$ are the punctured plane which we just analyzed (which is the same as the twice punctured sphere), as well as tori of genus one (corresponding to subgroups Γ with two generators). We will discuss this latter case in detail later on in this chapter; it is the same as the classical theory of elliptic functions.

Now suppose we remove *two points* from $\mathbb{C}$. Thus, consider the Riemann surface $S := \mathbb{C}\setminus\{0,1\}$; what is its universal cover? A remarkable theorem of Picard from 1879 states that S is covered by the disk; in other words, there exists a surjection $\pi : \mathbb{D} \to \mathbb{C}\setminus\{0,1\}$. We shall prove this result later in this chapter. An immediate application of this fact is the little Picard theorem (see Theorem 4.18): *Every entire function which omits two values is constant.* The exponential function shows that Picard's theorem is sharp: an entire function may omit one value (which for the exponential is 0). Later in this chapter we present a standard construction based on the geometry of the Poincaré disk, and the Riemann mapping theorem, to identify the disk as the universal cover of $\mathbb{C}$ with two or more points removed.

6) *Riemann surfaces defined as smooth (projective) algebraic curves*: Let $P(w,z)$ be an irreducible polynomial such that $dP \neq 0$ on

$$S := \{(z,w) \in \mathbb{C}^2 \mid P(z,w) = 0\}.$$

In other words, $(\partial_z P, \partial_w P)(z,w) \neq (0,0)$ when $P(z,w) = 0$ (such P are called *nonsingular*). By the implicit function theorem, $S \subset \mathbb{C}^2$ is a Riemann surface embedded in $\mathbb{C}^2$, called an affine *algebraic curve.* To define the complex structure on S, one can use either z or w as local coordinates depending on whether $\partial_z P \neq 0$ or $\partial_w P \neq 0$ on that neighborhood. The irreducibility of P implies that S is connected. By construction, any function of the form $\frac{f(z)}{g(w)}$ or $\frac{f(w)}{g(z)}$ where f, g are meromorphic on $\mathbb{C}$ and g is not identically zero, is a meromorphic function on S.

The Riemann sphere is obtained from the complex plane by compactification. Here, too, we wish to compactify the affine algebraic curve S. As in Example 2) above, we may carry out a projective construction.

To be specific, pass to the homogeneous version of P; thus, let $\nu \geq 1$ be the minimal integer for which

$$u^\nu P(z/u, w/u) =: Q(z, w, u) \tag{4.8}$$

has no negative powers of u. Then

$$\widetilde{S} := \{[z : w : u] \in \mathbb{C}P^2 \mid Q(z, w, u) = 0\} \tag{4.9}$$

is well-defined. Assuming that Q is nonsingular, i.e., $dQ \neq 0$ on $\widetilde{S}$, it follows just as before that $\widetilde{S}$ is a Riemann surface which is compact as a closed subset of the compact space $\mathbb{C}P^2$. $\widetilde{S}$ is called a *smooth projective algebraic curve*, whereas $S = \widetilde{S} \cap \{[z : w : 1] \,:\, (z, w) \in \mathbb{C}^2\}$ is called the *affine part* of $\widetilde{S}$. To be more precise, we use the three charts

$$\{[1 : w : u] \mid (w, u) \in \mathbb{C}^2\},\ \{[z : w : 1] \mid (z, w) \in \mathbb{C}^2\},\ \{[z : 1 : u] \mid (z, u) \in \mathbb{C}^2\}$$

to cover $\mathbb{C}P^2$. Differentiating the homogeneity equation $Q(\lambda z, \lambda w, \lambda u) = \lambda^\nu Q(z, w, u)$ in λ at $\lambda = 1$ one obtains Euler's relation

$$\begin{aligned} dQ(z, w, u)(z, w, u) &:= Q_z(z, w, u)z + Q_w(z, w, u)w + Q_u(z, w, u)u \\ &= \nu Q(z, w, u). \end{aligned}$$

In particular, on $\widetilde{S}$ we infer that $dQ(z, w, u)(z, w, u) = 0$. If we set any one of the coordinates equal to 1, then the polynomial in the two remaining variables is nonsingular in the affine sense (see above); otherwise, we violate $dQ \neq 0$ on $\widetilde{S}$. This allows us to define complex structures on $\widetilde{S}$ over each of the three projective charts which are of course compatible with each other. The meromorphic functions

$$\frac{f(z)}{g(w)} \quad \text{and} \quad \frac{f(w)}{g(z)}$$

extend to meromorphic functions on $\widetilde{S}$ provided f, g are rational. Being compact, $\widetilde{S}$ has finite genus.

An example of the affine part of a curve of genus $g \geq 1$ is given by

$$w^2 - \prod_{j=1}^{N} (z - z_j) = 0 \tag{4.10}$$

where $\{z_j\}_{j=1}^N \subset \mathbb{C}$ are distinct and $N = 2g+1$ or $N = 2g+2$. For any $z_0 \in \mathbb{C} \setminus \{z_j\}_{j=1}^N$ one has local coordinates

$$w(z) = \pm\sqrt{\prod_{j=1}^{N}(z - z_j)}$$

where the two signs correspond precisely to the two "sheets" locally near z_0; note that the square root is analytic. Near any z_ℓ, $1 \le \ell \le N$ one sets $z := z_\ell + \zeta^2$ so that

$$w(\zeta) = \zeta\sqrt{\prod_{j=1,\, j\neq\ell}^{N}(z_\ell - z_j + \zeta^2)}$$

where the ambiguity of the choice of sign can be absorbed into ζ (the square root is again analytic). It is clear that the transition maps between the charts are holomorphic. The reader will easily verify that the projective version of (4.10) with $N \ge 2$, i.e.,

$$Q(z, w, u) := w^2 u^{N-2} - \prod_{j=1}^{N}(z - uz_j),$$

is nonsingular for $N = 2, 3$ but **singular** when $N \ge 4$. Indeed, if $N \ge 4$ one has $dQ = 0$ "at the point at infinity" of $\widetilde{S}$ (which means for all points $u = 0$). Indeed, let $N \ge 3$. Then $u = 0$ means that

$$\widetilde{S} \cap \{[z : w : 0] \in \mathbb{C}P^2\} = \{[0 : w : 0] \in \mathbb{C}P^2\}.$$

On the other hand, on $\widetilde{S}$,

$$dQ(z, w, 0) = w^2\, du \neq 0$$

for $N = 3$, but $dQ = 0$ for $N = 4$.

Returning to the affine curve, one refers to the z_j as *branch points*, since the natural projection $w \mapsto z$, which is a covering map on $\mathbb{C}_\infty \setminus \{z_j\}_{j=1}^N$, ceases to be a covering map near each z_j. Rather, one refers to this case as a *branched covering map* and any algebraic curve is a *branched cover* of the Riemann sphere. The specific algebraic curves we just considered are called **elliptic curves** if $g = 1$ or **hyper-elliptic curves** if $g > 1$. We shall return to these important examples in the following chapters, where the reader will also find figures illustrating how the genus is determined through cuts in the extended complex plane.

A most remarkable fact in the theory of Riemann surfaces is this: *any compact Riemann surface is conformally equivalent to an algebraic curve defined by some irreducible polynomial* $P \in \mathbb{C}[z, w]$; see Chapter 5. For that construction, we no longer assume that P is nonsingular and therefore proceed differently with our construction of the Riemann surface of P; it is defined via all possible analytic continuations of a locally defined analytic solution $w = w(z)$ of $P(z, w(z)) = 0$.

If we accept this fact for now, then we are presented with a basic question which provided a major impetus in the development of complex function theory as well as analysis as a whole: *How do we find any nontrivial global meromorphic functions on a hyper-elliptic curve, or for that matter, any Riemann surface?*

The aforementioned vanishing $dQ = 0$ at infinity (i.e., $u = 0$) where Q is the homogeneous version of (4.10) provided $N \geq 4$, precludes us from using the naive construction $f(z)/g(w)$ and $f(w)/g(z)$. Recall that $N \geq 4$ means that the genus of the curve satisfies $g > 1$. In the *nonsingular* case we could simply use these functions. But this only makes sense if we may solve for one variable in terms of the other. This precisely fails in the singular case.

However, the reader should not be misled into thinking of $u = 0$ as being a genuine singularity, or there being some sort of a "problem" with this Riemann surface. This would be absurd, since the behavior of (4.10) at $z = \infty$ is clearly that of $w^2 - z^N = 0$ which does not present any problem whatsoever: if N is even, then there is no branching at $z = \infty$, otherwise there is. In Chapter 5 we shall develop the "affine construction" of the Riemann surfaces systematically, which is based on analytic continuation.

A much more serious mystery surrounding the question on the existence of meromorphic functions has to do with the abundance of Riemann surfaces that **a priori** have nothing to do with concrete algebraic or analytical expressions, as evidenced by the classes 3) and 4), above, of surfaces. In fact, a branched cover of S^2 of any genus *can be constructed by purely topological methods.* For any reader versed in basic algebraic topology, one possible construction proceeds as follows: fix three distinct points $p_1, p_2, p_3 \in S^2$. Then $\pi_1(S^2 \setminus \{p_1, p_2, p_3\})$ is isomorphic to the free group with two generators, F_2. Construct a homomorphism $\pi_1(S^2 \setminus \{p_1, p_2, p_3\}) \to \mathbb{Z}/k\mathbb{Z}$ where k is odd. Gluing small disks into the sphere around p_1, p_2, p_3 then shows that the Euler characteristic is $3 - k = 2(1 - g)$. From this sketch it is evident that the branched cover which we have constructed is in fact a compact Riemann surface.

As we shall see in our development of Hodge theory in later chapters, one does need a fair amount of 20th century mathematics to answer this problem of the 19th century, namely Hilbert spaces applied to existence problems in potential theory. In this respect, there are some similarities to the resolution of the Dirichlet problem and the "naive" usage of the Dirichlet principle in the 19th century; see the introduction to Chapter 3. We shall also see that the resolution of this question lies at the heart of much of the classical Riemann surface theory.

7) *Riemann surfaces defined via analytic continuation of an analytic germ*: Let $\mathcal{P}$ be the set of all power series

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n$$

around some point $z_0 \in C$ and with radius of convergence $0 < r_0 \le \infty$. We denote elements of $\mathcal{P}$ by $y(z_0, r_0)(z)$ or simply by $y(z_0, r_0)$. We say that $y(z_1, r_1)$ is a *direct continuation* of $y(z_0, r_0)$ if $|z_1 - z_0| < r_0$ and $y(z_0, r_0)(z) = y(z_1, r_1)(z)$ for all z in the intersections of the respective disks of convergence, i.e.,

$$|z - z_0| < r_0, \quad |z - z_1| < r_1.$$

By the uniqueness theorem from Chapter 1 (see Proposition 1.26), for every $|z_1 - z_0| < r_0$ there exists exactly one direct continuation of $y(z_0, r_0)$. We say that $y(\zeta, \rho)$ can be joined to $y(z_0, r_0)$ by a *chain* if there exist finitely many elements $y(z_k, r_k) \in \mathcal{P}$, $0 \le k \le N$ so that $z_n = \zeta$, $r_N = \rho$ and such that $y(z_{k+1}, r_{k+1})$ is a direct continuation of $y(z_k, r_k)$ for each $0 \le k < N$ (see Figure 2.6). This establishes an equivalence relation in $\mathcal{P}$. The equivalence classes are called *analytic functions*. This is Weierstrass' notion of *analytisches Gebilde*.

Every analytic function is completely determined by a single power series $y(z_0, r_0)$. Denote by $S(a)$ all distinct power series in $z - a$ which are equivalent to $y(z_0, r_0)$. The theorem of Poincaré-Volterra asserts that this set is countable. See the notes for more history of this theorem.

A typical example of this process would be $f(z) = \log z$, some branch of the logarithm, defined on the disk $\{|z-1| < 1\}$. We can analytically continue this branch to some neighborhood of any point other than the origin. In this way, we obtain the full complex logarithm. In Chapter 5 we will see that one can put a complex structure on an equivalence class of power series as defined above which renders it a Riemann surface.

4.3. Functions on Riemann surfaces

We already defined analytic functions between Riemann surfaces and also introduced the concept of a conformal isomorphism. Note that any conformal isomorphism has a conformal inverse. In Example 1) above, the Riemann surfaces with conformal structures induced by (Ω, z) and $(\Omega, \bar{z})$, respectively, do not have equivalent conformal structures but are conformally isomorphic (via $z \mapsto \bar{z}$). For the sake of completeness, we now demonstrate the conformal equivalence of the three models of the Riemann sphere. As usual,

$$\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\{\pm \mathrm{Id}\}.$$

Theorem 4.4. *The Riemann surfaces $S^2, \mathbb{C}_\infty$, and $\mathbb{C}P^1$ are conformally isomorphic. Furthermore, the group of automorphisms of these surfaces is* $\mathrm{PSL}(2,\mathbb{C})$.

Proof. As already noted before, the maps are straightforward: stereographic projection takes S^2 onto $\mathbb{C}_\infty$, and the quotient $[z_1 : z_2] \mapsto \frac{z_1}{z_2}$ equates $\mathbb{C}P^1$ with $\mathbb{C}_\infty$. The analyticity of these maps is immediate from the definition of the complex structures on these Riemann surfaces. In fact, those structures are defined precisely with these identifications in mind.

