Advanced Complex Analysis

Course Notes — Harvard University — Math 213a

C. McMullen

December 2, 2025

Contents

Forward

Complex analysis is a nexus for many mathematical fields, including:

1. Algebra (theory of fields and equations);

2. Algebraic geometry and complex manifolds;

3. Theory of Riemann surfaces;

4. Topology (notion of a manifold; surfaces, covering spaces, mapping–class groups; Schoenflies theorem);

5. Geometry (Platonic solids; flat tori; hyperbolic manifolds of dimensions two and three);

6. Lie groups, discrete subgroups and homogeneous spaces (e.g. $\mathbb{H}/\operatorname{SL}_2(\mathbb{Z})$;

7. Dynamics (iterated rational maps, Kleinian groups);

8. Number theory and automorphic forms (elliptic functions, zeta functions);

9. Several complex variables;

10. Real analysis and PDE (harmonic functions, elliptic equations and distributions).

This course covers some basic material on both the geometric and analytic aspects of complex analysis in one variable.

Prerequisites: Background in real analysis and basic differential topology (such as covering spaces and differential forms), and a first course in complex analysis.

Exercises

(These exercises are review.)

1. Let $T \subset \mathbb{R}^3$ be the spherical triangle defined by $x^2 + y^2 + z^2 = 1$ and $x, y, z \geq 0$. Let $\alpha = z\,dx\,dz$.

   (a) Find a smooth 1-form β on $\mathbb{R}^3$ such that $\alpha = d\beta$.

(b) Define consistent orientations for T and ∂T.

(c) Using your choices in (ii), compute $\int_T \alpha$ and $\int_{\partial T} \beta$ directly, and check that they agree. (Why should they agree?)

2. Let $f(z) = (az + b)/(cz + d)$ be a Möbius transformation. Show the number of rational maps $g : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that

$$g(g(g(g(g(z))))) = f(z)$$

is 1, 5 or ∞. Explain how to determine which alternative holds for a given f.

3. Let $\sum a_n z^n$ be the Taylor series for $\tanh(z)$ at $z = 0$.

   (a) What is the radius of convergence of this power series?

   (b) Show that $a_5 = 2/15$.

   (c) Give an explicit value of N such that $\tanh(1)$ and $\sum_0^N a_n$ agree to within $\epsilon = 10^{-1000}$. Justify your answer.

4. Let $f : U \to V$ be a proper local homeomorphism between a pair of connected open sets $U, V \subset \mathbb{C}$. Prove that f is a covering map. (Here *proper* means that $f^{-1}(K)$ is compact whenever $K \subset V$ is compact.)

5. Let $f : \mathbb{C} \to \mathbb{C}$ be given by a polynomial of degree 2 or more. Let

$$V_1 = \{f(z) \ : \ f'(z) = 0\} \subset \mathbb{C}$$

be the set of critical values of f, let $V_0 = f^{-1}(V_1)$, and let $U_i = \mathbb{C} - V_i$ for $i = 0, 1$. Prove that $f : U_0 \to U_1$ is a covering map.

6. Give an example where U_0/U_1 is a normal (or Galois) covering, i.e. where $f_*(\pi_1(U_0))$ is a normal subgroup of $\pi_1(U_1)$.

7. Let $H = \mathbb{R}[i, j, k] \cong \mathbb{R}^4$ denote the real quaternions, with $ij = k$ and $i^2 = j^2 = k^2 = -1$. True or false: For any nonzero $x \in H$, the transformation of $\mathbb{R}^4$ defined by $f(y) = xy$ is conformal.

(A conformal map is the composition of an isometry of $\mathbb{R}^4$ with multiplication by a real number.)

1 Basic complex analysis

Bitte vergisse alles, was Du auf der Schule gelernt hast; denn Du hast es nicht gelernt. — Landau

We begin with an overview of basic facts about the complex plane and analytic functions.

Some notation. The complex numbers will be denoted $\mathbb{C}$. We let $\Delta, \mathbb{H}$ and $\widehat{\mathbb{C}}$ denote the unit disk $|z| < 1$, the upper half plane $\mathrm{Im}(z) > 0$, and the Riemann sphere $\mathbb{C} \cup \{\infty\}$. We write $S^1(r)$ for the circle $|z| = r$, and S^1 for the unit circle, each oriented counter-clockwise. We also set $\Delta^* = \Delta - \{0\}$ and $\mathbb{C}^* = \mathbb{C} - \{0\}$.

1.1 Algebraic and analytic functions

The complex numbers are formally defined as the field $\mathbb{C} = \mathbb{R}[i]$, where $i^2 = -1$. They are represented in the Euclidean plane by $z = (x, y) = x + iy$. There are two square-roots of -1 in $\mathbb{C}$; the number i is the one with positive imaginary part.

An important role is played by the Galois involution $z \mapsto \overline{z}$. We define $|z|^2 = N(z) = z\overline{z} = x^2 + y^2$. (Compare the case of a real quadratic field, where $N(a + b\sqrt{d}) = a^2 - db^2$ gives an *indefinite* form.) Compatibility of $|z|$ with the Euclidean metric justifies the identification of $\mathbb{C}$ and $\mathbb{R}^2$. We also see that z is a field: $1/z = \overline{z}/|z|$.

It is also convenient to describe complex numbers by polar coordinates

$$z = [r, \theta] = r(\cos\theta + i\sin\theta).$$

Here $r = |z|$ and $\theta = \arg z \in \mathbb{R}/2\pi\mathbb{Z}$. (The multivaluedness of $\arg z$ requires care but is also the ultimate source of powerful results such as Cauchy's integral formula.) We then have

$$[r_1, \theta_1][r_2, \theta_2] = [r_1 r_2, \theta_1 + \theta_2].$$

In particular, the linear maps $f(z) = az + b$, $a \neq 0$, of $\mathbb{C}$ to itself, preserve angles and orientations.

This formula should be proved *geometrically*: in fact, it is a consequence of the formula $|ab| = |a||b|$ and properties of similar triangles. It can then

be used to derive the addition formulas for sine and cosine (in Ahlfors the reverse logic is applied).

Roots of unity. We let

$$\zeta_n = \exp(2\pi i/n)$$

denote the unique complex number of absolute value one with $\arg(z) = 2\pi/n$. Clearly $\zeta_n^n = 1$ by the formulas above, and no smaller power of ζ_n is one. Thus ζ_n is a *primitive* nth root of unity. (The other primitive roots of unity are given by ζ_n^k, $\gcd(k, n) = 1$.)

This simple formula shows that in $\mathbb{C}$ we can solve, not only $x^2 = -1$, but $x^n = -1$ for any value of n.

Algebraic closure. The previous example illustrates a critical feature of the complex numbers: they are an *algebraically closed* field. That is, every nontrivial polynomial $p(z) \in \mathbb{C}[z]$ has a root in z. (A proof will be reviewed below).

Classically, the complex numbers were introducing in the course of solving *real* cubic equations. Consider, for example, the equation

$$x^3 = 3x + a.$$

(A general cubic can be reduced to this form.) Note that if $x = u + 1/u$, then $x^3 = 3x + u^3 + 1/u^3$. Thus we can solve the cubic by first solving $v + 1/v = a$, then setting $x = u + 1/u$ where $u^3 = v$.

But the first step may already lead to the complex numbers, even when a is real. For example, when $a = 1$, we find $v = \exp(2\pi i/6) = (1 + \sqrt{-3})/2$. Taking the cube–root gives $u = \exp(2\pi i/18)$, and $x = 2\cos \pi/9$. (This result can also be verified using the triple angle formula.)

Analytic functions. Let U be an open set in $\mathbb{C}$ and $f : U \to \mathbb{C}$ a function. We say f is *analytic* if

$$f'(z) = \lim_{t \to 0} \frac{f(z+t) - f(z)}{t}$$

exists for all $z \in U$. It is crucial here that t approaches zero through arbitrary values in $\mathbb{C}$. Remarkably, this condition implies that f is a smooth (C^∞) function. For example, polynomials are analytic, as are rational functions away from their poles.

Note that any *real linear* function $\phi : \mathbb{C} \to \mathbb{C}$ has the form $\phi(v) = av + b\overline{v}$. The condition of analytic says that $Df_z(v) = f'(z)v$; in other words, the $\overline{v}$ part is absent.

To make this point systematically, for a general C^1 function $F : U \to \mathbb{C}$ we define

$$\frac{dF}{dz} = \frac{1}{2}\left(\frac{dF}{dx} + \frac{1}{i}\frac{dF}{dy}\right) \quad \text{and} \quad \frac{dF}{d\overline{z}} = \frac{1}{2}\left(\frac{dF}{dx} - \frac{1}{i}\frac{dF}{dy}\right).$$

We then have

$$DF_z(v) = \frac{dF}{dz}v + \frac{DF}{d\overline{z}}\overline{v}.$$

1 and 2-forms; dz and $d\overline{z}$. We can also write the complex-valued 1-form dF as

$$dF = \partial F + \overline{\partial}F = \frac{dF}{dz}\,dz + \frac{dF}{d\overline{z}}\,d\overline{z}$$

Thus F is analytic iff $\overline{\partial}F = 0$; these are the *Cauchy-Riemann* equations.

We note that $(d/d\overline{z})z^a\overline{z}^b = bz^a\overline{z}^{b-1}$; a polynomial $p(z,\overline{z})$ behaves as if these variables are independent. In fact, we have an isomorphism of $\mathbb{C}$-algebras,

$$\mathbb{C}[x, y] \cong \mathbb{C}[z, \overline{z}].$$

The operators ∂ and $\overline{\partial}$ extend naturally to maps from 1-forms to 2-forms; they satisfy, for any function f, $\partial(f(z)\,d\overline{z}) = (\partial f)\,d\overline{z}$, $\partial(f(z)\,dz) = 0$. Then $d = \partial + \overline{\partial}$ on 1-forms as well.

When using exterior algebra for computations, it is important to note that

$$dz \wedge d\overline{z} = (dx + i\,dy) \wedge (dx - i\,dy) = -2i\,dx \wedge dy.$$

Thus the integral of $f(z)$ with respect to the usual *area measure* on a region U is given by

$$\int f(z)\,dx \wedge dy = \frac{i}{2}\int f(z)\,dz \wedge d\overline{z}.$$

Sources of analytic functions.

1. **Polynomials and rational functions.** Using addition and multiplication we obtain naturally the polynomial functions $f(z) = \sum_0^n a_n z^n$: $\mathbb{C} \to \mathbb{C}$. The ring of polynomials $\mathbb{C}[z]$ is an integral domain and a unique factorization domain, since $\mathbb{C}$ is a field. Indeed, since $\mathbb{C}$ is algebraically closed, fact every polynomial factors into linear terms.

   It is useful to add the allowed value ∞ to obtain the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. Then rational functions (ratios $f(z) = p(z)/q(z)$ of relatively prime polynomials, with the denominator not identically zero)

determine rational maps $f : \mathbb{C} \to \mathbb{C}$. The rational functions $\mathbb{C}(z)$ are the same as the field of fractions for the domain $\mathbb{C}[z]$. We set $f(z) = \infty$ if $q(z) = 0$; these points are called the *poles* of f.

2. **Algebraic functions.** Beyond the rational and polynomial functions, the analytic functions include *algebraic* functions such that $f(z) = \sqrt{z^2 + 1}$. A general algebraic function $f(z)$ satisfies

$$P(f) = \sum_0^d a_n(z) f(z)^n = 0$$

for some rational functions $a_n(z)$; these functions occur natural when one considers an algebraic extension $\mathbb{C}(f)$ of $\mathbb{C}(z)$ of degree d. Such functions are generally *multivalued*, so to treat them as analytic functions of z, we need to choose a branch cut (e.g. a simply–connected region where the discriminant of $P(f)$ is never 0 or ∞).

The desire to remove the clumsiness of a branch cut leads to the theory of Riemann surfaces.

3. **Differential equations.** Analytic functions also arise when one solves *differential equations*. Even equations with constant coefficients, like $y'' + y = 0$, can give rise to transcendental functions such as $\sin(z)$, $\cos(z)$ and e^z. Here are some useful facts about these familiar functions when extended to $\mathbb{C}$:

$$
\begin{aligned}
|\exp(z)| &= \exp \operatorname{Re} z \\
\cos(iz) &= \cosh(z) \\
\sin(iz) &= i \sinh(z) \\
\cos(x + iy) &= \cos(x)\cosh(y) - i\sin(x)\sinh(y) \\
\sin(x + iy) &= \sin(x)\cosh(y) + i\cos(x)\sinh(y).
\end{aligned}
$$

In particular, the apparent *boundedness* of $\sin(z)$ and $\cos(z)$ fails badly as we move away from the real axis, while $|e^z|$ is actually very small in the halfplane $\operatorname{Re} z \ll 0$.

4. **Integration.** A special case of course is integration. While $\int (x^2 + ax + b)^{-1/2} \, dx$ can be given explicitly in terms of trigonometric functions, already $\int (x^3 + ax + b)^{-1/2} \, dx$ leads one into elliptic functions; and

higher degree polynomials lead one to hyperelliptic surfaces of higher genus. Note that the 'periodicity' of the function in increases from $\mathbb{Z}$ (trigonometric) to $\mathbb{Z}^2$ (elliptic) to $H_1(\Sigma_g, \mathbb{Z})$ (hyperelliptic).

5. **Power series.** Analytic functions can be given concretely, locally, by power series such as $\sum a_n z^n$. Conversely, suitable coefficients determine analytic functions; for example, $e^z = \sum z^n/n!$.

    For a more interesting example one can consider the partition function $p(n)$ (satisfying $p(5) = 7$ because $5 = 1+1+1+1+1 = 1+1+1+2 = 1+1+2+2 = 3+2 = 1+4 = 5$), whose generating function satisfies

$$\sum_0^\infty p(n)z^n = \prod_1^\infty (1 - z^n)^{-1},$$

    for $z \in \Delta$. We remark that $\prod(1 + a_n)$ converges if $\sum |a_n|$ converges. (It is known that $p(n) = O(\exp(\pi\sqrt{2n/3}).$ and this is the right order of exponential growth.)

6. **Riemann surfaces and automorphy.** Another natural source of complex analytic functions is functions that satisfy invariant properties such as $f(z + \lambda) = f(z)$ for all $\lambda \in \Lambda$, a lattice in $\mathbb{C}$; or $f(g(z)) = f(z)$ for all $g \in \Gamma \subset \operatorname{Aut}(\mathbb{H})$.

    The *elliptic modular functions* $f : \mathbb{H} \to \mathbb{C}$ with have the property that $f(z) = f(z + 1) = f(-1/z)$, and hence $f((az + b)/(cz + d)) = f(z)$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \operatorname{SL}_2(\mathbb{Z})$.

7. **Geometric function theory.** A complex analytic function can be specified by a domain $U \subset \mathbb{C}$; we will see that for simply-connected domains (other than $\mathbb{C}$ itself), there is an essentially unique analytic homeomorphism $f : \Delta \to U$. When U tiles $\mathbb{H}$ or $\mathbb{C}$, this is related to automorphic functions; and when ∂U consists of lines or circular arcs, one can also give a differential equation for f.

8. **Limits.** We can also define analytic functions by taking limits of polynomials or other known functions. (It is a remarkable fact, which we will prove, that a *uniform* limit of analytic functions is not just continuous, but analytic.)

For example, consider the formula:

$$e^z = \lim(1 + z/n)^n.$$

The triangle with vertices $0, 1$ and $1 + i\theta/n$ has a hypotenuse of length $1 + O(1/n^2)$ and an angle at 0 of $\theta + O(1/n^2)$. Thus one finds *geometrically* that $z_n = (1 + i\theta/n)^n$ satisfies $|z_n| \to 1$ and $\arg z_n \to \theta$; in other words,

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

In particular, $e^{\pi i} = -1$ (Euler, Kimbune).

1.2 Complex integration and power series

We now return to the general theory of analytic functions. A *region* $U \subset \mathbb{C}$ is a connected open set. When we integrate a function $f(z)$ over ∂U, we always assume that:

1. The region U has compact closure $\overline{U}$;

2. The boundary of U is a finite union of smooth arcs (it is *piecewise smooth*), and

3. The function $f(z)$ is continuous on $\overline{U}$.

For simplicity, we will say that $f : \overline{U} \to \mathbb{C}$ is *analytic* if $f|U$ is analytic and $f|\overline{U}$ is continuous. Note that U inherits a natural orientation from $\mathbb{C}$, and ∂U inherits an orientation from U. Thus we can naturally integrate differential forms over these objects.

We then have one of the central theorems in complex analysis:

Theorem 1.1 (Cauchy) *For any analytic function $f : \overline{U} \to \mathbb{C}$, we have $\int_{\partial U} f(z)\, dz = 0$.*

Remark. It is critical to know the *definition* of such a path integral. (For example, $f(z) = 1$ is analytic, its average over the circle is 1, and yet $\int_{S^1} 1\, dz = 0$; why is this?)

If $\gamma : [a, b] \to \mathbb{C}$ parameterizes an arc, then we define

$$\int_\gamma f(z)\, dz = \int_a^b f(\gamma(t))\, \gamma'(t)\, dt.$$

Alternatively, we choose a sequence of points $z_0, \ldots, z_n$ close together along γ, and then define

$$\int_\gamma f(z) = \lim \sum_1^n f(z_i)(z_i - z_{i-1}).$$

This should not be confused with the integral with respect to arclength:

$$\int_\gamma f(z) \, |dz| = \lim \sum f(z_i)|z_i - z_{i-1}|.$$

Note that the former *depends* on a choice of orientation of γ, while the latter does not.

Proof of Cauchy's formula: (i) observe that $d(f \, dz) = (\bar{\partial} f) d\bar{z} \, dz = 0$ and apply Stokes' theorem.

(ii — Goursat). Cut the region U into small squares, observe that on these squares $f(z) \approx az + b$, and use the fact that $\int_{\partial U} (az + b) \, dz = 0$. ∎

The first proof implicitly assumes f is C^1, while the second does not. More on the Cauchy–Riemann equations and with minimal smoothness assumptions can be found in [GM]. We will soon see that $f(z)$ analytic $\implies$ $f'(z)$ is analytic, and thus any analytic function is C^∞.

Goursat: Why rectangles? Goursat's argument just shows that Cauchy's theorem holds for a rectangle. Why is this sufficient? The reason is that, with this tool at hand, one can show — by integrating — that if f is analytic in a disk, then $f(z) = F'(z)$ for some analytic function F. That is, f has a (local) antiderivative — which is smoother than F. Now once F is smooth enough, we are allowed to apply Stokes' theorem, and then use Cauchy's integral formula (below) to show F can be expressed by a power series — and then the same is true for all of its derivatives. In particular f can be expressed by a power series, so it too is smooth.

Complement: periods of a holomorphic 1–form. One can also apply Cauchy's theorem to smooth paths or loops, instead of just boundaries of regions. The fact that the form $f(z) \, dz$ is closed leads to:

Theorem 1.2 *Let $f : U \to \mathbb{C}$ be an analytic function, let b be a basepoint in U, and let $\gamma : S^1 \to U$ be a smooth loop based on b. Then*

$$p(\gamma) = \int_\gamma f(z) \, dz$$

9

depends only on the homotopy class of γ, and defines a group homomorphism

$$p : \pi_1(U, b) \to \mathbb{C}$$

called the period map.

Proof. The main point is the first statement. Let $B = S^1 \times [0, 1]$. A smooth isotopy $\gamma_t : (S^1, *) \to (U, b)$, $t \in [0, 1]$, gives a smooth map $G : B \to U$, and $p(\gamma_0) - p(\gamma_1)$ gives the integral of $\alpha = G^*(f(z)\, dz)$ over ∂B. Since $f(z)$ is analytic, $\omega = f(z)\, dz$ is closed, and hence $\alpha = G * (\omega)$ is closed as well — we have $d(\alpha) = d(G^*\omega) = G^*(d\omega) = 0$. Thus $\int_{\partial B} \alpha = \int_B d\alpha = 0$, and thus $p(\gamma_0) = p(\gamma_1)$. ∎

Corollary 1.3 *If U is simply–connected, then $F(b) = \int_a^b f(z)\, dz$ is well–defined for all $b \in U$. Indeed, $F : U \to \mathbb{C}$ is the unique analytic function satisfying $F'(z) = f(z)$ and $F(a) = 0$.*

Example: the logarithm. For $f(z) = 1/z$ on $U = \mathbb{C}^*$, the period map gives an isomorphism, $\pi_1(U, 1) \cong 2\pi i\mathbb{Z}$. This reflects the fact that integrating dz/z gives the multi–valued function,

$$\log(z) = \log|z| + i\arg(z),$$

and $\arg(z)$ changes by 2π along a loop that winds $+1$ times around $z = 0$.

On the other hand, if $1 \in U$ and $U \subset \mathbb{C}^*$ is *simply–connected*, then there exists a *unique* analytic branch of the logarithm, $\log : U \to \mathbb{C}$, such that $\log(1) = 0$.

Cauchy's integral formula: Differentiability and power series. Because of Cauchy's theorem, only one integral has to be explicitly evaluated in complex analysis (hence the forgetability of the definition of the integral). Namely, setting $\gamma(t) = e^{it}$ we find, for any $r > 0$,

$$\int_{S^1(r)} \frac{1}{z}\, dz = 2\pi i.$$

By integrating between ∂U and a small loop around $p \in U$, we then obtain *Cauchy's integral formula:*

Theorem 1.4 *For any $p \in U$ we have*

$$f(p) = \frac{1}{2\pi i} \int_{\partial U} \frac{f(z)\,dz}{z - p}.$$

Aside: Distributions. We note that in the sense of distributions, we have

$$\frac{d}{d\bar{z}} \frac{1}{z} = \pi \delta_0, \tag{1.1}$$

and the evaluation of the integrals above can be based on this fact.

In more detail: recall that a distribution on $\mathbb{C}$ is an element of the dual of the space of compactly support smooth functions, with the pairing defined for $f \in C_0^\infty(\mathbb{C})$ and $u \in L_{\text{loc}}^1(\mathbb{C})$ by

$$\langle f, u \rangle = \int_{\mathbb{R}^2} f(x, y) u(x, y)\,dxdy.$$

The action of differentiation on distributions is defined by formally applying integration by parts, to move the operator from u to f.

Now for each $r > 0$ define

$$u_r(z) = \begin{cases} 1/z & \text{if } |z| > r, \text{ and} \\ \bar{z}/r^2 & \text{if } |z| \le r. \end{cases}$$

This function is continuous on $S^1(r)$, since $|z|^2 = z\bar{z} = r^2$ there. Note that $u_r \to u_0(z) = 1/z$ in $L^1(\Delta)$, and $v_r = du_r/d\bar{z}$ is the constant $1/r^2$ on $\Delta(r)$, so it has total mass π. Clearly $v_r \to \pi \delta_0$, justifying (1.1) .

Keeping in mind that $(i/2)dz \wedge d\bar{z} = dx\,dy$, we have

$$\bar{\partial}(dz/z) = \pi \delta_0 \, d\bar{z}\,dz = 2\pi i \delta_0 \, dx\,dy,$$

we obtain Cauchy's integral formula by the calculation:

$$\int_{\partial U} \frac{f(z)\,dz}{z - p} = \int_U f(z) \bar{\partial} \frac{dz}{z - p} = \int_U f(z) 2\pi i \delta_p \, dx\,dy = 2\pi i f(p).$$

Power series. Weierstrass approached analytic functions through power series rather than path integrals.

Let a_n be a sequence of complex numbers, and let

$$A = \limsup |a_n|^{1/n}.$$

11

This familiar quantity measures the growth rate of $|a_n|$; roughly speaking, $|a_n|$ grows at most like A^n. A precise statement is that for all $\epsilon > 0$, we have

$$|a_n| \le (A+\epsilon)^n \text{ for all } n \gg 0, \text{ and } \text{ and } |a_n| \ge (A-\epsilon)^n \text{ for infinitely many } n..$$

We then have:

Theorem 1.5 *The series $\sum a_n z^n$ defines an analytic function $f(z)$ on the disk $|z| < 1/A$, and it diverges for $|z| > A$.*

Here it is clear that $f'(z) = \sum n a_n z^n$ exists and is analytic on the same disk as $f(z)$.

Examples. The most important power series in mathematics is the geometric series:

$$1 + z + z^2 + \cdots = \sum_0^\infty z^n = \frac{1}{1-z},$$

convergent for $|z| < 1$. By integrating this series, with $1 - z$ replaced by $1 + z$, we obtain:

$$\log(1+z) 1 - z + z^2/2 + \cdots = \sum_0^\infty \frac{(-1)^k z^k}{k},$$

and replacing $1 - z$ with $1 + z^2$, we obtain:

$$\tan^{-1}(z) = z - z^3/3 + z^5/5 - \cdots = \sum_0^\infty \frac{(-1)^k z^{2k+1}}{2k+1}.$$

Derivatives via integration. We now connect Cauchy and Weierstrass to show that simply the assumption that $f'(z)$ exists in U implies that $f(z)$ can be locally expressed by a power series.

Note that the integrand in Cauchy's formula depends on p only through the rational function $1/(z - p)$, which is infinitely differentiable. (It is the *convolution* of $f|\partial U$ and $1/z$.) So we conclude that $f(p)$ itself is infinitely differentiable, indeed, it is approximated by a sum of rational functions with poles on ∂U. In particular, differentiating under the integral, we obtain:

$$\frac{f^{(k)}(p)}{k!} = \frac{1}{2\pi i} \int_{\partial U} \frac{f(z)\,dz}{(z-p)^{k+1}}. \tag{1.2}$$

12

If $d(p, \partial U) = R$, the length of ∂U is L and $\sup_{\partial U} |f| = M$, then this gives the bound:
$$|a_k| = \left| \frac{f^{(k)}(p)}{k!} \right| \le \frac{ML}{2\pi R^{k+1}}.$$

In particular, $\sum a_k(z-p)^k$ has radius of convergence at least R, since $\limsup |a_k|^{1/k} < 1/R$. This suggests that f is represented by its power series, and indeed this is the case:

Theorem 1.6 *If f is analytic on $B(p, R)$, then $f(z) = \sum a_k(z - p)^k$ on this ball.*

Proof. We can reduce to the case $p = 0$ and $R = 1$. Then for $w \in S^1 = \partial B(0, 1)$ and fixed z with $|z| < 1$, we have

$$\frac{1}{w - z} = \frac{1}{w} \left(1 + (z/w) + (z/w)^2 + \cdots \right),$$

converging uniformly on the circle $|w| = 1$. We then have:

$$f(z) = \frac{1}{2\pi i} \int_{S^1} \frac{f(w)\, dw}{w - z} = \sum z^k \frac{1}{2\pi i} \int_{S^1} \frac{f(w)\, dw}{w^{k+1}} = \sum a_k z^k$$

as desired. ∎

Corollary 1.7 *An analytic function has at least one singularity on its circle of convergence.*

That is, if f can be extended analytically from $B(p, R)$ to $B(p, R')$, then the radius of convergence is at least R'. So there must be some obstruction to making such an extension, if $1/R = \limsup |a_k|^{1/k}$.

Example: Fibonacci numbers. Let $(a_0, a_1, a_2, a_3, \ldots) = (1, 1, 2, 3, 5, 8, \ldots)$ denote the Fibonacci numbers, and let $\gamma = (1 + \sqrt{5})/2 = 1.618\ldots$ be the golden ratio (slightly more than the number of kilometers in a mile). As an illustration of the use of power series, we will show that there is a constant $\alpha > 0$ such that
$$|a_n - \alpha\gamma^n| \to 0$$
rapidly. Let $f(z) = \sum a_n z^n$. Since $a_n = a_{n-1} + a_{n-2}$, except for $n = 0$, we get
$$f(z) = (z + z^2)f(z) + 1$$

13

and so $f(z) = 1/(1 - z - z^2)$. This function has a singularity at $z = 1/\gamma$ and thus $\limsup |a_n|^{1/n} = \gamma$.

Note that $\sum(a_n - \alpha\gamma^n)z^n = f(z) - \alpha/(1 - \gamma z)$ has radius of convergence $|\gamma| > 1$, if we choose the constant α to cancel the pole of $f(z)$ at $z = 1/\gamma$. (In fact $\alpha = \gamma/\sqrt{5}$ will do this.) And then $|a_n - \alpha\gamma^n| \to 0$ (exponentially fast).

Fibonacci quarterly. Question: why are 10:09 and 8:18 such pleasant times? [Mon].

Pisot numbers: an open problem. One can show by a variant of the argument above that the powers of the golden ratio get close and closer to the integers; that is, $d(\gamma^n, \mathbb{Z}) \to 0$ (see the Exercises).

A real number $\lambda > 1$ is a *Pisot number* if (i) $d(\lambda^n, \mathbb{Z}) \to 0$ as $n \to \infty$; and (ii) λ is algebraic. It is unknown if (i) implies (ii). However, it is known that every Pisot number is an algebraic *integer*; all its Galois conjugates satisfy $|\lambda'| < 1$; and these two properties characterize the class.

Power series of rational functions. As we have see above, it is easy to prove:

Theorem 1.8 *A power series $\sum a_n z^n$ represents a rational function iff its coefficients satisfy a recurrence relation.*

Kronecker's theorem asserts that $\sum a_i z^i$ is a rational function iff the determinants of the matrices $a_{i,i+j}$, $0 \le i, j \le n$ are zero for all n sufficiently large [Sa, §I.3].

Aside: Pisot numbers. We have just seen that $d(\alpha\gamma^n, \mathbb{Z}) \to 0$ as $n \to \infty$. A similar argument shows $d(\gamma^n, \mathbb{Z}) \to 0$.

A *Pisot number* is an algebraic integer $\lambda > 1$ whose Galois conjugates satisfy $|\lambda'| < 1$. Since $\mathrm{Tr}(\lambda^n)$ is an integer, it is easy to see $d(\lambda^n, \mathbb{Z}) \to 0$ for any Pisot number.

It is an unsolved problem to prove the converse: show that if $d(\lambda^n, \mathbb{Z}) \to 0$, then λ is a Pisot number. It is sufficient to prove that λ is algebraic.

How to compute π (Euler's method) . The power series for the arctangent is easy to evaluate by relating it to $\int dx/(1 + x^2)$. Thus we get:

$$f(z) = \tan^{-1}(z) = z - z^3/3 + z^5/5 + z^7/7 + \cdots$$

which suggest correctly that

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

14

However the convergence is very slow, since the error after n terms is on the order of $1/n$.

In general, to accelerate the convergence of the sums $s_n = s_n(0) = \sum_0^n a_i$, we can set $s_n(1) = (s_n + s_{n+1})/2$ and take its limit instead. This will be especially efficacious for an alternating series. But why not repeat the process again and again? To this end we define inductively:

$$s_n(k+1) = \frac{s_n(k) + s_{n+1}(k)}{2}.$$

If we have the terms $(a_0, \ldots, a_n)$ at our disposal, we can then compute

$$e_n = s_0(n) = 2^{-n} \sum_0^n \binom{n}{k} s_k.$$

For $f(z) = \sum a_k z^k$ as above, we find that

$$4 \sum_0^{100} a_k = 3.121\ldots \quad \text{but} \quad 4 \sum_0^{100} 2^{-100} \binom{100}{k} s_k = 3.141592653589793273\ldots,$$

which agrees with π to 16 decimal places (!)

Here is a general result, that also explains why this works.

Theorem 1.9 *Let $f(z) = \sum a_n z^n$ have radius of convergence 1, but no singularity at $z = 1$, and $\sum a_n$ exists. Define*

$$F(y) = f(y/(1-y)) = \sum b_k y^k.$$

Then $F(y)$ has radius of convergence greater than $1/2$, and

$$\sum_0^n b_k (1/2)^k \to F(1/2) = \sum a_n$$

exponentially fast.

Proof. Note that if $z = y/(1-y)$ then $y = \phi(z) = z/(1+z)$. Since ϕ sends 0, 1 and -1 to 0, $1/2$ and ∞ respectively, we see that $\phi(\Delta)$ is the left halfplane $H = \{y : \operatorname{Re}(y) < 1/2\}$.

Since $f(z)$ is analytic in Δ, $F(y) = f(y/(1-y))$ is analytic in H. And since $f(z)$ is analytic near $z = 1$, $F(y)$ is analytic in a neighborhood of $y = 1/2$. Thus the nearest singularity of $F(y)$ lies at distance $r > 1/2$ from the origin, and hence the power series for $F(y)$ converges exponentially fast at $y = 1/2$, where (by continuity) it gives $f(1) = \sum a_n$. ∎

15

Theorem 1.10 *We have $e_n = \sum_0^n b_k (1/2)^k$, so this method of computing $\sum a_n$ agrees with the method sketched above.*

Example. The function $f(z) = \tan^{-1}(z)$ has singularities at $z = \pm i$ (and nowhere else). (In fact, $\tan : \mathbb{C} \to \widehat{\mathbb{C}} = \{-i, i\}$ is a covering map; one can write $\tan(z) = M(\exp(2iz))$ where $M(z)$ is a Möbius transformation sending $(0, \infty) = 0^{\pm 1}$ to $\pm i$).

Thus $F(z)$ has singularities at $\pm i/(1 + i)$, so its radius of convergence is $1/\sqrt{2}$ which is bigger than $1/2$. In fact we expect the error after summing n terms to be roughly of size $2^{-n/2}$, and $2^{-50} = 9 \times 10^{-16}$, consistent with the results above.

Proof of Theorem 1.10. First note that:

$$\frac{1}{(1-y)^{n+1}} = \sum_0^\infty \binom{n+k}{k} y^k.$$

This can be seen by either differentiating both sides, starting with $n = 0$, or by observing that the coefficient of y^k is the number of ways of writing $k = m_1 + \cdots m_n$ with $m_i \geq 0$. Plugging into $f(y/(1-y))$, we find that $b_0 = a_0$ and

$$b_{n+1} = \sum_0^n \binom{n}{k} a_{k+1}.$$

We can now show Theorem 1.10 by induction. We have $e_0 = b_0 = a_0$, and

$$
\begin{aligned}
e_{n+1} - e_n &= \frac{s_1(n) - s_0(n)}{2} = 2^{-(n+1)} \sum_0^n \binom{n}{k}(s_{k+1} - s_k) \\
&= 2^{-(n+1)} \sum_0^n \binom{n}{k} a_{k+1} = b_{n+1} 2^{-(n+1)},
\end{aligned}
$$

which implies Theorem 1.10 for $n + 1$. ∎

Remarks. The number e_n is the expected value of $s_{I(n)}$, where $I(n)$ is the sum of n independent random variables, each taking on the values 0 or 1 with probability $1/2$.

This method can also sum series that do not converge! For example, if we set $a_n = (-1)^n$, then we get $\sum a_n = 1/2$. In fact $f(z) = \sum a_n z^n = 1/(1+z)$,

so this value is not unexpected. The answer in this case agrees with the result of 'Abel summation', is defined by:

$$\sum a_n(A) = \lim_{r \to 1-} \sum a_n r^n.$$

For more on this and other summation methods, see [Har, §VIII].

Isolation of zeros. If $f(z) = \sum a_n(z-p)^n$ vanishes at p but is not identically zero, then we can factor out the leading term and write:

$$f(z) = (z-p)^n(a_n + a_{n+1}(z-p) + \cdots) = (z-p)^n g(z)$$

where g is analytic and $g(p) \neq 0$. This is the simplest case of the *Weierstrass preparation theorem*: it shows germs of analytic functions behave like polynomials times units (invertible functions).

In particular, we find:

Theorem 1.11 *The zeros of a nonconstant analytic function are isolated.*

Warning: we are assuming the domain is connected!

Proof. Let $U_0 \subset U$ be the largest open set with $f|U_0 = 0$. Let $U_1 = U - U_0$. Then by the factorization theorem above, the zeros of f in U_1 are isolated. Thus U_1 is open as well. So either $U = U_0$ — in which case f is constant — or $U = U_1$ — in which case f has isolated zeros. ∎

Corollary 1.12 *An analytic function which is constant along an arc, or even on a countable set with an accumulation point, is itself constant.*

Corollary 1.13 *The extension of a function $f(x)$ on $[a,b] \subset \mathbb{R}$ to an analytic function $f(z)$ on a connected domain $U \supset [a,b]$ is* unique — *if it exists.*

1.3 The maximum principle

In this section we develop the maximum principle and related to ideas that lead to compactness of spaces of analytic functions.

Mean value and maximum principle. On the circle $|z| = r$, with $z = re^{i\theta}$, we have $dz/z = i\,d\theta$. Thus Cauchy's formula gives

$$f(0) = \frac{1}{2\pi i} \int_{S^1(r)} f(z)\frac{dz}{z} = \frac{1}{2\pi} \int_{S^1(r)} f(z)\,d\theta.$$

In other words, analytic functions satisfy:

17

Theorem 1.14 (The Mean-Value Formula) *The value of $f(p)$ is the average of $f(z)$ over $S^1(p, r)$.*

Corollary 1.15 (The Maximum Principle) *A nonconstant analytic function does not achieve its maximum in U.*

Proof. Suppose $f(z)$ achieves its maximum at $p \in U$. Then $f(p)$ is the average of $f(z)$ over a small circle $S^1(p, r)$. Moreover, $|f(z)| \leq |f(p)|$ on this circle. The only way the average can agree is if $f(z) = f(p)$ on $S^1(p, r)$. But then f is constant on an arc, so it is constant in U. ∎

Corollary 1.16 *If $\overline{U}$ is compact, then $\sup_U |f| = \sup_{\partial U} |f|$.*

Cauchy's bound, *reprise.* Here is a variant of Cauchy's bounded that is more useful in practice:

Theorem 1.17 *Let $f : U \to \mathbb{C}$ be an analytic function. Then for each $n \geq 0$ and $p \in U$ we have*
$$\frac{|f^{(n)|}(p)}{n!} \leq \frac{\sup_U |f(z)|}{d(p, \partial U)}.$$

The proof is immediate by applying the usual bound to a suitable ball centered at p.

Algebraic completeness. Now suppose $f : U \to \mathbb{C}$ is an *entire function*; that is, suppose $U = \mathbb{C}$. Then the bound above shows:

Theorem 1.18 *A bounded entire function is a constant. More generally, if $|f(z)| = O(|z|^n)$, then f is a polynomial of degree at most n.*

Corollary 1.19 *Any polynomial $f \in \mathbb{C}[z]$ of degree 1 or more has a zero in $\mathbb{C}$.*

Proof. Otherwise $1/f(z)$ would be a nonconstant, bounded entire function. ∎

Alternatively, $1/f(z) \to 0$ as $|z| \to \infty$, so $f(z)$ must vanish identically, a contradiction.

Corollary 1.20 *There is no conformal homeomorphism between $\mathbb{C}$ and Δ.*

The Picard theorem says an entire function can omit at most one value in $\mathbb{C}$. Here is an easy form (cf. Weierstrass–Casorati, to be proved later):

Theorem 1.21 *Let $f : \mathbb{C} \to \mathbb{C}$ be an entire function. Then the closure of its image is either a single point, or $\mathbb{C}$.*

Proof. If $p \notin \overline{f(\mathbb{C})}$, then $1/(f(z) - p)$ is a bounded entire function, hence constant. ∎

Aside: quasiconformal maps. A diffeomorphism $f : U \to V$ between domains in $\mathbb{C}$ is *quasiconformal* if $\sup_U |\overline{\partial} f / \partial f| < \infty$. Many qualitative theorems for conformal maps also hold for quasiconformal maps. For example, there is no quasiconformal homeomorphism between $\mathbb{C}$ and Δ.

Parseval's theorem. The power series of an analytic function on the ball $B(0, R)$ also contains information about its L^2-norm on the circle $|z| = R$: namely if $f(z) = \sum a_n z^n$, then we have:

$$\sum |a_n|^2 R^{2n} = \frac{1}{2\pi} \int_{|z|=R} |f(z)|^2 d\theta.$$

This comes from the fact that the functions z^n are orthogonal in $L^2(S^1)$. It also gives another important perspective on holomorphic functions: they are the elements in $L^2(S^1)$ with *positive* Fourier coefficients, and hence give a half–dimensional subspace of this infinite–dimensional space.