As for the automorphism group, each

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{C})$$

defines an automorphism of $\mathbb{C}P^1$ via the Möbius transformation

$$[z : w] \mapsto [az + bw : cz + dw].$$

Note that A and $-A$ define the same map. On the other hand, if f is an automorphism of $\mathbb{C}_\infty$, then composing with a Möbius transformation we may assume that $f(\infty) = \infty$. Indeed, the Möbius transformations act transitively on $\mathbb{C}P^1$. Hence restricting f to $\mathbb{C}$ yields a map from $\mathrm{Aut}(\mathbb{C})$ which is of the form $f(z) = az + b$ by Proposition 2.10, and we are done. $\square$

We now state the uniqueness and open mapping theorems for analytic functions on Riemann surfaces. These are of course the analogues of the "local" theorems we encountered in the first two chapters of the book. In this section, M, N will denote Riemann surfaces.

Theorem 4.5 (Uniqueness theorem). *Let $f, g : M \to N$ be analytic mappings. Then either $f = g$ identically, or $\{p \in M \mid f(p) = g(p)\}$ is discrete in M.*

Proof. Define

$$A := \{p \in M \mid \text{locally at } p,\ f \text{ and } g \text{ are identically equal}\},$$
$$B := \{p \in M \mid \text{locally at } p,\ f \text{ and } g \text{ agree only on a discrete set}\}.$$

As usual, "locally" means in a chart of a maximal atlas. It is clear that both A and B are open subsets of M. We claim that $M = A \cup B$ which then finishes the proof since M is connected. If $p \in M$ is such that $f(p) \neq g(p)$, then by continuity $p \in B$. Suppose, on the other hand, that $f(p) = g(p)$. If

$\{f = g\}$ is not discrete, then we apply the standard uniqueness theorem in charts to conclude that $f = g$ locally around p. $\square$

As an obvious corollary, note that for any analytic map $f : M \to N$ each "level set" $\{f \in M \mid f(p) = q\}$ with $q \in N$ fixed, is either discrete or all of M (and thus f is constant). In particular, if M is compact and f is not constant, then $\{p \in M \mid f(p) = q\}$ is *finite*.

Theorem 4.6 (Open mapping theorem)**.** *Let $f : M \to N$ be an analytic map. If f is not constant, then $f(M)$ is an open subset of N. More generally, f takes open subsets of M to open subsets of N.*

Proof. By the uniqueness theorem, if f is locally constant around any point, then f is globally constant. Hence we can apply the usual open mapping theorem (see Proposition 1.27) in every chart to conclude that $f(M) \subset N$ is open. $\square$

Corollary 4.7. *Let M be a compact Riemann surface and $f : M \to N$ an analytic nonconstant map. Then f is onto and N is compact.*

Proof. First, $f(M)$ is closed, since it is a compact set and since N is a Hausdorff space. Second, $f(M)$ is open by Theorem 4.6. By connectivity of N it follows that $f(M) = N$ as claimed. $\square$

Recall from Definition 4.2 that the *holomorphic* functions on a Riemann surface M, denoted by $\mathcal{H}(M)$, are defined as all analytic $f : M \to \mathbb{C}$. The *meromorphic* functions on M, denoted by $\mathcal{M}(M)$, are defined as all analytic $f : M \to \mathbb{C}_\infty$.

In view of the preceding, the following statements are immediate.

Corollary 4.8. *Let M be a Riemann surface. Then the following properties hold:*

i) if M is compact, then every holomorphic function on M is constant.

ii) Every nonconstant meromorphic function on a compact Riemann surface is onto $\mathbb{C}_\infty$.

iii) If f is a nonconstant holomorphic function on a Riemann surface M, then $|f|$ attains neither a local maximum nor a positive local minimum on M.

To illustrate what we have accomplished so far, let us give a "topological proof" of Liouville's theorem: assume that $f \in \mathcal{H}(\mathbb{C}) \cap L^\infty(\mathbb{C})$. Then $f(1/z)$ has a removable singularity at $z = 0$. In other words, $f \in \mathcal{H}(\mathbb{C}_\infty)$ and is therefore constant. The analytical ingredient in this proof consists of the uniqueness and open mapping theorems as well as the removability theorem: the first two are reduced to the same properties in charts which then require power series expansions. But instead of using expansions that converge on all of $\mathbb{C}$ and Cauchy's estimate we relied on connectivity to pass from a local property to a global one.

It is a good exercise at this point to verify the following assertion: the meromorphic functions on $\Omega \subset \mathbb{C}$ in the sense of standard complex analysis coincide exactly with $\mathcal{M}(\Omega) \setminus \{\infty\}$ in the sense of Definition 4.2 up to the function which is constant equal to infinity. In particular,

$$\mathcal{M}(\mathbb{C}_\infty) = \Big\{ \frac{P}{Q} \mid P, Q \in \mathbb{C}[z],\ Q \not\equiv 0 \Big\} \cup \{\infty\}.$$

In other words, the meromorphic functions on $\mathbb{C}_\infty$ up to the function which is constant equal to ∞, are exactly the rational functions.

Note that we may prescribe the location of the finitely many zeros and poles of any meromorphic function $f \in \mathcal{M}(\mathbb{C}P^1)$ arbitrarily provided the combined order of the zeros exactly equals the combined order of the poles and provided the set of zeros is distinct from the set of poles.

4.4. Degree and genus

We shall now use the "normal form" from Chapter 1 (see Proposition 1.27) to define the valency of an analytic map.

Definition 4.9. Let $f : M \to N$ be an analytic and nonconstant map between Riemann surfaces. Then the *valency* of f at $p \in M$, denoted by $\nu_f(p)$, is defined to be the unique positive integer n with the property that in charts (U, ϕ) around p (with $\phi(p) = 0$) and (V, ψ) around $f(p)$ (with $\psi(f(p)) = 0$) we have $(\psi \circ f \circ \phi^{-1})(z) = (zh(z))^n$ where $h(0) \neq 0$. If M is compact, then the degree of f at $q \in N$ is defined as

$$\deg_f(q) := \sum_{p: f(p)=q} \nu_f(p)$$

which is a positive integer.

Locally around any point $p \in M$ with valency $\nu_f(p) = n \geq 1$ the map f is n-to-one; in fact, every point q' close to but not equal to $q = f(p)$ has exactly n pre-images close to p.

Let $f = \frac{P}{Q}$ be a nonconstant rational function on $\mathbb{C}_\infty$ represented by a reduced fraction (i.e., P and Q are relatively prime). Then for every $q \in \mathbb{C}_\infty$, the reader will easily verify that $\deg_f(q) = \max(\deg(Q), \deg(P))$ where the degree of P, Q is in the sense of polynomials. It is a general fact that $\deg_f(q)$ does not depend on $q \in N$.

Lemma 4.10. *Let $f : M \to N$ be an analytic and nonconstant map between two compact Riemann surfaces M and N. Then* $\deg_f(q)$ *does not depend on q. It is called* the degree *of f and is denoted by* $\deg(f)$. *The isomorphisms* *from M to N are precisely those nonconstant analytic maps f from M to N with* $\deg(f) = 1$.

Proof. We recall that, by Corollary 4.7, if M is a compact Riemann surface and $f : M \to N$ is as in the statement, then N is compact and $f(M) = N$. We shall prove that $\deg_f(q)$ is locally constant. Let $f(p) = q$ and suppose that $\nu_f(p) = 1$. As remarked before, f is then an isomorphism from a neighborhood of p onto a neighborhood of q. If, on the other hand, $n = \nu_f(p) > 1$, then each q' close but not equal to q has exactly n pre-images $\{p'_j\}_{j=1}^n$ and $\nu_f(p'_j) = 1$ at each $1 \leq j \leq n$. This proves that $\deg_f(q)$ is locally constant and therefore globally constant by connectivity of N. The statement concerning isomorphisms is evident. □

We remark that this notion of degree coincides with the one associated with general differentiable manifolds; see the appendix. Let us now prove the Riemann-Hurwitz formula for *branched covers*. The latter notion refers to any analytic nonconstant map $f : M \to N$ from a compact Riemann surface M onto another compact Riemann surface N. If necessary, the reader might wish to review the Euler characteristic and the genus on compact surfaces; see the appendix.

Theorem 4.11 (Riemann-Hurwitz formula)**.** *Let $f : M \to N$ be an analytic nonconstant map between compact Riemann surfaces. Define the* total branching number *to be*

$$B := \sum_{p \in M} (\nu_f(p) - 1).$$

Then

$$g_M - 1 = \deg(f)(g_N - 1) + \frac{1}{2}B \tag{4.11}$$

where g_M and g_N are the genera of M and N, respectively. In particular, B is always an even nonnegative integer.

Proof. Denote by $\mathcal{B}$ the set of all $p \in M$ with $\nu_f(p) > 1$ (the branch points). We shall now use a theorem of Radó which asserts that every Riemann surface admits a triangulation (see Chapter 6, in particular, Problem 6.5). Thus let $\mathcal{T}$ be a triangulation of N such that all $f(p)$, $p \in \mathcal{B}$ are vertices of $\mathcal{T}$. Lift $\mathcal{T}$ to a triangulation $\widetilde{\mathcal{T}}$ on M. If $\mathcal{T}$ has V vertices, E edges and F faces, then $\widetilde{\mathcal{T}}$ has $nV - B$ vertices, nE edges, and nF faces where $n = \deg(f)$. Therefore, by the Euler-Poincaré formula (A.1),

$$\begin{aligned} 2(1 - g_N) &= V - E + F, \\ 2(1 - g_M) &= nV - B - nE + nF = 2n(1 - g_N) - B, \end{aligned}$$

as claimed. □

The Riemann-Hurwitz formula is very useful in the computations of genera of various Riemann surfaces, such as for the algebraic curves defined in Example 6) of Section 4.2. We shall apply it on several occasions in the remainder of this text.

4.5. Riemann surfaces as quotients

Many Riemann surfaces M are generated as quotients of other surfaces N modulo an equivalence relation, i.e., $M = N/\sim$. We have already encountered several instances of this. A common way of defining the equivalence relation is via the action of a subgroup $G < \mathrm{Aut}(N)$. In this case, $q_1 \sim q_2$ in N if and only if there exists some $g \in G$ with $gq_1 = q_2$. Let us state a theorem to this effect where $N = \mathbb{C}_\infty$. Examples will follow immediately after the theorem.

Recall that a covering map is a local homeomorphism. For example, $\exp : \mathbb{C} \to \mathbb{C}^*$ is a covering map, as is $z^n : \mathbb{C}^* \to \mathbb{C}^*$ for each $n \geq 1$. Note that if $n \geq 2$, then the latter example does not extend to a covering map $\mathbb{C} \to \mathbb{C}$; rather, we encounter a branch point at zero and this extension is then referred to as a *branched cover*. Since we have defined the notion of branched cover only for compact surfaces we consider z^n as a map $\mathbb{C}P^1 \to \mathbb{C}P^1$.

Theorem 4.12. *Let $\Omega \subset \mathbb{C}_\infty$ and $G < \mathrm{Aut}(\mathbb{C}_\infty)$ with the property that*

- *$g(\Omega) \subset \Omega$ for all $g \in G$,*
- *for all $g \in G$, $g \neq \mathrm{id}$, all fixed points of g in $\mathbb{C}_\infty$ lie outside of Ω,*
- *for all $K \subset \Omega$ compact, the cardinality of $\{g \in G \mid g(K) \cap K \neq \emptyset\}$ is finite.*

Under these assumptions, the natural projection $\pi : \Omega \to \Omega/G$ is a covering map which turns Ω/G canonically into a Riemann surface.

Proof. By definition, the topology on Ω/G is the coarsest one that makes π continuous. In this case, π is also open; indeed, for every open set $A \subset \Omega$,

$$\pi^{-1}(\pi(A)) = \bigcup_{g \in G} g(A)$$

is open since $g(A)$ is open. Next, let us verify that the topology is Hausdorff. Suppose $\pi(z_1) \neq \pi(z_2)$ and define for all $n \geq 1$,

$$A_n := \left\{z \in \Omega \,\middle|\, |z - z_1| < \frac{r}{n}\right\} \subset \Omega,$$
$$B_n := \left\{z \in \Omega \,\middle|\, |z - z_2| < \frac{r}{n}\right\} \subset \Omega,$$

where $r > 0$ is sufficiently small. Define $K := \overline{A}_1 \cup \overline{B}_1$ and suppose that $\pi(A_n) \cap \pi(B_n) \neq \emptyset$ for all $n \geq 1$. Then for some $a_n \in A_n$ and $g_n \in G$ we have

$$g_n(a_n) \in B_n \qquad \forall\, n \geq 1.$$

Since, in particular, $g_n(K) \cap K \neq \emptyset$, we see that there are only finitely many possibilities for g_n and one of them therefore occurs infinitely often. Let us say that $g_n = g \in G$ for infinitely many n. Passing to the limit $n \to \infty$ implies that $g(z_1) = z_2$ or $\pi(z_1) = \pi(z_2)$, a contradiction. For all $z \in \Omega$ we can find a small pre-compact open neighborhood of z denoted by $K_z \subset \Omega$, so that

$$g(\overline{K_z}) \cap \overline{K_z} = \emptyset \qquad \forall\, g \in G,\ g \neq \mathrm{id}. \tag{4.12}$$

Note that at this point we require all three assumptions in the statement of the theorem. Then $\pi : K_z \to K_z$ is the identity and therefore we can use the K_z as charts. Note that the transition maps are given by $g \in \mathrm{Aut}(\mathbb{C}_\infty)$, which are Möbius transformations, and are therefore holomorphic. Finally, $\pi^{-1}(K_z) = \bigcup_{g \in G} g^{-1}(K_z)$ with pairwise disjoint open sets $g^{-1}(K_z)$. The disjointness follows from (4.12) and we are done. $\square$

We remark that any group G as in the theorem is necessarily *discrete* in the topological sense. First, we remark that $G < \mathrm{Aut}(\mathbb{C}) = \mathrm{PSL}(2, \mathbb{C})$ carries the natural topology; second, if G is not discrete, then the third requirement in the theorem will fail (since we can find group elements in G as close to the identity as we wish). There are many natural examples to which this theorem applies. In what follows, we use $\langle g_1, g_2, \ldots, g_k\rangle$ to denote the group generated by these k elements and, as usual, $\mathbb{H}$ is the upper half-plane.

1) The punctured plane and disk: the cylinder $\mathbb{C}/\langle z \mapsto z+1\rangle$ satisfies $\mathbb{C}/\langle z \mapsto z+1\rangle \simeq \mathbb{C}^*$ where the isomorphism is given by the exponential map $e^{2\pi i z}$. Here $\Omega = \mathbb{C}$, and $G = \langle z \mapsto z+1\rangle$. Similarly, $\mathbb{H}/\langle z \mapsto z+1\rangle \simeq \mathbb{D}^*$.

2) The tori: Let $\omega_1, \omega_2 \in \mathbb{C}^*$ be linearly independent over $\mathbb{R}$. Then

$$\mathbb{C}/\langle z \mapsto z + \omega_1, z \mapsto z + \omega_2\rangle$$

is a Riemann surface. It is the same as $\mathbb{C}/\Lambda$ with the lattice

$$\Lambda = \{n\omega_1 + m\omega_2 \mid n, m \in \mathbb{Z}\}. \tag{4.13}$$

In Figure 4.3 the lattice is generated by any distinct pair of vectors from $\{\omega_1, \omega_2, \omega_3\}$, showing that a pair of generators, or *basis*, is not unique. Furthermore, up to conformal equivalence of $\mathbb{C}/\Lambda$ we may always assume that $\omega_1 = 1$ and $\omega_2 = \tau$ where $\mathrm{Im}(\tau) > 0$.

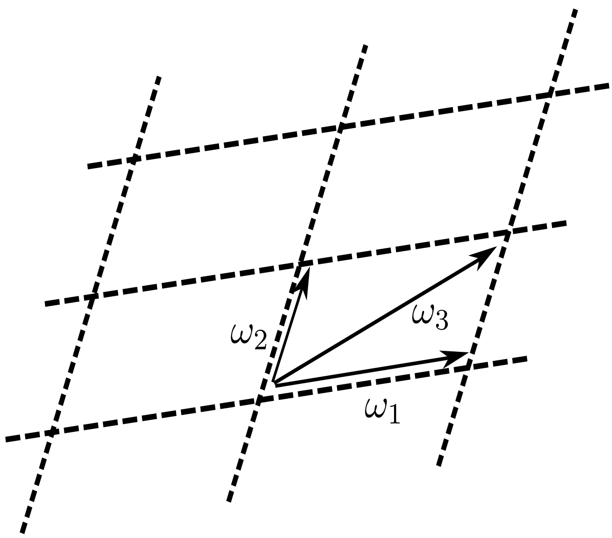

Figure 4.3. A lattice in $\mathbb{C}$

We may identify the surfaces

$$\mathbb{C}^*/\langle z \mapsto \lambda z\rangle \simeq \mathbb{C}/\langle z \mapsto z+1, z \mapsto z + \frac{1}{2\pi i}\log\lambda\rangle$$

where $\lambda > 1$. The surface on the left-hand side naturally has the doughnut or inner-tube shape that one associates with a torus of genus 1. The same exponential map induces the isomorphism in this case as in 1). A natural question is to determine the *space of all conformal equivalence classes of tori*. For this, see Section 4.8.

3) The annuli: consider $\mathbb{H}/\langle z \mapsto \lambda z\rangle$ with $\lambda > 1$. Then $\log z$ maps this quotient onto

$$\{w \in \mathbb{C} \,:\, 0 < \operatorname{Im} w < \pi,\ 0 \le \operatorname{Re} w \le \log\lambda\}$$

with the sides $\operatorname{Re} w = 0, \operatorname{Re} w = \log\lambda$ identified. Next, send this via the conformal map

$$w \mapsto \exp\big(2\pi i \frac{w}{\log\lambda}\big)$$

onto the annulus $\Delta_r := \{r < |z| < 1\}$ where $\log r = -\frac{2\pi^2}{\log\lambda}$.

It is well-known that there is no conformal isomorphism between Δ_r and Δ_s if $0 \le r < s < 1$. The reader is invited to try to establish this by elementary means, although this is perhaps somewhat tricky. Hence the space of conformal equivalence classes of annuli $\{z \in \mathbb{C}^* \,:\, r_1 < |z| < r_2\}$ with $0 < r_1 < r_2$ is the same as all $\frac{r_2}{r_1}$, i.e., $(1, \infty)$.

It is simple to prove the nonequivalence for the special case $r = 0$. Thus, assume φ is a conformal homeomorphism from $\Delta_0 = \mathbb{D}^*$ onto Δ_r for some $r > 0$. Then $z = 0$ is a removable singularity of φ since it remains uniformly bounded as we approach this point. Moreover, $\varphi'(0) \neq 0$ since otherwise φ would be n-to-1 near $z = 0$ for some $n \geq 2$. So near $z = 0$ the extended map φ is an isomorphism.

Clearly, $w_0 := \varphi(0) \in \bar{\Delta}_r$. If $\varphi(0) \in \Delta_r$, then $\varphi(0) = \varphi(z_0)$ for some $z_0 \in \mathbb{D}^*$. But this is impossible since then a neighborhood of w_0 would be hit both by a neighborhood of 0 and by one of z_0, respectively.

On the other hand, if $w_0 \in \partial\Delta_r$, then we obtain a contradiction by the maximum principle applied to either φ or $\frac{1}{\varphi}$; depending on whether $|w_0| = 1$ or $|w_0| = r$, respectively.

Problem 8.7 in Chapter 8 treats the general case within the context of the uniformization theorem.

This list of examples is relevant for a number of reasons. First, we remark that we have exhausted all possible examples with $\Omega = \mathbb{C}$. Indeed, we leave it to the reader to verify that all nontrivial discrete subgroups of $\mathrm{Aut}(\mathbb{C})$ *that have no fixed point* are either $\langle z \mapsto z + \omega \rangle$ with $\omega \neq 0$, or $\langle z \mapsto z + \omega_1, z \mapsto z + \omega_2 \rangle$ with $\omega_1 \neq 0, \omega_2/\omega_1 \notin \mathbb{R}$ (see Problem 4.1). Second, $\mathbb{C}^*, \mathbb{D}^*, \Delta_r$ and $\mathbb{C}/\Lambda$ where Λ is a lattice, is a complete list of Riemann surfaces (up to conformal equivalence, of course) whose fundamental group is nontrivial and abelian. See the notes to this chapter for references.

Theorem 4.12 leaves open the case where the action does exhibit fixed points. If this is so, the map $\pi : \Omega \to \Omega/G$ may or may not induce the structure of a Riemann surface on Ω/G. Problem 5.8 provides a tool by which one may obtain a Riemann surface in the presence of fixed points.

4.6. Elliptic functions

Throughout this section, we let

$$M = \mathbb{C}/\langle z \mapsto z + \omega_1, z \mapsto z + \omega_2 \rangle$$

be the torus of Example 2 from the above list. As usual, we refer to the group by which we quotient as the lattice Λ. We first remark that $\omega_1' := a\omega_1 + b\omega_2$, $\omega_2' := c\omega_1 + d\omega_2$ is another basis of the same lattice if and only if $a, b, c, d \in \mathbb{Z}$ and $ad - bc = \pm 1$. Indeed, since ω_1', ω_2' are in the lattice, it follows that the coefficients a, b, c, d are integers. If they are a basis, then we may express ω_1, ω_2 via the inverse transformation, which must also have integer coefficients and thus an integral determinant. So the determinant of the original transformation must have been ± 1 as claimed.

Thus, in Figure 4.3 we can pass to other bases as well as other **fundamental regions**.

Definition 4.13. A fundamental region is any closed connected set $P \subset \mathbb{C}$ with the property that:

i) every point in $\mathbb{C}$ is congruent modulo Λ to some point of P,

ii) no pair of points from the interior of P are congruent.

In Figure 4.3 the parallelogram spanned by ω_1, ω_2 is one such fundamental region, whereas the parallelogram spanned by ω_1, ω_3 is another. A natural choice of such a region is given by the *Dirichlet polygon* (see Problem 4.5).

Let us now turn to the study of meromorphic functions on the torus M. By definition

$$\mathcal{M}(M) = \{f \in \mathcal{M}(\mathbb{C}) \mid f(z) = f(z+\omega_1) = f(z+\omega_2)\} \tag{4.14}$$

where we ignore the function constant and equal to ∞. For the purists, we remark that (4.14) is not an alternative definition but rather a description. These functions are called *doubly-periodic* or *elliptic functions*. First, since M is compact the only holomorphic functions are the constants. Next, we claim that any nonconstant function $f \in \mathcal{M}(M)$ satisfies $\deg(f) \geq 2$. Indeed, suppose $\deg(f) = 1$. Then in the notation of the Riemann-Hurwitz theorem above, $B = 0$ and therefore $1 = g_M = g_{S^2} = 0$, a contradiction. The reader should check that we can arrive at the same conclusion by verifying that

$$\oint_{\partial P} f(z)\, dz = 0 \tag{4.15}$$

which implies that the sum of the residues inside P is zero. Here P is a fundamental region so that f has neither zeros nor poles on its boundary. To obtain a contradiction from (4.15) note that a function f of degree 1 would need to have a unique simple pole in P, for which the integral (4.15) would then give a vanishing residue.

An interesting question concerns the existence of elliptic functions of minimal degree, viz. $\deg(f) = 2$. From the Riemann-Hurwitz formula, any elliptic function f with $\deg(f) = 2$ satisfies $B = 4$ and therefore has exactly four branch points each with valency 2. We shall now present the classical Weierstrass function $\wp$ which is of this type. As we shall see later in this section, all elliptic functions can be expressed in terms of this one function; see Proposition 4.16.

Throughout this section, we shall use meromorphic functions both in terms of the compact surface M, as well as on the plane $\mathbb{C}$. In other words, we invoke the identification (4.14). This should not cause any confusion as it will be clear from the context in which sense we are viewing functions.

In the following proposition we use the term *group of periods* of a function f. This just refers to all $\omega \in \mathbb{C}$ with the property that $f(z+\omega) = f(z)$. It is clear that the periods form a group.

Proposition 4.14. *Let Λ be as in* (4.13) *and set $\Lambda^* := \Lambda \setminus \{0\}$. For any integer $n \geq 3$, the series*

$$f(z) = \sum_{w \in \Lambda} (z+w)^{-n} \tag{4.16}$$

defines a function $f \in \mathcal{M}(M)$ with $\deg(f) = n$. Furthermore, the Weierstrass function

$$\wp(z) := \frac{1}{z^2} + \sum_{w \in \Lambda^*} \big[(z+w)^{-2} - w^{-2}\big], \tag{4.17}$$

is an even elliptic function of degree two with Λ as its group of periods. The poles of $\wp$ are precisely the points in Λ and they are all of order 2.

Proof. If $n \geq 3$, then we claim that

$$f(z) = \sum_{w \in \Lambda} (z+w)^{-n}$$

converges absolutely and uniformly on every compact set $K \subset \mathbb{C} \setminus \Lambda$. It suffices to prove this on the closure of any fundamental region. There exists $C > 0$ such that

$$C^{-1}(|x| + |y|) \leq |x\omega_1 + y\omega_2| \leq C(|x| + |y|)$$

for all $x, y \in \mathbb{R}$. Hence when $z \in \{x\omega_1 + y\omega_2 \mid 0 \leq x, y \leq 1\}$, then

$$|z + (k_1\omega_1 + k_2\omega_2)| \geq C^{-1}(|k_1| + |k_2|) - |z| \geq (2C)^{-1}(|k_1| + |k_2|)$$

provided $|k_1| + |k_2|$ is sufficiently large. Since

$$\sum_{|k_1|+|k_2|>0} |k_1\omega_1 + k_2\omega_2|^{-n} < \infty$$

as long as $n > 2$, we conclude $f \in \mathcal{H}(\mathbb{C} \setminus \Lambda)$. Since

$$f(z) = f(z+\omega_1) = f(z+\omega_2)$$

for all $z \in \mathbb{C} \setminus \Lambda$, it is clear that $f \in \mathcal{M}(M)$. The degree of (4.16) is determined by noting that inside a fundamental region the series has a unique pole of order n.