1.4 Sequences of analytic functions

In this section we address the issue of limits of analytic functions. The most basic fact, which is already remarkable, is that analyticity is preserved under uniform limits.

Theorem 1.22 *Let $f_n \to f$ be a uniformly convergent sequence of analytic functions f_n on U with a continuous limit f. Then f is analytic.*

Proof. Since the statement is local, we may assume that $U = B$ is a ball and f_n and f are continuous on $\overline{B}$. Then by Cauchy's integral formula, we have, for all $p \in B$,

$$f_n(p) = \frac{1}{2\pi} \int_{\partial B} \frac{f_n(z) \, dz}{z - p}.$$

Fixing p and letting n tend to infinity, we find that $f(p)$ also satisfies this equation. But then $f'(z)$ can computed by differentiating under the integral sign:

$$f'(p) = \frac{1}{2\pi} \int_{\partial B} \frac{f(z) \, dz}{(z - p)^2}.$$

Thus $f(z)$ is analytic. ∎

Note: *Many fallacies in real analysis become* theorems *in complex analysis.* The proof exhibits the following general principle:

Theorem 1.23 *A continuous function $f(z)$ on U is analytic if and only if it satisfies Cauchy's formula for all balls B with $\overline{B} \subset U$.*

Convergence of $f'(z)$. It is tempting to think that something even better happens: if $f_n \to f$ uniformly, then $f'_n(z) \to f'(z)$ uniformly. But this is *false*: for example, $f_n(z) = z^n/n$ converges to zero uniformly on Δ, but $f'_n(z) = z^{n-1}$ does not converge to zero *uniformly*.

For a more complete discussion, it will be useful to introduce the notation

$$\|f\|_E = \sup\{|f(z)| \ : \ z \in E\}$$

for the sup norm of a complex–valued function f on a set $E \subset \mathbb{C}$. Then Cauchy's bound gives

$$|f'(p)| \leq \frac{\|f\|_{\partial U}}{d(p, \partial U)}, \tag{1.3}$$

which easily implies:

Theorem 1.24 *If $f_n \to f$ uniformly on U, then $f'_n \to f'$ uniformly on each compact subset $K \subset U$.*

The space of continuous functions $C(U)$. To formalize this notion of convergence, we begin with spaces of continuous functions. Given $E \subset \mathbb{C}$, we let

$$C(E) = \{f : E \to \mathbb{C} \ : \ f \text{ is continuous on } K\}.$$

If $K \subset \mathbb{C}$ is compact, then the vector space $C(K)$ is a Banach space with respect to the norm $\|f\|_K$. However when we replace K by an open domain U, we obtain a space $C(U)$ including functions that are *unbounded*, so the sup–norm is not defined.

Instead, we endow $C(U)$ with the unique topology such that $f_n \to f$ if and only if $\|f_n - f\|_K \to 0$ for each compact set $K \subset U$. This limiting process is called *uniform convergence on compact sets*.

In comes from the family of pseudo–norms $\|f_n\|_K$, and indeed this topology is metrizable and complete: if $f_n|K$ is a Cauchy sequence for each compact set K, then it has a limit $f \in C(U)$.

The space of holomorphic function $\mathcal{O}(U)$. Now let

$$\mathcal{O}(U) = \{f : U \to \mathbb{C} \ : \ f(z) \text{ is analytic.}\}$$

Since $\mathcal{O}(U) \subset C(U)$, it inherits a natural topology from $C(U)$, and the results above show:

Theorem 1.25 *The subspace $\mathcal{O}(U)$ is closed in $C(U)$.*

Moreover, we have:

Theorem 1.26 *Given $M > 0$, the set of $f \in \mathcal{O}(U)$ with $|f(z)| \le M$ is compact.*

Proof. For any compact set $K \subset U$, the restrictions $f_n|K$ are equicontinuous, since $d(K, \partial U) > 0$. Apply the Arzela–Ascoli theorem. ∎

1.5 Laurent series and singularities

In this section we aim to describe the behavior of analytic functions near an isolated point where they may not be defined.

Laurent series. Using again the basic series $1/(1-z) = \sum z^n$ and Cauchy's formula over the two boundary components of the *annulus* $A(r, R) = \{z \ : \ r < |z| < R\}$, we find:

Theorem 1.27 *If $f(z)$ is analytic on $A(r, R)$, then in this region we have*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n.$$

The positive terms converge for $|z| < R$, and the negative terms converge for $|z| > r$.

Corollary 1.28 *An analytic function on the annulus $r < |z| < R$ can be expressed as the sum of a function analytic on $|z| < R$ and a function analytic on $|z| > r$.*

This fact can be used to show that the sheaf cohomology group $H^1(\widehat{\mathbb{C}}, \mathcal{O}) = 0$.

Theorem 1.29 *For all n and $r < s < R$, we have*

$$a_n = \frac{1}{2\pi i} \int_{S^1(s)} \frac{f(z)\,dz}{z^{n+1}}.$$

Corollary 1.30 *If $|f(z)| \le M$ on the circle $|z| = s$, then*

$$|a_n| \le Ms^{-n}.$$

Isolated singularities. As a special case, if f is analytic on $U - p$, with $p \in U$, then near p we have a Laurent series expansion $f(z) = \sum_{-\infty}^{\infty} a_n(z - p)^n$. We say f has an *isolated singularity* at p.

Let

$$\mathrm{ord}_p(f) = \inf\{n \ : \ a_n \neq 0\}.$$

This is the *order* of f at p; the values $\pm\infty$ are allowed.

We now have 3 possibilities:

1. *Removable singularities.* If $\mathrm{ord}_p(f) \ge 0$, then the apparent singularity at p is removable and f has a *zero* of order n at p.

2. *Poles.* If $\mathrm{ord}(f, p) = -n < 0$, then $f(z)$ has a *pole* of order n at p, and we can write

$$f(z) = \frac{a_{-n}}{(z - p)^n} + \cdots + \frac{a_{-1}}{z - p} + \sum_{n=0}^{\infty} a_n z^n$$

   near p. The germs of functions at p with such finite *Laurent tails* form a local field, with $\mathrm{ord}(f, p)$ as its discrete valuation. (Compare $\mathbb{Q}_p$, where $v_p(p^n a/b) = n$.)

3. *Essential singularities.* If ord$(f, p) = -\infty$ we say f has an *essential singularity* at p. (Example: $f(z) = \sin(-1/z)$ at $z = 0$.)

Bounded functions and essential singularities.

Theorem 1.31 *Every isolated singularity of a bounded analytic function is removable.*

Proof. Suppose the singularity is at $z = 0$, and $|f(z)| \leq M$. Write $f(z)$ as a Laurent series $\sum a_n z^n$. Then for any $r > 0$ and $n > 0$ we have

$$|a_{-n}| \leq Mr^n.$$

Letting $r \to 0$, we find that all the negative terms in the Laurent series vanish, and hence $f(z)$ is analytic at $z = 0$. ∎

Corollary 1.32 (Weierstrass-Casorati) *If $f(z)$ has an essential singularity at p, then there exist $z_n \to p$ such that $f(z_n)$ is dense in $\mathbb{C}$.*

Proof. Otherwise there is a neighborhood U of p such that $f(U - \{p\})$ omits some ball $B(q, r)$ in $\mathbb{C}$. But then $g(z) = 1/(f(z) - q)$ is bounded on U, and hence analytic at p, with a zero of finite order. Then $f(z) = q + 1/g(z)$ has at worst a pole at p, not an essential singularity. ∎

Aside: several complex variables. A function $f(z_1, \ldots, z_n)$ is analytic if it is C^∞ and $df/d\bar{z}_i = 0$ for $i = 1, \ldots, n$. (There are several other equivalent definitions).

Using the same type of argument and some basic facts from several complex variables, it is easy to show that if $V \subset U \subset \mathbb{C}^n$ is an analytic hypersurfaces, and $f : U - V \to \mathbb{C}$ is a bounded analytic function, then f extends to all of U.

More remarkably, *every* analytic function extends across V if codim$(V) \geq 2$. For example, an isolated point is always a removable singularity in $\mathbb{C}^2$. Let us prove a stronger version of this fact, to illustrate the new phenomena that arise in several complex variables.

Theorem 1.33 (Hartogs) *Let $f(z, t)$ be analytic on $\Delta^2 - \Delta(r)^2$, with $r < 1$. Then f extends to an analytic function on Δ^2.*

Proof. Define $F(z, t) = (2\pi i)^{-1} \int_{S^1} f(\zeta, t)/(z - \zeta) \, d\zeta$. Since the integrand is holomorphic on as a function of (z, t), so is F. But $F(z, t) = f(z, t)$ for $|t| > r$, so it provides the desired extension. ∎

1.6 The residue theorem, the argument principle, and definite integrals

In this section we will discuss complex integrals for analytic functions $f(z)$ with isolated singularities: these are controlled by the *residues* of f. We will use the residue theorem to show that (nonconstant) analytic functions are *open maps*, and to evaluate definite integrals.

The residue. A critical role is played by the *residue* of $f(z)$ at p, defined by

$$\mathrm{Res}(f, p) = a_{-1}.$$

It satisfies

$$\int_\gamma f(z)\, dz = 2\pi i \, \mathrm{Res}(f, p)$$

for any small loop encircling the point p in U. Thus the residue is intrinsically an invariant of the 1-form $f(z)\, dz$, *not* the function $f(z)$. (If we regard $f(z)\, dz$ as a 1-form, then its residue is invariant under change of coordinates.)

Theorem 1.34 (The residue theorem) *Let $f : U \to \mathbb{C}$ be a function which is analytic apart from a finite set of isolated singularities. We then have:*

$$\int_{\partial U} f(z)\, dz = 2\pi i \sum_{p \in U} \mathrm{Res}(f, p).$$

Alternative definition. A more intrinsic definition of the derivative of a *1–form* ω at an isolated singularity $p \in U$ is

$$\mathrm{Res}_p(\omega) = \frac{1}{2\pi} \int_{\gamma_p} \omega,$$

where γ is a small loop encircling p (small enough that the only singularity of f inside γ is at p).

With this definition, the residue theorem follows immediately from Cauchy's theorem; at the same time, it is easy to see that

$$\mathrm{Res}_p(f(z)\, dz) = a_{-1},$$

consistent with the definition above. When discussing function on $\mathbb{C}$, we usually drop the dz and simply write $\mathrm{Res}_p(f)$.

Calculating the residue. The simplest case of the reside arises when $f(z) = g(z)/(z - p)$ and $g(z)$ is holomorphic in a neighborhood of p. In this case we have

$$\text{Res}_p(g(z)/(z - p)) = g(p)$$

by Cauchy's integral formula.

Similarly, if $f(z) = 1/g(z)$ and $g(z)$ has a *simple* zero at p, then

$$\text{Res}_p(f) = 1/g'(p).$$

For example, for $f(z) = 1/(x^n + 1)$ we have

$$\text{Res}(f, \zeta_{2n}) = \zeta_{2n}^{1-n}/n = -\zeta_{2n}/n.$$

Applications of the residue theorem. We will develop two applications of the residue theorem: the argument principle, and the evaluation of definite integrals.

The intermediate value theorem. For context, suppose $f : I = [a, b] \to \mathbb{R}$ is a continuous function, nonzero on ∂I. We can then define the *winding number* of $f|\partial I$ around $x = 0$ by 1 is $0 \in [f(a), f(b)]$, and 0 otherwise.

The *intermediate value theorem* then states that the number of zeros of $f(x)$ in I is bounded below by $w(f|\partial I)$.

We will now formulate a similar principle in the complex plane, and use it to get an exact count of the number of zeros of $f(z)$ in a region U.

The argument principle. We first observe that if f has at worst a pole at p, then its *logarithmic derivative*

$$d \log f = f'(z)/f(z) \, dz$$

satisfies

$$\text{Res}(f'/f, p) = \text{ord}(f, p).$$

We thus obtain:

Theorem 1.35 (Argument principle) *If f is analytic in U and $f|\partial U$ is nowhere zero, then the number of zeros of f in U is given by*

$$N(f, U) = \sum_{p \in U} \text{ord}(f, p) = \frac{1}{2\pi i} \int_{\partial U} \frac{f'(z) \, dz}{f(z)}.$$

25

Corollary 1.36 (Rouché's Theorem) *If $|f| > |g|$ along ∂U, then f and $f + g$ have the same number of zeros in U.*

Proof. The continuous, integer-valued function $N(f + tg, U)$ is constant for $t \in [0, 1]$. ∎

Example. Consider $p(z) = z^5 + 14z + 1$. Then all its zeros are inside $|z| < 2$, since $|z|^5 = 32 > 29 \geq |14z + 1|$ when $|z| = 2$; but only one inside $|z| < 3/2$, since $|z^5 + 1| \leq 1 + (3/2)^5 < 9 < |14z|$ on $|z| = 3/2$. (Intuitively, the zeros of $p(z)$ are close to the zeros of $z^5 + 14z$ which are $z = 0$ and otherwise 4 points on $|z| = 14^{1/4} \approx 1.93$.

Algebraic completeness revisited. Our previous proof that $\mathbb{C}$ is algebraically closed (by considering $1/p(z)$) was clever but nonconstructive; it did not tell us where to find a zero of $p(z)$. Here is a more precise version:

Corollary 1.37 *Let $p(z) = z^d + a_1 z^{d-1} + \ldots + a_d$, and suppose $|a_i| < R^i/d$. Then $p(z)$ has d zeros inside the disk $|z| < R$.*

Proof. Write $p(z) = f(z) + g(z)$ with $f(z) = z^d$, and let $U = B(0, R)$; then on ∂U, we have $|f| = R^d$ and $|g| < R^d$; now apply Rouché's Theorem. ∎

Geometric picture. More generally, provided $p \notin f(\partial U)$, we let

$$N_p(f, U) = \text{the number of solutions to } f(z) = p \text{ in } U,$$

counted with multiplicity. We have $N_p(f, U) = N(f(z) - p, U)$, and thus:

$$N_p(f, U) = \frac{1}{2\pi i} \int_{\partial U} d\log(f(z) - p) = \frac{1}{2\pi} \int_{\partial U} d\arg(f(z) - p).$$

This is nothing more than the *winding number* of $f(\partial U)$ around p.

Observing that $N_p(f, U)$ is a continuous function taking integral values, we conclude:

Theorem 1.38 *The number $N_p(f, U)$ is constant on each connected component of $\mathbb{C} - f(\partial U)$.*

Corollary 1.39 *If $p \notin f(\partial U)$, then $f(U)$ contains an open neighborhood of $f(p)$.*

26

Corollary 1.40 *A nonconstant analytic function is an open mapping.*

Proof. Suppose $f(p) = q$. Then p is an isolated zero of $f(z) - q$. Choose a small ball U such that $f(z) \neq q$ on ∂U. Then $f(U)$ contains the component of $\mathbb{C} - f(\partial U)$ in which q lies. ∎

The open mapping theorem gives an alternate proof of the maximum principle *and* its strict version:

Corollary 1.41 *If $|f|$ achieves its maximum in U, then f is a constant.*

Moreover, the geometric picture of the argument principle yields:

Theorem 1.42 *For any $p \in \mathbb{C} - f(\partial U)$, the number of solutions to $f(z) = p$ in U is the same as the winding number of $f(\partial U)$ around p.*

Using isolation of zeros, we have:

Corollary 1.43 *If $f'(p) \neq 0$, then f is a local homeomorphism at p.*

In fact for r sufficiently small and q close to $f(p)$, we have

$$f^{-1}(q) = \frac{1}{2\pi} \int_{B(p,r)} \frac{z f'(z)\, dz}{f(z) - q}.$$

This shows:

Corollary 1.44 *The local inverse of $f(z)$ is analytic wherever it exists.*

Formal inversion. The functional inverse of a power series can be readily computed using back substitution. For example, to invert

$$\sin(z) = z - z^3/6 + z^5/120 - \cdots,$$

we set $z = w + a_3 w^3 + a_5 w^5 + \cdots$ and compute

$$w = \sin(z) = (w + a_3 w^3 + \cdots) - (w + a_3 w^3 + \cdots)^3/6,$$

which gives, for the coefficient of w^3, $a_3 - 1/6 = 0$ and thus $a_3 = 1/6$. Then:

$$w = \sin(z) = (w + w^3/6 + a_5 w^5 \cdots) - (w + w^3/6 + \cdots)^3/6 + (w + \cdots)^5/120 + \cdots$$

gives, as the coefficient for w^5, $a_5 - 3/36 + 1/120 = 0$, and hence $a_5 = 3/40$. Thus

$$\sin^{-1}(z) = z + \frac{z^3}{6} + \frac{3z^5}{40} + \cdots$$

With some work, one can show that the formal inverse always has a positive radius of convergence, which gives another proof of the analyticity of $f^{-1}(z)$.

Moduli spaces $\mathcal{M}_{0,n}$. The universal formula for inverting $f(z) = z + \sum_2^\infty a_n z^n$ has a remarkable connection with the moduli spaces of n points on $\widehat{\mathbb{C}}$ [Mc].

The Weierstrass preparation theorem. The same idea can be used to analyze functions of two (or more) complex variables. First a definition: we say $f(z,t)$ is analytic on $\mathbb{C}^2$ if it is analytic in each variable separately. This implies that f is locally given by a convergent power series $\sum a_{ij}(z-z_0)^i(w-w_0)^j$.

A function of one complex variable looks like a polynomial: $f(z) = z^n g(z)$. In particular, its zero set *agrees* with the zero set of a polynomial. Similar statements hold in $\mathbb{C}^2$, but we must allow polynomials with analytic coefficients.

To be precise, suppose $f_t(z) = f(z,t)$ is analytic near $(0,0)$ on $\mathbb{C}^2$, $f_0(0) = 0$, but $f_0(z)$ is not identically zero.

Theorem 1.45 *There exists analytic functions $a_i(t)$ on the unit disk, and a nowhere vanishing analytic function $h_t(z)$, such that*

$$f_t(z) = (z^n + a_1(t)z^{n-1} + \cdots + a_n(t))h_t(z)$$

for (z,t) near $(0,0)$, and $a_i(0) = 0$.

For the proof, use the fact that $\int_{S^1(\epsilon)}(z^k f_t'(z)/f_t(z))\,dz$ is analytic in t and gives the sums of the powers of the zeros of f_t near the origin.

Algebraicity. It can also be shown that if $f(z,t) = 0$ has an isolated singularity at $(0,0)$, then one can make an analytic change of coordinates so the $f(Z,T)$ becomes a polynomial in (Z,T). See [Ad] and references therein.

Aside: the smooth case. These results also hold for *smooth mappings f* once one finds a way to count the number of solutions to $f(z) = p$ correctly. (Some may count negatively, and the zeros are only isolated for *generic* values of p.)

Aside: another proof of the Fundamental Theorem of Algebra.
Artin gave the following proof that $\mathbb{C}$ is algebraically closed using only the
real version of the argument principle, i.e. the intermediate value theorem.

1. Odd polynomials in $\mathbb{R}$ have roots in $\mathbb{R}$.

2. Using Galois theory, it follows that any algebraic extension $K/\mathbb{R}$ has
   degree 2^k for some k.

3. Consequently any algebraic extension $K/\mathbb{C}$ also has degree 2^k. It fol-
   lows that if $k > 0$, there exists an algebraic extension of $\mathbb{C}$ of degree
   2.

4. But by the quadratic formula, every quadratic polynomial has a root
   in $\mathbb{C}$.

**Aside: Linking numbers and intersection multiplicities of curves
in $\mathbb{C}^2$.** Counting the number of zeros of $y = f(x)$ at $x = 0$ is the same
as counting the multiplicity of intersection between the curves $y = 0$ and
$y = f(x)$ in $\mathbb{C}^2$, at $(0,0)$.

In general, to an analytic curve C defined by $F(x,y) = 0$ passing through
$(0,0) \in \mathbb{C}^2$, we can associate a *knot* by taking the intersection of C with the
boundary $S^3(r)$ of a small ball centered at the origin.

A pair of distinct, irreducible, curves C_1 and C_2 passing through $(0,0)$
(say defined by $f_i(x,y) = 0$, $i = 1, 2$), have a *multiplicity* of intersection
$m(C_1, C_2)$. This can be defined geometrically by intersection with a small
sphere $S^3(r)$ in $\mathbb{C}^2$ centered at $(0,0)$. Then we get a pair of *knots*, and their
linking number is the same as this multiplicity.

Examples of links. It is often simpler the use $S^3 = \partial \Delta^2$ instead of $|x|^2 +
|y|^2 = 1$. Then S^3 is the union of two solid tori, $\overline{\Delta} \times S^1$ and $S^1 \times \overline{\Delta}$. The
axes $x = 0$ and $y = 0$ are the core curves of these tori; they have linking
number one. In fact the variety $xy = 0$ gives the *Hopf link* in S^3.

The line $y = x$ is a $(1, 1)$ curve on $S^1 \times S^1$; more generally, for $\gcd(a,b) =
1$, $y^a = x^b$ is an (a, b) curve, parameterized by $(x, y) = e^{ita}, e^{itb}$). In particular,
the cusp $y^2 = x^3$ meets S^3 in a trefoil knot. It links $x = 0$ twice and $y = 0$
three times.

The simplest case is $y = 0$ — the core of the solid torus — and $y = x^n$
— a $(1, n)$–torus knot. These have linking number 1. For more details see
[Mil1].

Problem. Show that the figure-eight knot cannot arise from an analytic curve.

Functional factorization. Here is an alternate proof of the open mapping theorem. It is clear that $p_n(z) = z^n$ is an open map, even at the origin; and (by the inverse function theorem) that an analytic function is open at any point p where $f'(p) \neq 0$. The general case follows from these via:

Theorem 1.46 *Let f be an analytic map at p with $\operatorname{ord}(f', p) = n \geq 0$. Then up to a local chance of coordinates in domain and range, $f(z) = z^{n+1}$. That is, there exist analytic diffeomorphisms with $h_1(0) = p$, and $h_2(0) = f(p)$, such that*

$$h_2 \circ f \circ h_1^{-1} = z^{n+1}.$$

Proof. We may assume $p = f(p) = 0$ and $f(z) = z^n g(z)$ where $g(0) \neq 0$. Then $h(z) = g(z)^{1/n}$ is a well-defined analytic function near $z = 0$, once we have chosen a particular value for $g(0)^{1/n}$. It follows that $f(z) = (zh(z))^n = p_n(h_1(z))$ where $h_1'(z) \neq 0$. Thus h_1 is a local diffeomorphism, and $f \circ h_1^{-1}(z) = z^n$. ∎

The argument as the basic multivalued function. It is useful to have a general discussion of the sometimes confusing notion of 'branch cuts' and 'multivalued functions'. Here are 2 typical results.

Theorem 1.47 *Let $U \subset \mathbb{C}^*$ be a simply-connected region. Suppose $p \in U$, $\theta \in \mathbb{R}$ and $\arg(p) = \theta \bmod 2\pi$. Then there is a unique continuous function Q*

$$\operatorname{Arg} : U \to \mathbb{R}$$

such that $\operatorname{Arg}(p) = \theta$ and $\operatorname{Arg}(z) = \arg(z) \bmod 2\pi$ for all $z \in U$.

Proof. In fact we can explicitly write

$$\operatorname{Arg}(z) = \theta + \operatorname{Im} \int_p^z \frac{d\zeta}{\zeta};$$

the fact that U is simply-connected implies that the integral does not depend on the choice of a path from p to z. ∎

Since $\log(z) = \log|z| + i\arg(z)$, we find:

Corollary 1.48 *There is a single-valued branch of the logarithm on U.*

Since $z^\alpha = \exp(\alpha \log z)$, we have:

Corollary 1.49 *There is a single-valued branch of z^α on U.*

Residue calculus and definite integrals. The residue theorem can be used to systematically evaluate various definite integrals.

Definite integrals 1: rational functions on $\mathbb{R}$. Whenever a rational function $R(x) = P(x)/Q(x)$ has the property that $\int_{\mathbb{R}} |R(x)|\, dx$ is finite, we can compute this integral via residues: we have

$$\int_{-\infty}^{\infty} R(x)\, dx = 2\pi i \sum_{\operatorname{Im} p > 0} \operatorname{Res}(R, p).$$

(Of course we can also compute this integral by factoring $Q(x)$ and using partial fractions and trig substitutions.)

Example. Where does π come from? It emerges naturally from rational functions by integration — i.e. it is a *period*. Namely, we have

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = 2\pi i \operatorname{Res}(1/(1 + z^2), i) = 2\pi i(-i/2) = \pi.$$

Of course this can also be done using the fact that $\int dx/(1+x^2) = \tan^{-1}(x)$. More magically, for $f(z) = 1/(1 + z^4)$ we find:

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^4} = 2\pi i(\operatorname{Res}(f, (1+i)/\sqrt{2}) + \operatorname{Res}(f, (1+i)/\sqrt{2}) = \frac{\pi}{\sqrt{2}}.$$

Both are obtain by closing a large interval $[-R, R]$ with a circular arc in the upper halfplane, and then taking the limit as $R \to \infty$.

We can even compute the general case, $f(z) = 1/(1 + z^n)$, with n even. For this let $\zeta_k = \exp(2\pi i/k)$, so $f(\zeta_{2n}) = 0$. Let P be the union of the paths $[0, \infty)\zeta_n$ and $[0, \infty)$, oriented so P moves positively on the real axis. We can then integrate over the boundary of this pie-slice to obtain:

$$(1 - \zeta_n) \int_0^{\infty} \frac{dx}{1 + x^n} = \int_P f(z)\, dz = 2\pi i \operatorname{Res}(f, \zeta_{2n}) = 2\pi i/(n\zeta_{2n}^{n-1}),$$

which gives
$$\int_0^\infty \frac{dx}{1+x^n} = \frac{2\pi i}{n(-\zeta_{2n}^{-1} + \zeta_{2n}^{+1})} = \frac{\pi/n}{\sin \pi/n}.$$

Here we have used the fact that $\zeta_{2n}^n = -1$. Note that the integral tends to 1 as $n \to \infty$, since $1/(1+x^n)$ converges to the indicator function of $[0, 1]$.

Definite integrals 2: rational functions of $\sin(\theta)$ and $\cos(\theta)$. Here is an even more straightforward application of the residue theorem: for any rational function $R(x, y)$, we can evaluate
$$\int_0^{2\pi} R(\sin\theta, \cos\theta)\, d\theta.$$

The method is simple: set $z = e^{i\theta}$ and convert this to an integral of an analytic function over the unit circle. To do this we simple observe that $\cos\theta = (z + 1/z)/2$, $\sin\theta = (z - 1/z)/(2i)$, and $dz = iz\, d\theta$. Thus we have:
$$\int_0^{2\pi} R(\sin\theta, \cos\theta)\, d\theta = \int_{S^1} R\left(\frac{1}{2i}\left(z - \frac{1}{z}\right), \frac{1}{2}\left(z + \frac{1}{z}\right)\right) \frac{dz}{iz}.$$

For example, for $0 < a < 1$ we have:
$$\int_0^{2\pi} \frac{d\theta}{1 + a^2 - 2a\cos\theta} = \int_{S^1} \frac{i\, dz}{(z-a)(az-1)} = 2\pi i(i/(a^2 - 1)) = \frac{2\pi}{1 - a^2}.$$

Definite integrals 3: fractional powers of x. $\int_0^\infty x^a R(x)\, dx$, $0 < a < 1$, R a rational function.

For example, consider
$$I(a) = \int_0^\infty \frac{x^a}{1 + x^2}\, dx.$$

Let $f(z) = z^a/(1 + z^2)$. We integrate out along $[0, \infty)$ then around a large circle and then back along $[0, \infty)$. The last part gets shifted by analytic continuation of x^a and we find
$$(1 - 1^a)I(a) = 2\pi i(\text{Res}(f, i) + \text{Res}(f, -i))$$

and $\text{Res}(f, i) = i^a/(2i)$, $\text{Res}(f, -i) = (-i)^a/(-2i)$ (since $x^a/(1 + x^2) = x^a/(x - i)(x + i)$). Thus, if we let $i^a = \omega = \exp(\pi i a/2)$, we have
$$I(a) = \frac{\pi(i^a - (-i)^a)}{(1 - 1^a)} = \pi\frac{\omega - \omega^3}{1 - \omega^4} = \frac{\pi}{\omega + \omega^{-1}} = \frac{\pi}{2\cos(\pi a/2)}.$$

For example, when $a = 1/3$ we get

$$I(a) = \pi/(2\cos(\pi/6)) = \pi/\sqrt{3}.$$

Residues and infinite sums. The periodic function $f(z) = \pi\cot(\pi z)$ has the following convenient properties: (i) it satisfies $f(z) = f(z+1)$; (ii) It has simple poles with residue 1 for all $z \in \mathbb{Z}$, and no other poles; and (iii) it remains bounded as $\operatorname{Im} z \to \infty$. From these facts we can deduce some remarkable properties: by integrating over a large rectangle $S(R)$, we find for $k \geq 2$ even,

$$0 = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{S(R)} \frac{f(z)\,dz}{z^k} = \operatorname{Res}(f(z)/z^k, 0) + 2\sum_1^\infty 1/n^k,$$

and thus, remarkably, we can evaluate the infinite sum using residues to obtain:

$$\zeta(k) = \sum_0^\infty \frac{1}{n^k} = -\frac{1}{2}\operatorname{Res}(f(z)/z^k, 0).$$

Here $\zeta(s) = \sum n^{-s}$ is the Riemann zeta function. For explicit formulas, we compute the Laurent series:

$$\begin{aligned}
\cot(z) &= \frac{\cos(z)}{\sin(z)} = \frac{1 - z^2/2! + z^4/4! - \cdots}{z(1 - z^2/3! + z^4/5! - \cdots)}\\
&= z^{-1}(1 - z^2/2! + z^4/4! - \cdots)(1 + z^2/6 + 7z^4/360 + \cdots)\\
&= z^{-1}(1 - z^2/3 - z^4/45 - 2z^6/945 - \cdots),
\end{aligned}$$

(using the fact that $(1/(3!)^2 - 1/5! = 7/360)$, which gives

$$f(z) = \pi\cot(\pi z) = z^{-1}(1 - \pi^2 z^2/3 - \pi^4 z^4/45 - 2\pi^6 z^6/945 \cdots).$$

Thus $\operatorname{Res}(f(z)/z^2, 0) = -\pi^2/3$ and hence we obtain Euler's famous formula:

$$\sum \frac{1}{n^2} = \frac{\pi^2}{6},$$

as well as

$$\sum \frac{1}{n^4} = \frac{\pi^4}{90},$$

and a justification for the value

$$\zeta(0) = 1 + 1 + 1 + \cdots = -1/2$$

33

(which turns out to be correct, if $\zeta(0)$ is defined by analytic continuation). Little is known about $\zeta(2k+1)$. Apéry showed that $\zeta(3)$ is irrational, but it is believed to be transcendental.

We note that $\zeta(s) = \sum 1/n^s$ is analytic for $\mathrm{Re}\, s > 1$ and extends analytically to $\mathbb{C} - \{1\}$ (with a simple pole at $s = 1$). Because of the factorization $\zeta(z) = \prod(1 - 1/p^s)^{-1}$, the behavior of the zeta function is closely related to the distribution of prime numbers. The famous *Riemann hypothesis* states that any zero of $\zeta(s)$ with $0 < \mathrm{Re}\, s < 1$ satisfies $\mathrm{Re}\, s = 1/2$. It implies a sharp form of the prime number theorem, $\pi(x) = x/\log x + O(x^{1/2+\epsilon})$. The zeta function also has *trivial zeros* at $s = -2, -4, -6, \ldots$.

Power series for $\tan(x)$. The fact that the power series

$$\tan(x) = z + \frac{z^3}{3} + \frac{2z^5}{15} + \frac{17z^7}{315} + \frac{62z^9}{2835} + \cdots$$

is much harder to describe than the power series for $\sin(x)$ and $\cos(x)$ is almost always quietly skirted in calculus classes. In fact, the power series for $\tan(x)$ and $\cot(x)$, and the value of the ζ-function at positive even integers, are all expressible in terms of the *Bernoulli numbers*, defined by

$$\sum_0^\infty B_k \frac{t^k}{k!} = \frac{t}{e^t - 1}.$$

For example, we have

$$\cot(x) = \sum_0^k B_{2k} \frac{(-1)^k 4^k x^{2k-1}}{(2k)!}.$$

The Bernoulli numbers themselves arise in the calculation of the *finite* sums, $\sum_1^N n^k$.

Hardy's paper on $\int \sin(x)/x \, dx$. We claim

$$I = \int_0^\infty \frac{\sin x \, dx}{x} = \frac{\pi}{2}.$$

Note that this integral is improper, i.e. it does not converge absolutely. Also, the function $f(z) = \sin(z)/z$ has no poles — so how can we apply the residue calculus?

The trick is to observe that

$$-2iI = \lim_{r \to 0} \int_{r < |x| < 1/r} \frac{e^{ix}\, dx}{x}.$$

We now use the fact that $|e^{ix+iy}| \le e^{-y}$ to close the path in the upper halfplane, and conclude that

$$2iI = \lim_{r \to 0} \int_{S^1(r)_+} \frac{e^{iz}\, dz}{z}.$$

(Here the semicircle is oriented counter-clockwise, as usual.) Since $\mathrm{Res}(e^{iz}/z, 0) = 1$, we find $2iI = (2\pi i)(1/2)$ and hence $I = \pi/2$.

1.7 Harmonic functions

In this section we relate complex analytic functions — defined by the vanishing of a first order operator — to real harmonic functions, defined by the vanishing of a second order operator. Harmonic functions can be defined on $\mathbb{R}^n$ and indeed on any Riemannian manifold, and they play a central role in differential geometry and mathematical physics.

Harmonic functions on $\mathbb{R}^n$. A C^2 real-valued function u on $\mathbb{R}^n$ is *harmonic* if

$$\Delta u = \sum \frac{d^2 u}{dx_i^2} = 0.$$

Equivalently, we have

$$d * du = 0,$$

where $*$ is the Hodge star operator (satisfying $dx_i \wedge *dx_i = dx_1 \cdots dx_n$). (This formulation shows Δu is naturally a volume form).

In physical terms, $\nabla \cdot \nabla u = 0$, and $d^2 u = 0$ is equivalent to $\nabla \times \nabla u = 0$; i.e., u generates a *volume preserving flow* with *zero curl* (since it is a gradient flow). Laplaces's equation says that u locally minimizes the *Dirichlet energy* $\int |\nabla u|^2\, dV$

Theorem 1.50 *A harmonic function on a domain $\mathbb{R}^n$ satisfies the mean value property and the maximum principle.*

Proof. Let $u : U \to \mathbb{R}$ (or $\mathbb{C}$) be harmonic. Given $p \in U$ let $A(r)$ be the average of u over the sphere $S^{n-1}(p, r)$, assuming the ball it encloses is contained in U. It suffices to prove the mean value property, $A(r) = u(p)$; and for this, it suffices to show that $A'(r) = 0$.

But $A'(r)$ is proportional to the average of $(\nabla u) \cdot \widehat{n}$ over the sphere of radius r, where $\widehat{n}$ is the unit normal vector field to the sphere. By Stokes' theorem this average is constant multiple of the integral of $\text{div} \nabla u = \Delta u$ over $B(p, r)$, which is zero since u is harmonic. ∎

Harmonic functions on $\mathbb{C}$. In the case of $\mathbb{R}^2 = \mathbb{C}$, we have

$$4 \frac{d^2 u}{dz \, d\bar{z}} = \Delta u.$$

Thus Δu is a constant multiple of the $(1,1)$–form $\partial \bar{\partial} u = -\bar{\partial} \partial u$.

Basic facts:

1. If $f = u + iv$ is analytic, then f, $\bar{f}$, u and v are harmonic.

2. A function u is harmonic iff du/dz is holomorphic.

3. Any real-valued harmonic function u is locally the real part of a holomorphic function $f = u + iv$.

   Indeed, one can integrate ∂u to obtain an analytic function with $\partial f = \partial u$. Then $\overline{\partial f} = \bar{\partial} u$; thus $d(f + \bar{f}) = du$ and so $u = f + \bar{f}$ up to an additive constant.

   The function v is called a *harmonic conjugate* of u; it is well–defined up to an additive constant.

4. Thus any C^2 harmonic function is actually infinitely differentiable.

5. A harmonic function satisfies the mean-value theorem: $u(p)$ is the average of $u(z)$ over $S^1(z, p)$.

6. A harmonic function satisfies the maximum principle. (The real part of an analytic functions defines an open map to $\mathbb{R}$.)

7. If u is harmonic and f is analytic, then $u \circ f$ is also harmonic. (Since $\partial \bar{\partial}(u \circ f) = f^*(\partial \bar{\partial} u)$).

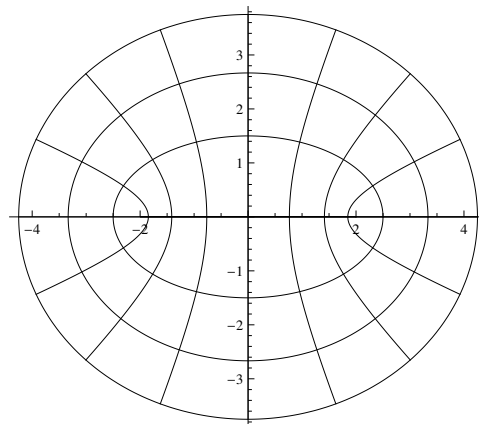8. A uniform limit of harmonic functions is harmonic.

Figure 1.1. Orthogonal level sets.

Examples. The function $\mathrm{Re}(z^3) = x^3 - 3xy^2$ is a harmonic polynomial. The function $\arg z$ is the harmonic conjugate of $\log|z|$. This shows the harmonic conjugate may be multivalued.

Orthogonality and harmonic conjugates. It is an important fact that the level sets of a pair of harmonic conjugates u and v are orthogonal. This follows, e.g. from the fact that (u, v) are just the pullbacks of (x, y) under the conformal map $f(z) = u + iv$. More formally, their gradients satisfy

$$2\overline{\partial}(u + iv) = 0 = \nabla u + i\nabla v,$$

so in fact $\nabla v = i\nabla u$.

The condition that u is harmonic is the same as saying that $i\nabla u = \nabla v$ for some v. Equivalently, it says that $d(\overline{\partial}u) = \partial\overline{\partial}u = 0$.

Flows. Since the level sets of u and v are orthogonal, the area-preserving flow generated by ∇u follows the level sets of v, and vice-versa. A simple example is provided by polar coordinates, which give the level sets of u and v where $\log|z| = u + iv$. The area-preserving flows are rotation around the origin, and the radial flow at rate $1/r$ through circles of radius r.

See Figure 1.1 for another example, this time on $U = \mathbb{C} - \{-2, 2\}$, where the level sets are conics. These conics are the images of radial lines and circles under $f(z) = z + 1/z$, so they are also locally level sets of harmonic functions.

Electricity. It is often useful to think of the physics behind harmonic functions, namely the electric potential u associated to a charge distribution

ρ satisfies $\Delta u = \rho$, and the corresponding electric field is given by $E = \nabla u$. Thus u is harmonic in the charge-free regions of space, and for a compact region U, the flux of E through ∂U gives the amount of charge in U.

The potential $u(z) = \log|z|$ represents a point charge at the origin, while $u(z) = \mathrm{Re}(1/z)$ represents a dipole. The level sets of this section function are tangent circles at $z = 0$, arising as the pullbacks of the level sets of $\mathrm{Re}(z)$ under $z \mapsto 1/z$.

Harmonic extension. Here is one of the central existence theorems for harmonic functions.

Theorem 1.51 *There is a unique linear map* $P : C(S^1) \to C(\overline{\Delta})$ *such that* $u = P(u)|_{S^1}$ *and* $P(u)$ *is harmonic on* Δ.