For the second part, we note that

$$\Big|(z+w)^{-2} - w^{-2}\Big| \leq \frac{|z||z+2w|}{|w|^2|z+w|^2} \leq \frac{C|z|}{|w|^3}$$

provided $|w| > 2|z|$ so that the series defining $\wp$ converges absolutely and uniformly on compact subsets of $\mathbb{C} \setminus \Lambda$. Clearly, $\wp(z)$ is an even function of z. For the periodicity of $\wp$, note that

$$\wp'(z) = -2 \sum_{w \in \Lambda} (z+w)^{-3}$$

is periodic relative to the lattice Λ. Thus, for every $w \in \Lambda$,

$$\wp(z+w) - \wp(z) = C(w) \quad \forall\, z \in \mathbb{C}$$

with some constant $C(w)$. Expanding around $z = 0$ yields $C(w) = 0$ as desired. Alternatively, one may note that

$$\wp(\omega_1/2) - \wp(-\omega_1/2) = 0$$

whence $C(w) = 0$. □

Likewise, one shows that

$$\wp'(z) = -2 \sum_{w \in \Lambda} (z+w)^{-3}$$

is an elliptic function with poles of order 3 at all points in Λ.

Another way of obtaining the function $\wp$ is as follows: let σ be defined as the Weierstrass product

$$\sigma(z) := z \prod_{\omega \in \Lambda^*} E_2(z/\omega) \tag{4.18}$$

with canonical factors E_2 as in Chapter 2 (see (2.28)). Thus σ is entire with simple zeros precisely at the points of Λ. Consider the logarithmic derivative of σ, i.e.,

$$\zeta(z) = \frac{\sigma'(z)}{\sigma(z)} = \frac{1}{z} + \sum_{\omega \in \Lambda^*} \Big[\frac{1}{z-\omega} + \frac{1}{\omega} + \frac{z}{\omega^2}\Big].$$

By inspection, $\wp = -\zeta'$. In particular, $\zeta'(z+\omega) = \zeta'(z)$ for all $\omega \in \Lambda$. Integrating, we obtain

$$\zeta(z+\omega) - \zeta(z) = C(\omega)$$

where the latter represents some constant. In particular,

$$\zeta(z+\omega_j) - \zeta(z) = \eta_j, \qquad j = 1, 2$$

or, in other words,

$$\frac{\sigma'(z+\omega_j)}{\sigma(z+\omega_j)} - \frac{\sigma'(z)}{\sigma(z)} = \eta_j, \qquad j = 1, 2.$$

Integrating this relation implies that

$$\sigma(z+\omega_j) = \sigma(z)\, e^{z\,\eta_j + \theta_j} \tag{4.19}$$

where θ_j is some complex constant. We may read off from (4.18) that $\sigma(z)$ is an odd function. Setting $z = -\omega_j/2$ in (4.19) and using the fact that $\sigma(\omega_j/2) \neq 0$ shows that

$$\sigma(z + \omega_j) = -\sigma(z)\, e^{\eta_j(z+\omega_j/2)}. \tag{4.20}$$

We shall make use of this relation in the construction of meromorphic functions on the tori (see Theorem 4.17 below).

The Weierstrass $\wp$ function has many remarkable properties, the most basic of which is the following differential equation.

Lemma 4.15. *With $\wp$ as before, one has*

$$(\wp'(z))^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3) \tag{4.21}$$

where $e_1 = \wp(\omega_1/2)$, $e_2 = \wp(\omega_2/2)$, and $e_3 = \wp((\omega_1 + \omega_2)/2)$ are pairwise distinct. Furthermore, one has $e_1 + e_2 + e_3 = 0$ so that (4.21) can be written in the form

$$(\wp'(z))^2 = 4(\wp(z))^3 - g_2\wp(z) - g_3 \tag{4.22}$$

with constants $g_2 = -4(e_1e_2 + e_1e_3 + e_2e_3)$ and $g_3 = 4e_1e_2e_3$.

Proof. In this proof it will be convenient to view the underlying torus as given by

$$S = \{x\omega_1 + y\omega_2 \mid -1/2 \le x, y < 1/2\}.$$

By inspection,

$$\wp'(z) = -2\sum_{z\in\Lambda}(z + w)^{-3}$$

is an odd function in $\mathcal{M}(M)$. Since it has a pole of order 3 at $z = 0$ but no other poles in the parallelogram S, it follows that $\wp'(z)$ is of degree 3. Thus,

$$\wp'(\omega_1/2) = -\wp'(-\omega_1/2) = -\wp'(\omega_1/2) = 0.$$

Similarly, $\wp'(\omega_2/2) = \wp'((\omega_1 + \omega_2)/2) = 0$. In other words, the three points

$$\frac{1}{2}\omega_1, \quad \frac{1}{2}\omega_2, \quad \frac{1}{2}(\omega_1 + \omega_2)$$

are the three zeros of $\wp'$, each simple, and thus also the unique points where $\wp$ has valency 2 apart from $z = 0$. The e_j are distinct, for otherwise $\wp$ would assume such a value four times in contradiction to its degree being 2.

Denoting the right-hand side of (4.21) by $F(z)$, this implies that

$$\frac{(\wp'(z))^2}{F(z)} \in \mathcal{H}(M)$$

and therefore equals a constant. Considering the expansion of $\wp'(z)$ and $F(z)$, respectively, around $z = 0$ shows that the value of this constant equals 1, as claimed. The final statement follows by observing from the Laurent series around zero that

$$(\wp'(z))^2 - 4(\wp(z))^3 + g_2\wp(z),$$

with g_2 as above, is analytic and therefore equals a constant. $\square$

The values e_1, e_2, e_3 are of great importance. Later in this chapter we will study the ratio

$$\lambda = \frac{e_3 - e_2}{e_1 - e_2}$$

as a function of ω_1, ω_2. It can be easily seen to be a function of ω_1/ω_2, and as such is known as the *modular function.*

The previous proof shows that 0, $\omega_1/2$, $\omega_2/2$, and $(\omega_1 + \omega_2)/2$ are precisely the branch points of $\wp$ in the fundamental region S. We are now able to establish the following property of $\wp$.

Proposition 4.16. *Every $f \in \mathcal{M}(M)$ is a rational function of $\wp$ and $\wp'$. If f is even, then it is a rational function of $\wp$ alone.*

Proof. Suppose that f is nonconstant and even. Then for all but finitely many values of $w \in \mathbb{C}_\infty$, the equation $f(z) - w = 0$ has only simple zeros (since there are only finitely many zeros of f'). Pick two such $w \in \mathbb{C}$ and denote them by c, d. Moreover, we can ensure that the zeros of $f - c$ and $f - d$ are distinct from the branch points of $\wp$. Thus, since f is even and with $2n = \deg(f)$, one has:

$$\{z \in M \ : \ f(z) - c = 0\} = \{a_j, -a_j\}_{j=1}^n,$$
$$\{z \in M \ : \ f(z) - d = 0\} = \{b_j, -b_j\}_{j=1}^n.$$

The elliptic functions

$$g(z) := \frac{f(z) - c}{f(z) - d}$$

and

$$h(z) := \prod_{j=1}^n \frac{\wp(z) - \wp(a_j)}{\wp(z) - \wp(b_j)}$$

have the same zeros and poles which are all simple. It follows that $g = \alpha h$ for some $\alpha \neq 0$. Solving this relation for f yields the desired conclusion.

If f is odd, then $f/\wp'$ is even so $f = \wp' R(\wp)$ where R is rational. Finally, if f is any elliptic function, then

$$f(z) = \frac{1}{2}\big(f(z) + f(-z)\big) + \frac{1}{2}\big(f(z) - f(-z)\big)$$

is a decomposition into even/odd elliptic functions whence

$$f(z) = R_1(\wp) + \wp' R_2(\wp)$$

with rational R_1, R_2, as claimed. □

For another result along these lines see Problem 4.10. It is interesting to compare the previous result for the tori to a similar one for the simply-periodic functions, i.e., functions on the surface $\mathbb{C}/\langle z \mapsto z+1\rangle \simeq \mathbb{C}^*$. These can be represented via Fourier series, i.e, infinite expansions in the basis $e^{2\pi i z}$ which plays the role of $\wp$ in this case. Observe the distinction between the *infinite expansions* for the cylinder on the one hand, and the finite ones for tori, on the other hand. This is of course rooted in the fact that the tori are compact whereas $\mathbb{C}^*$ is not.

We conclude our discussion of elliptic functions by turning to the following natural question: given disjoint finite sets of distinct points $\{z_j\}$ and $\{\zeta_k\}$ in M as well as positive integers n_j for z_j and ν_k for ζ_k, respectively, is there an elliptic function with precisely these zeros and poles and of the given orders? We remark that for the case of $\mathbb{C}_\infty$ the answer was affirmative if and only if the constancy of the degree was not violated, i.e.,

$$\sum_j n_j = \sum_k \nu_k. \tag{4.23}$$

For the tori, however, there is a new obstruction:

$$\sum_j n_j z_j = \sum_k \nu_k \zeta_k \pmod{\Lambda}. \tag{4.24}$$

To obtain this relation, we first observe that by the residue theorem one has

$$\frac{1}{2\pi i}\oint_{\partial P} z\frac{f'(z)}{f(z)}\,dz = \sum_j n_j z_j - \sum_k \nu_k \zeta_k \tag{4.25}$$

where ∂P is the boundary of a fundamental region P as in Figure 4.3 so that no zero or pole of f lies on that boundary. Second, comparing parallel sides of the fundamental region and using the periodicity shows that the left-hand side in (4.25) is of the form $n_1\omega_1 + n_2\omega_2$ with $n_1, n_2 \in \mathbb{Z}$ and thus equals 0 in Λ.

Indeed, consider the edges in ∂P given by $\gamma_1(t) := \{t\omega_1 \mid 0 \le t \le 1\}$ and $\gamma_2(t) := \{\omega_2 + (1-t)\omega_1 \mid 0 \le t \le 1\}$, respectively. Note that we have already reversed the orientation of the second relative to the first edge. Using the

ω_2-periodicity of $\frac{f'(z)}{f(z)}$ we infer that

$$\begin{aligned}&\int_{\gamma_1} z\frac{f'(z)}{f(z)}\,dz+\int_{\gamma_2} z\frac{f'(z)}{f(z)}\,dz\\&=-\omega_2\int_{\gamma_1}\frac{f'(z)}{f(z)}\,dz=-\omega_2\int_{\gamma_1} d\log f(z).\end{aligned}$$

The branch of the logarithm here is irrelevant, since the arbitrary constant is differentiated away. Recall that by assumption $f\neq 0$ on the edge along which we are integrating. Using periodicity, we see that f takes the same value at the endpoints of the edge γ_1. In conclusion

$$\omega_2\frac{1}{2\pi i}\int_{\gamma_1} d\log f(z)\in\omega_2\mathbb{Z}$$

as desired. The other edge pair gives an element of $\omega_1\mathbb{Z}$, whence (4.24).

It is a remarkable fact that (4.23) and (4.24) are also *sufficient* for the existence of an elliptic function with precisely these given sets of zeros and poles. Notice that these two conditions can only hold simultaneously if the total degree is two or more.

Theorem 4.17. *Suppose* (4.23) *and* (4.24) *hold. Then there exists an elliptic function which has precisely these zeros and poles with the given orders. This function is unique up to a nonzero complex multiplicative constant.*

Proof. Listing the points z_j and ζ_k with their respective multiplicities, we obtain sequences z_j' and ζ_k' of the same length, say n. Shifting by a lattice element if needed, one has

$$\sum_{j=1}^{n} z_j'=\sum_{k=1}^{n}\zeta_k'. \tag{4.26}$$

Now set

$$f(z)=\prod_{j=1}^{n}\frac{\sigma(z-z_j')}{\sigma(z-\zeta_j')}$$

where σ has been defined in (4.18). The function f has the desired zeros and poles. It remains to check the periodicity. This, however, follows immediately from (4.20) and (4.26). □

The reader may be wondering why *elliptic functions* are so-called. Thus, let $0<a<b$ and suppose we wish to compute the arc-length of a section of the ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

Setting $x = a\cos\theta$, $y = b\sin\theta$ where $0 \le \theta_0 < \theta < \theta_1 \le \pi$, say, we are thus lead to the integral

$$\begin{aligned}\int_{\theta_1}^{\theta_0} \sqrt{a^2 \sin^2\theta + b^2\cos^2\theta}\, d\theta &= \int_{t_0}^{t_1} \frac{\sqrt{a^2t^2 + b^2(1-t^2)}}{\sqrt{1-t^2}}\, dt \\ &= b\int_{t_0}^{t_1} \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\, dt \\ &= b\int_{t_0}^{t_1} \frac{dt}{\sqrt{(1-k^2t^2)(1-t^2)}} - bk^2 \int_{t_0}^{t_1} \frac{t^2\, dt}{\sqrt{(1-k^2t^2)(1-t^2)}}\end{aligned} \tag{4.27}$$

where we substituted $t = \sin\theta$ and with $k^2 = 1 - a^2/b^2$. In the 19th century evaluating or approximating integrals such as those—aptly named *elliptic integrals*, and classified by Legendre—was intensely researched by Abel, Legendre, Dirichlet, Monge, Dupin, Gauss, Weierstrass, Jacobi, Riemann, just to name a few. The reason for this lies with the perhaps not so surprising fact that many concrete problems in mechanics and geometry involve integrals of this type.