Proof. Uniqueness is immediate from the maximum principle. To see existence, observe that we must have $P(\overline{z}^n) = \overline{z}^n$ and $P(z^n) = z^n$. Thus P is well-defined on the span S of polynomials in z and $\overline{z}$, and satisfies there $\|P(u)\|_\infty = \|u\|_\infty$. Thus P extends continuously to all of $C(S^1)$. Since the uniform limit of harmonic functions is harmonic, $P(u)$ is harmonic for all $u \in C(S^1)$. ∎

Poisson kernel. The map P can be given explicit by the *Poisson kernel*. For example, $u(0)$ is just the average of u over S^1. We can also say $u(p)$ is the expected value of $u(z)$ under a random walk starting at p that exits the disk at z.

To find the Poisson kernel explicitly, suppose we have a δ-mass at $z = 1$. Then it should extend to a positive harmonic function u on Δ which vanishes along S^1 except at 1, and has $u(0) = 0$. In turn, u should be the real part of an analytic function $f : \Delta \to \mathbb{C}$ such that $f(0) = 1$ and $\mathrm{Re}\, f|S^1 = 0$ and f has a pole at $z = 1$. Such a function is given simply by the Möbius transformation $f : \Delta \to U = \{z : \mathrm{Re}\, z > 0\}$:

$$f(z) = \frac{1+z}{1-z} = 1 + \sum_1^\infty 2z^n.$$

Convolving, we find the analytic function with $\mathrm{Re}\, f = u$ for a given $u \in C(S^1)$ is given by

$$F(z) = \frac{1}{2\pi} \int_{S^1} f(z/t) u(t) |dt|,$$

38

and thus
$$u(r, \alpha) = \frac{1}{2\pi} \int_{S^1} P_r(\alpha - \theta) u(\theta) \, d\theta,$$

where, for $z = re^{i\theta}$, we have

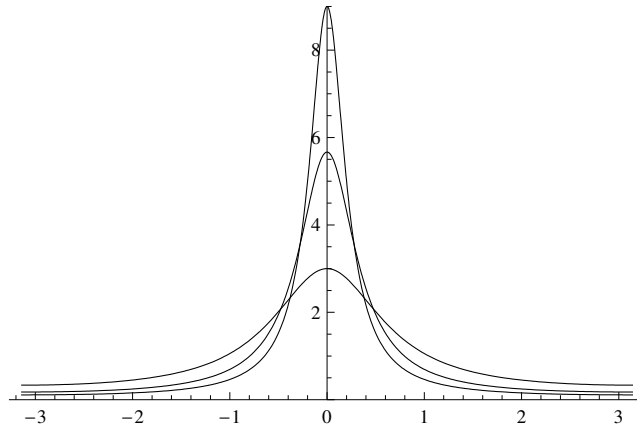$$P_r(\theta) = \operatorname{Re} f(z) = \frac{1 - |z|^2}{|1 - z|^2} = \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

Figure 1.2. The Poisson kernel $P_r(\theta)$ for $r = 0.5, 0.7, 0.8$.

Relation to Fourier series. The above argument suggests that, to define the harmonic extension of u, we should just write $u(z) = \sum_{-\infty}^{\infty} a_n z^n$ on S^1, and then replace z^{-n} by $\bar{z}^n$ to get its extension to the disk. This actually works, and gives another approach to the Poisson kernel.

In fact, the δ-function at $z = 1$ on S^1 is formally given by

$$\delta = \sum_{-\infty}^{\infty} z^n = 1 + 2\operatorname{Re} \sum_{1}^{\infty} z^n = \operatorname{Re} f(z),$$

and so its extension to Δ is given by

$$P_r(\theta) = \frac{1}{1 - z} + \frac{1}{1 - \bar{z}} - 1,$$

where $z = \exp(i\theta)$.

39

Given $u \in C(S^1)$, it is not true, in general, that its Fourier series $\sum a_n z^n$ converges for all $z \in S^1$. However, it *is* true that this series converges in the unit disk and defines a harmonic function there. As we have seen, this harmonic function provides a continuous extension of u. This shows:

Theorem 1.52 *If $u \in C(S^1)$ has Fourier coefficients a_n, then for all $z \in S^1$ we have*

$$u(z) = \lim_{r \to 1-} \sum_{-\infty}^{\infty} a_n r^{|n|} z^n.$$

(Here $a_n = (1/2\pi) \int_{S^1} u(\theta) \exp(-n\theta) |d\theta|$.)

Abel summation. In general, we say S is the *Abel sum* of the series $\sum b_n$ if $S = \lim_{r \to 1-} \sum b_n r^n$. For example,

$$1 - 2 + 3 - 4 + \cdots = \lim(1 - 2r + 3r^2 - \cdots) = \lim -1/(1+r)^2 = -1/4.$$

The result above shows the Fourier series of f is *Abel summable* to the original function f.

Liouville's theorem. Using the Poisson kernel, it is easy to see that a harmonic function u on $\mathbb{C}$ is the real part of an analytic function defined on all of $\mathbb{C}$ (this is true for any simply connected domain). From this one can deduce that a bounded harmonic function on $\mathbb{C}$ is constant. In fact we have:

Theorem 1.53 *Any positive harmonic function on $\mathbb{C}$ is constant.*

Proof. Write $u = \mathrm{Re}(f)$ with f analytic; then $f(\mathbb{C})$ is not dense in $\mathbb{C}$, so f is constant. ∎

These results generalize to $\mathbb{R}^n$, where other proofs are available.

Laplacian as a quadratic form, and physics. Suppose $u, v \in C_c^\infty(\mathbb{C})$ – so u and v are smooth, real-valued functions vanishing outside a compact set. Then, by integration by parts, we have

$$\int_\Delta \langle \nabla u, \nabla v \rangle = -\int_\Delta \langle u, \Delta v \rangle = -\int_\Delta \langle v, \Delta u \rangle.$$

To see this using differential forms, note that:

$$0 = \int_\Delta d(u * dv) = \int_\Delta (du)(*dv) + \int_\Delta u(d * dv).$$

In particular, we have

$$\int_\Delta |\nabla u|^2 = -\int_\Delta u\Delta u.$$

Compare this to the fact that $\langle Tx, Tx\rangle = \langle x, T^*Tx\rangle$ on any inner product space. Thus $-\Delta$ defines a positive-definite quadratic form on the space of smooth functions.

The extension of u from S^1 to Δ is a 'minimal surface' in the sense that it minimizes $\int_\Delta |\nabla u|^2$ over all possible extensions. Similarly, minimizing the energy in an electric field then leads to the condition $\Delta u = 0$ for the electrical potential.
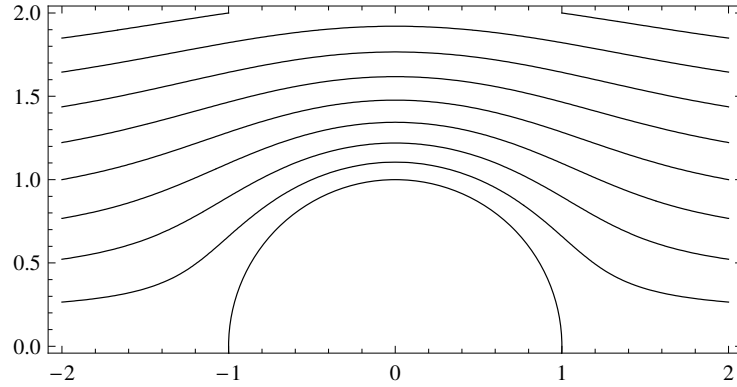

Figure 1.3. Streamlines around a cylinder (idealized).

Probabilistic interpretation. Brownian motion is a way of constructing random paths in the plane (or $\mathbb{R}^n$). It leads to the following simply interpretation of the extension operator P. Namely, given $p \in \Delta$, one considers a random path p_t with $p_0 = p$. With probability one, there is a first $T > 0$ such that $|p_T| = 1$; and then one sets $u(p) = E(u(p_T))$. In other words, $u(p)$ is the expected value of $u(p_T)$ at the moment the Brownian path exits the disk.

Using the Markov property of Brownian motion, it is easy to see that $u(p)$ satisfies the mean-value principle, which is equivalent to it being harmonic. It is also easy to argue that $|p_0 - p_T|$ tends to be small when p_0 is close to S^1, and hence $u(p)$ is a continuous extension of $u|S^1$.

The Poisson kernel $(1/2\pi)P_r(\theta)d\theta$ gives the *hitting density* on S^1 for a Brownian path starting at $(r, 0)$.

41

Hyperbolic geometry interpretation. Alternatively, $u(z)$ is the expected value of $u(p)$ and the endpoint of a random hyperbolic geodesic ray γ in Δ with one vertex at z. (The angle of the ray in $T_z\Delta$ is chosen at random in S^1.)

Fluid flow around a cylinder. We begin by noticing that $f(z) = z + 1/z$ gives a conformal map from the region $U \subset \mathbb{H}$ where $|z| > 1$ to $\mathbb{H}$ itself, sending the circular arc to $[-2, 2]$. Thus the level sets of $\operatorname{Im} f = y(1 - 1/(x^2+y^2))$ describe fluid flow around a cylinder. Note that we are modeling incompressible fluid flow with *no rotation*, i.e. we are assuming the curl of the flow is zero. This insures the flow is given by the gradient of a function.

Note that $f(z)$ tends to the identity at infinity, but $|f'(z)| = |1 - 1/z^2| \leq 1 + 1/|z|^2 \leq 2$ gets close to two near the top of the cylinder. Thus a flow line starting at the left at a small height h and moving at unit speed is compressed to height $h/2$ as it runs along the top of the cylinder, and moves at speed close to 2.

In this idealized model for fluid flow, there is no drag on the cylinder; this physically unreasonable conclusion is known as *d'Alembert's paradox.* In reality the velocity of the fluid should tend to zero as one approaches the cylinder.

Harmonic functions and the Schwarz reflection principle. Here is an application of harmonic functions that will be repeatedly used in geometric function theory.

Let $U \subset \mathbb{C}$ be a region invariant under $z \mapsto \bar{z}$. Suppose $f : U \to \mathbb{C}$ is continuous,

$$f(\bar{z}) = \overline{f(z)}, \tag{1.4}$$

and f is analytic on $U_+ = U \cap \mathbb{H}$. We can then conclude that f *is analytic on U.* In particular, f is analytic at each point of $U \cap \mathbb{R}$.

Here is a stronger statement:

Theorem 1.54 *Suppose $f : U_+ \to \mathbb{C}$ is analytic and $\operatorname{Im} f(z) \to 0$ as $\operatorname{Im} z \to 0$. Then f extends to an analytic function on U satisfying (1.4).*

Proof. The statement is local, so we can assume $U = B(p, r)$ where $p \in \mathbb{R}$. Let $f(z) = u(z) + iv(z)$; then v is harmonic on U_+ and v extends continuously to the real axis, with $v(x) = 0$. Extend v to U by $v(\bar{z}) = -v(z)$.

Now by the Poisson integral, $v|S^1(p, r)$ extends to a unique harmonic function h on $B(p, r)$. By uniqueness, $h(\bar{z}) = -h(z)$, and hence h also

vanishes on the real axis. Thus $h = v$ on ∂U_+. By the maximum principle, $h = v$ on U_+, and hence on U.

We are now done: by the existence of harmonic conjugates, there is *some* analytic function on U with $\operatorname{Im} F = v$. But then $\operatorname{Im} F = \operatorname{Im} f$ on U_+, so F and f differ by a real constant on U_+, which can be normalized to be zero.

∎

Example. Suppose $f(z)$ is analytic on $\mathbb{H}$ and $f(z) \to 0$ along an interval in $\mathbb{R}$. Then f is identically zero. This extends the 'isolated zero' principle to the boundary of a region.

Remark. One can replace $z \mapsto \overline{z}$ with reflection through any circle, or more generally with local reflection through a real-analytic arc. For example, we have:

Theorem 1.55 *Let $U \subset \Delta$ be a region and let $f : U \to \Delta^*$ be analytic. Suppose $|f(z)| \to 1$ as $|z| \to 1$ in U. Then f extends to $U \cup \rho(U)$ by Schwarz reflection, where $\rho(z) = 1/\overline{z}$.*

Proof. We can assume that $U \subset \Delta^*$ since the extension issue is local along S^1. Lift f to an analytic map $\widetilde{f} : \widetilde{U} \to \mathbb{H}$, using the universal covering map $\pi : \mathbb{H} \to \Delta^*$ given by $\pi(z) = e^z$ on the domain and range. Then $\operatorname{Im} \widetilde{f}(z) = \log|f(\pi(z))| \to 0$ as $|\operatorname{Im}(z)| \to 0$ in $\widetilde{U}$, so we can apply the reflection principle in $\mathbb{H}$ to $\widetilde{f}$. It then descends by periodicity to give the require extension of f. ∎

1.8 Additional topics

Here we briefly mention some other classical topics in complex analysis.

The Phragmen–Lindelöf Theorems. These theorems address the following question. Suppose $f(z)$ is an analytic function on the horizontal strip $U = \{x + iy : a < y < b\}$, and continuous on $\overline{U}$. Can we assert that $\sup_U |f| = \sup_{\partial U} |f|$?

The answer is no, in general. However, the answer is yes if $f(x + iy)$ does not grow too rapidly as $|x| \to \infty$. In fact, this is a property of harmonic functions $u(z)$. The point is that if we truncate the strip to a rectangle by cutting along the lines where $|x| = R$, then the harmonic measure of the ends

(as seen from a fixed point $z \in U$) tends to zero exponentially fast. Thus if $|u(x + iy)| = O(|x|^n)$ for some n, we get the desired control.

Runge's theorem. Here is an interesting and perhaps surprising application of Cauchy's formula.

Let $K \subset \mathbb{C}$ be a compact set, let $C(K)$ denote the Banach space of continuous functions with the sup-norm, and let $A(K), R(K)$ and $P(K) \subset C(K)$ denote the closures of the analytic, rational and polynomial functions. *Note that all three of these subspaces are algebras.*

Example. For $K = S^1$ we have $R(K) = A(K) = C(K)$ by Fourier series, but $P(K) \neq C(K)$, since $\int_{S^1} p(z)\, dz = 0$ for any polynomial. In particular, $1/z \notin P(S^1)$.

Remarkably, if we remove a small interval from the circle to obtain an arc $K = \exp[0, 2\pi - \epsilon]$, then $1/z$ *can* be approximated by polynomials on K.

Theorem 1.56 (Runge) *For any compact set $K \subset \mathbb{C}$ we have $R(K) = A(K)$, and $P(K) = A(K)$ provided $\mathbb{C} - K$ is connected.*

Proof. For the first result, suppose $f(z)$ is analytic on a smoothly bounded neighborhood U of K. Then we can write

$$f(z) = \frac{1}{2\pi i} \int_{\partial U} \frac{f(t)\, dt}{t - z} = \int_{\partial U} F_z(t)\, dt.$$

Since $d(z, \partial U) \geq d(K, \partial) > 0$, the functions $\{F_z\}$ range in a compact subset of $C(\partial U)$. Thus we can replace this integral with a finite sum at the cost of an error that is small independent of z. But the terms $f(t_i)/(t_i - z)$ appearing in the sum are rational functions of z, so $R(K) = A(K)$.

The second result is proved by pole–shifting. By what we have just done, it suffices to show that $f_p(z) = 1/(z - p) \in P(K)$ for every $p \notin K$. Let $E \subset \mathbb{C} - K$ denote the set of p for which this is true.

Clearly E contains all p which are sufficiently large, because then the power series for $f_p(z)$ converges uniformly on K. Also E is closed by definition. To complete the proof, it suffices to show E is open.

The proof that E is open is by 'pole shifting'. Suppose $p \in E$, $q \in B(p, r)$ and $B(p, r) \cap K = \emptyset$. Note that $f_q(z)$ is analytic on $\mathbb{C} - B(p, r)$, and tends to zero as $|z| \to \infty$. Thus $f_q(z)$ can be expressed as a power series in $1/(z - p)$:

$$(z - q)^{-1} = \sum_0^\infty a_n (z - p)^{-n} = \sum a_n f_p(z)^n,$$

convergent for $|z - p| > |z - q|$, and converging uniformly on K. (Compare the expression

$$\frac{1}{z-1} = \sum_{n=1}^{\infty} \frac{1}{z^n},$$

valid for $|z| > 1$.) Since $(z - q)^{-1} \to 0$ as $|z| \to \infty$, only terms with $n \geq 0$ occur on the right. But $f_p \in A(K)$ by assumption, and $A(K)$ is an algebra, so it also contains f_p^n. Thus $f_q \in A(K)$ as well. ∎

Aside: Lavrentiev's Theorem. It can be shown that if $\mathbb{C} - K$ is connected and K has no interior, then $A(K) = C(K)$. This is definitely false for a fat Swiss cheese: if ∂K is rectifiable and of finite length, then $\int_{\partial K} f(z)\,dz = 0$ for all f in $A(K)$, and so $A(K) \neq C(K)$. For more details see [Gam].

Applications: pointwise convergence. Runge's theorem can be used to show easily that there is a sequence of polynomials $f_n(z)$ that converge pointwise, but whose limit is not even continuous. Indeed, let $A_n = [0, n]$ and let $B_1 \subset B_2 \subset \cdots$ be an increasing sequence of compact sets such that $\bigcup B_n = \mathbb{C} - [0, \infty)$. Then every $z \in \mathbb{C}$ eventually belongs to A_n or B_n. Let $f_n(z)$ be a polynomial, whose existence is guaranteed by Runge's theorem, such that $|f_n(z)| < 1/n$ on A_n and $|f_n(z) - 1| < 1/n$ on B_n. Then clearly $\lim f_n(z) = 0$ on $[0, \infty)$ and $\lim f_n(z) = 1$ elsewhere.

Applications: embedding the disk into affine space. Runge's theorem can also be used to show there is a *proper* embedding of the unit disk into $\mathbb{C}^3$. See [Re, §12.3].

1.9 Exercises

1. For $a, b \in \mathbb{C}$, express $a\overline{b}$ in terms of the dot-product and cross-product of the vectors represented by a and b.

2. Let $v(z)$ be a smooth vector field in $\mathbb{C}$. Express dv/dz in terms of $\mathrm{div}(v)$ and $\mathrm{curl}(v)$.

3. Let $f(z)$ be one-to-one and analytic on a neighborhood of the unit circle $S^1 \subset \mathbb{C}$. Relate $\int_{S^1} \overline{f(z)} f'(z)\,dz$ to the area of the region enclosed by $f(S^1)$.

4. Compute df/dz and $df/d\overline{z}$ for the map $f : \mathbb{C} \to \mathbb{C}$ given in polar coordinates by $f(r, \theta) = (r^\alpha, n\theta)$, $\alpha > 0$, $n \in \mathbb{Z}$.

5. Suppose $f(z)$ is analytic. Compute $(d/dz)(d/d\bar{z})|f(z)|^2$.

6. When is a polynomial $p(x, y)$ expressible in the form

$$p(x, y) = f(x + iy) + g(x - iy),$$

where f and g are holomorphic?

7. Let $f_1, \ldots, f_n$ be analytic functions on Δ. Suppose $\sum |f_i(z)|^2 = 1$ for all $z \in \Delta$.

(i) Show all f_i are constant functions, by taking the Laplacian of both sides of this equation.

(ii) Show all f_i are constant functions, by applying the maximum principle (or open mapping theorem) to suitable linear combinations of these functions.

(iii) Let $H : \mathbb{C}^n \to \mathbb{R}$ be a strictly convex smooth function. Generalize these statements to show that if $H(f_1(z), \ldots, f_n(z))$ is constant on Δ, then each function $f_i(z)$ is constant. (Hint: use the supporting hyperplane for the level set of H through $f(0)$.)

8. Let $p(x) \in \mathbb{C}[x]$ be a nonconstant polynomial, and let $f(z) = \sum_0^\infty p(n)z^n$.

(i) Show that $f(z)$ represents a rational function, $P(z)/Q(z)$.

(ii) What is the radius of convergence of this series?

(Hint: differentiate $\sum_0^\infty z^n$.)

9. Consider the sequence $(a_0, a_1, \ldots) = (2, 1, 3, 4, 7, 11, 18, 29, 47, 76, \ldots)$ defined by $a_n = \mathrm{Tr}(A^n)$, where $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$.

(i) Show that a_n satisfies the Fibonacci recursion relation, $a_{n+2} = a_{n+1} + a_n$.

(ii) Find the rational function $P(z)/Q(z)$ represented by $f(z) = \sum_0^\infty a_n z^n$.

(iii) Show that $|a_n - \gamma^n| \to 0$, where $\gamma = (1 + \sqrt{5})/2$.

10. (i) Show that $\log(2) = 1 - 1/2 + 1/3 - 1/4 + \cdots$.

(ii) Show that $\log(2) = \sum_1^\infty 2^{-k}/k$. (Hint: this is obtained from (i) by Euler's method, but a much simpler proof can be given.)

(ii) Sum the first 5 terms of each series, and compare the results to the actual value of $\log(2)$.

11. Let ω be a smooth 1–form on a region $U \subset \mathbb{C}$. Formulate properties P1 and P2 of loops γ in U such that:

(i) ω is closed ($d\omega = 0$) if and only if $\int_\gamma \omega = 0$ whenever P1 holds; and

(ii) ω is exact ($\omega = df$) if and only if $\int_\gamma \omega = 0$ whenever P2 holds.

Justify your answer.

12. Let $f_n \to f$ uniformly on compact subsets of an open connected set $\Omega \subset \mathbb{C}$, where f_n is analytic, and f is not identically equal to zero.

Show if $f(w) = 0$ then we can write $w = \lim z_n$, where $f_n(z_n) = 0$ for all n sufficiently large.

13. Let $C(\overline{\Delta})$ denote the Banach space of continuous, complex–valued functions on the unit disk, with $\|f\| = \sup |f(z)|$. Let $f \in C(\overline{\Delta})$ be analytic on Δ.

(i) Prove that the $\mathbb{C}$–algebra $\mathbb{C}[f]$ generated by f is not dense in $C(\overline{\Delta})$.

(ii) Prove that $\mathbb{C}[f, \overline{f}]$ is dense in $C(\overline{\Delta})$ $\iff$ f is injective on $\overline{\Delta}$. (Here $\overline{f}(z) = \overline{f(z)}$.)

14. Evaluate:
$$\int_{-\infty}^{\infty} \frac{x^6}{(1 + x^4)^2} \, dx.$$

15. Let $f : \mathbb{C} \to \mathbb{C}$ be analytic and let $U \subset \mathbb{C}$ be a bounded region. Suppose $|f(z)|$ is constant on ∂U. Show that either f is constant, or f has a zero in U.

16. Compute the Laurent series centered at $z = 0$ such that
$$\sum_{-\infty}^{\infty} a_n z^n = \frac{1}{z(z - 1)(z - 2)}$$

in the region $1 < |z| < 2$.

17. Show for any polynomial $p(z)$ there is a z with $|z| = 1$ such that $|p(z) - 1/z| \geq 1$.

18. Let $U \subset \mathbb{C}$ be a region with $0 \in U$, and let

$$\mathcal{O}_M(U) = \{f \in \mathcal{O}(U) \ : \ \|f\|_U = \sup_U |f(z)| \leq M\}.$$

(i) Prove that for any sequence $f_n \in \mathcal{O}_M(U)$, the conditions

(a) $f_n \to 0$ uniformly on compact subset of U, and
(b) for each $k \geq 0$, $f_n^{(k)}(0) \to 0$,

are equivalent.

(b) Construct a sequence $f_n \in \mathcal{O}(\Delta)$ such that for each $k \geq 0$, $f_n^{(k)}(0) \to 0$, but $|f_n(z)| \to \infty$ for all $z \neq 0$.

19. Let $\mathcal{O}_p = \dir\lim \mathcal{O}(B(p, 1/n))$ denote the ring of *germs* of analytic functions defined near $p \in \mathbb{C}$. (An element of $\mathcal{O}_p$ is uniquely specified by a power series $\sum a_n(z - p)^n$ with positive radius of convergence.) Prove that $\mathcal{O}_p$ is a *local ring*, i.e. it has a unique maximal ideal m; and that $\mathrm{ord}_p(f) = \sup\{n \ : \ f \in m^n\}$. What is the residue field $\mathcal{O}_p/m$?

20. Let $f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a continuous function, and suppose that f does not vanish on the unit sphere S^n. Define a map $F : S^n \to S^n$ by $F(x) = f(x)/|f(x)|$. Prove that if F has nonzero degree, then f has a zero in the unit ball.

21. Let $p(n)$ be the number of partitions of n into *unequal* parts. (E.g. $p(7) = 5$ because 7 can be written as 7, $1+2+4$, $1+6$, $2+5$ and $3+4$.)

(i) Show that $f(z) = \prod_1^\infty (1 + z^n)$ defines an analytic function on Δ.
(ii) Show that $f(z) = \sum_0^\infty p(n)z^n$.

22. 'There is no analytic function $f : \mathbb{C} \to \mathbb{C}$ such that $f(S^1)$ traces out a figure 8 in $\mathbb{C}$.' Make this statement precise, and then prove it.

23. Let $p(z)$ be a polynomial of degree $d \geq 2$, with distinct roots $r_1, \ldots, r_d$. Show that $\sum 1/p'(r_i) = 0$.

24. Let $E \subset (-1, 1)$ be a compact set of zero linear measure, and let $f : \Delta - E \to \mathbb{C}$ be a bounded analytic function. Show that f extends to an analytic function on the whole disk.

25. Let $f_n(z)$ be a sequence of polynomials such that $f_n(z) \to f(z)$ point-wise in $\mathbb{C}$. Show there exists a dense, open set $U \subset \mathbb{C}$ such that $f|U$ is analytic. (You may quote the Baire Category theorem.)

26. *Find a sequence of analytic function $f_n : \mathbb{C} \to \mathbb{C}$ such that $f(z) = \lim f_n(z)$ exists for each $z \in \mathbb{C}$, but $f(z)$ is not analytic.

27. For each $n > 0$, give an example of an analytic function $f : \Delta^* \to \mathbb{C}$ such that (i) f is nowhere zero, (ii) f has an essential singularity at $z = 0$, and (iii) $\mathrm{Res}(f'/f, 0) = n$.

28. Consider the equation

$$\int_0^\infty \frac{x^\alpha \, dx}{(1 + x^2)^2} = \frac{\pi(1 - \alpha)}{4\cos\frac{1}{2}\pi\alpha}.$$

Find the set of real α such that the integral above is absolutely convergent, and prove the equation above holds for all such α.

29. Let $C(R)$ be the circular arc defined by $|z| = R$ and $\mathrm{Im}(z) > 0$. Prove that $\int_{C(R)} (e^{iz}/z) dz \to 0$ as $R \to \infty$. (This is used in our calculation of $\int_0^\infty \sin(x)/x \, dx$.)

30. Prove the identity

$$\int_0^{2\pi} \cos^{2n}\theta \, d\theta = 2\pi \frac{1 \cdot 3 \cdot 5 \cdots (2n - 1)}{2 \cdot 4 \cdot 6 \cdots (2n)}$$

by applying the residue theorem to $\int_{S^1} (z + 1/z)^{2n} \, dz/z$.

31. Given $w \in \mathbb{C}$, find an explicit sequence $z_n \to 0$ such that $e^{1/z_n} \to w$.

32. Let $p(z)$ be a polynomial of degree d with no zeros on the imaginary axis. (i) Prove that

$$N(p) = \lim_{R \to \infty} \frac{1}{2\pi i} \int_{-iR}^{iR} \frac{p'(z)}{p(z)} \, dz$$

exists (where the integral is taken along the imaginary axis). (ii) Prove that the number of zeros of $p(z)$ with $\mathrm{Re}(z) > 0$ is given by $d/2 - N(p)$.

33. Suppose $a, b, c \in \mathbb{R}$, and $b \leq 0$. How many zeros does

$$p(z) = z^4 + az^3 + bz^2 + cz + 1$$

have in the region $\text{Re}(z) > 0$?

(Hint: first prove that the answer does not depend on the values of a, b or c, by showing $p(iy) \neq 0$, $y \in \mathbb{R}$.)

34. Using residues, compute the average value of $1/(1 + \cos^2(\theta))$ (over the circle).

35. (a) Show that there exists a unique analytic function $f(z)$ on $\mathbb{C} - [-1, 1]$ such that (i) $f(z)^2 + z^2 = 1$ and (ii) $\text{Im} f(z) > 0$ for $z \in (1, \infty)$. (Hint: use the theory of covering spaces.)

(b) What is $\lim f(iy)$ as $y \to 0$ from above? From below?

(c) Use (a), (b) and the residue theorem to evaluate

$$\int_{-1}^{1} \frac{\sqrt{1 - x^2}}{1 + x^2} \, dx.$$

36. Compute the first four nonzero terms in the power series for $\tan(z)$ at $z = 0$ by formally inverting the power series for

$$\tan^{-1}(z) = \int \frac{dz}{1 + z^2}.$$

37. Suppose a pair of analytic functions on a region U satisfy $|f(z)| = |g(z)|$. Prove that $f(z) = ag(z)$ for some constant a with $|a| = 1$.

38. (Feynman's method.) Evaluate $J = \int_0^\infty \sin(x)/x \, dx$ as follows. First, for all $a \geq 0$, define

$$I(a) = \int_0^\infty \exp(-ax) \sin(x)/x \, dx.$$

Then, compute $I'(a)$ in closed form by differentiating under the integral sign. Prove that this computation is justified!

Finally, use the fact that $I(a) \to 0$ as $a \to \infty$ to evaluate $I(0) = J$.

39. Let $H_d \subset \mathbb{C}[x,y]$ be the space of harmonic polynomials p that are homogeneous of degree d. (Harmonic means $d^2p/dx^2 + d^2p/dy^2 = 0$.) Compute $\dim H_d$, give an explicit example of a nonzero polynomial $p(x,y) = \sum a_{ij}x^iy^j \in H_d$, and show the $\oplus H_d|S^1$ spans a dense subset of $C(S^1)$, the space of continuous functions on the circle $S^1 \subset \mathbb{C}$ with $\|f\| = \sup_{S^1}|f(z)|$. Explain these results in terms of analytic functions.

40. Let $f(z)$ be an analytic function on Δ with $f(0) = 0$ and $\int_\Delta |f(z)|^2 |dz|^2 < \infty$. Prove that:

$$\int_\Delta |\operatorname{Re}(f)|^2 |dz|^2 = \int_\Delta |\operatorname{Im}(f)|^2 |dz|^2.$$

41. Let $f : \Delta \to \mathbb{C}$ be an analytic whose power series $f(z) = \sum a_n z^n$ has integral coefficients, Suppose $\int_\Delta |f(z)|^2 |dz|^2 < \infty$. Prove that $f(z)$ is a polynomial.

42. Let $f(z)$ be analytic in a region U, and suppose $\int_U |f(z)| |dz|^2 < \infty$. Prove that there is a constant M such that for all $z \in U$ we have

$$|f(z)| \le \frac{M}{d(z, \partial U)^2}.$$

43. Use the preceding two problems to prove that a uniform limit of harmonic functions on a domain $U \subset \mathbb{C}$ is harmonic.

44. Find a bounded harmonic function $u : \Delta \to \mathbb{R}$ such that $\lim_{r \to 1} u(re^{i\theta}) = 1$ for $\theta \in (0, \pi)$ and $= -1$ for $\theta \in (\pi, 2\pi)$.

45. Give an example of a bounded harmonic function u on the unit disk whose harmonic conjugate v is unbounded.

46. Let $u : \overline{\Delta} \to \mathbb{R}$ be a smooth function such that $\Delta u(z) > 0$ for all $z \in \Delta$. Show that u satisfies the maximum principle.

47. Give an example of an unbounded analytic function on the unit disk with the property that $\lim_{z \to w} |f(z)| = 1$ for all $w \in S^1$ except $w = 1$.

48. Given $R > 1$ let $A_R = \{z : 1 < |z| < R\}$. Let $f : A_R \to A_S$ be a bijective holomorphic map. Prove that $R = S$. (Hint: apply Schwarz reflection to extend f to $\mathbb{C}^*$.)

49. Let $p(z)$ be a polynomial of degree $d > 1$. Show that the zeros of $p'(z)$ belong to the convex hull of the zeros of $p(z)$. (The convex hull of a set E is the intersection of all the convex sets containing E.)

   Hint: consider $p'(z)/p(z)$.

50. Let $p, q \in \mathbb{R}[x]$ be polynomials of degree $d > 0$ and assume pq has $2d$ simple zeros in $\mathbb{C}$. Show the following are equivalent:

   (i) The zeros of $p + iq$ all lie in the same component of $\mathbb{C} - \mathbb{R}$;

   (ii) The zeros of p and q are all real, and they are interlaced: a zero of q lies between any two consecutive zeros of p, and vice–versa.

51. Define $u(z)$ on S^1 by $u(z) = 1$ if $z^2 \in \mathbb{H}$ and $u(z) = 0$ otherwise. Find a formula for the extension of u to a harmonic function on the disk, and draw the locus where $u(z) = 1/2$. (The extension should be continuous on $\overline{\Delta}$ apart from finitely many points on S^1.)

52. Prove that for any $p \geq 1$, the set of analytic functions

$$\mathcal{F} = \left\{ f : \Delta \to \mathbb{C} \ : \ \int |f(z)|^p |dz|^2 \leq 1 \right\}$$

   is compact. That is, every sequence $f_n \in \mathcal{F}$ has a subsequence converging uniformly on compact subsets of Δ, and $f = \lim f_n$ belongs to $\mathcal{F}$ as well.

53. Let $f(z) = \sum a_n z^n$ be the power series for a bounded analytic function the unit disk. Prove that $\sum |a_n|^2 < \infty$.

54. Let $f(z) = \sum a_n z^n$ be a power series with coefficients $a_n \in \mathbb{Z}$. Suppose that $f(z)$ defines a bounded analytic function on the unit disk. Prove that $f(z)$ is a polynomial.

55. Let $U = \{z \ : \ 0 < |z| < 1 \text{ and } 0 < \arg(z) < \alpha\}$, where $0 < \alpha < 2\pi$. Find a formula for an analytic homeomorphism $f : U \to \mathbb{H}$.

56. Compute

$$\int_{|z|=1/2} \frac{dz}{z \sin^2(z)}.$$

57. Let $f_n : U \to \mathbb{C}$ be a sequence of analytic functions converging uniformly on compact sets to a nonconstant function $f : U \to \mathbb{C}$. Assume the mappings $f_n(z)$ are at most d-to-1. Show the same is true of $f(z)$.

58. Define an analytic function on the unit disk by $f(z) = \sum z^{n!}$. Prove that $f(z)$ has a singularity at every point of S^1.

59. Harmonic functions and analytic functions on $\mathbb{C}$ are each defined by the vanishing of a linear differential equation, $Pf = 0$, $P = \Delta$ or $d/d\bar{z}$. But the analytic functions form an algebra, while the harmonic functions do not. Explain this difference in terms of properties of P.

60. On $\mathbb{C}$ the function $u = \log|z|$ satisfies $\Delta u = C\delta_0$ for some constant C. Compute the value of C.

2 The simply-connected Riemann surfaces

In this section we discuss $\mathbb{C}$, $\widehat{\mathbb{C}}$ and $\mathbb{H} \cong \Delta$ from a geometric perspective.

Riemann surfaces. A *Riemann surface X* is a *connected* complex 1-manifold. This means X is a Hausdorff topological space equipped with a family of *charts* which are homeomorphisms

$$\phi_i : V_i \to \mathbb{C},$$

with $V_i \subset \mathbb{C}$ an open region, such that $\bigcup \phi_i(V_i) = X$ and the transition functions $\phi_i^{-1} \circ \phi_j$ are analytic where defined.

It then makes sense to discuss analytic functions on X, or on any open subset of X: we say f is analytic if $f \circ \phi_i$ is analytic for all i.

We then obtain a *sheaf of rings $\mathcal{O}_X$* with $\mathcal{O}_X(U)$ consisting of the analytic maps $f : U \to \mathbb{C}$. From a more modern perspective, a Riemann surface is a connected Hausdorff ringed space $(X, \mathcal{O}_X)$ that is locally isomorphic to $(\Delta, \mathcal{O}_\Delta)$.

Aside from $\mathbb{C}$ and *connected* open sets $U \subset \mathbb{C}$, the first interesting Riemann surface (and the basic example of a *compact* Riemann surface) is the Riemann sphere $\widehat{\mathbb{C}}$. The map $f : \widehat{\mathbb{C}} - \{0\} \to \mathbb{C}$ given by $f(z) = 1/z$ provides a chart near infinity. The second class of examples come from the *complex tori* $X = \mathbb{C}/\Lambda$, where $\Lambda \cong \mathbb{Z}^2$ is a discrete subgroup of $(\mathbb{C}, +)$. Charts are provided by the covering map $\mathbb{C} \to X$.

Maps. The space of Riemann surfaces forms a *category* with the analytic maps $f : X \to Y$ as morphisms. A map is analytic if it gives a locally analytic map on $\mathbb{C}$ in all charts; equivalent, f is analytic if f is continuous and $f^*(\mathcal{O}_Y)$ is a subsheaf of $\mathcal{O}_X$.

Meromorphic functions. A *meromorphic function* on X is an analytic map $f : X \to \widehat{\mathbb{C}}$ that is *not* identically ∞. Equivalently, $f(z)$ is meromorphic if it can be locally written as a quotient of two holomorphic functions, not both zero. The sheaf of meromorphic functions is denoted $\mathcal{M}$. Its construction from $\mathcal{O}$ is one of the most basic examples of *sheafification*.

Since X is *connected*, the ring $\mathcal{M}(X)$ (also denoted $\mathbb{C}(X)$) of all meromorphic functions on X forms a *field*, and the ring $\mathcal{O}(X)$ of all analytic functions forms an *integral domain*.

We will see that $\mathcal{M}(\widehat{\mathbb{C}}) = \mathbb{C}(z)$, while $\mathcal{O}(\widehat{\mathbb{C}}) = \mathbb{C}$, so not every meromorphic function is a quotient of holomorphic functions. We will see that $\mathcal{O}(\mathbb{C})$ is much wilder (it contains $\exp(\exp z)$, etc.), and yet $\mathcal{M}(\mathbb{C})$ *is* the field of fractions of $\mathcal{O}(\mathbb{C})$.

Proper maps. A continuous function $f : X \to Y$ between a pair of locally compact metric spaces is *proper* if $f^{-1}(K)$ is compact whenever K is compact. This is equivalent to the condition that $f(z_n) \to \infty$ whenever $z_n \to \infty$. A proper map has several good properties:

1. A proper map sends closed sets to closed sets.

2. If f is a local homeomorphism, then it is a covering map.

Proper analytic maps $f : X \to Y$ are the 'tamest' maps between Riemann surfaces. For example, they have the following properties:

1. If f is not constant, it is surjective. (Proof: $f(X)$ is both open and closed in the connected space Y.)

2. If $f'(x)$ never vanishes, then f is a covering map.

3. If $f'(x)$ never vanishes and Y is simply-connected, then f is an isomorphism.

Note that if X is *compact*, then *every* holomorphic map $f : X \to Y$ is proper, and hence every nonconstant $f : X \to Y$ is surjective.

Degree. For a proper holomorphic map $f : X \to Y$, the number of points in $f^{-1}(z)$, counted with multiplicity, is independent of z. This number is called

the (topological) *degree* of f. In the case $X = Y = \mathbb{C}$, it is the same as the degree of the polynomial f.

Example. A polynomial $f : \mathbb{C} \to \mathbb{C}$ is a proper map, with its topological degree equal to its algebraic degree. The entire function $\exp : \mathbb{C} \to \mathbb{C}$ is not proper, since it omits the point 0.

Classification. In principle, the classification of Riemann surfaces is completed by the following result:

Theorem 2.1 (The Uniformization Theorem) *Every simply-connected Riemann surface is isomorphic to* $\mathbb{H}$, $\mathbb{C}$ *or* $\widehat{\mathbb{C}}$.

Then by the theory of covering spaces, an arbitrary Riemann surface satisfies $X \cong \widehat{\mathbb{C}}$, $X = \mathbb{C}/\Gamma$, or $X = \mathbb{H}/\Gamma$, where Γ is a group of automorphisms. Such Riemann surfaces are respectively *elliptic*, *parabolic* and *hyperbolic*. Their natural metrics have curvatures $1, 0$ and -1.

In this section we will discuss each of the simply-connected Riemann surfaces in turn. We will discuss their geometry, their automorphisms, and their proper endomorphisms.

2.1 The complex plane

To say $f : \mathbb{C} \to \mathbb{C}$ is proper is the same as to say that f extends to a continuous function $F : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$.

Theorem 2.2 *An analytic function* $f : \mathbb{C} \to \mathbb{C}$ *is proper iff* $f(z)$ *is a non-constant polynomial.*

Proof 1. Clearly a polynomial of positive degree is proper. Conversely, suppose f is proper; then the preimage of the compact set $\overline{\Delta}$ is contained in $B(0, R)$ for some R. Thus $|f(z)| > 1$ if $|z| > R$.

Now consider the function on Δ^* defined by $g(z) = 1/f(1/z)$. Then $|g(z)| \leq 1$ provided $|z| < 1/R$. By the removable singularities theorem, $g(z)$ has a zero of finite order n at $z = 0$. Thus for small z, we have $|g(z)| \geq \epsilon|z|^n$, and thus for large z, we have $|f(z)| \leq M|z|^n$. Hence by Cauchy's bound, f is a polynomial. ∎

Proof 2. Write $f(z) = \sum a_n z^n$ as a power series, and observe that if $f(z)$ has an essential singularity at infinity, then (by Weierstrass–Casorati) the image under f of any neighborhood of $z = \infty$ must be dense in $\mathbb{C}$, contradicting properness. ∎

Corollary 2.3 *The automorphisms of $\mathbb{C}$ are given by the affine maps of the form $f(z) = az + b$, where $a \in \mathbb{C}^*$ and $b \in \mathbb{C}$.*

Thus $\mathrm{Aut}(\mathbb{C})$ is a *solvable group*. If $f \in \mathrm{Aut}(\mathbb{C})$ is fixed-point free, then it must be a translation. This shows:

Corollary 2.4 *Any Riemann surface covered by $\mathbb{C}$ has the form $X = \mathbb{C}/\Lambda$, where Λ is a discrete subgroup of $(\mathbb{C}, +)$.*

Example. We have $\mathbb{C}/(z \mapsto z + 1) \cong \mathbb{C}^*$; the isomorphism is given by $\pi(z) = \exp(2\pi i z)$.

Removability on Riemann surfaces. The first 2 proofs show, more generally:

Theorem 2.5 *Let $f : X \to Y$ be a continuous map between Riemann surfaces. If f is analytic outside a discrete set $E \subset X$, then f is analytic on X.*

Metrics. A *conformal metric* on a Riemann surface is given in local coordinates by $\rho = \rho(z) |dz|$. We will generally assume that $\rho(z) \geq 0$ and ρ is continuous, although metrics with less regularity are also useful.

A conformal metric allows one to measure lengths of arcs, by

$$L(\gamma, \rho) = \int_\gamma \rho = \int_a^b \rho(\gamma(t)) \, |\gamma'(t)| \, dt;$$

and areas of regions, by

$$A(U, \rho) = \int_U \rho^2 = \int_U \rho^2(z) \, |dz|^2 = \int_U \rho^2(z) \, dx \, dy.$$

Metrics pull back under analytic maps by the formula:

$$f^*(\rho) = \rho(f(z)) \, |f'(z)| \, |dz|,$$

and satisfy natural formulas such as

$$L(\gamma, f^*\rho) = L(f(\gamma), \rho)$$

(so long as $f(\gamma)$ is understood as the parameterized path $\gamma \circ f$).

A metric allows us to intrinsically measure the size of the derivative a map $f : (X_1, \rho_1) \to (X_2, \rho_2)$: it is given by

$$\|Df\| = \frac{|f'(z)|\rho_2(f(z))}{\rho_1(z)} = \frac{f^*\rho_2}{\rho_1}.$$

We have $\|Df\| \equiv 1$ iff f is a local isometry.

Flat metrics. The Euclidean metric on the plane is given by $\rho = |dz|$. Since $|dz|$ is Λ-invariant, we find:

Theorem 2.6 *Every Riemann surface covered by $\mathbb{C}$ admits a complete flat metric, unique up to scale.*

For example, on $\mathbb{C}^* \cong \mathbb{C}/\mathbb{Z}$ we have the *cylindrical metric* $\rho = |dz|/|z|$. Every circle $|z| = r$ has length 2π in this metric.

Orbifolds. An orbifold is a generalization of a manifold. Instead of being locally modeled on $\mathbb{R}^n$, it is modeled on $B^n(1)/G$ where $G \subset \mathrm{O}(n)$ is a finite group of isometries fixing the origin, and $B^n(1) \subset \mathbb{R}^n$ is the unit ball.

If (M^n, g) is a Riemannian manifold, and $G \subset \mathrm{Isom}(M)$ is a discrete group of isometries, then for all $p \in M$ the stabilizer G^p is a finite subgroup of $\mathrm{O}(T_pM)$, and hence $X = M/G$ has the structure of an orbifolds.

In this case the *metric g* on M also descends to give a metric on X, generally with singularities coming from the *orbifold locus*, namely the image of the subset of M where G^p is nontrivial.

In the category of orbifolds, the map $\pi : M \to X$ is considered a *covering map*; and if M is simply–connected, it is the *universal cover* of the orbifold X.

Orbifold surfaces. When M is a *surface* and G preserves *orientation*, the picture is rather simple: $X = M/G$ is topologically a surface, the orbifold locus $X_0 \subset X$ is discrete, and the orbifold structure is determined by a map $m : X_0 \to \{2, 3, 4, \ldots\}$ defined so that at $p \in X_0$, X is locally a quotient of a disk by a rotation of order n.

For example, for the quotient of S^2 by a rotation of order n – -which has 2 fixed points – the underlying topological space is $X = S^2$, but two points

are marked as orbifold points of order n. If we let the usual metric on the sphere descend to X, then (X, ρ) has 2 cone points with apex angles $2\pi/n$.

Warning. It is tempting to think that, at least for surfaces, every orbifold is the quotient of a manifold. However this is not true! The sphere with 2 orbifold points of orders p and q, with $\gcd(p, q) = 1$, has no manifold covering space.

Orbifold metrics. Let us now compute some of the metrics coming from orbifold quotients of $\mathbb{C}$. A prime example is provided by the group $\Gamma \cong \mathbb{Z}/n \subset \operatorname{Aut}(\mathbb{C})$ generated by $g(z) = \zeta_n z$, where $\zeta_n = \exp(2\pi i/n)$.

Theorem 2.7 *The quotient of $(\mathbb{C}, |dz|)$ by the action of $g(z) = \zeta_n$ is the flat orbifold $X = (\mathbb{C}, \rho)$, with the singular flat metric*

$$\rho = \frac{|z|^{1/n}|dz|}{n|dz|}.$$

The quotient map is given by $\pi(z) = z^n$.

Proof. Just check that pullback of ρ under the quotient map $\pi(z) = z^n$ is $|dz|$. ∎

Cone metrics. The orbifold above can be obtained, metrically, from the region $0 \le \arg(z) \le 2\pi/n$ by identifying opposite edges; thus it is metrically a cone with apex angle $2\pi/n$.

To see this directly, note that the region $|z| < r$ is, in the ρ metric, a ball of radius $r^{1/n}$ and circumference $2\pi r^{1/n}/n$. In fact, on the positive real axis the metric above is

$$(1/n)x^{1/n-1}dx = d(x^{1/n}).$$

(This is a good way to remember it).

More generally, the metric $\rho = |z|^\alpha |dz|/|z|$ makes $\mathbb{C}$ into a cone with total angle $2\pi\alpha$ at $z = 0$.

The quotient of $\mathbb{C}^*$ by $\mathbb{Z}/2$. Another important example comes from $\mathbb{Z}/2$ acting on a cylinder.

Theorem 2.8 *The quotient of $(\mathbb{C}^*, |dz|/|z|)$ by the action of $g(z) = 1/z$ is the orbifold $(\mathbb{C}, \rho)$, with orbifold points of order 2 at ± 2 and quotient metric:*

$$\rho = \frac{|dz|}{|4 - z^2|^{1/2}}.$$

The quotient map is given by $\pi(z) = z + 1/z$.

The map π sends circles to ellipse and radial lines to hyperboli, with foci $[-2, 2]$. In fact all conics with these foci arise in this way. (See Figure 1.1.)

Quotient by the dihedral group D_∞. We can also take the quotient of $\mathbb{C}$ by the infinite dihedral group, $D_\infty = \langle z + 2\pi, -z \rangle \subset \operatorname{Aut}\mathbb{C}$.

Theorem 2.9 *The quotient $(\mathbb{C}, |dz|)/D_\infty$ is the flat orbifold $(\mathbb{C}, \rho)$, with orbifold points of order 2 at $z = \pm 1$. The quotient metric is given by*

$$\rho = \frac{|dz|}{|1 - z^2|^{1/2}},$$

and the quotient map is $\pi(z) = \cos(z)$.

To check this, note we have:

$$\pi^*(\rho) = |\sin(z)|/|1 - \cos^2(z)|^{1/2} = 1.$$

Note also that the length of $[-1, 1]$ in the ρ-metric is given by $\int_{-1}^{1}(1 - x^2)^{-1/2}\,dx = \pi$.

Universal cover of $[-1, 1]$. The group D_∞ restricts to an action on $\mathbb{R}$, and gives $X = [-1, 1]$ as a quotient, where the endpoints are orbifold points. The map $\cos : \mathbb{R} \to [-1, 1]$ gives the universal cover of this one–dimensional orbifold.

Factorization through the cylinder. We note that for $C(z) = (z + 1/z)/2$, we can write

$$\cos(z) = C(\exp(iz)).$$

This illustrates the factorization of the quotient of $\mathbb{C}$ by D_∞ through the cylinder $\mathbb{C}^*$, which is the intermediate covering space $\mathbb{C}/\mathbb{Z}$,

Other examples. In general, by broadening the scope of covering spaces and deck groups to include orbifolds and maps with fixed points, we enrich the supply of Euclidean (and other) Riemann surfaces. For example, the $(3, 3, 3)$ orbifold is also Euclidean — it is the double of an equilateral triangle.

Chebyshev polynomials. Let us study the degree two map $C : \mathbb{C}^* \to \mathbb{C}$ in more detail.

Since $C(z^n) = C(z^{-n})$, there is a polynomial T_n such that

$$T_n(C(z)) = C(z^n) = C_n(z).$$

These are the *Chebyshev polynomials*; they are given by $T_1(z) = z$, $T_2(z) = 2z^2 - 1$, $T_3(z) = 4z^3 - 3z$, $P_4(z) = 8z^4 - 8z^2 + 1$, etc.

Cosine and trace. The Chebyshev polynomials give the multiple angle formulas for cosine: since for $z = e^{i\theta}$ we have $C(z) = \cos\theta$, it follows that

$$T_n(\cos\theta) = \cos(n\theta).$$

By the same token, we have

$$T_n(\operatorname{tr} A/2) = \operatorname{tr} A^n/2$$
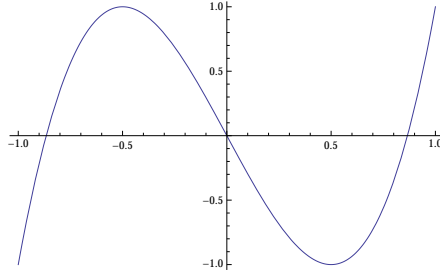
for all $A \in \mathrm{SL}_2(\mathbb{C})$.

Figure 2.1. The cubic Chebyshev polynomial $T_3(x)$.

Since z^n preserves circles and radial lines, and its Julia set is S^1, we find:

Theorem 2.10 *The polynomials $T_n(z)$ preserve the hyperbolas and ellipses with foci at ± 1, and its Julia set is $[-1, 1]$.*

The maps z^n and $T_n(z)$ are the only polynomials with smooth Julia sets.

Uniformizing the circle. It is natural also to introduce the rational function on $\mathbb{C}^*$ defined by

$$S(z) = \frac{z - z^{-1}}{2i},$$

so that $(C(z), S(z)) = (\cos\theta, \sin\theta)$ when $z = \exp(i\theta)$. Note that $S(1/z) = -S(z)$. Since $\sin^2(\theta) = \cos^2(\theta) = 1$, we have:

$$S(z)^2 + C(z)^2 = 1.$$

One readily finds:

Theorem 2.11 *The map $f(z) = (S(z), C(z))$ gives an isomorphism between $\mathbb{C}^*$ and the curve $X \subset \mathbb{C}^2$ defined by $x^2 + y^2 = 1$.*

Note that projection of X to the y–axis is given by the function C, which is clearly branched over the points $y = \pm 1$. Note also that in $\mathbb{P}^2$, the conic X is isomorphic to $\mathbb{P}^1$; it meets the line at infinity in $\pm i$, and thus its affine part is isomorphic to $\mathbb{C}^*$.

Transcendental functions. Passing to the universal cover, we conclude that the universal covering map

$$\pi : \mathbb{C} \to X$$

can be given by $\pi(z) = (\sin(z), \cos(z))$. This is the most basic example of the principle that *algebraic curves* can be uniformized by *transcendental functions*. Indeed, X is a degenerate example of an elliptic curve, which as we will later see is uniformized by $\pi(z) = (\wp(z), \wp'(z))$.

Multiple angle formulas for $\sin\theta$. It is straightforward to see that we can find a second sequence of polynomials $U_n(C)$ satisfying

$$U_n(C(z))S(z) = S(z^n).$$

For example, $U_2(z) = 2z$ and $U_3(z) = 4z^2 - 1$. (It is more common to index $U_n(z)$ by its degree, which is $n - 1$.) Since $\cos^2(n\theta) + \sin^2(n\theta) = 1$, we have:

Theorem 2.12 *For all $n \geq 1$ and all $C \in \mathbb{C}$, we have*

$$T_n(C)^2 - (C^2 - 1)U_n(C) = 1.$$

Units and Pell's equation. To explain the preceding equation more conceptually, let us consider the quadratic extension of rings B/A, where

$$A = \mathbb{C}[C] \quad \text{and} \quad B = \mathbb{C}[C, S]/(C^2 + S^2 = 1).$$

Note that $A \cong \mathcal{O}^{\mathrm{alg}}(\mathbb{C})$ and

$$B \cong \mathcal{O}^{\mathrm{alg}}(\mathbb{C}^*) \cong \mathbb{C}[z, 1/z],$$

where $z = C + iS$. The main difference in these rings is their units: we have $A^* = \mathbb{C}^*$ while

$$B^* = \mathbb{C}^* z^{\mathbb{Z}}.$$

We can regard B as the extension $A[\sqrt{D}]$, where $D = C^2 - 1$; then the preceding theorem shows that (T_n, U_n) are solutions to *Pell's equation*,

$$p^2 - Dq^2 = 1,$$

which gives the norm from B to A and describes the units in B of the form $\epsilon = p + q\sqrt{D}$. In our case, $\sqrt{D} = iS$, and

$$\epsilon^n = z^n = T_n(C) + iU_n(C)S.$$

(A similar discussion can be pursued for all hyperelliptic curves, using an algorithm due to Abel.)

The classical solutions to Pell's equation arise when $D \in \mathbb{Z}$ and describe the units of norm 1 in $\mathbb{Z}[\sqrt{D}]$, which form a cyclic group. For example, when $D = 3$ this group is generated by $\epsilon = 2 + \sqrt{3}$. These solutions can be very large, even when D has only moderate size, and are responsible for the enormous numbers arising in Archimedes' cattle problem.

Solving the cubic. Complex algebra finds its origins in the work of Cardano et al on solving cubic polynomial equations. Remarkably, complex numbers intervene even when the root to be found is real.

One can always make a simple transformation of the form $x \mapsto x + c$ to reduce to the form

$$x^3 + ax + b = 0.$$

One can further replace x with cx to reduce to the form

$$x^3 - 3x = b.$$

Thus the solution to the cubic involves inverting a *single* cubic function $P_3(z) = z^3 - 3z$.

Now observe that $P_3(z + 1/z) = z^3 + 1/z^3$; i.e., $P_3(z)$ is a rescaling of the Chebyshev polynomial $T_3(z)$. Thus, to solve $P_3(x) = a$, we just write $a = y + 1/y$ (by solving a quadratic equation), and then we have $x = y^{1/3} + y^{-1/3}$.

Classification of polynomials. Let us say $p(z)$ is equivalent to $q(z)$ is there are $A, B \in \mathrm{Aut}(\mathbb{C})$ such that $Bp(Az) = q(z)$. Then every polynomial is equivalent to one which is monic and centered (the sum of its roots is zero). Every quadratic polynomial is equivalent to $p(z) = z^2$.

The reasoning above shows, every cubic polynomial with distinct critical points is equivalent to $P_3(z) = z^3 - 3z$. Otherwise it is equivalent to z^3.

But for degree 4 polynomials we are in new territory: the cross-ratio of the 3 critical points, together with infinity, is an invariant.

It is a famous fact (proved using Galois theory) that a general quintic polynomial (with integral coefficients) cannot be solved by radicals.

2.2   The Riemann sphere

We now turn to the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. We will examine the sphere from several perspectives: as a Riemann surface, as a round sphere, as the projectively line, and as the boundary of hyperbolic space.

The projective line. We have a natural projection

$$\pi : \mathbb{C}^2 - \{(0,0)\} \to \widehat{\mathbb{C}} \cong \mathbb{P}^1$$

given by $\pi(z_0, z_1) = z_1/z_0$; it records the slope of each line. This gives a natural identification of $\widehat{\mathbb{C}}$ with the *projective line* $\mathbb{P}^1$, i.e. the space of lines in $\mathbb{C}^2$.

Aside: Some topology of projective spaces. The real projective plane $\mathbb{RP}^2$ is the union of a disk and a Möbius band. The natural map $p : \mathbb{C}^2 - \{(0,0)\} \to \widehat{\mathbb{C}}$ factors through the Hopf map $S^3 \to S^2$, whose fibers have linking number one. This map generates $\pi_3(S^2)$.

The topological degree of $f(z) = P(z)/Q(z)$, assuming $\gcd(P, Q) = 1$ in $\mathbb{C}[z]$, is $\max(\deg P, \deg Q)$.

$$f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$$

by $f(z) = P(z)/Q(z)$, as can be verified by looking in charts near ∞.

Möbius transformations. So long as $ad - bc \neq 0$, the map $f(z) = (az + b)/(cz + d)$ defines an automorphism of $\widehat{\mathbb{C}}$. Its inverse can be found by inverting the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. In fact, the map π above transports the linear action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{C}^2$ to the fractional linear action of $\mathrm{PSL}_2(\mathbb{C})$ on $\widehat{\mathbb{C}}$.

Theorem 2.13 *We have* $\mathrm{Aut}(\widehat{\mathbb{C}}) = \mathrm{PSL}_2(\mathbb{C})$.

Proof. Given $f \in \mathrm{Aut}(\widehat{\mathbb{C}})$, pick $g \in \mathrm{PSL}_2(\widehat{\mathbb{C}})$ so $g(f(\infty)) = \infty$. Then $g \circ f \in \mathrm{Aut}(\mathbb{C}) \subset \mathrm{PSL}_2(\mathbb{C})$, so $f \in \mathrm{PSL}_2(\mathbb{C})$. ∎

Corollary 2.14 *The action of* $\mathrm{Aut}(\widehat{\mathbb{C}})$ *is uniquely triply-transitive: any distinct triple of points can be sent to* $(0, 1, \infty)$ *by a unique Möbius transformation.*

Corollary 2.15 *The group* $\mathrm{PSL}_2(\mathbb{C})$ *is isomorphic, as a complex manifold, to the space of distinct triples of points on* $\widehat{\mathbb{C}}$.

Rational maps. Note that every analytic map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is *proper*, because $\widehat{\mathbb{C}}$ is compact. We can also describe these maps very simply.

Theorem 2.16 *Every analytic map* $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ *is given by a rational map* $f(z) = P(z)/Q(z)$.

Proof. The result is clear if f is constant. Otherwise, f has finitely many poles $p_1, \ldots, p_n$ in $\mathbb{C}$. Let $g_i(z) = a_i/(z - p_i)^{b_i} + \cdots$ be the Laurent tail of f at p_i. Then $f - \sum g_i(z)$ has no poles, hence it is equal to a polynomial (if $f(\infty) = \infty$) or a constant (if $f(\infty)$ is finite)). It follows that $f(z)$ is a rational function. $\blacksquare$

Classification of automorphisms up to conjugacy. There is a natural map $\mathrm{tr} : \mathrm{PSL}_2(\mathbb{C}) \to \mathbb{C}/\langle \pm 1 \rangle$. This is clearly a class function (it is constant on conjugacy classes), and in fact the conjugation class is almost determined by this map. Namely we have the following classes:

(1) The identity map, $\mathrm{tr}(A) = \pm 2$.
(2) Parabolics, $A(z) = z + a$; $\mathrm{tr}(A) = \pm 2$.
(3) Elliptics: $A(z) = e^{i\theta} z$; $\mathrm{tr}(A) = 2 \cos \theta/2 \in [-2, 2]$.
(4) Hyperbolics: $A(z) = e^t z$, $\mathrm{Re}\, t \neq 0$; $\mathrm{tr}(A) = 2 \cosh(t/2)$.

Note that a Möbius transformation (other than the identity) either has two simple fixed points, or a single fixed point of multiplicity two. The fixed points of A correspond to its eigenvectors on $\mathbb{C}^2$.

Note also that all these elements, except for irrational elliptics, generate discrete subgroups of $\mathrm{Aut}(\widehat{\mathbb{C}})$.

If A has eigenvalues $\lambda^{\pm 1}$, then $\mathrm{tr}(A) = \lambda + 1/\lambda$. Our previous analysis of this map shows, for example, that the traces of matrices a given value for $|\lambda|$ correspond to an ellipse with foci ± 2.

Cross-ratios. The *cross-ratio* is an invariant of ordered 4-tuples of points, characterized by the conditions that $[a : b : c : d] = [ga : gb : gc : gd]$

for all $g \in \mathrm{Aut}(\widehat{\mathbb{C}})$, and $[0 : 1 : \infty : \lambda] = \lambda \in \widehat{\mathbb{C}} - \{0, 1, \infty\}$. The cross-ratio gives an explicit isomorphism between the *moduli space* $\mathcal{M}_{0,4}$ and the triply-punctured sphere.

Stereographic projection. There is a geometric identification between the unit sphere $S^2 \subset \mathbb{R}^3$ and the Riemann sphere $\widehat{\mathbb{C}}$ in which the north pole N becomes ∞ and the rest of the sphere is projected linearly to $\mathbb{C} = \mathbb{R}^2 \times \{0\}$.

Theorem 2.17 *Stereographic projection is a conformal map that sends circles to circles.*

Proof of conformality. Consider two vectors at a point $p \neq N$ on S^2. Construct a pair of circles tangent to these vectors at p and passing through N. Then these circles meet in the same angles at p and N. On the other hand, each circle is the intersection of the sphere with a plane. These planes meet $\mathbb{C}$ in the same angle they meet a plane tangent to the sphere at the north pole N. Thus stereographic projection preserves angles. ∎
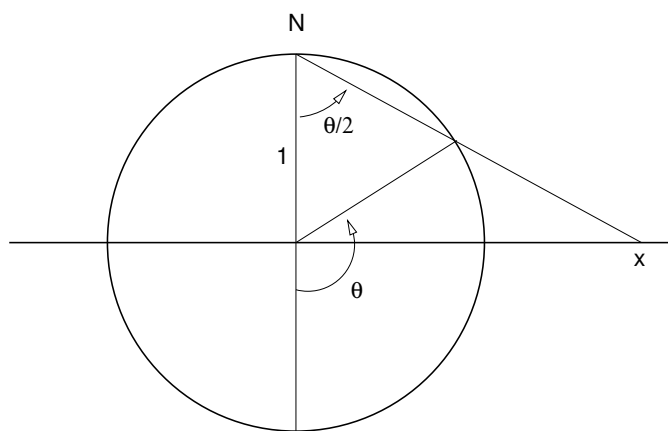

Figure 2.2. Stereographic projection on S^1.

Theorem 2.18 *The Euclidean metric on S^2 is transported, by stereographic projection, to the metric $2|dz|/(1 + |z|^2)$ on $\widehat{\mathbb{C}}$.*

Proof. Reducing dimensions by one, we obtain the stereographic projection $p : S^1 - N \to \mathbb{R}$. If $\theta \in [-\pi, \pi]$ is the angle on the sphere normalized so

65

$\theta(N) = \pi$, then we find $x = p(\theta) = \tan(\theta/2)$. Thus $dx = (1/2)\sec^2(\theta/2)\, d\theta = (1/2)(1 + x^2)\, d\theta$, which gives $d\theta = 2dx/(1 + x^2)$. The case of S^2 follows by conformality and rotation invariance. ■

Unit speed. As an example, we find that

$$z = \tan(t/2)$$

moves at unit speed in the spherical metric, because $1+\tan^2(t/2) = \sec^2(t/2) = 2(d/dt)\tan(t/2)$. This gives a map $z : [-\pi/2, \pi/2] \to \widehat{\mathbb{R}}$.

The unitary point of view. Here is an alternative perspective on the spherical metric. Let $\langle z, w\rangle = z_0\overline{w}_0 + z_1\overline{w}_1$ be the usual Hermitian metric on $\mathbb{C}^2$. Define the angle θ between two lines $\mathbb{C}z, \mathbb{C}w \subset \mathbb{C}^2$ by

$$\cos(\theta/2) = \frac{|\langle z, w\rangle|}{|z|\,|w|}. \tag{2.1}$$

This θ agrees with the spherical distance on $\widehat{\mathbb{C}}$. (To check this, consider the points 1 and $\exp(i\alpha)$ in S^1. Their spherical distance is α, and if we set $z = [1 : 1]$ and $w = [e^{-i\alpha/2} : e^{i\alpha/2}]$, then $|z||w| = 2$ and $\langle z, w\rangle = 2\cos(\alpha/2)$.)

Spherical distance. What is the distance, in the spherical metric, from 1 to $1 + i$? For $[1 : 1]$ and $[1 : 1 + i]$ we get

$$\cos\theta/2 = \sqrt{5}/\sqrt{6},$$

and hence $\theta = 2\arccos(\sqrt{5/6}) \approx 0.841069\ldots$.

Given this agreement, it is now clear that the group

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} : |a|^2 + |b|^2 = 1 \right\}$$

(consisting of matrices satisfying $AA^* = I$) acts *isometrically* on $\widehat{\mathbb{C}}$. In fact this group is the full group of orientation-preserving isometries.

Theorem 2.19 *The isometry group of $\widehat{\mathbb{C}}$ in the spherical metric is given by*

$$\mathrm{Isom}^+(S^2) = \mathrm{SU}(2)/(\pm I) \subset \mathrm{PSL}_2(\mathbb{C}).$$

Circles and lines. A *circle* $C \subset \widehat{\mathbb{C}}$ is either a line through ∞, or an ordinary Euclidean circle in $\mathbb{C}$.

Theorem 2.20 *Möbius transformations send circles to circles, and any two circles are equivalent.*

Proof 1. A circle $x^2 + y^2 + Ax + By + C = 0$ is also given by $r^2 + r(A\cos\theta + B\sin\theta) + C = 0$, and it is easy to transform the latter under $z \mapsto 1/z$, which replaces r by $1/r$ and θ by $-\theta$.

Proof 2. A circle is the same, projectively, as the set of null vectors for a Hermitian form on $\mathbb{C}^2$ of signature $(1,1)$. Any two such forms are, up to scale, related by an element of $\mathrm{SL}_2(\mathbb{C})$. ∎

Subgroups of SU(2). The stabilizer of an oriented circle gives a copy of $S^1 \subset \mathrm{SU}(2)$. For example, the stabilizer of the unit circle just corresponds to the diagonal matrices. On the other hand, the stabilizer of $\widehat{\mathbb{R}}$ corresponds to the real unitary matrices:

$$K = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}.$$

These matrices fix $\pm i$. For example, the transformation that cyclically permutes $(\infty, 1, 0, -1)$ is given by $g(z) = (z-1)/(z+1)$.

Area of triangles. The geodesics on S^2 are arcs of great circles. The angle sum of a spherical triangle always exceeds π, and in fact we have:

Theorem 2.21 *The area of a spherical triangle is equal to its excess angle, $\alpha + \beta + \gamma - \pi$.*

Proof. Since the sphere has area 4π, a lune of angle θ has area 4θ. The three lunes coming from a given triangle T cover the whole sphere, with points inside T and its antipode triply covered and the rest simply covered. Thus $4\pi + 4\,\mathrm{area}(T) = 4(\alpha + \beta + \gamma)$. ∎

This result is a special case of the Gauss-Bonnet formula:

$$2\pi\chi(X) = \int_X K + \int_{\partial X} k.$$

In the case of a polygon P, the boundary integral becomes a sum of external angles, and in the spherical case we get $\mathrm{area}(P) = \delta(P)$, where $\delta(P) > 0$ is the angle defect (equal to zero in the Euclidean case).

The general formula can be deduced from the constant curvature formulas (for $K = \pm 1$) by subdivision.

Hyperbolic 3-space. Finally we note that one can identify $\widehat{\mathbb{C}} = G/AN$ with the boundary of hyperbolic 3–space $\mathbb{H}^3 = G/K$, where $G = \mathrm{SL}_2(\mathbb{C})$ and $K = \mathrm{SU}(2)$. From this perspective, $\mathbb{H}^3$ is the space of round conformal metrics on S^2; these can be identified with certain probability measures, which in turns are compactified by δ–functions.

Finite quotients of $\widehat{\mathbb{C}}$. The discrete subgroups of $\mathrm{Aut}(\widehat{\mathbb{C}})$ are isomorphic to $\mathbb{Z}/n$, D_{2n} and A_4, S_4 and A_5, which quotient orbifolds of type (n, n), $(2, 2, n)$, $(2, 3, 3)$, $(2, 3, 4)$ and $(2, 3, 5)$ respectively. These (a_i) come from solutions to the equation $\sum_1^k 1/a_i > k - 2$.

2.3 The hyperbolic plane

We now turn to a discussion of the hyperbolic plane $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. This is the most versatile of the simply-connected Riemann surfaces. Note that $\mathbb{H}$ is isomorphic to Δ, e.g. by the Möbius transformation $z \mapsto \frac{z-i}{z+i}$ which sends $(i, 0, \infty)$ to $(0, -1, 1)$. One often uses these two models interchangeably.

First, observe that the Schwarz reflection principle implies:

Theorem 2.22 *Every automorphism of $\mathbb{H}$ or Δ extends to an automorphism of $\widehat{\mathbb{C}}$.*

Corollary 2.23 $\mathrm{Aut}(\mathbb{H})$ *corresponds to the subgroup* $\mathrm{SL}_2(\mathbb{R}) \subset \mathrm{SL}_2(\mathbb{C})$.

Note: A Möbius transformation with real coefficients always preserves $\widehat{\mathbb{R}}$. The condition $\det(A) > 0$ is equivalent to A preserving $\mathbb{H}$, rather than swapping it with $-\mathbb{H}$.

Hyperbolic geometry. We define the *hyperbolic metric* on $\mathbb{H}$ is by $\rho = |dz|/\mathrm{Im}\, z = |dz|/y$. This conformal metric has constant curvature -1.

Theorem 2.24 *Every automorphism of $\mathbb{H}$ is an isometry for the hyperbolic metric. Thus* $\mathrm{Aut}(\mathbb{H}) \cong \mathrm{Isom}^+(\mathbb{H})$.

Proof. This is clear for automorphisms of the form $g(z) = az + b$, and can be easily checked by $g(z) = -1/z$. These two types of automorphisms generate $\mathrm{Aut}(\mathbb{H})$. ∎

The unit disk. On the unit disk Δ, the hyperbolic metric becomes $\rho = 2|dz|/(1 - |z|^2)$. Thus the path

$$z = \tanh(t/2)$$

moves at unit speed in the hyperbolic metric, because $dz/dt = (1/2)\operatorname{sech}^2(t/2)$ and $1 - \tanh^2(t/2) = \operatorname{sech}^2(t/2)$. This gives a map $z : (-\infty, \infty) \to (-1, 1)$ that is an isometry from the Euclidean metric to the hyperbolic metric.

The unitary perspective. Consider the Hermitian inner product of signature $(1, 1)$ on $\mathbb{C}^2$ given by

$$\langle z, w \rangle = z_0 \overline{w}_0 - z_1 \overline{w}_1.$$

The symmetry group of this form is denote by $\mathrm{SU}(1,1) \subset \mathrm{SL}_2(\mathbb{C})$. It consists of matrices satisfying $AQA^* = Q$, where $Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Thus:

$$\mathrm{SU}(1,1) = \left\{ \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} : |a|^2 - |b|^2 = 1 \right\} \subset \mathrm{SL}_2(\mathbb{C}).$$

This group, unlike $\mathrm{SU}(2)$, is noncompact.

Note that for $Z = (1 : z)$, we have

$$\langle Z, Z \rangle = 1 - |z|^2.$$

A line $\mathbb{C} \cdot z$ is *isotropic* if $\langle z, z \rangle = 0$. For the form above, the set of isotropic lines gives the unit circle in $\mathbb{P}^1 \cong \widehat{\mathbb{C}}$. Similarly, the set of *positive lines* gives the unit disk Δ. It follows that every element of $\mathrm{SU}(1,1)$ lies in $\mathrm{Aut}(\Delta)$, and in fact we have:

Theorem 2.25 *The automorphism group of the unit disk is given by the subgroup*
$$\mathrm{Aut}(\Delta) = \mathrm{SU}(1,1)/\pm I \subset \mathrm{PSL}_2(\mathbb{C}).$$

Hyperbolic distance. The hyperbolic distance d between two lines $\mathbb{C}z, \mathbb{C}w \subset \mathbb{C}^2$ representing points in Δ is given, in analogy with the spherical formula (2.1), by

$$\cosh(d/2) = \frac{|\langle z, w \rangle|}{|z|\,|w|}. \tag{2.2}$$

69

Geodesics. The *geodesics* for the hyperbolic metric are given by circles orthogonal to the boundary (of $\mathbb{H}$ or Δ). For example, in the case of the imaginary axis $\gamma = i\mathbb{R}_+ \subset \mathbb{H}$, it is easy to see that the projection $\pi : \mathbb{H} \to \gamma$ given by $\pi(x, y) = (0, y)$ is distance-decreasing, and so γ gives a shortest path between any two of its points. The general case follows from this one, since any circle orthogonal to the boundary of $\mathbb{H}$ is equivalent, under $\mathrm{Aut}(\mathbb{H})$, to γ.

These geodesics satisfy all of Euclid's postulates except the fifth. Thus if we declare them to be straight lines, we find:

Theorem 2.26 *Euclid's fifth postulate cannot be deduced from the other axioms of geometry.*

Triangles. Gauss-Bonnet for hyperbolic triangles reads:

$$\mathrm{area}(T) = \pi - \alpha - \beta - \gamma.$$

For example, an *ideal triangle* in $\mathbb{H}$ has vertices at infinity and internal angles of 0. Its area is π.

Curvature. We remark that the Gauss curvature of a conformal metric $\rho(z)\,|dz|$ is readily computed in terms of the Laplacian of ρ: we have

$$K(\rho) = -\rho^{-2}\Delta \log \rho.$$

Note that this expression is invariant under change of coordinates, since an analytic function satisfies $\Delta \log |f'| = 0$. It is particularly easy to check that $\Delta(-\log y) = 1/y^2$, and hence $K(1/y) = -1$; that is, $(\mathbb{H}, |dz|/y)$ has constant curvature -1.

Classification of isometries. Let $f : \mathbb{H} \to \mathbb{H}$ be an automorphism. Then f is an isometry, and we can define its *translation distance* by

$$\tau(f) = \inf_{\mathbb{H}} d(x, f(x)).$$

This invariant is useful for the classification of isometries. Here are the possibilities:

1. (Elliptic) $\tau(f) = 0$, achieved. Then f is conjugate to a rotation fixing i, of the form
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

(Note that when $\theta = \pi$ above, the matrix is $-I$ and so $f(z) = z$.) In the disk model, we can simply conjugate so $f(z) = e^{2i\theta}z$.

2. (Hyperbolic) $\tau(f) > 0$. In this case $\tau(f)$ is achieved along a unique geodesic $\gamma \subset \mathbb{H}$ which is translated distance $\tau(f)$ by f. Up to conjugacy, $f(z) = e^t z$ where $t = \tau(f)$.

3. (Parabolic) $\tau(f) = 0$, not achieved. Then f is conjugate to $f(z) = z+1$.

The Schwarz Lemma. The following fundamental fact is central to the theory of analytic maps between hyperbolic Riemann surfaces.

Theorem 2.27 *Let $f : \Delta \to \Delta$ be an analytic function such that $f(0) = 0$. Then either (i) $f(z) = e^{i\theta}z$ is a rotation, or (ii) $|f'(0)| < 1$ and $|f(z)| < |z|$ for all z.*

Proof. Observe that by the maximum principle, the analytic function defined by $g(z) = f(z)/z$ when $z \neq 0$ and $g(0) = f'(0)$ satisfies $|g(z)| \leq 1$. If equality holds for some z, then g is constant. ∎

Corollary 2.28 *A holomorphic map $f : \Delta \to \Delta$ is either an isometry, or it is a contraction for the hyperbolic metric. (This means $|f'|_\rho < 1$ and $d(f(x), f(y)) < d(x, y)$ if $x \neq y$.)*

Remark. The Schwarz lemma, instead of Schwarz reflection, can also be used to show every automorphism of $\mathbb{H}$ or Δ extends to $\widehat{\mathbb{C}}$.

The hyperbolic metric on other Riemann surfaces. Now suppose, more generally, that $X = \mathbb{H}/\Gamma$ is a *hyperbolic Riemann surface*. (Almost all Riemann surfaces are of this type — only $\widehat{\mathbb{C}}, \mathbb{C}, \mathbb{C}^*$ and $\mathbb{C}/\Lambda$ not hyperbolic. In particular, every region $U \subset \mathbb{C}$ whose complement contains 2 points is hyperbolic.)

Since $\rho = |dz|/y$ is invariant under $\mathrm{Aut}(\mathbb{H})$, it *descends* to give a *hyperbolic metric* on X itself. (Also called the Poincaré metric.)

Example. The hyperbolic metric on the strip $S = \{z = x+iy : 0 < y < \pi\}$ is given by $\rho = |dz|/\sin(y)$. Indeed, the map $f(z) = e^z$ sends S conformally to $\mathbb{H}$, and so

$$\rho = f^*(|dz|/y) = \frac{|e^z|\,|dz|}{\mathrm{Im}\,e^z} = \frac{e^x\,|dz|}{e^x\sin(y)} = \frac{|dz|}{\sin(y)}.$$

71

The quotient of S by $z \mapsto z + R$ gives a Euclidean cylinder A with circumference $c = R$ and height $h = \pi$; its modulus $h/c = \pi/R$ is a conformal invariant. The unique geodesic in S invariant under this metric lies on the line $y = \pi/2$, where $\rho = |dz|$; thus it descends to a closed geodesic $\gamma_A \subset A$ of length R. This gives:

Theorem 2.29 *The hyperbolic length of the core geodesic of an annulus A satisfies $L(\gamma_A) = \pi/\operatorname{mod}(A)$.*

The modern Schwarz lemma. The most general version of the Schwarz Lemma (for Riemann surfaces) reads:

Theorem 2.30 *Let $f : X \to Y$ be an analytic map between hyperbolic Riemann surfaces. Then f is either a locally isometric covering map, or it is a contraction for the hyperbolic metric.*

Proof. Pass to the universal covers of the domain and range and apply the usual Schwarz lemma. ∎

Proper maps. A *Blaschke product* $B : \Delta \to \Delta$ is a rational map of the form
$$f(z) = e^{i\theta} \prod_{1}^{d} \frac{z - a_i}{1 - \overline{a}_i z}$$
with zeros satisfying $|a_i| < 1$. Using the fact that $1 = z\overline{z}$ on S^1, it is easy to show that $|B(z)| = 1$ on S^1, and thus $B : \Delta \to \Delta$ is proper. Conversely, we have:

Theorem 2.31 *Every proper analytic map $f : \Delta \to \Delta$ is a Blaschke product.*

Proof. A proper map is surjective, so f has at least one zero, say a. Let $M(z) = (z - a)/(1 - \overline{a}z)$. Then $g = f(z)/M(z) : \Delta \to \overline{\Delta}$ is analytic, with one fewer zero than f. The proof now follows by induction on the degree of f. ∎

Corollary 2.32 *Every proper map on the unit disk extends to a rational map on $\widehat{\mathbb{C}}$.*

Remark: Dynamics and the Schwarz Lemma. Here are two sample applications of the Schwarz lemma to iterated analytic maps.