Just to give another example, consider the pendulum equation (with normalized constants) $\ddot\theta + \frac{1}{2}\sin\theta = 0$ where θ is the angle between the pendulum and a vertical. The conserved energy is $\dot\theta^2 - \cos\theta = E$, which leads to the solution

$$\begin{aligned} t &= t_0 \pm \int \frac{d\theta}{\sqrt{E + \cos\theta}} \\ &= t_0 \mp \int \frac{du}{\sqrt{(E+u)(1-u^2)}}. \end{aligned} \tag{4.28}$$

The integrals in (4.27) and (4.28) are examples of expressions of the form

$$\int \sqrt{R(t)}\, dt \tag{4.29}$$

where $R(t)$ is a rational function. Another rather famous problem which leads to such integrals is the problem of finding the geodesics on a generic triaxial ellipsoid, which occupied many mathematicians and which was eventually solved by Jacobi.

The absolutely remarkable, as well as "abstract", realization that emerged in the 19th century was that the integrals (4.29) are best viewed in the complex domain and, in fact, of the Riemann surface associated with the rational function $R(t)$. What this exactly means will become more transparent in the next chapter. However, let us emphasize at this point that the revolution was to view the endpoints of integration as a function of the integral, rather than the other way around; indeed, the endpoints do not lie in the complex plane as was of course the rule for the longest time, but on possibly different sheets of the surface defined by the integral.

To be more specific, let us solve (4.22) as follows: integrating

$$\frac{d\wp(z)}{\sqrt{4(\wp(z))^3 - g_2\wp(z) - g_3}} = dz$$

where we choose some branch of the root, yields

$$z - z_0 = \int_{\wp(z_0)}^{\wp(z)} \frac{d\zeta}{\sqrt{4\zeta^3 - g_2\zeta - g_3}}. \tag{4.30}$$

In other words, the Weierstrass function $\wp$ is the inverse of an elliptic integral. The integration path in (4.30) needs to be chosen to avoid the zeros and poles of $\wp'$, and the branch of the root is determined by $\wp'$.

The reader may wish to compare what we have just observed to the fact that

$$\int_{w_0}^{w} \frac{d\zeta}{\sqrt{1-\zeta^2}} = z - z_0$$

is satisfied by $w = \sin z$ (the restrictions on the path and the choice of branch again being determined appropriately). In other words, we obtain a periodic function with one period, whereas in the more complicated elliptic integral (4.30) we obtain two periods.

In more modern terminology, these integrals need to be viewed as those of differential forms on Riemann surfaces such as the hyper-elliptic ones (see Section 5.6 and Section 6.7). There are two algebraic equations implicit in (4.27) and (4.28):

$$w^2 - (1-k^2z^2)(1-z^2) = 0, \quad w^2 - (E+z)(1-z^2) = 0$$

where $0 < k < 1$, and $-1 < E < 1$. The latter condition means that the pendulum performs the usual motion as in a grandfather clock since $\dot{\theta}$ needs to vanish at the highest points of the pendulum. These equations are closely related from the point of view of Riemann surfaces: indeed, while the first one exhibits four distinct branch points in $\mathbb{C}P^1$, namely $\pm k^{-1}$ and ± 1, the second one exhibits three distinct ones in $\mathbb{C}$, namely $-E, \pm 1$, as well as one at infinity. So they in fact determine isomorphic compact Riemann surfaces which we shall identify in Chapter 5 as a torus. The reader familiar with the Schwarz-Christoffel formula from conformal mapping (see for example (2.30)) may recognize a connection with the elliptic integrals of this section. See Problem 4.14 for more on this topic.

4.7. Covering the plane with two or more points removed

We now provide the details of a construction to which we had alluded in Example 5 of Section 4.2. To be specific, we shall present the standard "geometric" procedure by which the covering map $\pi : \mathbb{D} \to \mathbb{C} \setminus \{p_1, p_2\}$ may be obtained. Here p_1, p_2 are distinct points. This then leads naturally to the "little" and "big" Picard theorems, which are fundamental results of classical function theory.

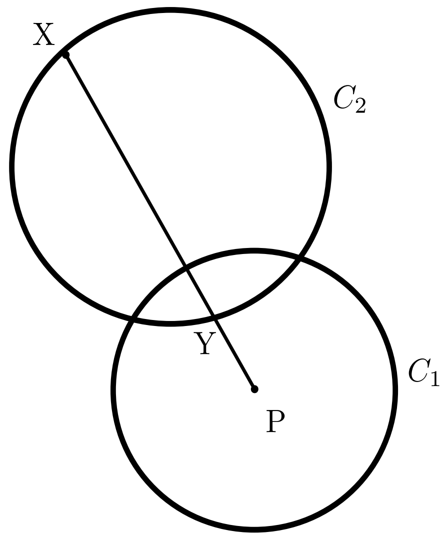

Figure 4.4. Reflection of one circle about another

The construction takes place in the Poincaré disk that we encountered in Chapter 1. We may assume that $p_1 = 0, p_2 = 1$. The main analytical tools we shall rely on are the Riemann mapping theorem and the Schwarz reflection principle (see Problem 2.6). As for the latter, recall from Chapter 1 the reflection about a circle.

Figure 4.4 shows the circle C_2 being reflected about C_1. The configuration is such that C_2 does not pass through the center of C_1, and that it intersects C_1 at a right angle. Since every point of C_1 is fixed, the two intersection points are fixed. Since the angle of intersection is also invariant, we see that C_2 gets mapped onto itself. However, the two arcs of C_2 relative to C_1 are interchanged: the one inside of C_1 is mapped onto the one outside and vice versa.

To construct the map, we start with a triangle Δ_0 inside the unit circle consisting of circular arcs that intersect the unit circle at right angles. In other words, a geodesic triangle in the Poincaré disk with vertices "at infinity" (which means that the interior angles are 0). Now reflect Δ_0 across each of its sides. The result are three more triangles with circular arcs intersecting the unit circle at right angles. Figure 4.5 shows how the unit disk is partitioned by triangles as a result of iterating these reflections indefinitely. To obtain the sought after covering map, we start from the Riemann mapping theorem from Chapter 2 which gives us a conformal isomorphism

$f : \Delta_0 \to \mathbb{H}$, the upper half-plane. We may also achieve that the three vertices of Δ_0 get mapped onto $0, 1, \infty$, respectively. Moreover, the map extends as a homeomorphism to the boundary; see Theorem 2.30. Thus, the three circular arcs of Δ_0 get mapped to the intervals $[-\infty, 0], [0, 1], [1, \infty]$, respectively. By the Schwarz reflection principle (see Problem 2.6) the map f extends analytically to the region obtained by reflecting Δ_0 across each of its sides as just explained above. In this way we obtain a conformal map onto $\mathbb{C} \setminus \{0, 1\}$ which is defined on the entire disk. By construction, it is a local isomorphism and a covering map.

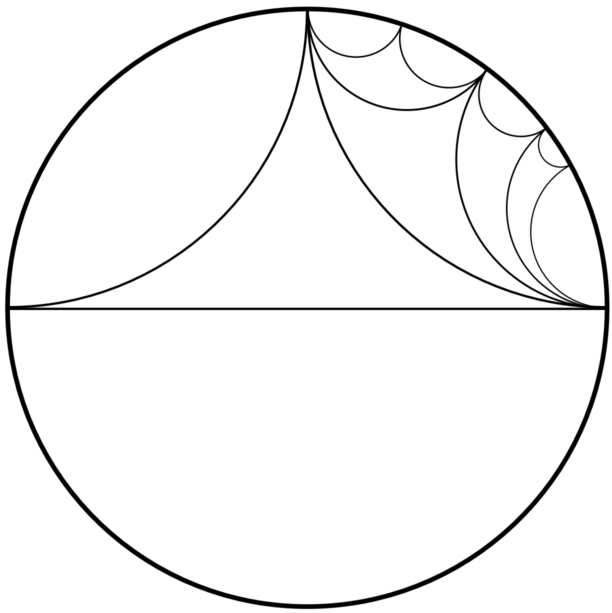

Figure 4.5. Successive reflection of triangles across their sides

If we remove three or more points (but finitely many) from $\mathbb{C}$, then we may apply the same procedure but starting with a polygon instead of a triangle.

The little Picard theorem is an immediate corollary.

Theorem 4.18. *Every entire function which omits two values is constant.*

Proof. Indeed, if f is such a function, we may assume that it takes its values in $\mathbb{C} \setminus \{0, 1\}$. But then we can lift f to the universal cover of $\mathbb{C} \setminus \{0, 1\}$ to obtain an entire function F into $\mathbb{D}$. By Liouville's theorem, F is constant. $\square$

For the relation between the map constructed here and the modular group, we refer the reader to Section 4.8. Another corollary is the following compactness theorem due to Montel. It uses the notion of a *normal family* that we encountered in Chapter 2. Let us recall it: we say that $\mathcal{F} \subset \mathcal{H}(\Omega)$ is **normal** if and only if for every compact $K \subset \Omega$ every sequence in $\mathcal{F}$ has a subsequence which converges uniformly on K. We also allow uniform convergence to ∞. Since we can always pass to finite subcovers of open covers of compact sets we deduce that normality is a *local property.*

Theorem 4.19. *Any family of functions $\mathcal{F}$ in $\mathcal{H}(\Omega)$ which omits the same two distinct values in $\mathbb{C}$ is a normal family.*

Proof. Let $\mathcal{F}$ omit the distinct values $a, b \in \mathbb{C}$. First, passing to $\frac{f-a}{b-a}$ if needed, we may assume that $a = 0$ and $b = 1$. This allows us to pass to the universal cover of $\mathbb{C} \setminus \{0, 1\}$ as before. We may lift $\mathcal{F}$ to another family $\widetilde{\mathcal{F}} \subset \mathcal{H}(\Omega)$, each element of which takes its values in the unit disk. In particular, this family is bounded, and therefore normal. However, it is not immediate that $\mathcal{F}$ inherits the normality. To be specific, let $\{f_n\}_{n=1}^{\infty} \subset \mathcal{F}$ be any sequence and fix some $z_0 \in \Omega$. Passing to a subsequence, we may assume that $f_n(z_0) \to w_\infty \in \mathbb{C}_\infty$. Clearly, the issue is now to distinguish the "degenerate cases" $w_\infty \in \{0, 1, \infty\}$ from the others. Let us start with the latter.

Case 1: $w_\infty \notin \{0, 1, \infty\}$. By normality of $\widetilde{\mathcal{F}}$ we can assume that the lifted sequence converges: $\widetilde{f_n} \to \widetilde{f_\infty}$. Clearly, $|\widetilde{f_\infty}| \le 1$. If equality is attained here anywhere in Ω, then $\widetilde{f_\infty}$ is a constant of absolute value 1. But then $|\widetilde{f_n}(z_0)| \to 1$ which implies that $f_n(z_0) \to w_\infty \in \{0, 1, \infty\}$—a contradiction. So we see that $|\widetilde{f_\infty}| < 1$ everywhere in Ω. Thus, on any compact $K \subset \Omega$ one has $|\widetilde{f_n}| \le \gamma(K) < 1$ for all large n where $\gamma(K)$ is some constant. But now we may pass back down from the cover to the twice punctured plane $\mathbb{C} \setminus \{0, 1\}$ to conclude that

$$\sup_K |f_n| \le M < \infty \quad \text{for all large } \ n.$$

By the "little Montel theorem" (see Proposition 2.22), we may pass to a uniformly convergent subsequence of $\{f_n\}$ as desired.

Case 2: $w_\infty = 1$. Since f_n misses 0, and we may assume that Ω is simply-connected (a disk, in fact) we may define the square root $g_n = \sqrt{f_n}$. We pick the branch for which $g_n(z_0) \to -1$ as $n \to \infty$. But then g_n misses $0, 1$ and Case 1 applies. Normality of g_n implies that of $f_n = g_n^2$.

Case 3: $w_\infty = 0$. In that case define $g_n = 1 - f_n$, which again misses $0, 1$, and for which we obtain a modified w_∞, denoted by $\widetilde{w}_\infty = 1 - w_\infty = 1$. So the previous case applies.

Case 4: Finally, suppose $w_\infty = \infty$. Then let $g_n = 1/f_n$ which is analytic and misses $0, 1$. Moreover, we obtain a new limit $\widetilde{w}_\infty = 0$ for this modified sequence, and so Case 3 applies. Thus, $\{g_n\}_{n=1}^\infty$ is a normal family whence $g_n \to g_\infty$ on a given compact set $K \subset \Omega$. By Lemma 2.20 we further see that $g_\infty \equiv 0$. Indeed, none of the g_n vanishes on Ω and $g_\infty(z_0) = 0$. Thus, we must have $g_\infty = 0$ everywhere. But this means that $f_n \to \infty$ uniformly in compact sets and we are done. $\square$

Montel's theorem 4.19 is referred to as *fundamental normality test.* Note that this fails for families which omit one value:

$$\mathcal{F} = \{e^{nz} \mid n \in \mathbb{Z}\} \subset \mathcal{H}(\mathbb{C})$$

omits 0 but is not normal. By passing to the Riemann sphere, one can establish similar results for meromorphic functions. One then obtains a normality test for functions which omit three values. As an application of the fundamental normality test, we now prove **Picard's big theorem** which is a substantial strengthening of the Casorati-Weierstrass principle (see Proposition 2.6).