Theorem 2.33 *Let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational map of degree $d > 1$. Then the immediate basin of any attracting cycle contains a critical point.*

Corollary 2.34 *A rational map f of degree d has at most $2d - 2$ attracting cycles.*

Theorem 2.35 (Denjoy-Wolff) *Let $f : \Delta \to \Delta$ be a holomorphic map which is not conjugate to $z \mapsto e^{i\theta}z$. Then there exists a point $p \in \overline{\Delta}$ such that $f^n(z) \to p$ for all $z \in \Delta$.*

2.4 Kähler potentials

We have seen that the spherical metric on $\widehat{\mathbb{C}}$ connects with $SU(2)$, and the hyperbolic metric on Δ connects with $SU(1,1)$. To conclude we will briefly sketch how to make the relationship between a metric on $\widehat{\mathbb{C}}$ and an inner product on $\mathbb{C}^2$ more concrete.

The idea is that a suitable real–valued subharmonic function ϕ on a region $U \subset \widehat{\mathbb{C}}$ naturally determines a metric ρ by the equation

$$\rho^2(z) = \Delta\phi(z).$$

(More generally, on a complex manifold one uses the symplectic form $\omega = i\partial\overline{\partial}\phi$ and takes the associated Kähler metric.)

We will be interested in the case where $\phi(z) = \log\psi$ for some locally–defined function ψ. Note that if we replace $\psi(z)$ by $\psi(z)|f(z)|^2$, where $f(z)$ is holomorphic, then the value of ρ is unchanged, since $\log|f(z)|^2$ is harmonic.

Here is the construction of ρ from an inner product on $\mathbb{C}^2$. First, we *locally* choose a section of the map $\pi : \mathbb{C}^2 - 0 \to \widehat{\mathbb{C}}$, say $Z(z) = (Z_0(z), Z_1(z))$. Then, we set

$$\psi(z) = \|Z(z)\|^2 = \langle Z(z), Z(z)\rangle.$$

Finally, we set $\rho^2 = \Delta\log\psi$.

Note that if we choose a different section $W(z)$, then it differs from the first by a nowhere vanishing holomorphic function: $W(z) = f(z)Z(z)$. This

changes ψ by a factor of $|f(z)|^2$, so the value of ρ^2 remains the same. Consequently the construction is natural, and hence the orthogonal group of the inner product maps into the isometry group of the metric.

Let us check the formula in the two main cases. In both cases we can take $Z(z) = (1, z)$. For the positive–definite inner product with $\|Z\|^2 = |Z_0|^2 + |Z_1|^2$, we obtain

$$\rho^2(z) = 4\frac{d^2}{dz\,d\bar{z}}\log(1 + |z|^2) = \frac{4}{(1 + |z|^2)^2},$$

which agrees with our previous definition of the spherical metric. For the indefinite inner product $\|Z\|^2 = |Z_1|^2 - |Z_0|^2$, and z in the unit disk, we need to throw in a -1 to account for negative curvature, and we get:

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

also consistent with our formula for the hyperbolic metric.

In fact, this construction works for any Hermitian line bundle over a Riemann surface, and the 'metric' is actually the curvature form for the induced connection on the line bundle.

2.5 Exercises

1. Compute the area of the region $T \subset \mathbb{H}$ defined by $\mathrm{Re}(z) \in [-1, 1]$ and $|z| > 1$, with respect to the metric $\rho = |dz|/\mathrm{Im}(z)$.

2. Prove there is no proper holomorphic map $f : \Delta \to \mathbb{C}$.

3. (i) Find all the proper holomorphic maps $f : \mathbb{C}^* \to \mathbb{C}$. (ii) Find all the proper holomorphic maps $g : \mathbb{C} \to \mathbb{C}^*$.

4. (i) Find all the discrete subgroups $\Gamma \subset \mathrm{Aut}\,\mathbb{C}$, up to conjugacy.

   (ii) Describe all of the corresponding orbifolds $\mathbb{C}/\Gamma$.

5. A rational map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a *quotient map* for a finite group $\Gamma \subset \mathrm{Aut}(\widehat{\mathbb{C}})$ if $f(x) = f(y)$ iff $g(x) = y$ for some $g \in \Gamma$.

   Let $\Gamma \subset \mathrm{Aut}(\widehat{\mathbb{C}})$ be the subgroup leaving the set $\{0, 1, \infty\}$ invariant.

   (i) Show that Γ is isomorphic to the symmetric group S_3.

74

(ii) Find a quotient map for Γ, normalized so that $f(\infty) = \infty$, $f(\zeta_6) = 0$ and $f(-1) = 1$.

6. (i) Verify that the spherical metric $\rho = 2|dz|/(1 + |z|^2)$ is invariant under $h(z) = (z - 1)/(z + 1)$. (ii) Show that the rotations $g_\theta(z) = e^{i\theta} z$, together with h, generate the full group $\mathrm{SU}(2)/(\pm I) \subset \mathrm{PSL}_2(\mathbb{C})$.

7. The quaternions are the algebra $H = \mathbb{R}[i, j, k]$ subject to the relations $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$. As a vector space, we have $H \cong \mathbb{R}^4$, and the *unit norm quaternions* U are a multiplicative subgroup isomorphic to S^3.

   Prove that $\mathrm{SU}(2)$ is isomorphic, as a group, the unit norm quaternions U. (Hint: find an embedding of H into $\mathrm{M}_2(\mathbb{C})$ as an $\mathbb{R}$–algebra, by choosing of matrices $(I, J, K) \in \mathrm{SU}(2)$ that satisfy the same relations as (i, j, k).)

8. Let $p(z) = a_0 z^3 + a_1 z^2 + a_2 z + a_3$ be a cubic polynomial. Assume $a_0 \neq 0$ and $p'(z)$ has distinct zeros.

   (a) Show that we can write $A(p(B(z))) = C(z) = z^3 - 3z$ for suitable $A, B \in \mathrm{Aut}\,\mathbb{C}$. (Hint: adjust p so its critical points are at ± 1.)

   (b) Show how to solve the equation $z^3 - 3z = w$ by radicals, using the fact that if $z = a + a^{-1}$, then $z^3 - 3z = a^3 + a^{-3}$.

   (c) Combining the previous calculations, solve the equation $z^3 + az + b = 0$ by radicals.

9. (a) Show that the metric $\rho = |dz|/|z|^2$ on $\widehat{\mathbb{C}}$ has no singularity at $z = \infty$ (i.e. its cone angle is 2π). (Hint: examine its behavior near $z = 0$ after the change of coordinates $z \mapsto 1/z$.)

   (b) Show that the singular flat metric $\rho = |dz|/|1 - z^3|^{2/3}$ on $\widehat{\mathbb{C}}$ has cone angle $2\pi/3$ at the cube roots of unity, and no singularity at infinity. (In fact $(\widehat{\mathbb{C}}, \rho)$ is isometric to the double of an equilateral triangle in the plane.)

10. Find the zeros, critical points and critical values of the Chebyshev polynomial $T_n(z)$, $n \geq 1$.

11. Prove that any proper holomorphic map $f : \Delta \to \Delta$ of degree two can be written in the form $f(z) = A(S(B(z)))$, with $A, B \in \operatorname{Aut}\Delta$ and $S(z) = z^2$.

12. Consider the half-infinite rectangle $S = \{z = x+iy : |x| < 1 \text{ and } y > 0\}$. Give an explicit conformal map (analytic homeomorphism) from S to the upper halfplane $\mathbb{H}$.

13. Give a formula for the stereographic projection $z \in \mathbb{C}$ of a point on S^2 with latitude and longitude (α, β).

14. Find the distance between Boston (42 N, 71 W) and Rio (23 S, 43 W), assuming the circumference of the earth is 40,000 km.

15. Find the hyperbolic distance between i and $i + 1$ in $\mathbb{H}$.

16. Prove that any proper holomorphic map $f : \mathbb{H} \to \mathbb{H}$ can be written in the form
$$f(z) = a_0 z + b_0 + \sum_1^n \frac{a_i}{b_i - z},$$
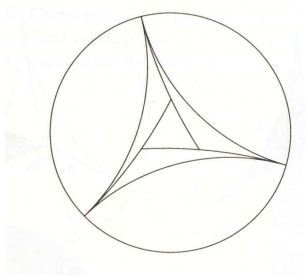with $a_i \geq 0$ and $b_i \in \mathbb{R}$.

17. Let $f : \Delta \to \Delta$ be a holomorphic map such that $f(0) = 0$. Prove that either f is a rotation, or $f^n(z) \to 0$ as $n \to \infty$. (Here $f^n = f \circ f \circ \cdots \circ f$).

18. Let $a, b \in \Delta \subset \mathbb{P}^1$ be represented by vectors $A, B \in \mathbb{C}^2$, and let δ be the distance between a and b in the hyperbolic metric $\rho_\Delta = 2|dz|/(1-|z|^2)$. Prove that
$$\cosh(\delta/2) = \frac{|A \cdot B|}{\sqrt{|A^2||B^2|}},$$
where $A \cdot B = A_0 \overline{B}_0 - A_1 \overline{B}_1$, $A^2 = A \cdot A$, and $B^2 = B \cdot B$.

19. Let $T(a, b, c) \subset \mathbb{H}$ be a hyperbolic triangle with interior angles a, b, c. Prove that area $T(a, b, c) = \pi - a - b - c$ as follows.

(i) Note that $T(0, 0, 0)$ is an 'ideal triangle' with all vertices at infinity. Show all ideal triangles have area π. (ii) Show *geometrically* that $A(a) = T(\pi - a, 0, 0)$ satisfies $A(a + a') = A(a) + A(a')$, and conclude $A(a) = a$. (iii) Extend the sides of $T(a, b, c)$ to rays as below:

and use the corresponding vertices at infinity to relate the area of $T(a, b, c)$ to the area of $T(\pi - a, 0, 0)$, $T(\pi - b, 0, 0)$ and $T(\pi - c, 0, 0)$.

20. Compute the hyperbolic metric on $\Delta^* = \Delta - \{0\}$.

21. Let $A(R) = \{z : 1 < |z| < R\}$. (i) Give an explicit analytic covering map $\pi : \mathbb{H} \to A(R)$. (ii) Compute the hyperbolic metric on $A(R)$. (iii) Show that $|z| = \sqrt{R}$ is a closed geodesic in $A(R)$, and compute its hyperbolic length.

22. Let $f : \Delta \to \mathbb{C}$ be an analytic function such that $\operatorname{Re} f(z) \geq 0$. Show that $|f(z)| \leq |f(0)|(1 + |z|)/(1 - |z|)$.

23. Observe that any proper analytic map $f : \Delta \to \Delta$ extends to a continuous function $F : \overline{\Delta} \to \overline{\Delta}$. Suppose $f(0) = 0$. Show that for all continuous functions $\phi : S^1 \to \mathbb{R}$, ϕ and $\phi \circ F$ have the same average over S^1. (Hint: use the harmonic extension of ϕ.)

24. Let $\mathcal{U}$ be set of *positive* harmonic function $u(z)$ on the unit disk Δ, normalized so that $u(0) = 1$. Show that $\mathcal{U}$ is compact; that is, any sequence has a subsequence converging uniformly on compact subsets of Δ.

25. Prove that there is no conformal metric $\rho = \rho(z)|dz|$ on $\widehat{\mathbb{C}}$ that is invariant under the full automorphism group, $\operatorname{Aut}(\widehat{\mathbb{C}})$.

26. Consider the statement:

> *Any discrete subgroup $\Gamma \subset \operatorname{Aut}(U)$ acts properly discontinuously on U, and thus U/Γ is a complex orbifold.*

For each of $U = \mathbb{C}, \widehat{\mathbb{C}}$ and $\mathbb{H}$, either prove this statement or given a counterexample. (Here $\mathrm{Aut}(U)$ is given the topology of uniform convergence on compact sets, or equivalently the topology it inherits as a subgroup of $\mathrm{SL}_2(\mathbb{C})/(\pm I)$.)

27. Pick a basis for the Lie algebra of $\mathrm{SL}_2(\mathbb{C})$, and show how each basis element can be canonically interpreted as a holomorphic vector field on $\widehat{\mathbb{C}} = \mathbb{PC}^2$. Check that Lie bracket corresponds to bracket of vector fields, up to sign.

28. Let $f_n : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be an unbounded sequence in $\mathrm{Aut}(\widehat{\mathbb{C}})$. Prove that after passing to a subsequence, there are $p, q \in \widehat{\mathbb{C}}$ such that $f_n(z) \to p$ uniformly on compact subsets of $\widehat{\mathbb{C}} - \{q\}$. Give an example of a sequence f_n where $p = q = 0$.

29. Give an example of a sequence of degree 3 rational maps $f_n : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $f_n(S^1) \subset S^1$, and $f_n(z) \to z$ uniformly on S^1. Compute $\int_{S^1} f_n'(z)/f_n(z)\, dz$ for your example.

30. A *circular triangle* is a region $T \subset \widehat{\mathbb{C}}$ bounded by three circular arcs or lines. Let $T_1, T_2 \subset \widehat{\mathbb{C}}$ be two circular triangles with the same interior angles in the same cyclic order. Prove there exists a Möbius transformation sending T_1 to T_2.

31. Which circles in $\widehat{\mathbb{C}}$ correspond to great circles on S^2? That is, which circles have length 2π in the metric $2|dz|/(1 + |z|^2)$? (Hint: how does a great circle meet the equator?)

32. Prove that the Chebyshev polynomials are orthogonal in $L^2([-1, 1])$ with respect to this density $\rho(x) = (1 - x^2)^{-1/2}$. That is, show that if $a \neq b$ then $\int T_a(x)T_b(x)\rho(x)\, dx = 0$. (Hint: use orthogonality of z^a and z^b on S^1).

33. Give a Kähler potential ϕ for the hyperbolic metric on $\mathbb{H}$. (This means you should have $\Delta \log \phi = 1/y^2$, $z = x + iy$.)

34. Find a Hermitian inner product of signature $(1, 1)$ on $\mathbb{C}^2$ such that its light cone is given by the 'circle' $\widehat{\mathbb{R}} \subset \widehat{\mathbb{C}}$. (This means $\langle Z, Z \rangle = 0 \iff Z_1/Z_0 \in \widehat{\mathbb{R}}$.)

3   Entire and meromorphic functions

This section discusses general constructions of functions on $\mathbb{C}$ with given zeros and poles, and analyzes special cases such as the trigonometric functions and $\Gamma(z)$.

The study of zeros leads to expressions for entire functions as *infinite products*, and the study of poles leads to expressions for meromorphic functions as *infinite sums*.

3.1   Zeros of entire functions

An entire function is simply an analytic map $f : \mathbb{C} \to \mathbb{C}$. An entire function has *finite order* if $|f(z)| \le A \exp(|z|^\rho)$ for some $A, \rho \ge 0$. The infimum of all such ρ is the *order* $\rho(f)$.

In this section we will study the relationship between f and its set of zeros $Z(f)$. The latter, when infinite, form a sequence $a_n \to \infty$ in $\mathbb{C}$. The *critical exponent* of a_n is infimum of those $\alpha > 0$ such that $\sum 1/|a_n|^\alpha < \infty$. In this sum we ignore the finitely many terms where $a_n = 0$.

We will see that $Z(f)$ is close to determining f provided f has finite order. Indeed, we will be able to express f in terms of an infinite product of functions $E_p(z/a_n)$, where $E_p(z)$ is a carefully chosen entire function vanishing only at $z = 1$, generalizing $E_0(z) = (1 - z)$.

In summary we will show:

1. Every sequence $a_n \to \infty$ arises as the zero set of some entire function. Consequently every meromorphic function on $\mathbb{C}$ can be written as a ratio of two entire functions. Thus $\mathcal{M}(\mathbb{C})$ is the field of fractions of $\mathcal{O}(\mathbb{C})$.

2. If $f(z)$ has order ρ, then the critical exponent of its zeros a_n satisfies $\alpha(a_n) \le \rho(f)$.

3. Conversely, if $\sum |a_n|^{p+1} < \infty$ but $\sum |a_n|^p = \infty$, then

$$P(z) = \prod_1^\infty E_p(z/a_n)$$

is a function of order $\rho(P) = \alpha(a_n)$ with the given zeros.

4. Any function $f(z)$ of finite order can be expressed canonically as an infinite product:
$$f(z) = z^m e^{Q(z)} \prod E_p(z/a_n),$$
where $Q(z)$ is a polynomial in z. (The product is finite when f is a polynomial.) The order $\rho(f)$ is the maximum of $\alpha(a_n)$ and the degree of Q.

Infinite products. We now turn to a detailed discussion. Our first task is to construct functions with prescribed zeros as infinite products. For example, we should like to construct 'by hand' a function similar to $\sin(\pi z)$ with zeros at the integers $\mathbb{Z} \subset \mathbb{C}$.

Recall that $\prod(1+a_n)$ converges (to an element of $\mathbb{C}^*$) if $\sum |a_n|$ converges. (Note: we say $\prod(1 - 1/n)$ *diverges to zero*.) Similarly, $\prod(1 + f_n(z))$ defines an entire function on $\mathbb{C}$ if $\sum |f_n(z)|$ converges uniformly on compact sets. For example, we have:

Theorem 3.1 *If $\sum 1/|a_n|$ is finite, then $f(z) = \prod(1 - z/a_n)$ defines an entire function with zero set $Z(f) = (a_n)$.*

Multiplicities of zeros can also be specified – they just correspond to repetitions of the same number in the sequence a_i.

However this result is still too weak to address problem of constructing an entire function with $Z(f) = \mathbb{Z}$.

Use of the logarithmic derivative. For later application, we record some useful properties of the logarithmic derivative
$$L(f) = d(\log f) = f'(z)/f(z)\, dz.$$
Since we are working in the complex plane, we often conflate f'/f and $L(f)$, just as we conflate the residue of $f(z)\, dz$ and the 'residue' of f.

1. We have $L(fg) = L(f) + L(g)$.

2. If $L(f) = L(g)$, then $f = Cg$ for some constant $C \neq 0$.

3. We have $L(f \circ g) = g^*(L(f))$.

4. We have $\mathrm{Res}_p(L(f)) = \mathrm{ord}_p(f)$.

(The last familiar identity was used in our discussion of the argument principle.)

Another useful identity is:

$$\left(\log \left(\frac{1}{1-z} \right) \right)' = \frac{1}{1-z}.$$

Put differently, $f(z) = 1/(1-z)$ satisfies $L(f) = f$. This identity underlies the power series

$$\log(1/1 - z) = z + z^2/2 + z^3/3 + \cdots ,$$

valid for $|z| < 1$, which is obtained by integrating the geometric series for $1/(1-z)$.
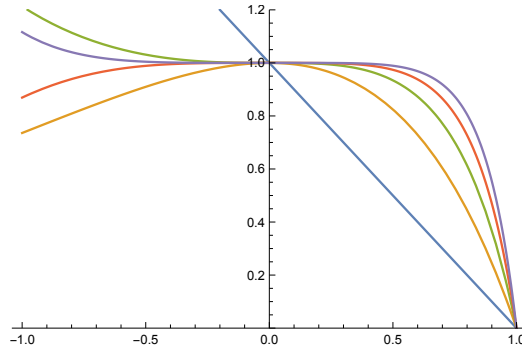

Figure 3.1. Weierstrass factors $E_p(z)$, $0 \leq p \leq 4$.

Weierstrass factors. Here is an elegant expression for an entire function with a zero only at $z = 1$, which is also close to 1 for $|z| < 1$. It is called the *Weierstrass factor* of order p:

$$E_p(z) = (1 - z) \exp \left(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p} \right).$$

By convention, $E_0(z) = (1 - z)$.

The idea behind this expression is that $\log(1/(1-z)) = z + z^2/2 + z^3/3 + \cdots$, and hence the two terms 'almost cancel' to give $(1 - z)/(1 - z) = 1$. (For more insight, compute the logarithmic derivative E_p'/E_p or see the proof below.)

Since we have truncated at the term z^p, it is easy to see that for $|z| < 1/2$ (say) we have
$$|E_p(z) - 1| = O(|z|^{p+1}).$$
For many purposes this bound is sufficient. However when p varies in the product, it is *essential* to have a bound with an *explicit constant*. It is also useful if the bound is uniform over $|z| < 1$. Such a bound is provided by:

Theorem 3.2 *For $|z| < 1$, we have $|E_p(z) - 1| \leq |z|^{p+1}$.*

Proof. We wish to control $f(z) = 1 - E_p(z)$ with satisfies $f(0) = 0$, $f(1) = 1$, and
$$f'(z) = (-E'_p(z)/E_p(z))E_p(z).$$
Taking the negative of the logarithmic derivative of E_p, we obtain:
$$-E'_p(z)/E_p(z) = 1/(1-z) - 1 - z - \ldots - z^{p-1} = z^p/(1-z).$$
Hence for all z we have
$$f'(z) = z^p \exp(z + z^2/2 + \cdots z^p/p) = z^p \sum_0^\infty a_k z^k,$$
with $a_k \geq 0$ for all k. Integrating term by term and using the fact that $f(0) = 0$ and $f(1) = 1$, we find
$$f(z) = z^{p+1} \sum_0^\infty b_k z^k$$
with $b_k \geq 0$ and $\sum b_k = 1$. Hence for $|z| < 1$ we have
$$|f(z)| \leq |z|^{p+1} \sum b_k = |z|^{p+1}.$$
∎

Theorem 3.3 *For any sequence of nonzero complex numbers $a_n \to \infty$, the formula $f(z) = \prod_1^\infty E_n(z/a_n)$ converges for all z and defines an entire analytic function with zero set exactly (a_n).*

Proof. In this and later discussions we set $r_n = |a_n|$. The previous estimate yields convergence of the tail of the series: for all z with $|z| < r$ we have:
$$\sum_{|a_n|>2r} |1 - E_n(z/a_n)| \leq \sum_{r_n>2r} (r/r_n)^{n+1} \leq \sum_1^\infty (1/2)^{n+1} < \infty.$$
∎

82

Note that in the estimate above, it was essential to have a bound on $|1 - E_n(z)|$ as n varies.

Corollary 3.4 *Every meromorphic function on $\mathbb{C}$ is the ratio of two entire functions.*

The Weierstrass factor for large z. For $|z| \geq 1$, and especially for large values of $|z|$, $E_n(z)$ grows very rapidly; however we at least have the uniform estimate:

$$\log |E_n(z)| \leq O(|z|^n).$$

This says slightly more than $\rho(E_n(z)) = n$, e.g. because the exponent is n instead of $n + \epsilon$.

Growth rates: max, min, average and zeros. Our next task is to develop a relationship between several important invariants of an entire function $f : \mathbb{C} \to \mathbb{C}$ at a given radius $r > 0$. We begin by define these functions.

The *maximum—* and *minimum—* functions are given by:

$$M(r) = \sup_{|z|=r} |f(z)| \quad \text{and} \quad m(r) = \inf_{|z|=r} |f(z)|.$$

We let

$$A(r) = \frac{1}{2\pi} \int_{S^1(r)} \log |f(z)| \frac{|dz|}{|z|}$$

denote the *average* of $\log |f(z)|$ over $|z| = r$. We always have

$$A(r) \leq \log M(r),$$

The *counting function* measures the number of zeros (a_n) of $f(z)$ (repeated with multiplicities) in the disk of radius r; it is defined by

$$N(r) = |\{n \ : \ |a_n| < r\}|$$

This function is also be defined for any discrete (a_n) in $\mathbb{C}$.

We will study the interplay of these functions, and then use some of these ideas to prove a *canonical factorization theorem* for functions of finite order. (This factorization, like the functions above, is not translation invariant.)

Hadamard's 3–circles theorem. We begin by studying the behavior of $M(r)$. The following result applies not just to entire functions, but also to functions analytic in an annulus of the form $r_1 < |z| < r_2$.

83

Theorem 3.5 *For any analytic function $f(z)$, the quantity $\log M(r)$ is a convex function of $\log r$.*

Proof. A function $\phi(s)$ of one real variable is convex if and only if $\phi(s) + ar$ satisfies the maximum principle for any constant a. This maximum function holds for $\phi(s) = \log M(\exp(s))$ by the usual maximum principle, and it follows for $\phi(s) + ar$ by considering $f(z)z^a$ locally. ∎

Corollary 3.6 *We have $M(\sqrt{rs}) \leq \sqrt{M(r)M(s)}$.*

The convex function defined by $F(\log r) = \log M(r)$ looks roughly linear for polynomials, e.g. like $F(x) = \deg(f)x + c$, and looks roughly exponential for functions of finite order, e.g. $F(x) = \exp(\rho(f)x)$.

Functions of finite order. For the remainder of this section we will focus on functions of finite order. We remark that the order of f is given by

$$\rho(f) = \limsup_{r \to \infty} \frac{\log \log M(r)}{\log r}. \qquad (3.1)$$

Examples. Polynomials have order 0; $\sin(z)$, $\cos(z)$ and $\exp(z)$ have order 1; $\cos(\sqrt{z})$ has order $1/2$; $E_p(z)$ has order p; $\exp(\exp(z))$ has infinite order.

Note that $\rho(fg)$ and $\rho(f+g)$ are both bounded above by $\max\{\rho(f), \rho(g)\}$. This implies:

Proposition 3.7 *The entire functions $\mathcal{O}_\rho(\mathbb{C})$ with $\rho(f) \leq \rho$ form a ring.*

In fact $\rho(fg) = \max\{\rho(f), \rho(g)\}$, provided $\rho(f) \neq \rho(g)$ (see bounds on the minimum modulus below). For example, if g is a polynomial then $\rho(fg) = \rho(f)$. Roughly speaking, adding or subtracting finitely many zeros to f does not change its order.

To see that functions of finite order have controlled behavior, we first characterize those with no zeros.

Theorem 3.8 *An entire function $f(z)$ with no zeros has finite order if and only if $f(z) = \exp(Q(z))$ for some polynomial $Q(z)$; and $\rho(f) = \deg(Q)$.*

To prove it we strengthen our earlier characterization of polynomials by $M(r) = O(r^d)$.

Lemma 3.9 *Let $Q(z)$ be an entire function satisfying $\operatorname{Re} Q(z) \le A|z|^d + B$ for some $A, B > 0$. Then Q is a polynomial of degree at most d.*

Proof. Applying the Schwarz Lemma to the hyperbolic metrics on the disk $\Delta(R)$ and the halfplane $H = \{z : \operatorname{Re}(z) < 0\}$, we find that if $f : \Delta(2R) \to H$ is an analytic function, then $\sup_{\Delta(R)} |f(z)| = O(|f(0)|)$. (Here we have used the fact that the hyperbolic diameter of $\Delta(R)$ in $\Delta(2R)$ is a constant, independent of R, so the image in H is contained in a hyperbolic ball centered at $f(0)$ of bounded radius.)

By assumption, $f(z) = Q(z) - A(2R)^d - B$ has negative real part on $\Delta(2R)$, so

$$\sup_{\Delta(R)} |Q(z)| = O(|R|^d),$$

which implies that Q is a polynomial of degree at most d. ∎

Proof of Theorem 3.8. Since f has no zeros, $f(z) = e^{Q(z)}$ for some entire function $Q(z)$. Since f has finite order, $|f(z)| = O(e^{|z|^\rho})$ for some ρ, and thus $\operatorname{Re} Q(z) \le |z|^\rho + O(1)$; now apply the Lemma above. ∎

The critical exponent. We now turn to the *zeros* of functions of finite order.

Let α denote the critical exponent of a sequence of nonzero complex numbers $a_n \to \infty$. Recall the critical exponent is the borderline between convergence and divergence for $\sum |a_n|^\alpha$, and $N(r)$ counts the number of zeros of modulus $< r$.

Theorem 3.10 *For any discrete set (a_n) in $\mathbb{C}$, we have:*

$$\alpha(a_n) = \limsup_{r \to \infty} \frac{\log N(r)}{\log r}. \tag{3.2}$$

In other words, the critical exponent is the same as the infimum of those α such that $N(r) = O(r^\alpha)$. Note the similarity to the formula for $\rho(f)$.

Proof. It is useful to observe that, at least formally, $N'(r) = \sum \delta_{r_n}$. Integrating by parts, we find that for any $\mu \ge 0$ we have:

$$\sum |a_n|^{-\mu} = \int_0^\infty N'(r) r^{-\mu} \, dr = \int_0^\infty N(r) \mu r^{-\mu} \frac{dr}{r}. \tag{3.3}$$

In particular, if $N(r) < r^\beta$ for all $r \gg 0$, then for any $\epsilon > 0$ we have $\sum |a_n|^{-\beta-\epsilon} < \infty$, and hence $\alpha(a_n) \le \beta$. On the other hand, if $N(\rho_i) > \rho_i^\beta$ along a sequence $\rho_i \to \infty$, then we can pass to a subsequence such that the intervals $[\rho_i, 2\rho_i]$ are disjoint and use the equation above to show that $\sum |a_n|^{-\beta} = \infty$, and hence $\alpha(a_n) \ge \beta$. This demonstrates equation (3.2). ■

Jensen's formula. We can now show that the zeros of f satisfy $\alpha(a_n) \le \rho(f)$. To this end we will relate the zeros to the log average function $A(r)$ by an *exact formula*.

Theorem 3.11 (Jensen's formula) *Let $f(z)$ be a holomorphic function on $B(0, r)$ with zeros $a_1, \ldots a_n$. Then the average of $\log |f(z)|$ over $S^1(r)$ satisfies*

$$A(r) = \log |f(0)| + \sum \log \frac{r}{|a_i|}.$$

Proof. We first note that if f has no zeros, then $\log |f(z)|$ is harmonic and the formula holds. Moreover, if the formula holds for f and g, then it holds for fg; and the case of general r follows from the case $r = 1$, since both sides are invariant under replacing z with rz and a_i with ra_i.

Next we verify that the formula holds when $f(z) = (z - a)/(1 - \bar{a}z)$ on the unit disk, with $|a| < 1$. Indeed, in this case $\log |f(z)| = 0$ on the unit circle, and $\log |f(0)| + \log(1/|a|) = \log |a/a| = 0$ as well.

The general case now follows, since a general function $f(z)$ on the unit disk can be written in the form $f(z) = g(z) \prod (z - a_i)/(1 - \bar{a}_i z)$. ■

Jensen's theorem can be restated as follows:

Corollary 3.12 *We have*

$$A(r) = \int_0^r N(s) \frac{ds}{s} + \log |f(0)|. \tag{3.4}$$

The integral over $[r/2, r]$ is at least $N(r/2) \log 2$, which gives

Corollary 3.13 *We have $N(r/2) = O(|A(r)| + 1) = O(\log M(r))$*

Corollary 3.14 *For any entire function $f(z)$ with zeros (a_n), we have $\alpha(a_n) \le \rho(f)$.*

Proof. We have

$$\alpha = \limsup \frac{\log N(r)}{\log r} \le \limsup \frac{\log \log M(2r)}{\log(r)} = \rho.$$

■

86

Behavior of the average of $\log |f|$. By equation 3.4, the quantity $N(r/2)$ is a good approximation to $A(r)$, in the sense that the two differ by a factor of at most $\log r$.

On the other hand, $A(r)$ is definitely only a lower bound for $\log M(r)$. For example, when $f(z) = \exp(Q(z))$ with Q a polynomial of degree d, we have $\log M(r) \asymp r^d$, while

$$A(r) = \mathrm{avg}_{S^1(R)} \operatorname{Re} Q(z) = \operatorname{Re} Q(0)$$

is *constant*! Since f is large over much of the circle, it must also be close to zero somewhere on the same circle. (For example, e^z is very large over half the circle $|z| = R$, and very small over the rest.) This observation is the beginning of *value distribution theory.*

The Hadamard factorization theorem will show that, for entire functions of finite order, *either* f has a factor of $\exp(Q(z))$, *or* the size of $M(r)$ is accounted for by the zeros of f (or both).

Canonical products. Let (a_n) be a sequence with critical exponent $\alpha < \infty$. Choose the least $p \geq 0$ such that $\sum 1/|a_n|^{p+1} < \infty$. Note that $\alpha \in [p, p+1]$.

The *canonical product* associated to a_n is given by

$$P(z) = \prod E_p(z/a_n).$$

To see that it converges, set $r_n = |a_n|$ as usual and suppose $|z| \leq r$; then we have

$$\sum_{|a_n|>r} |E_p(z/a_n) - 1| \leq \sum_{r_n>r} |r/r_n|^{p+1} < \infty.$$

Thus $P(z)$ is a function of finite order whose zero set is exactly (a_n). Moreover we have:

Theorem 3.15 *The canonical product $P(z)$ associated to (a_n) has order $\rho(f) = \alpha(a_n)$.*

Proof. Let $r_n = |a_n|$ and $r = |z|$ as usual, and let $\log^+ x = \max(0, \log x)$. Fix $\epsilon > 0$. We wish to show that $\log^+ |P(z)| = O(r^{\alpha+\epsilon})$. Recall that $p \leq \alpha \leq p+1$ and $\sum 1/r_n^{p+1} < \infty$.

We have seen that for $|z| \leq 1$ we have $|1 - E_p(z)| \leq |z|^{p+1}$ and hence

$$\log^+ |E_p(z)| = O(|z|^{p+1}).$$

This shows that

$$\sum_{|a_n|>|z|} \log^+|E_p(z/a_n)| = O\left(\sum_{r_n>r}(r/r_n)^{p+1}\right).$$

We claim the final sum is $O(r^{\alpha+\epsilon})$. This is clear if $\alpha = p + 1$; otherwise $\alpha < p + 1$, and for ϵ less than the difference we have

$$\sum_{r_n>r}(r/r_n)^{p+1} \le \sum_{r_n>r}(r/r_n)^{\alpha+\epsilon} = O(r^{\alpha+\epsilon}).$$

Here we have used the fact that $r/r_n < 1$ for the first step, and the fact that $\sum 1/r_n^{\alpha+\epsilon} < \infty$ for the second.

Now for $|z| > 1$ we have

$$\log^+|E_p(z)| = O(|z|^p + \log^+|z|) = O(|z|^{p+\epsilon})$$

for any $\epsilon > 0$. This gives

$$\sum_{|a_n|\le|z|} \log^+|E_p(z/a_n)| = O\left(\sum_{r_n\le r}(r/r_n)^{p+\epsilon}\right).$$

Since $\alpha \ge p$, we have

$$\sum_{r_n\le r}(r/r_n)^{p+\epsilon} \le \sum_{r_n\le r}(r/r_n)^{\alpha+\epsilon} = O(r^{\alpha+\epsilon}).$$

Here we have used the fact that $r/r_n > 1$ for the first step, and again the fact that $\sum 1/r_n^{\alpha+\epsilon} < \infty$ for the second. Putting these two bounds together, we get

$$\log^+|P(z)| = O(r^{\alpha+\epsilon})$$

for all $\epsilon > 0$, and hence $\rho(f) \le \alpha$. ∎

Hadamard's factorization theorem. We can now state a formula that describes every entire function of finite order in terms of its zeros and an additional polynomial.

88

Theorem 3.16 (Hadamard) *A entire function $f(z) \neq 0$ of finite order ρ can be uniquely expressed in the form*

$$f(z) = z^m e^{Q(z)}, \prod E_p(z/a_n),$$

where $m = \mathrm{ord}(f, 0)$, (a_n) are the other zeros of f, $p \geq 0$ is the least integer such that $\sum 1/|a_n|^{p+1} < \infty$, and $Q(z)$ is a polynomial of degree q. We have $\rho(f) = \max(q, \alpha(a_n)) \geq p$.

The number p is called the *genus* of f. Note that ordinary polynomials arise as a special case, with $p = q = \rho = 0$.

Remark. This theorem shows that the zeros of f determine f up to finitely many additional constants, namely the coefficients of $Q(z)$. It is tempting to conclude then that if $f(z)$ has no zeros, it is determined by its values at any $\lfloor \rho + 1 \rfloor$ points. This is not quite true, however, since $f(z)$ only determines $Q(z) \bmod 2\pi i\mathbb{Z}$; for example, $\exp(2\pi i z)$ and the constant function 1 agree on the integers.

It *is*, however, true that $f'/f = Q'$ in this case, so knowing the logarithmic derivative at enough points almost determines Q. This shows:

Corollary 3.17 *Suppose $f(z)$ and $g(z)$ are entire functions of order ρ with the same zeros, and $f'/f = g'/g$ at $\lfloor \rho \rfloor$ distinct points (where neither function vanishes). Then f is a constant multiple of g.*

Here is another fast Corollary that would be hard to prove directly.

Corollary 3.18 *If $\rho(f) < \infty$ is not an integer, then $f(z)$ assumes every value in the complex plane infinitely often.*

Proof. Use the fact that $\rho(f) = \rho(f(z) - a)$ and that when f has only finitely many zeros, its order is an integer. ∎

The minimum modulus. To control the result of division by a canonical product, we now estimate its minimum modulus. As an example — for $\sin(z)$, we have $m(r) \asymp 1$ for infinitely many r.

Theorem 3.19 *The minimum modulus of the canonical product $P(z)$ associated to the sequence (a_n) with critical exponent α satisfies, for each $\epsilon > 0$,*

$$m(r) \geq \exp(-r^{\alpha + \epsilon})$$

for all r outside a set of finite measure.

Proof. Fix $\epsilon > 0$, and set $\log^-(x) = \max(0, \log(1/x))$. We must show that $\log^- m(r) = O(r^{\alpha+\epsilon})$ for arbitrarily large r.

We must be careful about which values of r to consider, since $m(r) = 0$ whenever $r = |a_n|$. To this end, we fix $N > \alpha$ and exclude from consideration the balls B_n defined by $|z - a_n| < 1/r_n^N$. Since the sum of the radii of the excluded balls in finite, there are plenty of large circles $|z| = r$ which avoid $\bigcup B_n$.

To complete the proof, we will show that for z on such a circle of radius r, we have $\log^- |f(z)| = O(r^{\alpha+\epsilon})$. Our preceding arguments already control the exponential parts of the Weierstrass factors, so it remains to show that:

$$\sum_{|z-a_n|<r} \log^- |1 - z/a_n| = O(r^{\alpha+\epsilon}).$$

(For $|z - a_n| \geq r$ we have $\log^- |1 - z/a_n| \geq \log^- 1 = 0$.) Note that we must have $r_n \leq 2r$ for a_n to appear in the sum above. Thus the number of terms in the sum is at most $N(2r)$. Because we have kept z away from a_n, we also have

$$\log^- |1 - z/a_n| = O(\log r).$$

Consequently

$$\sum_{|z-a_n|<r} \log^- |1 - z/a_n| = O(N(2r) \log r) = O(r^{\alpha+\epsilon}),$$

as desired. ∎

Proof of the Hadamard factorization theorem. Let $f(z)$ be an entire function of order ρ with zeros (a_i), and let $P(z)$ be the corresponding canonical product. Then P also has order ρ, as we have just seen. Since f and P have the same zeros, the quotient f/P is an entire function with no zeros. Therefore $f/P = \exp(Q(z))$ for some entire function $Q(z)$. The lower bound on $m(r)$ just established implies that $\operatorname{Re} Q(z) = \log |f/P| = O(|z|^{\rho+\epsilon})$ on arbitrarily large circles. Using the Schwarz lemma, one can then deduce that $|Q(z)| = O(r^{\rho+\epsilon})$ on arbitrarily large circles. By Cauchy's bound, $Q(z)$ is then a polynomial, and clearly these estimates imply that $\deg(Q) \leq \rho$. ∎

3.2 The gamma function and trigonometric functions

There are 3 natural periodic or almost periodic zero sets whose canonical products lead to entire functions of central importance: $\mathbb{Z}$, $\mathbb{Z} \oplus \tau\mathbb{Z}$, and $\mathbb{Z}_- = \{-1, -2, -3, \ldots\}$. The first and last lead to the trigonometric functions and the gamma function, which we will discuss below. The second leads into the theory of elliptic functions, which form the subject of a later chapter.

Trigonometric functions. As a first application of Hadamard's theorem, we determine the canonical factorization of the sine function:

Theorem 3.20 *We have*

$$\frac{\sin(\pi z)}{\pi z} = \prod_{n > 0} \left(1 - \frac{z^2}{n^2} \right).$$

Proof. Indeed, the right hand side is a canonical product with $p = 1$; the exponential terms for $\pm n$ have canceled. On the left hand side, $\sin(\pi z)$ has order one, so the formula is correct up to a factor $\exp Q(z)$ where $Q(z)$ has degree one. But since $\sin(\pi z)$ is odd, we conclude Q has degree zero, and by checking the value of both sides at $z = 0$ we get $Q = 0$. ∎

Sine, cotangent and zeta. The product formula above gives, under logarithmic differentiation,

$$\frac{(\sin(\pi z))'}{\sin(\pi z)} = \pi \cot(\pi z) = \frac{1}{z} + \sum_{1}^{\infty} \frac{1}{z - n} + \frac{1}{z + n}.$$

This formula is useful in its own right: it shows $\pi \cot(\pi z)$ has simple poles at all points of $\mathbb{Z}$ with residue one. This property can be used, for example, to evaluate $\zeta(2k) = \sum_{1}^{\infty} 1/n^{2k}$ and other similar sums by the residue calculus — see the examples in Chapter 1.

We note that the product formula for $\sin(z)$ can also be used to prove $\zeta(2) = \pi^2/6$, by looking at the coefficient of z^2 on both sides of the equation. Similarly, by equating the coefficients of z^4, z^6, etc., the sine formula shows that $\sum_{a<b} 1/(ab)^2 = \pi^4/5!$, $\sum_{a<b<c} 1/(abc)^2 = \pi^6/7!$, etc., so with some more work it can be used to evaluate $\zeta(2k)$. For example, we have

$$\zeta(4) = \left(\sum \frac{1}{a^2} \right) \left(\sum \frac{1}{b^2} \right) - 2 \left(\sum_{a<b} \frac{1}{(ab)^2} \right) = \frac{\pi^4}{36} - \frac{\pi^4}{60} = \frac{\pi^4}{90}.$$

Translation and duplication formulas. Many of the basic properties of the sine and cosine functions can be derived from the point of view of the uniqueness of an odd entire function with zeros at $\mathbb{Z}\pi$. For example, one can prove the equations $\sin(z + \pi) = -\sin(z)$ and $\sin(2z) = 2\sin(z)\sin(z + \pi/2)$ by first observing that they hold up to a factor of $\exp(az+b)$, as a consequence of the fact that both sides have the same zero sets.

Boundaries. In preparation for the study of the Γ function, we remark more generally that we have a group homomorphism

$$\delta : \mathcal{M}(\mathbb{C})^* \to \mathcal{M}(\mathbb{C})^*$$

given by

$$\delta(f) = f(z+1)/f(z).$$

Its kernel is the functions that satisfy $f(z) = f(z+1)$. These are all of the form $f(z) = g(\exp(2\pi i z))$, where $g(z)$ is meromorphic.

   We emphasize that $\delta(fg) = \delta(f)\delta(g)$. The simplest examples are the exponentials:

$$\delta(a^z) = a.$$

The Γ function arises when we progress to linear polynomials and try to solve $\delta(f) = z$. Because we already know how to solve $\delta(f) = a$, it suffices to solve this new equation up to a constant, and then correct by an exponential factor.

The gamma function. We now turn to a discussion of the extension of the factorial function $n \mapsto n!$ to the complex numbers. We will describe three approaches to $\Gamma(z)$, involving (i) infinite products, (ii) binomial coefficients and (iii) the Mellin transform, respectively.

Half of a trigonometric function. We begin by studying the canonical product associated to the *negative integers*: namely

$$G(z) = \prod_1^\infty \left(1 + \frac{z}{n}\right) e^{-z/n}.$$

It has half the terms that appear in the factorization of $\sin(\pi z)$; indeed, we have

$$G(z)G(-z) = \frac{\sin(\pi z)}{\pi z}. \tag{3.5}$$

Euler's constant. We have $G(0) = 1$, but what is $G(1)$? To answer this, we define *Euler's constant* by:

$$\gamma = \lim_{n \to \infty} (1 + 2 + \cdots + 1/n) - \log(n+1).$$

This expression gives the limiting error obtained when one approximates $\int_1^{n+1} dx/x$ by n unit rectangles lying above the graph of $y = 1/x$. (The sum is finite because these areas can all be slid horizontally to lie disjointly inside a fixed rectangle of base one.)

Then, using the fact that $(1+1)(1+1/2) \cdots (1+1/n) = (n+1)$, we find:

Theorem 3.21 *We have $G(1) = \exp(-\gamma)$.*

Functional equation. The functions $G(z-1)$ and $zG(z)$ have the same zeros. How are they related? By Hadamard's theorem, we have $G(z-1) = z\exp(az+b)G(z)$ for a, b. In fact:

Theorem 3.22 *We have $G(z-1) = ze^\gamma G(z)$.*

Proof. We only know the value of $G(z)$ at 0 and 1, making a direct calculation of the exponential factor slightly tricky. Instead, we will compare the derivative of both sides of the equation, or rather the logarithmic derivative, which is easier to compute.

The logarithmic derivative of $zG(z)$ is given by:

$$\frac{1}{z} + \sum_1^\infty \frac{1}{z+n} - \frac{1}{n},$$

while the logarithmic derivative of $G(z-1)$ is just

$$\sum_1^\infty \frac{1}{z+n-1} - \frac{1}{n} = \frac{1}{z} - 1 + \sum_1^\infty \frac{1}{z+n} - \frac{1}{n+1}.$$

Since $\sum_1^\infty (1/n - 1/(n+1))$ telescopes to 1, these series are equal, and hence $G(z-1)$ and $zG(z)$ agree up to a constant. The value of this constant is determined by setting $z = 1$. ∎

Correcting $1/G(z-1)$. We are aiming to define a function with $\delta(f) = z$. By the preceding result, we have

$$\delta(1/G(z-1)) = \frac{G(z-1)}{G(z)} = ze^\gamma.$$

Now recall that

$$\delta(e^{-\gamma z}) = e^{-\gamma}.$$

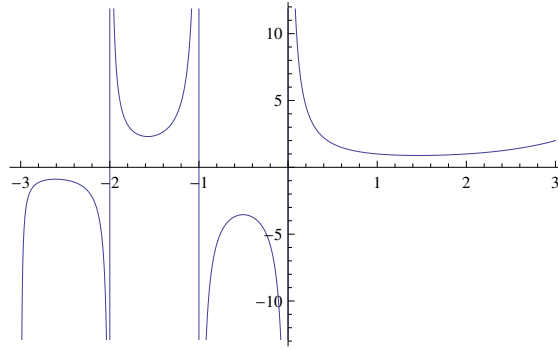Thus the product of these two functions satisfies $f(z+1) = zf(z)$.

Figure 3.2. The Γ function.

The Γ function. Motivated by the reasoning above, we now define

$$\Gamma(z) = \frac{e^\gamma}{e^{\gamma z}G(z-1)} = \frac{1}{ze^{\gamma z}G(z)}.$$

Here we have used the equation $G(z-1) = ze^\gamma G(z)$ to rewrite the denominator. This equation is sometimes written more explicitly as

$$\Gamma(z) = \frac{e^{-\gamma z}}{z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{z/n}. \tag{3.6}$$

Its most important properties are given by:

Theorem 3.23 *We have $\Gamma(z+1) = z\Gamma(z)$ and $\Gamma(1) = 1$.*

Proof. By Theorem 3.22 we have $G(z-1) = ze^\gamma G(z)$, and thus

$$\Gamma(z) = \frac{1}{e^{\gamma(z-1)}G(z-1)} = (z-1)\Gamma(z-1),$$

so $z\Gamma(z) = \Gamma(z+1)$; and $\Gamma(1) = 1$ since $G(1) = \exp(-\gamma)$. $\blacksquare$

We can then deduce, using this result and properties of $G(z)$:

1. $\Gamma(n+1) = n!$ for $n \geq 0$; equivalently:

2. $n! = n\Gamma(n)$;

3. $\Gamma(z)$ has no zeros; and

4. $\Gamma(z)$ has simple poles at $0, -1, -2, -3, \ldots$, and nowhere else.

5. In particular $\Gamma(z) > 0$ on the positive real axis.

6. Since
$$\Gamma(z-n) = \frac{\Gamma(z+1)}{(z-n)\cdots(z-1)z},$$
we find that
$$\mathrm{Res}_{-n}(\Gamma(z)) = \frac{(-1)^n}{n!}.$$

7. From (3.5) or (3.6) we obtain *Euler's supplement*
$$\frac{\pi}{\sin(\pi z)} = \Gamma(z)\Gamma(1-z).$$

8. What is the value of $(-1/2)!$? By the equation above, we have which implies, for example:
$$\Gamma(1/2) = \sqrt{\pi}.$$

9. Variants of the sine formula are:
$$\Gamma\left(\frac{1}{2}+z\right)\Gamma\left(\frac{1}{2}-z\right) = \frac{\pi}{\cos\pi z} \quad \text{and} \quad \Gamma(z)\Gamma(-z) = -\frac{\pi}{z\sin(\pi z)}.$$

10. Using the fact that $\Gamma(\bar{z}) = \overline{\Gamma(z)}$, we find from this last that
$$|\Gamma(iy)|^2 = \frac{\pi}{y\sinh(\pi y)}.$$

11. We also note that by the location of its zeros and poles, $\Gamma(z)$ has constant sign on each interval of the form $(n, n+1)$; in fact the sign is positive for $n > 0$ and $(-1)^n$ for $n < 0$.

95

12. Unlike the functions $\sin(z)$ and $\cos(z)$, the behavior of $\Gamma(z)$ is controlled by its behavior on the real axis. Namely, we have:

$$|\Gamma(z)| \le |\Gamma(\operatorname{Re} z)| \qquad (3.7)$$

for all z. In this way it behaves like $\exp(z)$. For the proof, just observe that the same inequality holds for each term in the product formula (3.6).

Characterizing $\Gamma(z)$. Equation (3.7) gives us good estimates on $\Gamma(z)$: it is *uniformly bounded* in any region of the form $0 < a < \operatorname{Re}(z) < b$, and for $-n < \operatorname{Re} z < b$ the function $z(z+1)\cdots(z+n)\Gamma(z)$ is uniformly bounded.

In fact these properties are enough to characterize the Γ function.

Note that $\exp(2\pi i z)\Gamma(z)$ satisfies the same functional equation as $\Gamma(z)$, and its reciprocal still has order 1; but it is no longer bounded in the vertical strip $[1,2] \times (-\infty, \infty)$.

Theorem 3.24 (Wielandt) *If $F(z+1) = zF(z)$ for $\operatorname{Re} z > 0$, $F(1) = 1$ and $F(z)$ is bounded on the strip $\{z : \operatorname{Re} z \in [1,2]\}$, then $F(z) = \Gamma(z)$.*

Proof. The functional equation allows one to extend $F(z)$ to a meromorphic function on the whole plane, whose poles and their residues agree with those of $\Gamma(z)$. Thus $G(z) = F(z) - \Gamma(z)$ is entire, $G(0) = G(1) = 0$ and $G(z+1) = zG(z)$. Our boundedness assumptions now imply that $G(z) = G(z+1)/z$ is bounded in the strip $S = \{\operatorname{Re} z \in [0,1]\}$, since it has a removable singularity at $z = 0$. Thus $H(z) = G(z)G(1-z)$ is also bounded in S. The functional equation for G implies $H(z+1) = -H(z)$ (as in the sine formula), and thus H is a constant, which must be zero. ∎

A duplication formula. As an application one can prove 'multiple angle' formulas for $\Gamma(nz)$. The simplest is:

Corollary 3.25 *We have $2\sqrt{\pi}\,\Gamma(2z) = 2^{2z}\Gamma(z)\Gamma(z+1/2)$.*

Proof. Let $F(z) = \Gamma(z/2)\Gamma(z/2 + 1/2)$. Then we have

$$F(z+1) = \Gamma(z/2 + 1/2)\Gamma(z/2 + 1) = zF(z)/2.$$

So if write instead $F(z) = 2^z\Gamma(z/2)\Gamma(z/2 + 1/2)$, then $F(z+1) = zF(z)$. Since $F(z)$ is also bounded for $\operatorname{Re} z \in [1,2]$, we have $F(z) = C\Gamma(z)$ for some C; and indeed, $C = F(1) = 2\Gamma(1/2) = 2\sqrt{\pi}$. The result above now follows, upon replacing z with $2z$. ∎

96

The integral representation: the Mellin transform. We now turn to a second motivation for introducing the Γ function.

Theorem 3.26 *For* $\mathrm{Re}(z) > 0$, *we have*

$$\Gamma(z) = \int_0^\infty e^{-t} t^z \frac{dt}{t}.$$

In other words, $\Gamma(z)$ is the *Mellin transform* of the function e^{-t} on $\mathbb{R}_+$, which is the restriction of the additive character $e : \mathbb{R} \to \mathbb{C}^*$.

Proof. Apply the uniqueness theorem above. ∎

Gauss sums. The Mellin transform is an integral against characters $\chi : \mathbb{R}^* \to \mathbb{C}^*$ (given by $\chi(t) = t^z$), and as such it can be compared to the Fourier transform (for the group $\mathbb{R}$ under addition) and to Gauss sums. Indeed the Gauss sum

$$\sigma(\chi) = \sum_{(n,p)=1} \chi(n) \exp(2\pi i n / p),$$

which pairs a multiplicative character $\chi : (\mathbb{Z}/p)^* \to S^1$ with the additive character $\exp(2\pi i n / p)$, is the analogue of the Gamma function for the group $(\mathbb{Z}/p)^*$.

Gauss's formula. The formulas above for the Γ function might not be easy to discover from scratch. Here is a formula that could also be taken as a simpler definition of the Gamma function.

Theorem 3.27 *We have*

$$\Gamma(z) = \lim_{n \to \infty} \frac{n! n^z}{z(z+1) \cdots (z+n)}.$$

As motivation, we note that for integers z we have $z! = \lim n^z \binom{z+n}{n}^{-1}$, which is consistent with the formula above if we multiply both sides by z. Indeed, formally we have

$$\binom{z+n}{n} = \frac{(z+1) \cdots (z+n)}{n!} \approx \frac{(n+1) \cdots (n+z)}{z!} \approx \frac{n^z}{z!},$$

and solving for $z! = z\Gamma(z)$ and taking the limit gives the formula above.

Proof. By definition, we have:

$$\frac{1}{\Gamma(z)} = \lim_{n\to\infty} ze^{\gamma z} \prod_{k=1}^{n} \left(1 + \frac{z}{k}\right) e^{-z/k}.$$

Now $\gamma - 1 - 1/2 - \cdots - 1/n \approx -\log(n)$, and $(1 + z/k) = (k + z)/k$, so this becomes

$$\frac{1}{\Gamma(z)} = \lim_{n\to\infty} \frac{z(z+1)\cdots(z+n)n^{-z}}{n!},$$

which gives the formula above. ∎

Stirling's formula. Here is a very brief introduction to the method of steepest descent, with the aim of explaining Stirling's formula: as $s \to \infty$, we have:

$$\Gamma(s) \sim \sqrt{\frac{2\pi}{s}} \left(\frac{s}{e}\right)^s.$$

The idea is to model $\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt$ by the more easily understood Gaussian integral:

$$\int_{-\infty}^{\infty} e^{a - b(t-t_0)^2/2}\, dt = e^a \sqrt{2\pi/b},$$

which is itself proved using the identity

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right)^2 = \int e^{-r^2/2} r\, dr\, d\theta = 2\pi.$$

To this end we rewrite $\Gamma(s) = \int_0^\infty e^{\phi(t)}\, dt$, where

$$\phi(t) = -t + (s-1)\log t.$$

Then $\phi'(t) = -1 + (s-1)/t$ vanishes at $t_0 = s-1$, with $\phi''(t_0) = -(s-1)/t_0^2 = -1/(s-1)$. So we have

$$\phi(t) \approx \phi(t_0) + \phi''(t_0)(t - t_0)^2/2 = a - b(t - t_0)^2/2$$

with $a = 1 - s + (s-1)\log(s-1)$ and $b = 1/(s-1)$. This gives

$$\Gamma(s) \sim e^{1-s}(s-1)^{s-1}\sqrt{2\pi(s-1)}.$$

98

Using the fact that $(s-1)^s \sim s^s/e$ and $\sqrt{s-1} \sim \sqrt{s}$, this gives Stirling's formula.

As a simple check, we note that Stirling's formula implies:

$$\lim (n!/n^n)^{1/n} = 1/e.$$

This can be seen by interpreting $(1/n)(2/n)(3/n)\cdots(n/n)$ as the exponential of an integral, and using the fact that $\int_0^1 \log x \, dx = -1$.

Error term. To fully justify the calculation above, we need to check that the main term in Stirling's formula dominates the error, so we actually get the asymptotic behavior of $\Gamma(z)$. To this end, when writing the integral as $\exp(\phi(t))$, we observe that $\phi''(t) < 0$ for all value of t, i.e. it is a *concave function*. Thus the integrand behaves like $e^{-|t-t_0|}$ away from its maximum, which gives a small error that can be controlled.

Steepest descent, stationary phase, etc. The name *steepest descent* comes from the fact that the real axis passes through the saddle pass of the absolute value of the integrand at its critical point, and proceeds in the direction where the absolute value decreases fastest. Compare [MH, p. 438].

3.3 Meromorphic functions: The Mittag-Leffler theorem

We now turn to the problem of describing meromorphic functions with *prescribed principal parts*. This means we specify a sequence of *distinct points* $a_n \to \infty$, and a finite *Laurent tail*

$$p_n(z) = \frac{b_k}{(z-a_n)^k} + \cdots + \frac{b_1}{z-a_n}$$

at each point. (The values of k and $b_1, \ldots, b_k$ depend on n.) We say $p_n(z)$ is the principal part of a meromorphic function $f(z)$ if $f(z) - p_n(z)$ is holomorphic at $z = a_n$.

Theorem 3.28 *For any sequence of points $a_n \to \infty$ and principal parts $p_n(z)$, there exists a meromorphic function $f(z)$ with poles exactly at a_n, and with the prescribed principal parts.*

Proof. Let $q_n(z)$ be the polynomial given by truncating the power series for $p_n(z)$ so that $|p_n(z) - q_n(z)| < 2^{-n}$ when $|z| < |a_n|/2$. Then $f(z) = \sum(p_n(z) - q_n(z))$ is the desired function. ∎

Singly-periodic functions. (This is a hint of the theory of elliptic or doubly-periodic functions to come.)

Any meromorphic function satisfying $f(z+1) = f(z)$ has the form $f(z) = F(e^{2\pi i z})$ for some meromorphic function $F : \mathbb{C}^* \to \widehat{\mathbb{C}}$. This leads to formulas such as:

$$\sum_{-\infty}^{\infty} \frac{1}{(z-n)^2} = \frac{\pi^2}{\sin^2(\pi z)}.$$

Note that the right hand side is a sum of principal parts.

To verify this equation, first note that $|z - n|^2 \geq |\operatorname{Im} z|^2 + n^2 - O(1)$ for $0 \leq \operatorname{Re} z \leq 1$, and hence the sum on the left converges *uniformly* in this strip so long as $|\operatorname{Im} z| \geq 1$. Thus both sides tends to 0 as $|\operatorname{Im} z| \to \infty$. But their difference is holomorphic and periodic, and so it must vanish identically.

As usual, there is a connection to $\zeta(2)$: by evaluating the constant term in the Laurent series on both sides, we obtain the formula $\sum_1^\infty 1/n^2 = \pi^2/6$.

Cotangent revisited. As a more interesting example of a Mittag–Leffler sum, we can write:

$$\pi \cot(\pi z) = \frac{1}{z} + \sum_{n \neq 0} \frac{1}{z - n} + \frac{1}{n}.$$

Note that this sum would not converge with the correcting terms $1/n$. To verify this formula, note that $\pi \cot(\pi z)$ is the logarithmic derivative of $\sin(\pi z)/\pi = z \prod_{n \neq 0}(1 - z/n) \exp(z/n)$.

From Mittag–Leffler to Weierstrass. The Weierstrass Theorem — that there exists a function $g(z)$ with prescribed zeros — is a Corollary of the Mittag–Leffler theorem (which historically came later).

Indeed, to construct an entire function with zeros of multiplicities m_n at distinct points $a_n \to \infty$, we can simply first construct a meromorphic function $f(z)$ with principle part $p_n(z) = m_n/(z - a_n)$ at a_n, and then take $g(z) = \exp(\int f(z)\, dz)$. Since the residues of $f(z)$ are integers, the integral is only well-defined modulo $2\pi i\mathbb{Z}$, but this is sufficient to make its exponential well-defined.

> *...this topsy–turvy way of doing things (*die Dinge auf dem Kopf zu stellen*) should be not sanctioned by anyone who sees mathematics as something other than a disordered heap of mathematical results.*
>
> —A. Pringsheim, 1915.

(See [Re, p.131].)

3.4 Exercises

1. Prove that for any nonzero polynomial $p(z)$, the function $f(z) = e^z - p(z)$ has infinitely many zeros. (Hint: use Theorem 3.8).

2. Let $M(r) = \sup_{|z|=r} |f(z)|$ where $f : \mathbb{C} \to \mathbb{C}$ is an analytic function, not identically equal to zero. Suppose $M(r^2)^2 = M(r)M(r^3)$ for some $r > 0$, $r \neq 1$. Prove that $f(z) = az^n$ for some $a \neq 0$ and integer $n \geq 0$.

3. Prove that in general, the function $M(r)$ itself need not be convex. (Hint: show that for $f(z) = z^2 + iz + 1$, we have $M(r) = f(-ir) = 1 + r - r^2$ for all r sufficiently small.)

4. Prove there is no nonzero analytic function $f : \Delta \to \Delta$ with zeros at the points $a_n = 1 - 1/(n+1)$, $n = 1, 2, 3, 4, \ldots$. (Thus in contrast to $f : \mathbb{C} \to \mathbb{C}$, the zeros of a map $f : \Delta \to \Delta$ cannot be an arbitrary discrete set. Hint: consider $f(0)/B_n(0)$, where $B_n : \Delta \to \Delta$ is a proper map of degree n with zeros at $a_1, \ldots, a_n$.)

5. Formulate and prove an infinite product formula for $\sin^2(\sqrt{z})$.

6. Given an example of an entire function of order $1/4$. (Hint: recall that $\cos(\sqrt{z})$ has order $1/2$.)

7. Give an example of a (convergent) canonical product $f(z) = \prod_1^\infty (1 - z/a_n)$ that has order exactly 1.

8. Show that if $\tau \in \mathbb{H}$ is fixed then

$$\theta(z) = \sum_{n=-\infty}^{\infty} \exp(2\pi i n^2 \tau) \exp(2\pi i n z)$$

is an entire function of order 2.

9. Prove that for any sequences $a_n, b_n \in \mathbb{C}$ with $a_n \to \infty$, there exists an entire function such that $f(a_n) = b_n$. (Hint: write $f(z) = \sum g_n(z)$ where g_n is a polynomial constructed by induction, such that $g_n(a_i) = 0$ for $i < n$, $g_n(a_n) = b_n - \sum_1^{n-1} g_i(a_n)$, and $|g_n(z)| < 2^{-n}$ when $|z| < |a_n|/2$.)

10. Show that if $f, g \in \mathcal{O}(\mathbb{C})$ have no common zeros, then $(f, g) = (1)$; i.e. show there exist $r, s \in \mathcal{O}(\mathbb{C})$ such that $fr + gs = 1$.

    Hint: write $1/(fg) = F + G$, where F only has poles at the zeros of f and G has zeros only at the zeros of g. (Why is this possible?)

11. Let $\mathcal{O}(\mathbb{C})$ and $\mathcal{M}(\mathbb{C})$ be the rings of analytic and meromorphic functions on $\mathbb{C}$. Show that $\mathcal{O}(\mathbb{C})$ is integrally closed in $\mathcal{M}(\mathbb{C})$. (The means any $f \in \mathcal{M}(\mathbb{C})$ satisfying a monic polynomial $p(X) \in \mathcal{O}(\mathbb{C})[X]$ is actually in $\mathcal{O}(\mathbb{C})$.)

12. Is $\mathcal{M}(\mathbb{C})$ algebraically closed?

13. Fix $\rho \geq 0$ and let $\mathcal{O}_\rho(\mathbb{C})$ be the set of entire functions of order $\leq \rho$. Show that $\mathcal{O}_\rho(\mathbb{C})$ is a ring, and that $\mathcal{O}_\rho(\mathbb{C})$ is integrally closed in $\mathcal{M}(\mathbb{C})$.

14. Prove that any finitely-generated ideal $I \subset \mathcal{O}(\mathbb{C})$ is principal.

15. Let $I = (f_1, f_2, \ldots) \subset \mathcal{O}(\mathbb{C})$ be the ideal generated by the sequence of functions $f_n(z) = \sin(z/n)/z$. Prove that $I \neq \mathcal{O}(\mathbb{C})$ and I is not contained in any proper principal ideal. Conclude that $\mathcal{O}(\mathbb{C})$ is not a PID and that $\mathcal{O}(\mathbb{C})$ contains maximal ideals that are not of the form $(z - a)$.

16. Express $f(z) = \sum_{n=-\infty}^{\infty} (z - n)^{-4}$ in terms of trigonometric functions, and use your result to evaluate $\sum_1^\infty 1/n^4$.

17. Prove:
$$\frac{\coth(\pi)}{1^7} + \frac{\coth(2\pi)}{2^7} + \frac{\coth(3\pi)}{3^7} + \ldots = \frac{19\pi^7}{56700}.$$
    (Hint: you may use the fact that the residue of $\cot(\pi z) \coth(\pi z)/z^7$ at $z = 0$ is $-19\pi^6/14175$.)

18. Find the orders of the entire functions $\sin^2(\sqrt{z})$ and $\exp(\sin(z))$.

19. Let $f(z)$ be an entire function of finite order with simple zeros at the points $z = n + im$, $(n, m) \in \mathbb{Z}^2$. Show there are polynomials P and Q such that $f(z + 1) = e^{P(z)} f(z)$ and $f(z + i) = e^{Q(z)} f(z)$. Prove that at least one of P and Q is nonzero.

20. Show that for any $\alpha > 0$ we have:

$$\int_0^\infty \exp(-x^\alpha)\,dx = \alpha^{-1}\Gamma(\alpha^{-1}).$$

(This generalizes the formula $\int_0^\infty \exp(-x^2)\,dx = \sqrt{\pi}/2$.)

21. State and prove an asymptotic formula for $\Gamma'(x)$ as $x \to +\infty$ along the real axis. (In other words, find a more elementary function $A(x)$ such that $\Gamma'(x)/A(x) \to 1$.)

22. Prove or disprove: $\Gamma'(3/2) = 0$.

23. Show that the entire function $1/\Gamma(z)$ has order one, *but* there is no constant $C > 0$ such that $1/\Gamma(z) = O(\exp(C|z|))$.

24. Evaluate $\Gamma(1/3)\Gamma(2/3)$. Then find a formula for $\Gamma(3z)$ in terms of Γ at z, $z + 1/3$ and $z + 2/3$.

25. Prove that $\int_0^1 \log\Gamma(t)dt = \log\sqrt{2\pi}$, using the duplication formula for $\Gamma(2z)$.

26. Show that for any polynomial $p(z)$, there exists a meromorphic function $f(z)$ such that $f(z + 1) = p(z)f(z)$.

27. State and prove a necessary and sufficient condition for a meromorphic 1-form $\omega = \omega(z)\,dz$ on $\mathbb{C}$ to be the logarithmic derivative, $\omega = d\log f = f'(z)/f(z)\,dz$, of a meromorphic function $f(z)$.

28. Let $\langle F_p(z)\rangle$ be a sequence of analytic functions with $F_p(1) = 0$, with no other zeros, and with $F_p(z) \to 1$ uniformly on compact subsets of $\mathbb{C} - \{1\}$. Let $0 \neq a_n \to \infty$ in $\mathbb{C}$. Show there exist p_n such that $G(z) = \prod F_{p_n}(z/a_n)$ converges absolutely on $\mathbb{C}$, giving an analytic function with zeros at $\langle a_n\rangle$ and nowhere else.

29. Prove Wallis's formula

$$\frac{\pi}{2} = \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots}{1\cdot 3\cdot 3\cdot 5\cdot 5\cdot 7\cdots}.$$

(Hint: use the fact that $\sin(\pi/2) = 1$.)

30. Let f and g be nonconstant entire functions, and suppose that $f \circ g = f$. Prove that $g(z) = z + t$ for some $t \in \mathbb{C}$.

31. Give an example of an entire function of finite order ρ such that $(\log \log M(r))/\log(r)$ does not converge to ρ as $r \to \infty$.

32. (Arnold problem.) Estimate the average value of $\cos(x)^{100}$ by hand. Do not use a calculator. You do not need to give an error estimate. (Hint: use the fact that $(1 - x^2/100)^{100} \approx \exp(-x^2)$.)

33. Let $M(x) > 0$ be a continuous function on $\mathbb{R}$.

    (i) Prove there exists an entire function with $|f(x)| > M(x)$ for all $x \in \mathbb{R}$.

    (ii) Prove there exists an entire function with $0 < |g(x)| < M(x)$ for all $x \in \mathbb{R}$.

    (iii) Prove there does not exist an entire function with $|f(z)| > |z|$ for all $z \in \mathbb{C}$.

34. Show there is a constant $\alpha \in [0, 1]$ such that $\sum_1^n 1/(k + \alpha) = \log(n) + o(1)$.

35. Given two proofs that $|1/\sin(z)| \le |1/\sin(\operatorname{Re} z)|$ for all $z \in \mathbb{C}$, using (i) the addition formula for $\sin(z)$, and (ii) the product formula for $\sin(\pi z)$.

36. Suppose a_n, b_n are sequence of real numbers, both tending to infinity, and that the product $f(z) = \prod_1^\infty (z - b_n)^{-1} \exp(a_n z)$ converges to define a meromorphic function on $\mathbb{C}$. Prove that $|f(z)| \le f(\operatorname{Re} z)$ for all $z \in \mathbb{C}$.

37. Let $f(z)$ be an entire function with no zeros, such that $|f(z)| = O(\exp(\exp(|z|^n)))$. Prove that $f(z) = \exp(g(z))$ for some entire function $g(z)$ *of finite order*.

38. Prove that $\prod_2^\infty \cos(\pi/2^n) = (2/\pi)$.

4 Conformal mapping

We now turn to the theory of analytic functions as mappings. Here the dominant operation is composition, rather than addition or multiplication.

4.1 The Riemann mapping theorem

A conformal homeomorphisms between plane domains is an isomorphism in the category of Riemann surfaces. The first main result classifies these domains when they are simply-connected.

Theorem 4.1 (Riemann mapping theorem) *Let $U \subset \mathbb{C}$ be a simply connected domain, other than $\mathbb{C}$ itself. Then there exists a conformal homeomorphism $f : U \to \Delta$.*

The same result holds for any simply-connected region $U \subset \widehat{\mathbb{C}}$ such that $|\widehat{\mathbb{C}} - U| \geq 1$. Note that $\pi_1(U) = (1)$ iff $\widehat{\mathbb{C}} - U$ is connected, so the complement is either empty, one point or a nontrivial continuum.

Uniqueness. Since $\mathrm{Aut}(\Delta)$ is large, the Riemann map is not unique. It can be made unique by picking a point $p \in \Delta$ and requiring $f(p) = 0$ and $f'(p) > 0$.

Uniformization. The Riemann mapping theorem is a special case of the *uniformization* theorem; it shows that uniformization is true for *any simply–connected region* $U \subset \mathbb{C}$. We will also prove that uniformization holds for double–connected regions.

Examples of Riemann maps.

1. We first note that $(\Delta,) \cong (\mathbb{H}, i)$ by the Möbius transformation $A(z) = i(1 - z)/(1 + z)$. We also remark that $A : S^1 \to \widehat{\mathbb{R}}$ becomes, in angular coordinates on S^1,

$$ x = A(\theta) = \frac{2}{2i} \frac{e^{i\theta/2} - e^{-i\theta/2}}{e^{i\theta/2} + e^{-i\theta/2}} = \tan(\theta/2). $$

   This is the change of variables used in calculus to integrate rational functions of $\sin(\theta)$ and $\cos(\theta)$, by turning them in to rational functions of x.

2. We also remark that $(\Delta, 0) \cong (\Delta, p)$, and $(\mathbb{H}, i) \cong (\mathbb{H}, p)$, as can be seen using the Möbius transformations $B(z) = (z + p)/(1 + \overline{p}z)$ and $C(z) = az + b$ where $p = ai + b$.

3. Of course any region in $\widehat{\mathbb{C}}$ bounded by a circle (or line) is also isomorphic to Δ.

4. The pie-slice $P(\alpha) \subset \mathbb{H}$ defined by $0 < \arg(z) < \alpha < \pi$ is isomorphic to $\mathbb{H}$ by the map $f(z) = z^{\pi/\alpha}$.

5. The map $f(z) = z + 1/z$ sends the region $\Omega = \mathbb{H} - \overline{\Delta}$ to $\mathbb{H}$.

6. Let $L \subset \widehat{\mathbb{C}}$ be a *lune* – the region between two circular arcs. Apply a Möbius transformation so the vertices of L are at 0 and ∞; then after a rotation, we can assume $L = \{z \ : \ 0 < \arg z < \alpha\}$, and $f(z) = z^{\pi/\alpha}$ send L to $\mathbb{H}$. (This illustrates the idea that the Riemann map $U \cong \Delta$ is often constructed as a composition of several simple transformations.)

   The previous two examples were also lunes.

7. The map $f : \mathbb{H} \to \mathbb{C} - [0, \infty)$ given by $f(z) = z^2$ is conformal. This is an extreme example of a lune.

8. The map $f(z) = z/(1-z)^2 = z + 2z^2 + 3z^3 + \cdots$ send Δ bijectively to $\mathbb{C} - (-\infty, -1/4]$. Up to change of coordinates in domain and range, this is the same as the previous example. It is famous for its connection to the Koebe $1/4$–theorem and the Bierberbach conjecture.

9. The strip $S = \{z \ : \ 0 < \operatorname{Im} z < \pi\}$ is a lune with angle zero; it is sent to $\mathbb{H}$ by $f(z) = e^z$. The region between any two tangent circles is isomorphic to S by a Möbius transformation.

   This is related to the problem of extending the harmonic function $u(z) = 1$ for $\operatorname{Im} z > 0$ and 0 for $\operatorname{Im} z < 0$ from S^1 to Δ; we can take $u = \operatorname{Re} f(z)/\pi$ where $f : \Delta \to S$ sends ± 1 to the two ends of S.

10. Let us write $S = S_+$ and S_- according to the sign of $\operatorname{Re}(z)$. Note that $f(z) = e^z$ sends the imaginary axis to S^1. Thus f sends S_- to $\Delta \cap \mathbb{H}$, and sends S_+ to $\mathbb{H} - \overline{\Delta}$. Both of these regions are lunes. The map $g(z) = (e^z + e^{-z})/2 = \cosh(z)$ sends $f(S_-)$ to $-\mathbb{H}$ and $f(S_+)$ to $+\mathbb{H}$.

11. As a variation on the discussion above, we observe that $f(z) = \cos z$ maps the half-strip $T \subset \mathbb{H}$ defined by $0 < \operatorname{Re} z < \pi$ to $-\mathbb{H}$. Indeed, $\cos(z) = (e^{iz} + e^{-iz})/2$ is simply the composition of the map $z \mapsto e^{iz}$, which sends T to $\Delta \cap \mathbb{H}$, with the map $g(z) = (z + 1/z)/2$.

   If we reflect T through its sides, we obtain a tiling of $\mathbb{C}$. The map $\cos(z)$ sends half the tiles to $\mathbb{H}$ and the other half to $-\mathbb{H}$, as can be seen by Schwarz reflection.

12. One last trigonometric example: the map $f(z) = \sin(z)$ sends $T = \{z \in \mathbb{H} : 0 < \operatorname{Re} z < \pi/2\}$ to the first quadrant. As an alternative way to see this, use the fact that $\sin(iy) = \sinh(iy)$ and $\sin(\pi/2 + iy) = \cosh(y)$.

    Consequently $\sin^2(z)$ maps T to $\mathbb{H}$. This is consistent with the previous example, because $\sin^2(z) = (1 - \cos(2z))/2$.

13. A compact set $K \subset \mathbb{C}$ is *full* if K and $\mathbb{C} - K$ are both connected. This is equivalent to the requirement that $\widehat{\mathbb{C}} - K$ is simply–connected.

    We note that Δ^* is isomorphic to $\mathbb{C} - \overline{\Delta}$ by $z \mapsto 1/z$. Thus Riemann mapping theorem shows there is a conformal map

    $$\mathbb{C} - \overline{\Delta} \to \mathbb{C} - K$$

    for any compact, full set $K \subset \mathbb{C}$ with $|K| > 1$.

Conformal radius. The Riemann mapping theorem shows the *hyperbolic metric* $\rho_U(z)\,|dz|$ is well–defined on any simply–connected region $U \subset \mathbb{C}$ (other than $\mathbb{C}$ itself). It also shows there is a natural conformal invariant, the *conformal radius* $R(U, p)$, characterized by the following properties:

1. If $f : (U, p) \to (V, q)$ is a conformal isomorphism, then

$$R(V, q) = |f'(p)| \cdot R(U, p); \quad \text{and}$$

2. We have $R(\Delta, 0) = 1$.

Indeed, it is easy to check that $R(U, p) = 2/\rho_U(p)$. By the Schwarz lemma, it follows that:

$$U \subset V \implies R(U, p) \le R(V, p),$$

with equality only when $U = V$. Since $R(\Delta(r), 0) = r$, we have:

$$R(U, p) \ge d(p, \partial U).$$

Examples. We have $R(\mathbb{H}, i) = 2$, and $R(\mathbb{C} - (-\infty, 0], 1) = 4$.

By the Koebe 1/4 theorem, to be proved below, this last example is extremal; we have

$$R(U, p) \le 4d(p, \partial U)$$

for any simply–connected region U.

Univalence and compactness. To begin the proof of the Riemann mapping theorem, we recall a few fundamental facts.