Theorem 4.20. *If f has an isolated essential singularity at z_0, then in any small neighborhood of z_0 the function f attains every complex value infinitely often, with one possible exception.*

Proof. The idea is of course to "zoom into" z_0. By the normality test we may pass to a limit and thus conclude that z_0 is either removable or a pole.

To be specific, let $z_0 = 0$ and define $f_n(z) = f(2^{-n}z)$ for an integer $n \geq 1$. We take n so large that f_n is analytic on $0 < |z| < 2$. Then $f_{n_k}(z) \to F(z)$ uniformly on $1/2 \leq |z| \leq 1$ where either F is analytic or $F \equiv \infty$. In the former case, we infer from the maximum principle that f is bounded near $z = 0$, which is therefore removable. In the latter case, $z = 0$ is a pole. $\square$

4.8. Groups of Möbius transforms

Automorphism groups of regions Ω in the complex plane, or for that matter of any Riemann surface, are natural objects to study. By automorphism group of course we mean all conformal homeomorphisms from some Riemann surface to itself. Prominent examples are $\mathrm{Aut}(\mathbb{D})$ (see Proposition 1.32) or $\mathrm{Aut}(\mathbb{H}) = \mathrm{PSL}(2,\mathbb{R})$, which are subgroups of $\mathrm{PSL}(2,\mathbb{C})$, the group of all fractional linear transformations. Example 5) from Section 4.2, as well as all of Section 4.5, show the importance of subgroups of $\mathrm{Aut}(\Omega)$ which are free of fixed-points and act *properly discontinuously* on Ω (see, however, Problem 5.8 for a construction in the presence of fixed points).

Let us determine the possible fixed points of elements in $\mathrm{Aut}(\mathbb{D})$. Problem 4.2 shows that this group is precisely given by maps

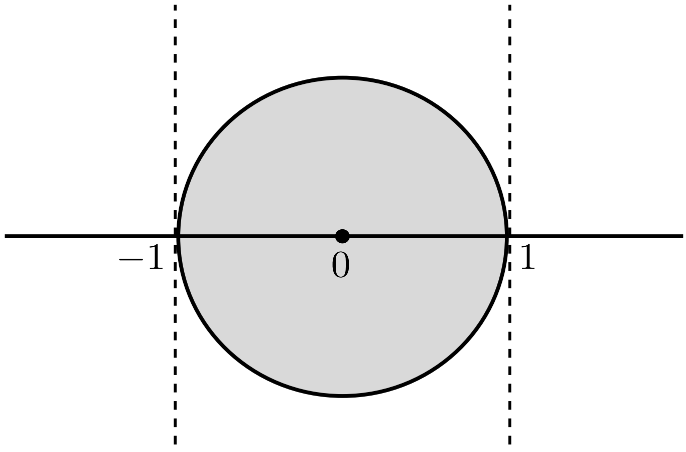

Figure 4.6. SU(1, 1) and fixed points in Aut($\mathbb{D}$)

$$z \mapsto \frac{az + \bar{b}}{bz + \bar{a}}, \qquad a, b \in \mathbb{C}, \;\; |a|^2 - |b|^2 = 1. \tag{4.31}$$

Thus, $|a| \geq 1$. If $|a| = 1$, then $b = 0$ and the map is a rotation about the origin which has precisely two fixed points, $0, \infty$ (unless it is the identity). Figure 4.6 shows the relevant geometry in the a plane. The unit circle is shaded to indicate it needs to be excluded. So let us assume that $|a| > 1$ which means $b \neq 0$. Then the fixed points are given by

$$\zeta_{\pm} := \frac{i \operatorname{Im} a \pm \sqrt{(\operatorname{Re} a)^2 - 1}}{b}. \tag{4.32}$$

Note that $|\zeta_+ \zeta_-| = 1$. These points coincide precisely if $|\operatorname{Re} a| = 1$, which also means that

$$|b| = |\operatorname{Im} a|.$$

We may assume that $\operatorname{Im} a > 0$ (since the sign is factored out in (4.31)) and set

$$b = ie^{-i\varphi} \operatorname{Im} a.$$

Thus, $\zeta_{\pm} = e^{i\varphi}$ is the unique fixed point on the unit circle (the so-called *parabolic case*, cf. Problem 4.3). Next, suppose we are in

Case 1: $|\operatorname{Re} a| > 1$. Then the root in (4.32) is real-valued and $|\zeta_+ \zeta_-| = 1$ implies that $|\zeta_{\pm}|^2 = 1$. We have a pair of distinct fixed points on the unit circle. Moreover,

$$|\zeta_+ - \zeta_-| = 2 \frac{\sqrt{(\operatorname{Re} a)^2 - 1}}{\sqrt{|a|^2 - 1}}$$

which can be anything in the interval $(0, 2]$. The endpoint 2 here means that $\zeta_+ = -\zeta_-$, which happens if and only if $a > 1$ or $a < -1$.

Case 2: $|\operatorname{Re} a| < 1$. Then $\operatorname{Im} a \neq 0$ and one checks by direct calculation that $|\zeta_-| < 1$ if $\operatorname{Im} a > 0$, and $|\zeta_+| < 1$ if $\operatorname{Im} a < 0$, respectively. In other words, if $|\operatorname{Re} a| < 1$, then we always have a fixed point inside of the unit disk, which is not allowed for the purposes of Theorem 4.12.

Conversely, suppose we are given two distinct points $\zeta_\pm$ on the unit circle. One wants to find $a, b \in \mathbb{C}$ so that $\zeta_\pm$ are the fixed points of (4.31). By the above analysis, the condition $|\operatorname{Re} a| > 1$ needs to be true. Setting

$$b = re^{i\varphi} = e^{i\varphi}\sqrt{|a|^2 - 1},$$

and taking into account (4.32), we are first lead to finding $\varphi \in \mathbb{R}$ so that $e^{i\varphi}(\zeta_+ - \zeta_-)$ is real-valued. Then we have

$$e^{i\varphi}\zeta_\pm = \pm\beta + i\alpha$$

with

$$\alpha, \beta \in \mathbb{R}, \quad \alpha^2 + \beta^2 = 1.$$

In fact, one has

$$\alpha = \frac{\operatorname{Im} a}{\sqrt{|a|^2 - 1}}, \quad \beta = \frac{\sqrt{(\operatorname{Re} a)^2 - 1}}{\sqrt{|a|^2 - 1}}.$$

A special case arises when $\alpha = 0, \beta = 1$. In that case, $\zeta_\pm = \pm e^{-i\varphi}$. We again read off from (4.32) that

$$a > 1, \quad b = e^{i\varphi}\sqrt{a^2 - 1}$$

is the 1-parameter subgroup of maps which fix the points $\pm e^{-i\varphi}$ on the unit circle. The case $a = 1, b = 0$ is special, since it corresponds to the identity.

When $|\beta| < 1$, then again from (4.32) one sees that the 1-parameter subgroup which fixes $\zeta_\pm$ is determined by the hyperbola

$$\xi^2 - \frac{\beta^2}{1 - \beta^2}\eta^2 = 1, \quad a = \xi + i\eta,\ \xi \geq 1. \tag{4.33}$$

The reader may want to sketch the location of these hyperbolas in Figure 4.6. The value of b is determined by $b = e^{i\varphi}\sqrt{|a|^2 - 1}$. Moreover, the pair (a, b) is uniquely determined by specifying a point on the hyperbola (4.33); but $(-a, -b)$ generates the same element in $\operatorname{Aut}(\mathbb{D})$.

To the partition of $\mathbb{D}$ generated from one geodesic triangle by successive reflections as in Figure 4.5, we can now associate the subgroup of $\operatorname{Aut}(\mathbb{D})$ that leaves that partition invariant. It is exactly the subgroup that takes the vertices of one geodesic triangle of the partition onto the vertices of another such triangle. Suppose we wish to keep two vertices fixed. As we have just seen there is a 1-parameter subgroup of $\operatorname{Aut}(\mathbb{D})$ that fixes these two points; the one-dimensional freedom corresponding exactly to sending the third vertex to an arbitrary point on the circle, with the exception of the two fixed points. By construction, we are sending that vertex to its reflection across the circular arc whose endpoints are fixed. Similarly, fixing only one point leaves us with two degrees of freedom in $\mathrm{Aut}(\mathbb{D})$ which we use to move the other two vertices.

By means of a fractional linear transformation, we may move our observations about $\mathbb{D}$ to the upper half-plane $\mathbb{H}$. Let us now introduce a special subgroup of $\mathrm{Aut}(\mathbb{H})$, namely $\mathrm{PSL}(2,\mathbb{Z}) \subset \mathrm{PSL}(2,\mathbb{R})$, the so-called **modular group**. We already encountered this subgroup in Section 4.6, where we noted that it is precisely the one which transforms any basis of a lattice Λ into another basis of the same lattice, and which, moreover, preserves the orientation. The role of fractional linear transformations is not immediately apparent here. It arises by the usual change of coordinates associated with an oriented basis (ω_1, ω_2), viz. $z \mapsto \omega_2^{-1} z$. This sends the basis into $(\tau, 1)$ where $\tau = \frac{\omega_1}{\omega_2}$ and the orientation is such that $\mathrm{Im}\, \tau > 0$. So if we pass from a basis (ω_1, ω_2) in Λ to another one, $(\omega_1', \omega_2') := (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$, then τ transforms according to

$$\tau' = \frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2} = \frac{a\tau + b}{c\tau + d}. \tag{4.34}$$

This shows that any two lattices Λ, Λ' whose normalized representations $(\tau, 1)$ and $(\tau', 1)$ are related by (4.34) are conformally equivalent.

Problem 4.9 asks the reader to verify that the converse holds, too. We are thus lead to the problem of understanding the action of $\mathrm{PSL}(2,\mathbb{Z})$ on $\mathbb{H}$. Note that elements of the modular group may have fixed points and thus $\mathbb{H}/\mathrm{PSL}(2,\mathbb{Z})$ is not a Riemann surface (cf. Theorem 4.12); for example, $\left[\begin{smallmatrix} 1 & -2 \\ 1 & -1 \end{smallmatrix}\right]$ has $1 \pm i$ as fixed points. Using the notation of the theorem, fixed points in Ω of maps in the group G create singularities of Ω/G. This will become clear to the reader after drawing the pictures associated with elliptic fixed points as in Problem 4.3. Nevertheless, this quotient exists as a topological space, in fact, a Hausdorff one provided the action is properly discontinuous—which it is for the modular group.

Examples of maps in $\mathrm{PSL}(2,\mathbb{Z})$ are $z \mapsto -\frac{1}{z}$ and $z \mapsto z+1$. We shall now show that these maps, which we call S, T, respectively, generate the whole modular group.

Lemma 4.21. *The modular group is generated by S and T.*

Proof. The underlying procedure, which is a form of the Euclidean algorithm, is best demonstrated by means of an example. First, if $Uz = az + b$ lies in the modular group, then we need $a = 1$ and thus $U = T^b$. Consider $Uz = \frac{7z+5}{4z+3}$. We first replace z with $z - 1$. This makes the constant in the

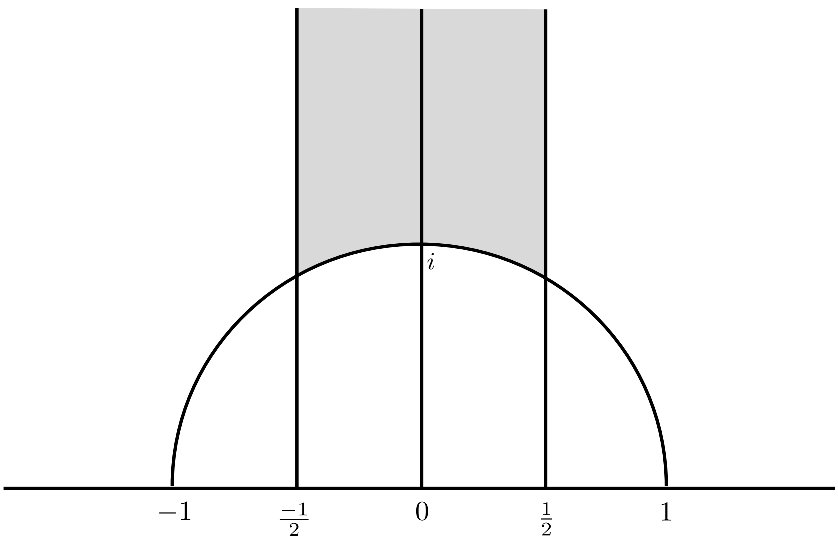

Figure 4.7. The fundamental domain of the modular group

denominator the smallest in absolute value:

$$UT^{-1}z = \frac{7z-2}{4z-1}.$$

Next, compose with S:

$$UT^{-1}Sz = \frac{2z+7}{z+4}.$$

Translate again to make the constant in the denominator as small as possible:

$$UT^{-1}ST^{-4}z = \frac{2z-1}{z}.$$

Finally,

$$UT^{-1}ST^{-4}ST^{-2}z = z$$

and we are done.

We leave it to the reader (see Problem 4.16) to show that this algorithm *always* produces the identity in a finite number of steps. The essential ingredient is of course that the determinant is 1 to begin with, and always remains so. Otherwise, we would end up with a transformation of the form $w = nz + m$ with $n, m \in \mathbb{Z}$ but $n \neq 1$ being the determinant we started with. $\square$

We can now easily describe the fundamental region of the modular group.

Proposition 4.22. *The fundamental region is the one shown in Figure* 4.7*, i.e., it is bounded by* $\operatorname{Re} z = \pm\frac{1}{2}$*, and the unit circle. In particular, the modular group acts properly discontinuously on* $\mathbb{H}$*.*

Proof. The map S is the composition of a reflection about the imaginary axis, and an inversion on the unit circle ($z \mapsto 1/\bar{z}$). The shaded region Ω is symmetric with respect to the former, and the latter maps Ω onto a geodesic triangle with one vertex at 0 and with the circular arc remaining fixed. The action of the translations T is clear. It follows that all images of Ω under the modular group tile $\mathbb{H}$ as in Figure 4.7.

It remains to show that no two points in the interior of Ω are mapped onto each other. Thus, let $w = \frac{az+b}{cz+d}$ be *unimodular*, i.e., an element of the modular group.

(1) $c = 0$. Then $w = z+b$ is a translation by an integer and the interior of Ω is mapped outside of itself.
(2) $c = \pm 1$. Then $|w-ac| = 1/|z+d|$ where $a, d \in \mathbb{Z}$. But for $z \in \Omega\backslash\partial\Omega$ we have $|z + d| > 1$ and thus $|w - a| < 1$ whence $w \notin \Omega$.
(3) $|c| \geq 2$. Then $|w - a/c| = 1/(c^2|z + d/c|)$. Now $\operatorname{Im} z > 1/2$ implies
$$1/(c^2|z + d/c|) < 2/c^2 \leq 1/2,$$
which means that $w \notin \Omega$.

This covers all cases and we are done. □

Problem 4.9 asks the reader to apply these observations to the problem of classifying conformally inequivalent tori.

Definition 4.23. A group of Möbius transformations which acts properly discontinuously on a disk and leaves the boundary circle invariant is called a *Fuchsian group.*

The modular group is an example of a Fuchsian group. These groups are of fundamental importance in the classification of Riemann surfaces (see Sections 8.6 and 8.7). Problem 4.6 shows the properly discontinuous action is equivalent to the group being discrete in the natural topology on matrices. Problem 4.7 introduces some elementary geometric properties of Fuchsian groups, namely the existence of a fundamental region, which tessellates the underlying disk under the action of the group. Furthermore, we may choose the boundary to consist of geodesic arcs.

Section 4.6 studied meromorphic functions on tori, whereas here we have been concerned with the symmetries of the tori. We conclude this chapter by exploring a connection between these two points of view. Returning to Lemma 4.15, consider the ratio as a function of the pair (ω_1, ω_2),

$$\lambda := \frac{e_3 - e_2}{e_1 - e_2},$$

which is well-defined and never vanishes since the finite complex numbers e_j are distinct. Inspection of the defining series (4.17) and the definition of e_j in terms of $\omega_1/2, \omega_2/2, (\omega_1 + \omega_2)/2$ shows that λ is in fact a function of $\tau = \frac{\omega_1}{\omega_2}$, analytic in the upper half-plane. Moreover, $\lambda(\tau)$ never assumes the values 0 or 1.

If we replace the basis (ω_1, ω_2) with another one (ω_1', ω_2') by means of a unimodular transformation, then $\wp$ does not change in view of the series expansion (4.17). From the *differential equation* (4.21) we infer therefore that the three points e_j can be at most permuted. If the unimodular transformation satisfies

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mod 2, \tag{4.35}$$

then $\frac{1}{2}\omega_j' \equiv \frac{1}{2}\omega_j$ in the lattice for $j = 1, 2$. Thus, the e_j are fixed under such a transformation whence

$$\lambda\Big(\frac{a\tau + b}{c\tau + d}\Big) = \lambda(\tau)$$

under maps in the subgroup (4.35) known as the *congruence subgroup modulo* 2. In other words, λ is invariant under this subgroup and therefore an example of an *automorphic function* relative to this subgroup (also known as *elliptic modular function*). By direct calculation we find that the entire modular group reduces to the group of six transformations modulo 2:

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\}. \tag{4.36}$$

From Lemma 4.21 or otherwise one concludes that this group is generated by the second and third matrices in this list. The second matrix interchanges e_1, e_3 and takes τ to $\tau + 1$. Thus,

$$\lambda(\tau + 1) = \frac{1}{\lambda(\tau)}. \tag{4.37}$$

Similarly, from the third matrix which corresponds to $z \mapsto -\frac{1}{z}$, we obtain the function equation

$$\lambda\Big(-\frac{1}{\tau}\Big) = 1 - \lambda(\tau). \tag{4.38}$$

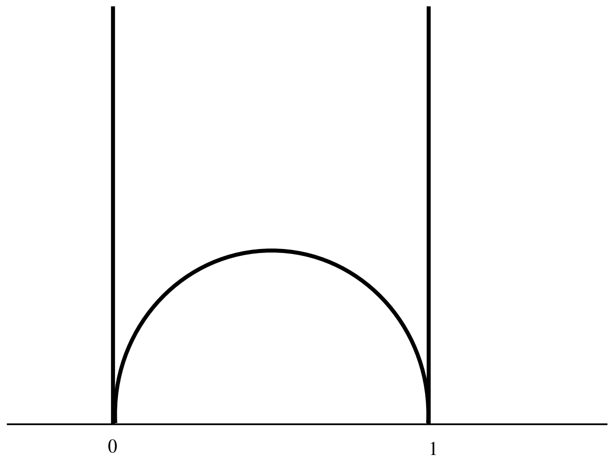

Figure 4.8. The geodesic triangle associated to λ

As far as the action on λ is concerned (meaning the right-hand sides of these relations), the full group in (4.36) generates the so-called *anharmonic group*

$$\Big\{\lambda,\ \frac{1}{1-\lambda},\ \frac{\lambda-1}{\lambda},\ \frac{1}{\lambda},\ \frac{\lambda}{\lambda-1},\ 1-\lambda\Big\}.$$

We leave it to the reader to work out the left-hand sides of the corresponding functional equations. For example, combining (4.37) and (4.38) yields the expression

$$\lambda\big(1-\frac{1}{\tau}\big)=\frac{1}{1-\lambda(\tau)}. \tag{4.39}$$

Proposition 4.24. *The modular function λ takes the geodesic triangle in $\mathbb{H}$ with vertices $0, 1, \infty$ (see Figure 4.8) bijectively onto the upper half-plane. It is continuous on the closure of this triangle and the boundary is mapped to the real axis. The points $0, 1, \infty$ are mapped onto $0, \infty, 1$, respectively.*

Proof. We may assume that $\omega_1 = \tau$ and $\omega_2 = 1$. Then from (4.17) we conclude that

$$\begin{aligned} e_1 - e_2 &= \sum_{n,m\in\mathbb{Z}} \Big[\big((n+\tfrac{1}{2})\tau+m\big)^{-2} - \big(n\tau+m+\tfrac{1}{2}\big)^{-2}\Big], \\ e_3 - e_2 &= \sum_{n,m\in\mathbb{Z}} \Big[\big((n+\tfrac{1}{2})\tau+m+\tfrac{1}{2}\big)^{-2} - \big(n\tau+m+\tfrac{1}{2}\big)^{-2}\Big]. \end{aligned} \tag{4.40}$$

These sums are absolutely and uniformly convergent in any region of the form $\operatorname{Im}\tau > \delta > 0$. Moreover, we may freely shift n, m by any fixed integer amount without changing anything. The sums are even in τ, thus, they remain the same if we replace τ with $-\tau$. In particular, on $\operatorname{Re}\tau = 0$ these expression are purely real-valued, and so is $\lambda(\tau)$. From (4.37) it follows that the same holds on $\operatorname{Re}\tau = 1$. Similarly, $\tau \mapsto \frac{\tau}{1+\tau}$ takes the upper half-plane to itself and $\operatorname{Re}\tau = 0, \operatorname{Im}\tau > 0$ onto the circular arc in Figure 4.8, and one checks that

$$\lambda\Big(\frac{\tau}{1+\tau}\Big) = 1 - \frac{1}{\lambda(\tau)},$$

so that the circular arc is indeed mapped onto the real axis again. So care, however, is needed in dealing with the points $0, 1, \infty$ with regard to the series (4.40) (the first two cannot be plugged into these series since it is illegitimate to evaluate them for real τ, whereas ∞ involves an interchange of limits).

Recall (2.24), i.e.,

$$\frac{\pi^2}{\sin^2(\pi z)} = \sum_{n\in\mathbb{Z}} (z-n)^{-2} \qquad \forall\, z \in \mathbb{C}\setminus\mathbb{Z}.$$

Thus,

$$\begin{aligned} e_1 - e_2 &= \pi^2 \sum_{n\in\mathbb{Z}} \Big[\frac{1}{\sin^2((n+\frac{1}{2})\tau\pi)} - \frac{1}{\cos^2(n\tau\pi)}\Big], \\ e_3 - e_2 &= \pi^2 \sum_{n\in\mathbb{Z}} \Big[\frac{1}{\cos^2((n+1/2)\tau\pi)} - \frac{1}{\cos^2(n\tau\pi)}\Big], \end{aligned} \tag{4.41}$$

and we read off from $n = 0$ that both of these sums converge to $-\pi^2$ as $\operatorname{Im}\tau \to \infty$. This settles the claim that $\lambda(\tau) \to 1$ as $\operatorname{Im}\tau \to \infty$. But by the functional equations for λ the limits $\tau \to 0$ and $\tau \to 1$ along the vertical direction now also follow. This settles all the claims apart from the bijection onto the upper half-plane. The latter follows by means of the argument principle and we leave this to the reader (see Problem 4.11). □

By the proposition, the modular function λ furnishes us with an explicit Riemann map from the geodesic triangle in Figure 4.8 to the upper half-plane. Moreover, if we reflect this map across the imaginary axis by means of Schwarz reflection, then we obtain a map which is a bijection from the union of the two geodesic triangles $\Delta(-1, 0, \infty) \cup \Delta(0, 1, \infty)$ onto $\mathbb{C}\setminus\{0, 1\}$. We have thus carried out an explicit construction of the map which we encountered in Section 4.7 (in the upper-half plane instead of the Poincaré disk, which, however, makes no difference).

Notes

This chapter sets the stage for the remainder of this textbook. Of particular importance is the question about nonconstant global meromorphic functions on any given Riemann surface. As we pointed out by means of the hyper-elliptic curves (which furnish examples of branched cover of the Riemann sphere of genus at least 2), the answer to this question is far from obvious. To appreciate the difficulty, note that branched covers of S^2 of any genus can be constructed by elementary topology, without any reference to complex variables. There is no obvious way of placing a meromorphic function on such an object. We shall devote a substantial part of the remainder of this text to this very question. It naturally leads into Hodge theory, which we develop to the extent needed in this text in Chapter 6.

Radó's countability theorem, which we mentioned after the definition of Riemann surfaces, has an interesting history. Its precursor is the Poincaré-Volterra theorem. This result says that we can only generate countably many different power series with center $z = a$ by all possible analytic continuations from a given one at $z = a$. See Example 7) in Section 4.2 above. For a fascinating account of the history of this result, see the papers [**45**] and [**82**]. These articles further explain the wider context in which the Poincaré-Volterra theorem developed. For example, Poincaré himself regarded Riemann surfaces as more of a heuristic geometric device, invaluable for building intuition and for experimentation, but not rigorous and acceptable for proofs. For rigorous verification he regarded Weierstrass' theory based on power series as the only safe and acceptable method. The paper by Majstrenko [**62**] carries out such a Weierstrassian approach to the Poincaré-Volterra theorem. Today of course this seems strange, but the modern reader needs to keep in mind that the rigorous development of Riemann surfaces and of the concept of a manifold only occurred in the 20th century. See pages 185–190 in Forster's book [**29**] for a modern formulation and proof of both the Poincaré-Volterra and Radó theorems.

Isothermal coordinates, which played a central role in Example 4) of Section 4.2, have been investigated and used in differential geometry since Gauss, who constructed them for analytic metrics. For those metrics we may solve the PDEs (4.5), (4.6), by means of power series; in other words, we can invoke the Cauchy-Kovalevskaya theorem (see Folland's book [**27**], for example). A beautiful timeless paper on the construction of isothermal coordinates for much more general metrics is the one by S. S. Chern [**11**].

Example 6) of Section 4.2 barely scratches the surface. For concrete examples of algebraic curves such as the Fermat curve, rational curves, and conics, cubics, see Walker's book [**85**].