Let us say $f : U \to \mathbb{C}$ is *univalent* if f is injective and analytic. By injectivity, its inverse is analytic, and f provides a homeomorphism between U and $V = f(U)$. (It need *not* provide a homeomorphism between their closures.) We then have:

Lemma 4.2 *Let $f_n : U \to \mathbb{C}$ be a sequence of univalent maps, converging locally uniformly to $f : U \to \mathbb{C}$. If f is nonconstant, it too is univalent.*

Proof. Suppose $f(a) = f(b)$ with $a \neq b$, but f is nonconstant. Then $g(z) = f(z) - f(b)$ has a zero at $z = a$ and $g(z) = \lim g_n(z) = f_n(z) - f_n(b)$. It follows by Rouché's theorem that for $n \gg 0$ we have $g_n(a_n) = 0$ with $a_n \to a$. But then $f_n(a_n) = f_n(b)$, contrary to univalence of f_n. ∎

Lemma 4.3 *The space of analytic maps $f : U \to \overline{\Delta}$ is compact in the topology of locally uniform convergence.*

Proof. This is a consequence of Arzela-Ascoli and the fact that $|f'(z)| \leq 1/d(z, \partial U)$ by Cauchy's integral formula. ∎

Lemma 4.4 *If $U \subset \mathbb{C}$ is simply-connected and $0 \notin U$, then there exists a univalent map $f : U \to \mathbb{C}$ such that $f(z)^2 = z$.*

Proof of the Riemann mapping theorem. Pick $p \in U$ and let $\mathcal{F}$ denote the space of univalent maps $f : (U, p) \to (\Delta, 0)$. For convenience, we will also impose the condition that $f'(p) > 0$.

We first observe that $\mathcal{F}$ is nonempty. This is clear if U is bounded. It is also easy to see when $\overline{U} \neq \mathbb{C}$, for in the case the image of U under the Möbius transformation $1/(z - q)$, $q \notin \overline{U}$, is bounded.

For the general case, recall we assume that $U \neq \mathbb{C}$. Pick $q \in \mathbb{C} - U$. Since U is simply–connected, there is a single–valued branch of $f(z) = \sqrt{z - q}$ defined U. Let $V = f(U)$. Since $V \cap -V = \emptyset$, we have $\overline{V} \neq \mathbb{C}$, and we are reduced to the case treated above.

Note that if $B(p, r) \subset U$, then by the Schwarz lemma we have $|f'(p)| \leq 1/r$ for all $f \in \mathcal{F}$. Thus $M = \sup_{\mathcal{F}} f'(p)$ is finite. By compactness, the bound M is actually realized: there exists an $f \in \mathcal{F}$ such that $f'(p) = M$.

To complete the proof, we need to show $f(U) = \Delta$. If not, there is an $a \in \Delta - f(U)$. We will construct a proper map $B : (\Delta, 0) \to (\Delta, 0)$ of degree 2, such that the unique critical *value* of $B(z)$ is at $z = a$, and $B'(0) > 0$. Moreover, if we let $V = B^{-1}(U)$, then $B : V \to U$ is a covering map. Since U is simply–connected, V has a unique component $V_0 Q$ containing zero and $B : V_0 \to U$ is an isomorphism. Thus

$$g = B^{-1} \circ f \in \mathcal{F}.$$

But by the Schwarz Lemma (or a direct computation), $B'(0) < 1$, so $g'(p) > f'(p)$, a contradiction. Thus f has surjective and we are done.

It remains to construct B. This is easy. Let $s(z) = z^2$ and choose $A_1 \in \text{Aut}(\Delta)$ such that $A_1(0) = a$. Then a is the unique critical value of $A_1(s(z))$. The same is true for $B = A_1 \circ s \circ A_2$, for any $A_2 \in \text{Aut}(\Delta)$. Choosing A_2 such that $B(0) = 0$ and $B'(0) > 0$, we are done. (An explicit formula can also be given.) ∎

Boundary behavior and exotic domains. It is now convenient to reverse domain and range and consider a Riemann map $f : \Delta \to U \subset \mathbb{C}$. We will investigate the question of extending f at least to a *continuous* map on ∂U.

To fully appreciate the power of the Riemann mapping theorem, it is useful to contemplate how exotic target of a conformal $f : \Delta \to U$ might be. For example, every if U is a Jordan domain, its boundary might be fractal (like the Koch snowflake curve) or it might even have positive area. On the other hand, the boundary might contain features like a comb or the topologists sine curve, that render some boundary points inaccessible.

The next few results show that nevertheless, the Riemann map *does* extend continuously to $S^1 = \partial \Delta$ whenever ∂U is tame enough for this to happen.

Jordan domains. Here is one of the main results. Recall that a compact set $J \subset \mathbb{C}$ is a *Jordan curve* if it is homeomorphic to a circle.

Theorem 4.5 *If ∂U is a Jordan curve, then any conformal map $f : \Delta \to U$ extends to a homeomorphism between $\overline{\Delta}$ and $\overline{U}$.*

Length–area. Here is the basic argument, exploiting the fact that f stretches all directions by the same factor, $|f'(z)|$.

Let $R(a, b) = [0, a] \times [0, b] \subset \mathbb{C}$. For each $y \in [0, b]$, let $L_y = [0, a] \times \{y\}$. Then we have

Lemma 4.6 *For any conformal map $f : R(a, b) \to U \subset \mathbb{C}$, there exists a $y \in [0, b]$ such that*

$$\text{length}(f(L_y))^2 \leq (a/b)\,\text{area}(U).$$

In other words, the area of $f(U)$ controls the length of $f(L_y)$.

Proof of Lemma 4.6 Use Cauchy–Schwarz. We have

$$\left(\int_0^b \text{length}(f(L_y))\, dy \right)^2 = \left(\int_{R(a,b)} |f'(z)|\, |dz|^2 \right)^2$$
$$\leq \int 1^2 \int |f'|^2 = ab\,\text{area}(U).$$

Dividing both sides by b^2, we find that the square of the average length of $f(L_y)$ is bounded by $a/b\,\text{area}(U)$, and the average bounds the minimum. ∎

Stretching. The length–area method strongly depends on conformality. The map $f(x, y) = (Ax, A^{-1}y)$ preserves the area of the unit square, but it sends all horizontal lines in $R(1, 1)$ to lines of length A.

Jordan curves. Here is a basic fact about a Jordan curve $J \subset \mathbb{C}$. Given $a, b \in J$, let $[a, b] \subset J$ be the subarc joining these two points with smallest diameter. Then $\text{diam}[a, b] \to 0$ if $|a - b| \to 0$.

Proof of Theorem 4.5. Given a point $z \in \partial\Delta$, map Δ to an infinite strip, sending z to one end. Then there is a sequence of disjoint squares in the strip tending towards that end. The images of these squares have areas tending to zero, so there are cross-cuts $\gamma_n \subset \overline{\Delta}$ enclosing z such that $\text{length}(f(\gamma_n)) \to 0$. Thus the endpoints of $f(\gamma_n)$ converge to points $a_n, b_n \in J$, with $z \in [a_n, b_n]$ for $n \gg 0$. Since $\text{diam}[a_n, b_n] \to 0$, the disk bounded by $f(\gamma_n) \cup [a_n, b_n]$ shrinks to a point p, and this implies $f(z_n) \to p$ whenever $z_n \to z$ in Δ.

Consequently f extends to a continuous map $S^1 \to J$ with $f(z) = p$. We claim f is injective. Indeed, if $f(a) = f(b)$ for $a \neq b$ in S^1, then we can join a, b by an arc γ in Δ which maps to an embedded loop $f(\gamma)$ meeting J in the single point $p = f(a) = f(b)$. Then one of the two components of $J - \{a, b\}$ must also map to p, since $f|\Delta$ is a homeomorphism. But if f is constant along a subarc of J, then f is constant by Schwarz reflection, a contradiction. ∎
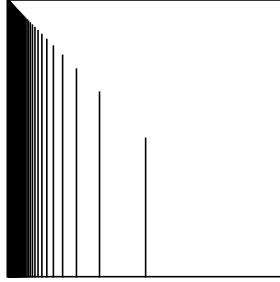
Figure 4.1. A comb in the boundary destroys local connectivity.

Local connectivity. The key point in the proof above is the following property of a Jordan domain $\overline{U}$: if $a, b \in \partial U$ are joined by a short arc $\alpha \subset U$, then the disk cut off by α has small diameter. This principle is called 'short dam, small lake'.

A compact set K is *locally connected* if for every open set $U \subset K$ and $x \in U$ there is a connected open set with $x \in V \subset U$.

For example, if we delete from the open unit square $S = (0, 1) \times (0, 1)$ the *comb* of intervals $\{1/n\} \times [0, 1 - 1/n]$, $n = 2, 3, 4, \ldots$, the boundary of the resulting simply–connected region U is *not* locally connected. See Figure 4.1.

Continuous extension. The argument in the proof of Theorem 4.5 furnishes short cross-cuts for any Riemann map, so it also shows:

Theorem 4.7 *The Riemann map extends continuously to S^1 iff ∂U is locally connected.*

Finally, using Schwarz reflection one has the useful statement:

Theorem 4.8 *The Riemann map extends analytically across an arc $\alpha \subset S^1$ if $f(\alpha)$ is a real-analytic arc in ∂U.*

The converse is almost, but not quite, true — the map $f : \mathbb{H} \to \mathbb{C}$ given by $f(z) = z^2$ extends analytically over the real axis, but the image has an endpoint.

Remark. The length–area method also easily shows:

Theorem 4.9 *Any Riemann mapping $f : \Delta \to \mathbb{C}$ has radial limits almost everywhere; that is, $\lim_{r \to 1} f(re^{i\theta})$ exists for almost every θ.*

111

Proof. We may assume the image of f is bounded; then $\int |f'|^2 < \infty$, which implies by Cauchy-Schwarz that $\int |f'| < \infty$ along almost every ray. ∎

More generally, it is known that any bounded analytic function has radial limits a.e. By [Ko], not much more can be said — given a $G_{\delta\sigma}$ set $A \subset S^1$ of measure zero, there exists a bounded analytic function that fails to have radial limits exactly on the set A. For *Riemann mappings*, radial limits exist outside a very small set (which is often empty) — the exceptional set A has capacity zero and in particular H. $\dim(A) = 0$.

Harmonic functions and conformal maps. Recalling that if u is harmonic and f is analytic then $u \circ f$ is also harmonic, it is now a simple matter to solve the Dirichlet problem (at least implicitly) on any Jordan domain.

Theorem 4.10 *For any Jordan domain $U \subset \mathbb{C}$, there exists a unique map $P : C(\partial U) \to C(\overline{U})$ such that $Pu|\partial U = u$ and Pu is harmonic in U.*

Proof. Let $f : \overline{\Delta} \to \overline{U}$ be the Riemann map, and let $Pu = P_0(u \circ f) \circ f^{-1}$, where P_0 is given by the Poisson kernel on the unit disk. ∎

It is interesting to note that analogous theorems do hold in $\mathbb{R}^n$, where almost no conformal mappings are available. The more general results concern *elliptic boundary value problems* and require new techniques.

Annuli. A domain $U \subset \widehat{\mathbb{C}}$ is an *annulus* if $\widehat{\mathbb{C}} - U = K_1 \sqcup K_2$ has exactly two connected components. Equivalently, $\pi_1(U) \cong \mathbb{Z}$. Note: in general $H_0(S^2 - U) \cong H_1(U)$; this is an example of *Alexander duality*.

The *standard annulus* is $A(R) = \{z : 1 < |z| < R\}$. Other examples of annuli are $\mathbb{C}^*$ and Δ^*. Exercise: none of these annuli are isomorphic.

Theorem 4.11 *The universal cover of any annulus $U \subset \widehat{\mathbb{C}}$ is isomorphic to $\mathbb{C}$ or $\mathbb{H}$.*

Proof. We may suppose the two complementary components of U contain 0 and ∞ respectively. Then the map $U \hookrightarrow \mathbb{C}^*$ induces an isomorphism $\pi_1(U) \cong \pi_1(\mathbb{C}^*) \cong \mathbb{Z}$. The universal cover of $\mathbb{C}^*$ is given by $\pi : \mathbb{C} \to \mathbb{C}^*$, $\pi(z) = e^z$, so

$$V = \log U = \pi^{-1}(U) \subset \mathbb{C}$$

gives the universal cover of U. By the Riemann mapping theorem, V itself is isomorphic to $\mathbb{H}$ or $\mathbb{C}$. ∎

Remark. Clearly $U \cong \Delta^*$ or $\mathbb{C}^*$ iff one or both of its complementary components are singletons. This can be proved e.g. by applying the removable singularities theorem to $f : \Delta^* \to U$.

Theorem 4.12 *Any annulus U is isomorphic to $\mathbb{C}^*$, Δ^*, or $A(R)$ for a unique $R > 1$.*

Proof. Let $g : V \to V$ denote a generator for $\pi_1(U) \cong \mathbb{Z}$ acting on its universal cover. If $V \cong \mathbb{C}$ then g is conjugate to $g(z) = z + 1$ and we have $U \cong \mathbb{C}/\langle g \rangle \cong \mathbb{C}^*$ by the map $\pi(z) = \exp(2\pi i z)$. The same reasoning shows $U \cong \Delta^*$ if $V \cong \mathbb{H}$ and g is parabolic.

Otherwise $V \cong \mathbb{H}$ and we can assume $g(z) = \lambda z$, $\lambda > 1$. This means the core curve of U is a geodesic of length $L = \log \lambda$. Then $\pi(z) = z^\alpha$ maps V to $A(R)$ if we choose α correctly. We want π to map $[1, \lambda]$ onto the unit circle, so we want $\pi(\lambda) = \lambda^\alpha = \exp(\alpha L) = 1$; so we take $\alpha = -2\pi i/L$. Note that this is a purely imaginary number, so π sends $\mathbb{R}_+$ to the unit circle and $\mathbb{R}_-$ to a circle of radius $R = (-1)^\alpha = \exp(\alpha \pi i) = \exp(2\pi^2/L)$. (We put a minus sign in α so that $R > 1$.) $\blacksquare$

Modulus of an annulus. Two natural invariants of an annulus are the number R such that $U \cong A(R)$, and the hyperbolic length L of its core geodesic. These are related by $\log R = 2\pi^2/L$ as we have just seen. For a direct proof, consider the metric $|dz|/|z|$ on $A(R)$. This makes it into a right cylinder with height over circumference given by $h/c = \log R/2\pi$. On the other hand, the same metric makes $\mathbb{H}/(z \mapsto e^L z)$ into a right cylinder with $h/c = \pi/L$. Equating these two expressions gives the desired relation.

The quantity h/c is often called the *modulus* of A, written $\mathrm{mod}(A)$.

The space of all Riemann mappings. It is traditional to denote the set of univalent mappings $f : \Delta \to \mathbb{C}$ such that $f(0) = 0$ and $f'(0) = 1$ by S, for *schlicht* (plain, simple). So

$$f(z) = z + a_2 z^2 + z_3 z^3 \cdots$$

We give S the topology of uniform convergence on compact sets. We will soon show:

Theorem 4.13 *The space S is compact.*

As a consequence, the coefficients a_n are bounded. The deeper Bieberbach Conjecture, now a theorem, asserts that $|a_n| \leq n$. This is also more than sufficient to show that S is compact.

Remark: no nesting. There are no nesting relations among the images of $f \in S$; that is, the Schwarz lemma implies that if $f(\Delta) = g(\Delta)$ then $f = g$. So if $f(\Delta)$ contains some points outside the unit disk, it must also omit some points inside the unit disk.

Remark: an extremal map. The map $f(z) = \sum n z^n = z/(1-z)^2$ lies in S and maps the unit disk to the complement of the arc $(-\infty, -1/4]$. This map simultaneously maximizes $|a_n|$ over all maps in S, so it is in some sense the univalent map farthest from the identity.

The map $f(z)$ is also closely related to the map $g_\alpha(z) = z^\alpha$ on the upper halfplane, at $\alpha = 2$. The map sends $\mathbb{H}$ to the complement of $[0, \infty)$, sending i to -1. Up to a Möbius transformation in domain and range, f and g_2 are the same. Clearly $g_{2+\epsilon}$ is not injective, so this is on the boundary of S.

Maps to the outside. It is somewhat easier to start with the space Σ of all univalent maps $F : \widehat{\mathbb{C}} - \overline{\Delta} \to \widehat{\mathbb{C}}$ with $F(\infty) = \infty$, normalized so that their Laurent series has the form:

$$F(z) = z + \sum_{n=1}^{\infty} \frac{b_n}{z^n}.$$

Since F is an open mapping near ∞, the complement of its image is a compact set $K(F) \subset \mathbb{C}$.

Example. If we set $b_1 = 1$, and $b_n = 0$ for $n \geq 2$, we get $F(z) = z + 1/z$ which satisfies $K(F) = [-2, 2]$.

Theorem 4.14 *The area of $K(F)$ is given by $A = \pi(1 - \sum n|b_n|^2)$.*

Proof. Integrate $\overline{F}dF$ over the unit circle and observe that, since $|dz|^2 = (-i/2)d\bar{z}\,dz$, the area A of $K(F)$ is given by:

$$A = -\frac{i}{2} \int_{S^1} \overline{F}dF = -\frac{i}{2}\left(1 - \sum n|b_n|^2\right) \int_{S^1} \frac{dz}{z} = \pi\left(1 - \sum n|b_n|^2\right).$$

■

Corollary 4.15 *We have $\sum n|b_n|^2 \le 1$. In particular, we have $|b_n| \le 1/\sqrt{n}$.*

Corollary 4.16 *The space Σ is compact.*

Proof. This follows from the fact that $|b_n| \le 1$. So if we fix $R > 1$, then for $|z| \ge R$ we have

$$|F(z) - z| \le \sum |b_n|/R^n \le \frac{R}{R-1} < \infty.$$

This uniform bound implies every sequence has a subsequence converging uniformly on compact sets. As usual, univalence is also preserved in the limit. ∎ Note: a small modification of this argument shows that

$|F(z) - z| = O(1)$ for all $F \in \Sigma$, since the image of F on the circle of radius 2 encloses the image of F on any smaller circle.

Remark. Little is known about optimal bounds for $|b_n|$ over Σ. It is conjectured that $|b_n| = O(1/n^{3/4})$; see [CJ].

We can now use the statement $|b_1| \le 1$ to prove the first case of the Bieberbach conjecture. Note that the case of equality, we must have all other $b_n = 0$, and hence up to a rotation, $F(z) = z + 1/z$ and $K(F) = [-2, 2]$.

Theorem 4.17 *For all $f \in S$ with have $|a_2| \le 2$.*

Proof. We first note that $F(z) = 1/f(1/z) + a_2$ is in Σ. Indeed, we find:

$$
\begin{aligned}
1/f(1/z) &= \frac{z}{1 + a_2/z + a_3/z^2 + \cdots} \\
&= z\left(1 - (a_2/z + a_3/z^2 + \cdots) + (a_2/z + a_3/z^2 + \cdots)^2 - \cdots\right) \\
&= z - a_2 + \frac{a_2^2 - a_3}{z} + \cdots
\end{aligned}
$$

and thus $F(z)$ has $b_1 = a_2^2 - a_3$. So we find $|a_2^2 - a_3| \le 1$, which is sort of interesting, but not what we are aiming for yet.

Next we use a nice trick to make $f(\Delta)$ more symmetrical: we consider, instead of $f(z)$, the new map $g(z) = \sqrt{f(z^2)}$. This map is also in S and it is given by

$$g(z) = z\sqrt{1 + a_2 z^2 + a_3 z^4 + \cdots} = z + (a_2/2)z^3 + \cdots$$

Thus $|a_2/2| \le 1$ and hence $|a_2| \le 2$. ∎

Corollary 4.18 *S is compact.*

Proof. Suppose $f_i \in S$ and let $F_i(z) = 1/f(1/z) + a_2(i) \in \Sigma$. Pass to a subsequence so $F_i(z)$ converges in Σ, and so $a_2(i)$ converges. Then $G_i(z) = F_i(z) - a_2(i)$ converges outside the disk, and hence $f_i(z) = 1/G_i(1/z)$ converges on the unit disk. ∎

Universal properties. Here is a useful geometric consequence of the bound on a_2.

Theorem 4.19 (Koebe 1/4 theorem) *For any $f \in S$ we have $B(0, 1/4) \subset f(\Delta)$.*

Proof. Suppose $p \notin f(\Delta)$. Note that $A(z) = z/(1 - z/p)$ has $A(0) = 0$ and $A'(0) = 1$. Thus $g(z) = A(f(z)) \in S$ as well. But we have

$$g(z) = (z + a_2 z^2 + \cdots)(1 + z/p + \cdots) = z + (a_2 + 1/p)z^2 + \cdots,$$

so $2 \geq |a_2 + 1/p| \geq |1/p| - 2$ and hence $|p| \geq 1/4$. ∎

From the Koebe theorem we get an important comparison between the hyperbolic metric ρ and the '1/d' metric $\delta = \delta(z)|dz| = |dz|/d(z, \partial U)$.

Corollary 4.20 *The hyperbolic metric $\rho(z)\,|dz|$ on a simply-connected region $U \subset \mathbb{C}$ is comparable to the $1/d$ metric: we have $\rho(z)/\delta(z) \in [1/2, 2]$ for all $z \in U$. Equivalent, $\rho(z)d(z, \partial U) \in [1/2, 2]$.*

Proof. We will check this at a given $p \in U$. Let $f : (\Delta, 0) \to (U, p)$ be a Riemann mapping. We may suppose $p = 0$ and $f'(0) = 1$; then $f \in S$, and $\rho_U(p) = \rho_\Delta(p) = 2$. By the Schwarz lemma we have $d(p, \partial U) \leq 1$ and by Koebe we have $d(p, \partial U) \geq 1/4$; hence the result. ∎

Corollary 4.21 *The conformal radius r of (U, p) satisfies $d(p, \partial U) \leq r \leq 4d(p, \partial U)$.*

116

Comparison of S and Σ. The most important mapping in Σ is $F(z) = z + 1/z$; the most important one in S is

$$f(z) = \sum n z^n = z \frac{d}{dz} \frac{1}{1-z} = \frac{z}{(1-z)^2}.$$

These maps are equivalent: $F(z) = 1/f(1/z) + 2$. We have $K(F) = [-2, 2]$ and $K(f) = (-\infty, -1/4]$. The map $f(z)$ shows the Bieberbach conjecture is sharp. We also note the problem with trying to find a 'Bieberbach conjecture' for Σ: there is no map which simultaneously maximizes all the b_n's. Indeed, by the area theorem, if $b_1 = 1$ then the rest of the b_n's are zero.

The distortion theorems. Given $f \in S$, think of $f(\Delta)$ as a splattered egg; then one finds that no matter what, the yolk $f(B(0, r))$, $r < 1$, is still good (not too distorted). For example, the curvature of $f(S^1(r))$ is bounded by a constant K_r independent of f. Also, $f(S^1(r))$ is convex if r is small enough.

Qualitative theorems of this type can be easily deduced from compactness of S. The Koebe distortion theorems make these results more precise. They state:

Theorem 4.22 *For all $f \in S$ and $z \in \Delta$ with $|z| = r$, we have*

$$\frac{(1-r)}{(1+r)^3} \le |f'(z)| \le \frac{(1+r)}{(1-r)^3}$$

and

$$\frac{r}{(1+r)^2} \le |f(z)| \le \frac{r}{(1-r)^2}. \tag{4.1}$$

The proof can be made rather conceptual. Let $U \subset \mathbb{C}$ be a proper simply-connected region, and let $f : U \to \mathbb{C}$ be a univalent map. We will use the hyperbolic metric $\rho(z) |dz|$ on U and the Euclidean metric $|dz|$ on $\mathbb{C}$. It is then natural to study how these metrics compare under f. To this end we define

$$\delta(z) = \log(|f'(z)|/\rho(z))$$

Note that if we replace $f(z)$ with $af(z) + b$, it only changes $\delta(z)$ to $\delta(z) + \log|a|$.

Lemma 4.23 *The gradient of $\delta(z)$ in the hyperbolic metric on U satisfies $|d\delta|/\rho \le 2$.*

Proof. We can assume $U = \Delta$, $z = 0$ and $f \in S$ by the Riemann mapping theorem and by the observations above. Then $\rho(z) = 2/(1-|z|^2)$ is stationary at $z = 0$. Thus $|d\delta| = |f''(0)|/|f'(0)| = |2a_2| \le 4$. Since $\rho(0) = 2$ we get the bound above. ∎

Proof of Theorem 4.22. We continue to assume $U = \Delta$ and $f \in S$. Let $D(z) = \exp \delta(z) = |f'(z)|/\rho(z)$. Integrating along the hyperbolic geodesic from 0 to z, we find

$$|\delta(z) - \delta(0)| \le \int_0^z |d\delta| \le \int_0^z 2\rho = 2\, d(0, r),$$

and thus $D(z)/D(0) \le \exp(2d(0, r))$, $r = |z|$. But $\exp(d(0,r)) = (1 + r)/(1 - r)$, as can be seen by using $i(1 - z)/(1 + z)$ to map to $\mathbb{H}$. Thus $D(z)/D(0) \le (1+r)^2/(1-r)^2$. It follows that for $f \in S$, since $f'(0) = 1$, we have:

$$|f'(z)| = \frac{|f'(z)|}{|f'(0)|} \le \frac{(1+r)^2}{(1-r)^2} \cdot \frac{\rho(r)}{\rho(0)} = \frac{(1+r)}{(1-r)^3}.$$

The reverse inequality is similar: we have

$$|f'(z)| \ge \frac{(1-r)^2}{(1+r)^2} \cdot \frac{\rho(r)}{\rho(0)} = \frac{(1-r)}{(1+r)^3}.$$

Integrating these bounds gives (4.1). ∎

These results also show that S is compact.

Multiply-connected regions. We briefly mention one of the standard forms for a region that has 'more than two holes'.

Suppose $U = \mathbb{C} - (K_1 \cup \cdots \cup K_n)$, where the K_i are disjoint compact connected sets, none of which is a single point. We then have:

Theorem 4.24 *There exists a unique conformal map $F : U \to \mathbb{C}$ of the form $F(z) = z + b_1/z + \cdots$ such that $K(F)$ consists of n disjoint horizontal segments.*

Remark: smoothing the boundary. As a preliminary remark, we note that by applying the Riemann mapping to $\widehat{\mathbb{C}} - K_i$ for each i, one can arrange (if desired) that each K_i is a Jordan domain with real-analytic boundary.

An extremal problems for slits. For the proof of Theorem 4.24, it is useful to introduce the family $\mathcal{F}$ of all univalent conformal maps $f : U \to \mathbb{C}$ of the form above. We then have the following complement:

The map F maximizes $\operatorname{Re} b_1(f)$ over all $f \in \mathcal{F}$.

The proof depends on the following observations. (Cf. [Gol, V.2].)

1. The theorem is true in the case $n = 1$. Indeed, in this case we can assume $U = \mathbb{C} - \overline{\Delta}$, and then $\mathcal{F} = \Sigma$, and $F(z) = z + 1/z$. This is the unique map maximizing $\operatorname{Re} b_1$, by the area theorem.

2. For any pair of maps defined near ∞ by $z + b_1/z + \ldots$, we have $b_1(f \circ g) = b_1(f) + b_1(g)$.

3. Thus the complement is also true in the case $n = 1$. For in this case we can assume (after translation) that U is the image of $g \in \Sigma$. Then $F = f \circ g^{-1}$, and $\operatorname{Re} b_1(F) = \operatorname{Re} b_1(f) - b_1(g) = 1 - \operatorname{Re} b_1(g)$.

   But even more is true! Unless U is already a slit domain, $\operatorname{Re} b_1(F) > 0$, by the area theorem applied to g — since $|b_1(g)| \le 1$.

4. Now return to the case of general U, and suppose $F \in \mathcal{F}$ maximizes $\operatorname{Re} b_1$. We have $K(F) = L_1 \cup \cdots \cup L_n$. Suppose one of these components — say L_1 — is not a horizontal slit. Then we can find a map $G : \mathbb{C} - L_1 \to \mathbb{C} - [a, b]$ such that $\operatorname{Re} b_1(G) > 0$ by what we have just observed. But then $G \circ F \in \mathcal{F}$ and $b_1(G \circ F) > b_1(F)$, contrary to the extremal property of F.

5. We should also check uniqueness. For this it is useful to think of U as the interior of a compact, smoothly bounded domain in $\mathbb{C}$, and $f_1, f_2 : U \to \widehat{\mathbb{C}}$ as normalized maps of the form $f_i(z) = 1/(z-p) + g_i(z)$ whose images are horizontal slit domains. Then the bounded analytic function $h(z) = f_1(z) - f_2(z)$ on U also sends each component of ∂U into a horizontal line. Thus $h(\partial U)$ has winding number zero about points not on these lines, so h must be constant.

Number of moduli. The space of n-slit regions in $\mathbb{C}$, i.e. those of the form $S = \mathbb{C} - \bigcup_1^n [z_i, z_i + a_i]$ with $a_i \in \mathbb{R}$, has real dimension $3n$. We can normalize by an affine transformation so the first slit is, say, $[-2, 2]$; so up to isomorphism, the space of such slit regions has dimension $3n - 3$.

Now to take an n-connected region U and produce a slit region, we need to choose a point $p \in U$ to send to infinity and we need to choose a 'horizontal' direction at p. That gives 3 more real parameters. But of course these choices

119

are only relevant up to the action of the automorphism group of U, which has dimension 3, 1 and 0 for $n = 1, 2$ and $n \geq 3$. Altogether we find:

$$\begin{aligned} d_n &= \dim\{\text{moduli space of } n\text{-connected regions}\} \\ &= 3n - 6 + \dim \text{Aut}(U) \end{aligned}$$

and hence $d_1 = 0$, $d_2 = 1$ (the modulus of an annulus), $d_3 = 3$, $d_4 = 6$, etc.

Rigidity. We remark that although an n-connected region U can have automorphisms, it cannot have a positive–dimensional set thereof, when $n > 2$. To see this, we can assume U is a plane region bounded by real-analytic Jordan curves. Then $\dim \text{Aut}(U) > 0$ would imply there is a holomorphic vector field $v = v(z)(d/dz)$ on U tangent to ∂U. If $v \neq 0$, then the Euler characteristic of U can be expressed as sum of positive contributions, one for each zero of v on ∂U or in U. But $\chi(U) < 0$, so $v = 0$.

A more geometric proof can be given once one knows that $U \cong \Delta/\Gamma$ (a proof will be sketched below). Then any automorphism must be an isometry, and hence send geodesics to geodesics. By looking at intersections of geodesics one can easily conclude that f is the identity.

4.2 Conformal mappings of polygons

What is the conformal radius of the unit square?

To get explicit forms for Riemann mappings, it is useful to initially try to determine, not f itself, but the deviation of f from a simpler class of mappings. This deviation will be a form in general, so we first make some remarks on forms.

Let $f : X \to \widehat{\mathbb{C}}$ be a meromorphic function on a Riemann surface. Then $\omega = df = f'(z)\, dz$ is naturally a 1-form. That is, if we change coordinates by setting $z = \phi(w)$, then in these new coordinates we have

$$\phi^*(df) = d(f(\phi(w)) = f'(\phi(w))\phi'(w)\, dw.$$

We can similarly define quadratic differentials $q(z)\, dz^2$ which transform by $\phi^* q = q(\phi(z))\phi'(z)^2\, dz^2$.

In more intrinsic terms, these forms are sections of the complex cotangent bundle T^*X and its tensor product with itself.

For a meromorphic 1-form, the *residue* $\text{Res}_p(\omega)$ is also coordinate independent, since it can be expressed as $(2\pi i)^{-1} \int_\gamma \omega$ for a small loop around p. Stokes' theorem implies:

Theorem 4.25 (The Residue Theorem) *If ω is a nonzero meromorphic 1-form on a compact Riemann surface X, then $\sum_{p \in X} \mathrm{Res}_p(\omega) = 0$.*

Let's check this in an example: if we take $\omega(z) = dz/P(z)$ where $P(z)$ is a polynomial of degree $d \geq 2$, then it says:

$$\sum_{P(z)=0} \frac{1}{P'(z)} = 0.$$

This can be verified using partial fractions: we have $P(z) = \sum a_i/(z - b_i)$ and $\sum a_i = 0$ because $|P(z)| = O(|z|^{-2})$ for large z. It can also be proved by integrating $dz/P(z)$ around a large circle and taking the limit.

If $\omega = \omega(z)\,dz$ is a meromorphic 1-form on the sphere, we can set $\pi(z) = 1/z$ and form $\pi^*\omega$ to find:

$$\mathrm{Res}_\infty(\omega) = \mathrm{Res}_0(-\omega(1/z)\,dz/z^2).$$

In particular $\mathrm{Res}_\infty(dz/z) = -1$.

Theorem 4.26 *Every meromorphic 1-form on the sphere has 2 more poles than zeros.*

Proof. This is true for $\omega = dz/z$, which has a simple poles at 0 and ∞ and no zeros. It then follows for any other 1-form η, since $f = \eta/\omega$ is a rational function on $\widehat{\mathbb{C}}$, which must have the same number of poles as zeros. ∎

Corollary 4.27 *A holomorphic 1-form on $\widehat{\mathbb{C}}$ must be zero. In particular, if ω_1 and ω_2 are 1-forms with the same principal parts, then $\omega_1 = \omega_2$.*

Remark: one forms and vector fields in higher genus. The first assertion can also be seen by integrating ω to obtain a global holomorphic function on $\widehat{\mathbb{C}}$. On a Riemann surface of genus g, a meromorphic 1-form has $2g - 2$ more zeros than poles, and $\dim \Omega(X) = g$. The space of holomorphic vector fields, on the other hand, satisfies $\dim \Theta(\widehat{\mathbb{C}}) = 3$, $\dim \Theta(E) = 1$ on a torus, and $\dim \Theta(X) = 0$ in higher genus. This is because a meromorphic vector field has $2g - 2$ more poles than zeros.

One can use the vanishing of $\Theta(X)$ is see an n-connected plane region U has a zero-dimensional automorphism group for $n \neq 3$; if not, there would

be a boundary-parallel holomorphic vector field on U, but then the double X of U would satisfy $\dim \Theta(X) > 0$; while its genus is given by $g = n - 1$.

Measuring distortion: the 3 great cocycles. Now let $C(f)$ be a differential operator that sends meromorphic functions to meromorphic forms. We say C is a *cocycle* if it satisfies:

$$C(f \circ g) = C(g) + g^*C(f).$$

This formula implies the maps satisfying $C(f) = 0$ form a group.

There are 3 important cocycles in complex analysis:

1. The *derivative* $Df(z) = \log f'(z)$.

2. The *nonlinearity* $Nf(z)\,dz = d\,Df = (f''/f')(z)\,dz$.

3. The *Schwarzian derivative*

$$Sf(z)\,dz^2 = (Nf)' - \frac{1}{2}(Nf)^2 = \left[\left(\frac{f''}{f'}\right)' - \frac{1}{2}\left(\frac{f''}{f'}\right)^2\right]dz^2.$$

Their values are functions, 1-forms and 2-forms respectively. The groups they annihilate are transformations of the form $f(z) = z + a$, $f(z) = az + b$ and $f(z) = (az + b)/(cz + d)$.

Let us check this for Sf: if $f(z) = (az + b)/(cz + d)$ and $ad - bc = 1$, then $f'(z) = 1/(cz + d)^2$, hence $Nf(z) = -2c/(cz + d)$, which satisfies

$$(Nf)' = \frac{2c^2}{(cz + d)^2} = \frac{1}{2}(Nf)^2.$$

Higher cocycles? There are in fact operators $T_d f$ generalizing Sf with the property that $T_d f = 0$ iff f is a rational map of degree $d > 1$. But these cannot be cocycles, because the rational maps of degree $d > 1$ do not form a group.

To get some insight into Sf (and T_d), note that if $f(z) = \sum a_i z^i = (az+b)/(cz+d)$, then $(cz+d)\sum a_i z_i = az+b$ and hence there must be a linear relation between most of the coefficients of $zf(z)$ and $f(z)$. More precisely, the coefficients of degree 2 or more must be proportional; in particular, we have

$$\det \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix} = a_2^2 - a_1 a_3 = 0,$$

which just says that $f'f''' - (3/2)(f'')^2$ vanishes. This expression is also the numerator of $Sf = (f'f''' - (3/2)(f'')^2)/(f')^2$. Similarly a rational map of degree two is characterized by the property that

$$\det \begin{pmatrix} a_2 & a_3 & a_4 \\ a_3 & a_4 & a_5 \\ a_4 & a_5 & a_6 \end{pmatrix} = 0$$

at every point. It is natural to divide this expression by a_1^3 just as we have divided Sf by $a_1^2 = (f')^2$, so we have $T_d(af + b) = T_d(f)$.

The Schwarz–Christoffel formula. We can now use the nonlinearity to derive a formula for the Riemann mapping to a polygonal region. It is based on the fact that

$$N(z^\alpha) = (\alpha - 1)\frac{dz}{z}.$$

Theorem 4.28 *Let $f : \mathbb{H} \to P$ be the Riemann mapping to a polygon with vertices p_i, $i = 1, \ldots, n$ and exterior angles $\pi\mu_i$, normalized so that $f(\infty)$ is not a vertex. Then*

$$f(z) = \alpha \int \frac{d\zeta}{\prod_1^n (\zeta - q_i)^{\mu_i}} + \beta,$$

where $f(q_i) = p_i$.

Proof. We will show that:

$$Nf = \sum \frac{-\mu_i \, dz}{(z - q_i)}.$$

We first observe that Nf extends to a holomorphic 1-form on $\mathbb{C} - \{p_1, \ldots, p_n\}$. This follows by Schwarz reflection across each complementary interval on the real axis. These reflected maps g_i do not agree, but they differ by linear maps: $g_j = A_{ij} \circ g_i$, and thus $N(g_i) = N(g_j)$. It remains to show Nf has simple poles with the given residues at the p_i, and at infinity.

Let $\alpha_i = 1 - \mu_i$. The idea of the proof is that $f(z)$ behaves like $(z - q_i)^{\alpha_i}$ near q_i, and thus Nf has the same residue, which is $\alpha_i - 1 = -\mu_i$. To check this, we note that $g(z) = (f(z) - p_i)^{1/\alpha_i}$ extends by Schwarz reflection across q_i, to give a conformal map $g(z)$ near q_i. Then we can locally write

$$g(z)^{\alpha_i} = f(z) - p_i.$$

Taking the nonlinearity of both sides, and using the fact that the residue is preserved under pullback, we find $\text{Res}_{q_i}(Nf) = 1 - \alpha_i = -\mu_i$.

Near infinity, $f(z)$ behaves like $1/z$ which has nonlinearity $-2dz/z$ and hence residue 2 at infinity. This is consistent with the residue theorem, because $\sum \pi \mu_i = 2\pi$. Since a 1-form is determined by its singularities, we have justified the formula for Nf.

Integrating, we find

$$\log f' = \sum -\mu_i \log(z - q_i) + C;$$

by exponentiating and integrating again, we get the desired formula for f. ∎

Two Examples. This formula explains the geometric connection between Riemann maps to polygons and transcendental functions we have seen in two examples, namely:

$$\log(z) = \int \frac{dz}{z}$$

maps $\mathbb{H}$ to a bigon with external angles of π, namely the strips $0 < \text{Im } z < \pi$; and

$$\sin^{-1}(z) = \int \frac{dz}{\sqrt{1 - z^2}}$$

maps $\mathbb{H}$ to a triangle with external angles $\pi/2$, $\pi/2$ and π, namely the half-strip defined by $\text{Im } z > 0$, $|\text{Re}(z)| < \pi/2$.

Unit disk version. The key point in the proof above was that the various results of extending f by Schwarz reflection through $\partial\mathbb{H}$ agree up to composition with linear maps. The same is true if $\mathbb{H}$ is replaced by the unit disk. Thus if we choose the domain of f to be Δ, with $q_i \in S^1$, we find exactly the same formula for f.

Regular polygons. As a particular example, we find:

Theorem 4.29 *The map $f(z) = \int (1 - z^n)^{-2/n}\, dz$ maps the unit disk to a regular n-gon, sending 0 to its center and the nth roots of unity to its vertices.*

Note that $f'(0) = 1$, and the distance from the center of the polygon to one of its vertices is given by

$$R_n = \int_0^1 (1 - z^n)^{-2/n}\, dz.$$

124

When $n = 4$, the image of f is actually a diamond, with two vertices on the real axis.