This chapter also has the purpose to place some classical topics such as elliptic functions and the Picard theorems into a more geometric context, rather than including them in Chapter 2, say. Sometimes these results are presented from a purely computational point of view (which is perhaps how they were discovered), without reference to their basic and simple geometrical underpinnings. At least to the author's mind this leads to a substantial loss of transparency.

The classical point of view of Riemann surfaces, which is connected with the practical problem of computing or integrals involving algebraic functions, is beautifully represented in Hancock's timeless book [**41**], as well as Felix Klein's wonderful exposition [**51**]. See, in particular, Article 153 in [**41**] where part of Dirichlet's obituary for Jacobi is reproduced. Dirichlet explicitly writes there that Abel and Jacobi had the fruitful and most important idea to view the integration limits as function of the elliptic integral rather than the other way around. While sources such as these are mostly forgotten today, they not only reveal the concrete motivations often deriving from practical problems in mechanics or geometry, but also present calculations and geometrical arguments which ultimately lead to the modern abstractions.

At least to the author's mind much of the "antiquated" material in [**41, 51**] has not lost any of its value. It would seem that the student of today, who is, more often than not, forced to absorb mathematics in the reverse historical order (for example, by learning complex line bundles and Serre duality without ever being asked to compute an integral of some algebraic function), would greatly benefit from occasionally returning to the concrete foundations on which the abstract machinery was built.

It was mentioned in this chapter that $\mathbb{C}^*, \mathbb{D}^*, \Delta_r$ and $\mathbb{C}/\Lambda$ where Λ is a lattice, is a complete list of Riemann surfaces whose fundamental group is nontrivial and abelian. For details about this assertion, see Chapter 8 as well as [**23**], Chapter IV.6.

For more on the fundamental normality test (FNT), and its version for meromorphic functions which involves omitting three values, see Schiff [**74**]. The relation between constancy of entire functions omitting two values and the FNT, which is also formulated in terms of omission of two values, is not accidental. The *Bloch principle* is a general manifestation of this relation; see loc. cit.

Fuchsian and Kleinian groups comprise a large body of mathematics, not to mention modular and automorphic functions. The problems below introduce the reader to some easy results on Fuchsian groups, which will play an essential role in the uniformization theorem; see Chapter 8. Some of the problems below introduce geometric concepts relevant to the study of these groups, and ask the reader to verify certain simple properties. An excellent introduction to this topic is the book by Katok [**50**]. See also Ford's classical but timeless book [**28**].

4.9. Problems

Problem 4.1. Show that all nontrivial discrete subgroups of $\mathrm{Aut}(\mathbb{C})$ *that have no fixed point* are either $\langle z \mapsto z + \omega \rangle$ with $\omega \neq 0$, or $\langle z \mapsto z + \omega_1, z \mapsto z + \omega_2 \rangle$ with $\omega_1 \neq 0, \omega_2/\omega_1 \notin \mathbb{R}$.

Problem 4.2. This exercise revisits fractional linear transformations.

(a) Prove that

$$G = \left\{ \begin{bmatrix} a & \bar{b} \\ b & \bar{a} \end{bmatrix} \;:\; a, b \in \mathbb{C},\ |a|^2 - |b|^2 = 1 \right\}$$

is a subgroup of $\mathrm{SL}(2, \mathbb{C})$ (it is known as $\mathrm{SU}(1,1)$). Establish the group isomorphism $G/\{\pm I\} \simeq \mathrm{Aut}(\mathbb{D})$ in two ways: (i) by showing that each element of G defines a fractional linear transformation which maps $\mathbb{D}$ onto

$\mathbb{D}$; and conversely, that every such fractional linear transformation arises in this way uniquely up to the signs of a, b. (ii) By showing that the map

$$e^{2i\theta}\frac{z-z_0}{1-\bar{z}_0 z} \mapsto \begin{bmatrix} \frac{e^{i\theta}}{\sqrt{1-|z_0|^2}} & \frac{-z_0 e^{i\theta}}{\sqrt{1-|z_0|^2}} \\ -\frac{\bar{z}_0 e^{-i\theta}}{\sqrt{1-|z_0|^2}} & \frac{e^{-i\theta}}{\sqrt{1-|z_0|^2}} \end{bmatrix} \tag{4.42}$$

leads to an explicit isomorphism.

(b) We have established in the text that $\mathrm{Aut}(\mathbb{C}_\infty)$ is the group of all fractional linear transformations, i.e.,

$$\mathrm{Aut}(\mathbb{C}_\infty) = \mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\{\pm\mathrm{id}\}$$

and that each such transformation induces a conformal homeomorphism of S^2 (indeed, the stereographic projection is conformal). The purpose of this exercise is to identify the subgroup G_{rig} of those transformations in $\mathrm{Aut}(\mathbb{C}_\infty)$ which are *isometries* (in other words, rigid motions) of S^2 (viewing $\mathbb{C}_\infty$ as the Riemann sphere S^2). Prove that

$$G_{\mathrm{rig}} \simeq \mathrm{SO}(3) \simeq \mathrm{SU}(2)/\{\pm I\}$$

where

$$\mathrm{SU}(2) = \left\{ \begin{bmatrix} a & -\bar{b} \\ b & \bar{a} \end{bmatrix} \;:\; a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1 \right\}.$$

Find the fractional linear transformation which corresponds to a rotation of S^2 of angle $\frac{\pi}{2}$ about the x_1-axis. For the latter recall how we defined the stereographic projection.

(c) Show that the quaternions can be viewed as the four-dimensional real-vector space spanned by the basis

$$e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad e_2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad e_4 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix},$$

and with the algebra structure being defined via the matrix products of the e_j's (typically, one writes $1, i, j, k$ instead of e_1, e_2, e_3, e_4). Show that in this representation the unit quaternions are nothing but $\mathrm{SU}(2)$ and exhibit a homomorphism Q of the unit quaternions onto the group $\mathrm{SO}(3)$ so that $\ker(Q) = \{\pm 1\}$.

Which rotation does the unit quaternion $\xi_1 + \xi_2 i + \xi_3 j + \xi_4 k$ represent (i.e., what are the axis and angle of rotation)?

Problem 4.3. Show that all Möbius transforms with two distinct fixed points $z_1, z_2 \in \mathbb{C}_\infty$ are of the form

$$\frac{w-z_1}{w-z_2} = K\frac{z-z_1}{z-z_2}, \tag{4.43}$$

where $K \in \mathbb{C} \setminus \{0\}$ is a complex parameter. For the case where $K > 0$ (called *hyperbolic*) demonstrate the action of such a map by means of the family of circles passing through the points z_1, z_2, as well as by means of the orthogonal family of circles.

Do the same when $K = e^{i\theta}$ (called *elliptic*). If K does not fall into either of these classes we call the map *loxodromic*; these maps are compositions of a hyperbolic and an elliptic transformation (with the same pair of fixed points).

The *parabolic* maps are the ones with a unique fixed point, such as translations. Draw figures that demonstrate the action of such maps.

To which of these classes do the maps

$$w = \frac{z}{2z-1}, \quad w = \frac{2z}{3z-1}, \quad w = \frac{3z-4}{z-1}, \quad w = \frac{z}{2-z}$$

belong? If $w = \frac{az+b}{cz+d}$ with $ad - bc = 1$, give a criterion in terms of $a + d$ which determines the class. Explain the relation between eigenvalues and eigenvectors of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ on the one hand, and fixed points $w = z$ on the other hand.

Problem 4.4. This problem combines the analysis of fixed points of transformations in $\mathrm{Aut}(\mathbb{D})$ as presented at the beginning of Section 4.8 with the classification given in the previous problem. Show that the maps with one fixed point inside $\mathbb{D}$ are elliptic, those with a unique fixed point on the unit circle are parabolic, and those with two on the circle are hyperbolic. Reformulate also for the upper half-plane, that is, for $\mathrm{Aut}(\mathbb{H})$.

Problem 4.5. Let $\Lambda \subset \mathbb{C}$ be the lattice (i.e., the discrete subgroup) generated by ω_1, ω_2 which are independent over $\mathbb{R}$. Show that the *Dirichlet polygon*

$$\{z \in \mathbb{C} \,:\, |z| \le |z - \omega| \quad \forall\, \omega \in \Lambda\}$$

is a fundamental region.

Problem 4.6. Let $G < \mathrm{PSL}(2, \mathbb{C})$ be a **Fuchsian** group; see Definition 4.23.

- Prove that a Fuchsian group G operating on the upper half-plane $\mathbb{H}$ is the same as a *discrete subgroup* of $\mathrm{PSL}(2, \mathbb{R})$. The topology here is natural one on 2×2 matrices.
- Show further that this discreteness is the same as each point having a locally finite orbit. This means that for every $z \in \mathbb{H}$ the set $\{gz \mid g \in G\}$ interests any compact subset of $\mathbb{H}$ in only finitely many points.

Problem 4.7. Prove the following properties of Fuchsian groups.

- Show that we may assume that the invariant circle is the real line.

- Show that each transformation in G either has a fixed point in $\mathbb{H}$, which is then necessarily elliptic, or a unique one on the line, which is parabolic, or two fixed points on the line which are the hyperbolic. Thus, there can be no loxodromic fixed points (why?).
- Show that G possesses a fundamental region, and that the boundary of any such region consists of a finite or infinite number of congruent curves.
- What can you say about fixed points of maps in G relative to any fundamental region?
- Give several distinct examples of Fuchsian groups (with the line as fixed circle) and of associated fundamental regions. Do these regions always necessarily lie above the real axis?
- Show that we may always select a fundamental region in such a way that the boundary consists entirely of circular arcs whose centers lie on the real axis.

We remark that **Kleinian groups** are those without a fixed circle.

Problem 4.8. Show that the tori $\mathbb{C}/\Lambda_1$ and $\mathbb{C}/\Lambda_2$ for lattices Λ_j, each generated by two $\mathbb{R}$-independent translations, are conformally equivalent if and only if Λ_1 and Λ_2 are conjugate subgroups in $\mathrm{Aut}(\mathbb{C})$.

Problem 4.9. Show that the classes of tori under conformal equivalence can be naturally identified with the Hausdorff space $\mathbb{H}/\mathrm{PSL}(2,\mathbb{Z})$. As usual, by torus we mean $\mathbb{C}/\Lambda$ where Λ is the group generated by two nontrivial translations $z \mapsto z+\omega_j$, $j=1,2$, with $\omega_1/\omega_2 \notin \mathbb{R}$.

In particular, we see that each lattice is represented by $\tau = \omega_1/\omega_2$ belonging to the closure of the shaded region shown in Figure 4.7. Show that this representation is unique if we include the portion of the boundary given by $\mathrm{Re}\, z = \frac{1}{2}$ and $|z| = 1$ restricted to $0 \le \mathrm{Re}\, z \le \frac{1}{2}$. Show further that to each τ in that region corresponds a choice of two, four, or six bases. Explain the special role of i and $(\pm 1 + i\sqrt{3})/2$.

Problem 4.10. Let $M = \mathbb{C}/\Lambda$ where Λ is the lattice generated by the vectors $\omega_1, \omega_2 \in \mathbb{C}^*$ with $\frac{\omega_1}{\omega_2} \notin \mathbb{R}$. As usual $\wp$ denotes the Weierstrass function on M. Suppose that $f \in \mathcal{M}(M)$ has degree two. Prove that there exists

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{C})$$

and $w \in \mathbb{C}$ such that

$$f(z) = \frac{a\wp(z-w)+b}{c\wp(z-w)+d}.$$

Problem 4.11. Finish the proof of Proposition 4.24, i.e., use the argument principle to prove that λ takes the region in Figure 4.8 bijectively onto the upper half-plane. Also, show that the map is monotone along the boundary.

Problem 4.12. Suppose N is a Riemann surface, and let N_1 be a compact manifold with boundary so that $N \subset N_1$ and the closure of N in N_1 is N_1. Thus, for every $p \in \partial N$ there exists a neighborhood U of p in N_1 and a map $\phi : U \to \mathbb{R}^2_+$ such that ϕ takes U homeomorphically onto $\mathbb{D} \cap \{\operatorname{Im} z \geq 0\}$. Moreover, we demand that the transition maps between such charts are conformal on $\operatorname{Im} z > 0$. Prove that then $N_1 \subset M$ where M is a Riemann surface. In other words, N can be extended to a strictly larger Riemann surface.

Problem 4.13. Let M be a compact Riemann surface and suppose $S \subset M$ is discrete. Assume $f : M \setminus S \to \mathbb{C}$ is analytic and nonconstant. Show that the image of $M \setminus S$ under f is dense in $\mathbb{C}$.

Problem 4.14. Prove that the special case of the Schwarz-Christoffel formula (2.30) takes the upper half-plane onto a rectangle. Explain how this fits together with the doubly-periodic elliptic integrals discussed in this chapter.

Problem 4.15. Let M, N be compact Riemann surfaces and suppose

$$\Phi : M \setminus \mathcal{S} \to N \setminus \mathcal{S}'$$

is an isomorphism where $\mathcal{S}, \mathcal{S}'$ are finite sets. Show that Φ extends to an isomorphism from $M \to N$.

Problem 4.16. Provide the details for the general case in Lemma 4.21.