The Beta function and conformal radius. The quantity R_n can be computed explicitly. To this end, we introduce the Euler beta-function

$$B(\alpha, \beta) = \int_0^1 u^\alpha (1-u)^\beta \frac{du}{u(1-u)} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

It is elementary to prove the integral above satisfies $B(\alpha, 1) = 1/\alpha$, and integration by parts gives $B(\alpha, \beta) = ((\beta - 1)/\alpha)B(\alpha + 1, \beta - 1)$; this easily implies

$$B(\alpha + 1, \beta + 1) = \frac{\alpha!\beta!}{(\alpha+\beta+1)!} = \frac{1}{\alpha+\beta+1}\binom{\alpha+\beta}{\alpha}^{-1}$$

for integral values of α, β. We also note that $|dx|/x$, which appears in the integral definition of the Γ function, is the hyperbolic metric on $\mathbb{R}_+$, and $|dx|/(x(1-x))$ is the hyperbolic metric on $(0, 1)$. Note that formally these one forms have residue 1 at the endpoints of the interval, just as $2|dx|/(1 - |x|^2)$ does on $(-1, 1)$.

Making the substitution $u = 1 - z^n$, $du = -nz^{n-1}dz$, $dz = -(1/n)u^{1/n-1}\, du$, we find

$$R_n = \frac{1}{n}\int_0^1 u^{-2/n}(1-u)^{1/n-1}\, du = \frac{\Gamma(1-2/n)\Gamma(1/n)}{n\Gamma(1-1/n)}.$$

Note that $R_n \to 1$ as $n \to \infty$, since $\Gamma(1/n) \sim n$, and $R_n > 1$ by the Schwarz lemma. For the square, we find:

$$R_4 = \frac{\pi^4}{2\sqrt{2}\Gamma(3/4)^2}.$$

Here we have used the fact that $\Gamma(1/4)\Gamma(3/4) = \pi/\sin(\pi/4) = \pi\sqrt{2}$.

Now if we let P_n be the perimeter of the polygon, then $P_n = 2n\sin(\pi/n)R_n$. But $\Gamma(1/n)\Gamma(1 - 1/n) = \pi/\sin(\pi/n)$, so we get

$$P_n = \frac{2\pi\Gamma(1-2/n)}{\Gamma(1-1/n)^2}.$$

Note that $P_n \to 2\pi$ as $n \to \infty$, since $\Gamma(1) = 1$. The factor $P_n/(2\pi)$ by which the perimeter exceeds the perimeter of the unit circle is given by $(P_3, P_4, P_5, P_6) \approx (1.461, 1.180, 1.098, 1.043)$.

125

Conformal radius of the unit square. In particular, using the fact that $\Gamma(1/2) = \sqrt{\pi}$, the conformal radius of the unit square (sides of length 2) is

$$R([-1,1] \times [-1,1], 0) = 8/P_4 = \frac{4\Gamma[3/4]^2}{\pi^{3/2}} = 1.078705\ldots.$$

(It is clear that the conformal radius of the unit square lies between 1 and $\sqrt{2}$.)

Remark: Beta and statistics. In statistics, unlike probability theory, the observations are known but the governing probabilities are not. Or, we may model p itself as a random variable.

For example, suppose we have a bag of coins whose chances p of coming up heads are uniformly distributed between 0 and 1. We pick a coin at random, flip it n times, and see a heads and b tails. Call this event $E(a, b)$.

What is the probability of $E(a, b)$? The answer is given by $B(a+1, b+1)$; it is the average value of $p^a(1-p)^b$ over $[0, 1]$.

Now once we have seen this event, we have more information that can be used to predict what p is. If we fix an interval I, then from the fact that $P(E|F)P(F) = P(E)$, we deduce that

$$P(p \in I | E(a,b)) = P(E(a,b) | p \in I)P(p \in I)/P(E(a,b)).$$

This implies our updated guess for the value of p is a random variable U_{ab} with distribution $u^a(1-u)^a \, du/B(a+1, b+1)$. This distribution is peaked at $a/(a+b)$ but it is still spread out; the peak becomes sharper as a and b increase.

Of course in these cases we only need the value of $B(a, b)$ when a and b are integers, in which case it is (the inverse of) a binomial coefficient.

Quadrilaterals. A *quadrilateral Q* is a Jordan domain with 4 distinguished points on its boundary. A conformal map between quadrilaterals is required to preserve these points as a set.

Theorem 4.30 *Any quadrilateral is conformally equivalent to a rectangle of the form $R(a) = [0, a] \times [0, 1]$, $a > 0$. This rectangle is unique up to $a \mapsto 1/a$.*

Proof. By the Riemann mapping theorem, any quadrilateral is isomorphic to $\mathbb{H}$ with 4 distinguished points, $q_1, \ldots, q_4 \in \mathbb{R}$. By the Schwarz–Christoffel formula, $f(z) = \int \prod (z - q_i)^{-1/2} \, dz$ maps $\mathbb{H}$ to a Euclidean rectangle, sending q_i to its vertices. ∎

Besikovitch lemma. As an exercise, one can now use the method of extremal length to show, for any quadrilateral $Q \subset \mathbb{C}$, we have $\text{area}(Q) \geq AB$ where A and B are the minimum distances between opposite sides of Q.

The Schwarzian derivative and hypergeometric functions. Just as the nonlinearity can be used to find a formula for the Riemann mapping to a polygon with linear sides, the Schwarzian derivative can be used to find the Riemann mapping to a polygon with circular sides.

To describe a result in this direction, we first recall that for $f(z) = z^\alpha$ we have $Nf(z) = (\alpha - 1)/z$, and thus

$$Sf(z) = (Nf)' - \frac{(Nf)^2}{2} = \frac{1 - \alpha^2}{2} \frac{dz^2}{z^2}.$$

The quantity $1 - \alpha^2$ is a version of the residue for the quadratic differential $Sf(z) \, dz^2$. Now given α, β, γ there is a unique quadratic differential with double poles at $0, 1, \infty$ with the corresponding residues, and no other singularities. It is given by

$$Q(\alpha, \beta, \gamma) = \frac{1}{2} \left[\frac{1 - \alpha^2}{z^2} + \frac{1 - \beta^2}{(z-1)^2} + \frac{1 - \gamma^2 - (2 - \alpha^2 - \beta^2)}{z(z-1)} \right] dz^2.$$

The third term is chosen to that $Q \sim (1 - \gamma^2)/z^2$ as $z \to \infty$.

Theorem 4.31 *Let P be a circular triangle with interior angles $\pi\alpha$, $\pi\beta$ and $\pi\gamma$. Then the Riemann mapping $f : \mathbb{H} \to P$ sending 0, 1 and ∞ to these vertices satisfies $Sf = Q(\alpha, \beta, \gamma)$.*

Proof. By Schwarz reflection, Sf extends to a meromorphic quadratic differential on $\widehat{\mathbb{C}}$ whose residues at $0, 1$ and ∞ are determined by the angles of P. ∎

Ideal triangles. As an example, if P is an ideal triangle, then we have

$$Q = Q(0, 0, 0) = \frac{z^2 - z + 1}{2z^2(z-1)^2}. \tag{4.2}$$

This differential is symmetric under the action of S_3 on $\mathbb{H}$, and it has zeros at the primitive sixth roots of unity.

Second order differential equations. To find f itself, we remark that any function f satisfying $Sf = Q$ can be expressed as the ratio $f = u_1/u_2$ of two independent solutions to the differential equation

$$u'' + (Q/2)u = 0.$$

In the case at hand, these solutions in turn can be expressed in terms of solutions to the *hypergeometric equation*

$$z(1 - z)u'' + Au' + Bu = 0$$

for suitable constants A and B.

4.3 The Picard theorems

We conclude this section with two further results concerning entire functions that can be related to geometric function theory.

Theorem 4.32 (Little Picard Theorem) *An entire function $f : \mathbb{C} \to \mathbb{C}$ which omits two values must be constant.*

Corollary 4.33 *A meromorphic function on $\mathbb{C}$ can omit at most two values on $\widehat{\mathbb{C}}$.*

The Little Picard Theorem is equivalent to the assertion that there is no solution to the equation $e^f + e^g = 1$ where f and g are nonconstant entire functions. Similarly, it implies there is no solution to Fermat's equation $f^n + g^n = 1$, $n \geq 3$, unless the entire functions f and g are constant.

Theorem 4.34 (Great Picard Theorem) *An analytic function $f : U \to \mathbb{C}$ takes on every value in $\mathbb{C}$, with at most one exception, in every neighborhood of an essential singularity p.*

The first theorem follows from the second by considering the essential singularity at $z = 0$ of $f(1/z)$. These results generalize Liouville's theorem and the Weierstrass-Casorati theorem respectively.

Rescaling arguments. Here is a third result that initially seems unrelated to the first two.

Theorem 4.35 (Bloch's Theorem) *There exists a universal $R > 0$ such that for any $f : \Delta \to \mathbb{C}$ with $|f'(0)| = 1$, not necessarily univalent, there is an open set $U \subset \Delta$ (perhaps a tiny set near S^1) such that f maps U univalently to a ball of radius R.*

Corollary 4.36 *For any analytic map on Δ, the image $f(\Delta)$ contains a ball of radius $R|f'(0)|$.*

Note that the ball usually *cannot* be centered at $f(0)$; for example, $f(z) = \exp(nz)/n$ satisfies $f'(0) = 1$ but the largest ball about $f(0) = 1/n$ in $f(\Delta) \subset \mathbb{C}^*$ has radius $1/n$.

The optimal value of R is known as *Bloch's constant*. It satisfies $0.433 < \sqrt{3}/4 \le R < 0.473$. The best-known upper bound comes from the Riemann surface branched with order 2 over the vertices of the hexagonal lattice.

These apparently unrelated theorems can both be proved using the same idea. (Cf. [BD] and references therein; for another way to relate these theorems, see [Re, Ch. 10].)

Proof of Bloch's theorem. Given $f : \Delta \to \mathbb{C}$, let

$$\|f'(z)\| = \|f'(z)\|_{\Delta,\mathbb{C}} = (1/2)|f'(z)|(1 - |z|^2)$$

denote the norm of the derivative from the hyperbolic metric to the Euclidean metric. By assumption, $\|f'(0)\| = 1/2$. We can assume (using $f(rz)$) that f is smooth on S^1; then $\|f'(z)\| \to 0$ as $|z| \to 1$, and thus $\sup \||f'(z)\|$ is achieved at some $p \in \Delta$.

Now replace f with $f \circ r$ where $r \in \mathrm{Aut}(\Delta)$ moves p to zero. Replacing f with $af + b$ with $|a| < 1$, we can also arrange that $f(0) = 0$ and $\|f'(0)\| = 1$; this will only decrease the size of its unramified disk. Then $\|f'(z)\| \le \|f'(0)\| = 1$, and thus $f|\Delta(1/2)$ ranges in a compact family of nonconstant analytic functions. Thus the new f has an unramified disk of definite radius; but then the old f does as well. ∎

Rescaling proof of Picard's theorem. The proof will use the following remarkably general rescaling theorem. This argument is related to Bloch's proof, to Brody's reparameterization theorem and to other results in complex analysis.

Theorem 4.37 *Let $f_n : \mathbb{C} \to \mathbb{C}$ be a sequence of nonconstant entire functions. There after passing to a subsequence, there is a sequence of Möbius*

129

transformations A_n and a nonconstant entire function $g : \mathbb{C} \to \mathbb{C}$ such that $g = \lim f_n \circ A_n$ uniformly on compact subsets of $\mathbb{C}$.

We note that the A_n need not fix infinity, so $f_n \circ A_n$ is undefined at some point $p_n \in \widehat{\mathbb{C}}$, but we will have $p_n \to \infty$.

Example. For $f_n(z) = z^n$ we can take $f_n \circ A_n(z) = (1 + z/n)^n \to e^z$.

Metrics. As a preliminary to the proof, for $g : \mathbb{C} \to \widehat{\mathbb{C}}$ we define

$$\|g'(z)\|_\infty = \frac{|g'(z)|}{(1 + |g(z)|^2)},$$

and $\|g'\|_\infty = \sup \|g'(z)\|$ over all $z \in \mathbb{C}$. This is the norm of the derivative from the scaled Euclidean metric $\rho_\infty = 2|dz|$ to the spherical metric. Note that $g(z) = \exp(z)$ has $\|g'\|_\infty = 1/2$; a function with bounded derivative can be rather wild.

Similarly, for $g : \Delta(R) \to \widehat{\mathbb{C}}$, we define

$$\|g'(z)\|_R = |g'(z)| \frac{1 - |z/R|^2}{1 + |g(z)|^2}.$$

This is the derivative from a suitably rescaled hyperbolic metric ρ_R on $\Delta(R)$ to $\widehat{\mathbb{C}}$. Clearly $\rho_R \to \rho_\infty$ uniformly on compact sets. Its key property is that $\|(g \circ A)'\|_R = \|g'\|_R$ for all $A \in \operatorname{Aut}(\widehat{\mathbb{C}})$ stabilizing $\Delta(R)$.

We also note that the set of maps with uniformly bounded derivatives in one of these norms is compact.

Proof of Theorem 4.37. Let us first consider an arbitrary nonconstant analytic function $f(z)$ and a radius $R > 0$. We claim there exists an $S \geq R$ and an $A \in \operatorname{Aut}(\widehat{\mathbb{C}})$ such that $g = f \circ A$ is analytic on $\Delta(S)$, and

$$\|g'(0)\|_S = \|g'\|_S = 1.$$

Indeed, by replacing f with $f(az + b)$, we can assume $\|f'(0)\|_R = 1$. Then $\|f'\|_R \geq 1$. On the other hand, the R-norm of the derivative of f tends to zero at the boundary of $\Delta(R)$. Thus we can choose $B \in \operatorname{Aut}(\Delta(R))$ such

$$M = \|(f \circ B)'(0)\|_R = \|(f \circ B)'\|_R \geq 1.$$

Now just let $g(z) = (f \circ B)(z/M)$, and $S = RM$.

Applying this claim to f_n and a sequence $R_n \to \infty$, we obtain $S_n \to \infty$ and maps $g_n = f_n \circ A_n$ with $\|g_n'(0)\|_\infty = 1$ and $\|g_n'\|_{S_n} \leq 1$. Now pass to a convergent subsequence. ∎

130

To complete the proof of Picard's theorem, we observe:

Lemma 4.38 *If $f_n \to f$ and f is nonconstant, then any value omitted by all f_n is also omitted by f.*

Proof of the Little Picard Theorem. Suppose if $f : \mathbb{C} \to \mathbb{C}$ is nonconstant and omits 0 and 1. Then $f_n(z) = f_n^{1/n}(z)$ omits more and more points on the unit circle. We can of rescale in the domain so the spherical derivative satisfies $\|f_n'(0)\|_\infty \to \infty$. Passing to a subsequence and reparameterizing, we obtain in the limit a nonconstant entire function that omits the unit circle. This contradicts Liouville's theorem. ∎

Classical proof. The classical proof of the Little Picard Theorem is based on the fact that the universal cover of $\mathbb{C} - \{0, 1\}$ can be identified with the upper halfplane.

To see this, it is useful to start by considering the subgroup $\Gamma_0 \subset \text{Isom}(\Delta)$ generated by reflections in the sides of the ideal triangle T with vertices $\{1, i, -1\}$. For example, $z \mapsto \bar{z}$ is one such reflection, sending T to $-T$. By considering billiards in T, one can see that its translates tile the disk and thus T is a fundamental domain for Γ_0. Thus the quadrilateral $F = T \cup (-T)$ is a fundamental domain for the orientation-preserving subgroup $\Gamma \subset \Gamma_0$, and the edges of $-T$ are glued to the edges of T to give a topological triply-punctured sphere as quotient.

Now let $\pi : T \to \mathbb{H}$ be the Riemann mapping sending T to $\mathbb{H}$ and its vertices to $\{0, 1, \infty\}$. Developing in both the domain and range by Schwarz reflection, we obtain a covering map $\pi : \Delta \to \widehat{\mathbb{C}} - \{0, 1, \infty\}$.

Given this fact, we lift an entire function $f : \mathbb{C} \to \mathbb{C} - \{0, 1\}$ to a map $\widetilde{f} : \mathbb{C} \to \mathbb{H}$, which is constant by Liouville's theorem.

Uniformization of planar regions. Once we know that $\widehat{\mathbb{C}} - \{0, 1, \infty\}$ is uniformized by the disk, it is straightforward to prove:

Theorem 4.39 *The universal cover of any region $U \subset \widehat{\mathbb{C}}$ with $|\widehat{\mathbb{C}} - U| \geq 3$ is isomorphic to the unit disk.*

Sketch of the proof. Consider a basepoint p in the abstract universal cover $\pi : \widetilde{U} \to U$, and let $\mathcal{F}$ be the family of all holomorphic maps

$$f : (\widetilde{U}, p) \to (\Delta, 0)$$

that are covering maps to their image. Using the uniformization of the triply-punctured sphere, we have that $\mathcal{F}$ is nonempty. It is also a closed, normal family of functions in $\mathcal{O}(\widetilde{U})$; and by the classical square-root trick, it contains a surjective function (which maximizes $|f'(p)|$). By the theory of covering spaces, this extremal map must be bijective. ∎

Proof of the Great Picard Theorem. Let $f : \Delta^* \to \widehat{\mathbb{C}} - \{0, 1, \infty\}$ be an analytic function. We will show f does not have an essential singularity at $z = 0$.

Consider a loop γ around the puncture of the disk. If f sends γ to a contractible loop on the triply-punctured sphere, then f lifts to a map into the universal cover $\mathbb{H}$, which implies by Riemann's removability theorem that f extends holomorphically over the origin.

Otherwise, by the Schwarz lemma, $f(\gamma)$ is a homotopy class that can be represented by an arbitrarily short loop. Thus it corresponds to a puncture, which we can normalize to be $z = 0$ (rather than 1 or ∞). It follows that f is bounded near $z = 0$ so again the singularity is not essential. ∎

4.4 Exercises.

Notation as in this chapter: S denotes the space of normalized univalent maps on Δ, Σ the space of normalized univalent maps on $\mathbb{C} - \overline{\Delta}$, and $A(R) = \{z : 1 < |z| < R\}$.

1. Let $f : [0, 1] \to \mathbb{C}$ be a continuous map, and let K be its image. Prove that K is locally connected.

2. Prove the converse: if $K \subset \mathbb{C}$ is a connected, locally connected compact set, then there exists a surjective map $f : [0, 1] \to K$.

3. Show that if $f, g \in S$ and $f(\Delta)$ contains $g(\Delta)$ then $f = g$.

4. Let $S = \{z : |\operatorname{Im}(z)| < 1\}$. Compute the conformal radius $R(S, 0)$.

5. Suppose $f \in S$ satisfies $\sup_\Delta |f(z)| \leq M$. Show that $|a_2| \leq 2\sqrt{1 - 1/M}$.
   (Hint: $g(z) = f(z^{-2})^{-1/2} \in \Sigma$ omits the ball $B(0, M^{-1/2})$ which has area π/M.)

6. Show that if $f \in \Sigma$ and w is not in the image of f, then $|w| \leq 2$.

   (Hint: apply the bound $|b_1| \leq 1$ to the function $\sqrt{f(z^2) - w}$.)

7. Prove the uniformization theorem for plane regions. That is, fill in the details of the sketch proof of the Theorem 4.39 given above.

8. Let $U \subset \mathbb{C}$ be a connected open set (which need not be simply-connected), and let $p \in U$. Let $\mathcal{F} \subset \mathcal{O}(U)$ be the family of all univalent maps $f : U \to \mathbb{C}$ with $f(p) = 0$ and $f'(p) = 1$. Show that $\mathcal{F}$ is compact. (You may use the fact that $S \subset \mathcal{O}(\Delta)$ is compact.)

9. For $a > b > 0$ consider the ellipse $E \subset \mathbb{C}$ with major axis $[-a, a]$ and minor axis $[-ib, ib]$. Let $I \subset [-a, a]$ be the segment joining the foci of E. Finally let B be the annular region between E and I.

   Find an explicit conformal map $f : A(R) \to B$ for some $R > 1$. (Hint: think about $z + 1/z$.)

10. Let $B \subset \mathbb{C}$ be an annulus bounded by a pair of Jordan curves C_1 and C_2. Prove that there exists an arc $\alpha \subset B$ joining C_1 to C_2, and a loop $\beta \subset B$ separating C_1 from C_2, such that their lengths satisfy

$$L(\alpha)L(\beta) \leq \operatorname{area}(B).$$

   (Hint: let $f : A(R) \to B$ be a conformal map from a standard annulus to B, and consider $\int_{A(R)} |f'(z)| \cdot |z|^{-1} |dz|^2$.)

11. Let $f : \Delta \to \mathbb{C}$ be analytic and suppose $\int_\Delta |f'(z)|^2 |dz|^2 < \infty$. Prove that $F(z) = \lim_{r \to 1} f(rz)$ exists and is finite for almost every $z \in S^1$.

12. Give an example of a homeomorphism $f : \Delta \to \Delta$ that is distance-decreasing for the hyperbolic metric, but does not extend continuously to S^1. (Hint: try a map of the form $f(r, \theta) = (R(r), T(r) + \theta)$, where $T(r) \to \infty$ as $r \to 1$.)

13. Compute the conformal radii of the following pointed regions:

    (a) $(\mathbb{H}, i)$;

    (b) $(\{z : -1 < \operatorname{Re} z < 1\}, 0)$;

    (c) $(\{z \in \mathbb{H} : -\pi < \operatorname{Re} z < \pi\}, i)$;

(d) $(\{z \in \mathbb{C} : z \notin (-\infty, -1] \cup [1, \infty)\}, 0)$; and

(e) (S_α, r), where $r > 0$ and $S_\alpha = \{z : \arg(z) \in (-\alpha, \alpha)\}$.

14. Let $A = \{z \in S^1 : \mathrm{Im}(z) \geq 0\}$. Find a Riemann map $f : \mathbb{C}-\overline{\Delta} \to \mathbb{C}-A$.

15. Show that the set of univalent maps whose images are Jordan domains is dense in S (in the topology of uniform convergent on compact subsets of Δ).

16. Find a conformal homeomorphism from $\mathbb{C} - \overline{\Delta}$ to $\mathbb{C} - E(a, b)$, where $E(a, b)$ is the ellipse $\{x + iy : x^2/a^2 + y^2/b^2 \leq 1\}$.

17. For $t \geq 0$, let $U_t = \mathbb{H} \cup (-\mathbb{H}) \cup (-t, t)$. Let $f_t : (\mathbb{H}, i) \to (U_t, i)$ be the Riemann mapping, normalized so that $f_t'(i) > 0$.

(i) Find $f_t(z)$ explicitly for $t > 0$.

(ii) Show that as $t \to 0$, $f_t(z) \to f_0(z) = z$ in $\mathcal{O}(\mathbb{H})$.

(iii) Note that $-\mathbb{H} \subset f_t(\mathbb{H})$ for all $t > 0$, but $-\mathbb{H}$ is disjoint from $f_0(\mathbb{H})$. How is this possible, if $f_t \to f_0$?

18. Let $U \subset \mathbb{C}$ be a simply–connected region such that $p = 0 \in U$ but $1 \notin U$. Suppose that conformal radius R of (U, p) is 2. Is U unique? What about if $R = 4$?

19. Is there a univalent map $f : \mathbb{C}-\overline{\Delta} \to \mathbb{C}$ of the form $f(z) = z + \sum_1^\infty b_n/z^n$ with $b_2 = 1/\sqrt{2}$?

20. Show that the set of Riemann mappings whose images are polygons is dense in S, in the topology of uniform convergence on compact sets.

21. Suppose $f \in S$ satisfies $f(iz) = if(z)$. Show that $f(\Delta)$ contains $B(0, 1/\sqrt{2})$.

22. Give an explicit bound $|a_n| \leq M_n$ valid for all $f \in S$. (Use only elementary methods, such as the area theorem or the distortion theorems.)

23. Prove $|a_n| \leq n$ for all n provided $f(z) = \sum a_n z^n \in S$ is *real*, meaning $f(\overline{z}) = \overline{f(z)}$ for all $z \in \Delta$.

(Hint: one may assume that $f(z)$ is smooth on S^1. The hypothesis implies that $\Omega = f(\Delta)$ is symmetric about the real axis, that $v(\theta) = \operatorname{Im} f(e^{i\theta}) \geq 0$ for $\theta \in [0, \pi]$, and that

$$a_n = \frac{2}{\pi} \int_0^\pi v(\theta) \sin(n\theta) \, d\theta$$

for all $n \geq 1$. Now use the fact that $|\sin(n\theta)| \leq n \sin(\theta)$ in this range to show $|a_n| \leq n a_1$.)

24. Prove that the Beta function satisfies

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

for a, b with $\operatorname{Re} a, \operatorname{Re} b > 0$. (Hint: express $\Gamma(a)\Gamma(b)$ as a double integral over $s, t \geq 0$; then write this in terms of the coordinates $u \in [0, 1]$ and $v \geq 0$ defined by $s + t = v$, $s = uv$ and $t = (1 - u)v$.)

25. What is the radius of the largest unramified disk for the map $f : \Delta \to \mathbb{C}$ given by $f(z) = e^{z^2}$? (Recall a disk $B(p, r) \subset \mathbb{C}$ is *unramified* if there is a holomorphic map $h : (p, r) \to \Delta$ such that $f(h(z)) = z$ for all $z \in B(p, r)$.)

26. Prove the nonlinearity $Nf = d \log f'$ satisfies $N(f \circ g) = Ng + g^*(Nf)$. (Keep in mind that Nf is a 1-form.)

27. Prove that the Schwarzian derivative $Sf = Sf(z) \, dz^2$ satisfies $S(f \circ g) = Sg + g^*(Sf)$.

28. Let $P_n \subset \mathbb{C}$ be a regular polygon with n sides, and let $f : P_n \to \Delta$ be a Riemann mapping (a bijective holomorphic map).

(1) Prove that if $n = 3$, then f extends to a meromorphic function on all of $\mathbb{C}$.

(2) Prove that this is false if $n = 6$.

29. Let $f : \Delta \to \mathbb{C}$ be a univalent function. What is the maximum possible value of the nonlinearity $f''(0)/f'(0)$?

30. Prove that if $f : \mathbb{H} \to \mathbb{C}$ is given by the Schwarz–Christoffel formula, but $S = \sum \mu_i \neq 2$, then the image $P = f(\mathbb{H})$ is a polygon with a vertex at $f(\infty)$ with external angle $(2 - S)\pi$.

31. Let $\phi(z)$ be an analytic function, and let u_1, u_2 be two linearly independent solutions to the differential equation $d^2u/dz^2 + \phi u = 0$.

    (i) Show $W(z) = \det \left(\begin{smallmatrix} u_1 & u_2 \\ u_1' & u_2' \end{smallmatrix} \right)$ is a constant.

    (ii) Show that if $f(z) = u_1(z)/u_2(z)$, then $Sf(z) = C\phi(z)$ and find C. Why does the value of Sf not depend on the choice of u_1 and u_2?

32. Find a formula (which may involve an integral) for a covering map $f : \mathbb{H} \to \mathbb{C} - T$ whose image is the complement of a square T.

33. Prove that the Little Picard Theorem implies directly that there is a constant $M > 0$ such that any holomorphic map $f : \Delta \to \mathbb{C} - \{0, 1\}$ satisfies $\|f'\| \leq M$, where the derivative is measured using the hyperbolic metric on the domain and the spherical metric on the range. (Hint: supposing the derivative tends to infinity, construct a nonconstant entire function $f : \mathbb{C} \to \mathbb{C} - \{0, 1\}$.)

34. Let $T \subset \mathbb{R}^2$ be a (closed) Euclidean triangle, and let $G \subset \text{Isom}(\mathbb{R}^2)$ be the group generated by reflections in the sides of T. (i) Show that the tiles $T_g = \{g(T) : g \in G\}$ cover $\mathbb{R}^2$. (ii) Give an example where every point in $\mathbb{R}^2$ belongs to the interior of at most one tile. (iii) Give an example where every point in $\mathbb{R}^2$ belongs to the interior of infinitely many tiles. (Hint: the closure of G is a Lie subgroup of $\text{Isom}(\mathbb{R}^2)$.)

35. Let f and g be entire functions solving Fermat's equation, $f^n + g^n = 1$, with $n > 2$. Prove that f and g are constant. (Hint: consider f/g).

36. Let $f(z)$ be an entire function such that $f(z)$ is never zero and $f^{-1}(1)$ is finite. Prove that f is constant.

37. Prove there is a unique quadratic differential $q = q(z)\,dz^2$ with double poles at $0, 1, \infty$ with residues A, B, C respectively, and no other singularities. (Recall if $q = (a/z^2 + O(1/z))\,dz^2$ then $\text{Res}_0(q) = a$.)

38. Fix a sequence of real numbers $A_n \geq 0$, and let $\mathcal{F} \subset \mathcal{O}(\Delta)$ be the set of $f(z) = \sum a_n z^n$ such that $|a_n| \leq A_n$.

    (i) Prove that if $A_n = 2^n$, then $\mathcal{F}$ is not compact.

    (ii) Give a necessary and sufficient criterion for compactness of $\mathcal{F}$ in terms of A_n. (You do not need to prove it).

39. Find a group law $x * y$ on $G = (0,1) \subset \mathbb{R}$ such that the measure $du/(u(1-u))$ (appearing in the definition of the $B(\alpha, \beta)$) is G–invariant. (Hint: make a comparison to the measure dt/t on $(\mathbb{R}_+, \cdot)$.)

5 Elliptic functions and elliptic curves

Branch points are mathematics; branch cuts are sociology.

In this section we move beyond the Riemann sphere and discuss the theory of functions on compact Riemann surfaces of genus 1. This theory can be viewed from several different perspectives.

1. *Periodic functions.* The *periods* of a meromorphic function $f : \mathbb{C} \to \widehat{\mathbb{C}}$ are the additive group

$$\mathrm{Per}(f) = \{\lambda \in \mathbb{C} \; : \; f(z + \lambda) = f(z)\}.$$

   Provided f is nonconstant, $\mathrm{Per}(f)$ is a discrete group isomorphic to $\mathbb{Z}^a$ for some $0 \le a \le 2$. We have previously discussed the case where $\mathrm{Per}(f) = \mathbb{Z}$ — in this case we can write $f(z) = g(\exp(2\pi i z))$ for some meromorphic function g on $\mathbb{C}^*$.

   A *doubly–periodic function* on $\mathbb{C}$ is a meromorphic functions such that $\Lambda = \mathrm{Per}(f)$ has rank two. In this case we can write $\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$, and the compact parallelogram with vertices $(0, \alpha, \beta, \alpha + \beta)$ forms a fundamental domain for the action of Λ on $\mathbb{C}$. We can assume this basis (α, β) is *positively oriented*, meaning $\beta/\alpha \in \mathbb{H}$.

2. *Compact tori.* We also attach to Λ the compact Riemann surface $E = \mathbb{C}/\Lambda$, and consider doubly–periodic functions as elements of the function field $\mathrm{M}(E)$. The study of these *elliptic functions* is then a natural entry point into the general theory of compact Riemann surfaces.

3. *Algebraic curves.* For a third perspective, consider a smooth cubic curve $V \subset \mathbb{P}^2$. It can be shown that V is topologically a torus, and hence we expect its universal cover to be the complex plane. We can then aim to explicitly *uniformize* V by finding meromorphic functions f, g such that
$$P(f, g) = 0,$$

where $P(x, y)$ is a cubic equation defining V in affine coordinates. Suitably chosen, these functions will give a *covering map* $F : \mathbb{C} \to V$ with deck group Λ, providing an explicit isomorphism between V and $\mathbb{C}/\Lambda$.

4. *Branched coverings of $\widehat{\mathbb{C}}$.* A *hyperelliptic Riemann surface* is a compact surface X that admits a degree 2 holomorphic map $\pi : X \to \widehat{\mathbb{C}}$. It turns out that π is almost a covering map. That is, if we let $C \subset X$ denote the set of *critical point* of π, and $B = \pi(C) \subset \widehat{\mathbb{C}}$ the set of critical values, then

$$\pi : (X - C) \to (\widehat{\mathbb{C}} - B)$$

is a degree two covering map. Moreover, $|B| = 2g + 2$ must be even, and B itself determines X up to isomorphism! Indeed, there is a unique homomorphism

$$p : \pi_1(\widehat{\mathbb{C}} - B) \to \mathbb{Z}/2$$

that sends a small loop around each $b \in B$ to 1, and $\mathrm{Ker}(p)$ determines the covering space $X - C$, from which we can recover X.

When $|B| = 2$, it turns out $X = \widehat{\mathbb{C}}$, as can be seen by considering the map $\pi(z) = z^2$. But when $|B| = 4$, it turns out — remarkably — that X is isomorphic to a complex torus $E = \mathbb{C}/\Lambda$; and indeed, in this case we can normalize so $B = \{e_1, e_2, e_3, \infty\}$, and $\pi(z) = \wp(z)$ is the Weierstrass $\wp$–function.

5. *Field extensions.* From an algebraic perspective, a hyperelliptic Riemann surface X, equipped with a degree 2 map $\pi : X \to \widehat{\mathbb{C}}$, is the same as a degree 2 field extension $K/\mathbb{C}(z)$. To see this, one takes $K = \mathcal{M}(X)$. Such an extension is *ramified* over a finite set of places v of $\mathbb{C}(z) = \mathcal{M}(\widehat{\mathbb{C}})$, corresponding to the branch locus B above. Indeed, when $B = \{b_1, \ldots, b_{2g+2}\} \subset \mathbb{C}$, we can write

$$K = \mathbb{C}[x, y]/(y^2 - (x - b_1) \cdots (x - b_{2g+2})).$$

Again, the case $g = 1$ gives the theory of elliptic functions.

6. *Elliptic integrals.* Perhaps the most classical perspective on elliptic functions arises from the theory of indefinite integrals. Let $Q(x)$ be a polynomial. We can compute $\int dx/Q(x)$ explicitly using partial fractions; it involves logarithms, in general. Similarly, if $Q(x)$ is quadratic,

we can compute $\int dx/\sqrt{Q(x)}$ explicitly; the solution involves the inverse trigonometric functions.

But what about the function $f(x) = \int dx/\sqrt{Q(x)}$ where $\deg(Q) = 3$ or 4? In the case where Q is a quartic with real zeros, it is easy to see (by the Schwarz–Christoffel formula) that the analytic continuation of f maps $\mathbb{H}$ to a rectangle in the plane. In fact the same is true for cubics. Thus the *inverse* of f can be developed, by Schwarz reflection, to give a doubly–period function. In summary, there is a close relationship between *elliptic functions* and *elliptic integrals*.

In this section we will develop and connect all of these perspectives, starting for concreteness with the first.

5.1   Doubly–periodic functions

Let $\Lambda \subset \mathbb{C}$ be a lattice – meaning a discrete subgroup such that $E = \mathbb{C}/\Lambda$ is compact. Then $\Lambda \cong \mathbb{Z}^2$ as an abstract group. We can choose a basis such that

$$\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta.$$

Every point of E is represented by an essentially unique point in the *period parallelogram* with vertices $0, \alpha, \beta$ and $\alpha + \beta$.

Meromorphic functions $F : E \to \widehat{\mathbb{C}}$ are the same thing as *doubly periodic functions* $f : \mathbb{C} \to \widehat{\mathbb{C}}$, i.e. meromorphic functions satisfying

$$f(z + \alpha) = f(z + \beta) = f(z)$$

(and hence $f(z + \lambda) = f(z)$ for all $\lambda \in \Lambda$. For a fixed Λ, these functions form a *field*, denoted $\mathcal{M}(E)$.

Note that E forms a group under addition, and hence $E \subset \mathrm{Aut}(E)$. We also have an involution $\eta : E \to E$ given by $\eta(z) = -z$. This is the full automorphism group of E provided Λ is not a square or hexagonal lattice. In the latter cases, $\mathrm{Aut}(E)$ also contains a cyclic group $\mathbb{Z}/4$ or $\mathbb{Z}/6$.

Because of these automorphisms, if $F \in \mathcal{M}(E)$ then so is $F(-z)$ and $F(z + p), p \in E$.

We note that a holomorphic doubly-periodic function must be constant, by Liouville's theorem, because the period parallelogram is compact. So to

find other functions we must allow poles. The simplest are constructed for each $k \geq 3$ by setting

$$\zeta_k(z) = \sum_\Lambda \frac{1}{(z - \lambda)^k}.$$

This sum does not quite converge for $k = 2$; it can be made to converge by writing

$$\wp(z) = \frac{1}{z^2} + {\sum_\Lambda}' \frac{1}{(z - \lambda)^2} - \frac{1}{\lambda^2}.$$

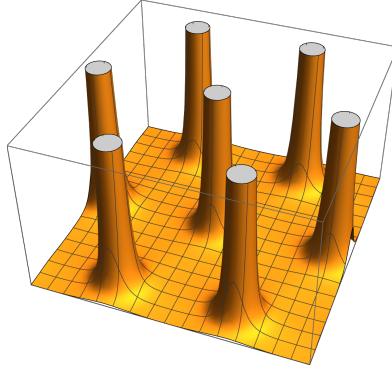The critical properties of $\wp(z)$ are that is it *even*, and it has a unique pole of order 2 on E.

Figure 5.1. The Weierstrass $\wp$-function for the hexagonal lattice.

Theorem 5.1 *The Weierstrass $\wp$-function is doubly-periodic.*

Proof. Note that $\wp'(z) = -2\zeta_3(z)$ is doubly-periodic. Thus $\wp(z + \alpha) = \wp(z) + A$ and $\wp(z + \beta) = \wp(z) + B$ for some constants A and B. But we also have $\wp(-z) = \wp(z)$; setting $z = \alpha/2$ and $\beta/2$, we find $A = B = 0$. ∎

Uniformization of cubic curves. We can now state the main theorems regarding elliptic functions.

Theorem 5.2 *The map $\pi : \mathbb{C} \to \mathbb{P}^2$ given by*

$$\pi(z) = (\wp(z), \wp'(z))$$

gives an isomorphism between the Riemann surface $E = \mathbb{C}/\Lambda$ and a smooth cubic curve V of the form $y^2 = (4x^3 + ax + b)$.

Here $\pi(0) = [0 : 1 : 0]$.

Theorem 5.3 *The field of all doubly-periodic functions for a given lattice Λ satisfies*

$$\mathcal{M}(\mathbb{C}/\Lambda) \cong \mathbb{C}[x, y]/(y^2 - (4x^3 + ax + b)),$$

where $(x, y) = (\wp, \wp')$. In particular, every doubly-periodic function is a rational function of $\wp$ and $\wp'$.

Irreducibility vs. smoothness. The quotient ring on right is a field because the given cubic polynomial is *irreducible*. This fact is independent of the choice of x and y. On the other hand, the cubic curve is *smooth* only when $4x^3 + ax + b$ has distinct zeros.

The projective plane and the line at infinity. Recall that $\mathbb{P}^2$ denotes the space of lines in $\mathbb{C}^3$; the line through (x, y, z) is denoted by $[x : y : z] \in \mathbb{P}^2$. There is a natural inclusion $\mathbb{C}^2 \to \mathbb{P}^2$ given by $(x, y) \mapsto [x : y : 1]$. The locus $z = 0$ defines the *line at infinity* $\mathbb{P}^1_\infty \subset \mathbb{P}^2$. We have $(x_n, y_n) \to s \in \mathbb{P}^1_\infty$ if $(x_n, y_n) \to \infty$ in $\mathbb{C}^2$ and $y_n/x_n \to s$.

Regarding the cubic above, it can be alternatively described as the locus in $\mathbb{P}^2$ where the *homogeneous* cubic equation

$$F(x, y, z) = y^2 z - x^3 + axz^2 + bz^3 = 0.$$

Setting $z = 0$ we find this cubic contains one point on $\mathbb{P}^1_\infty$, namely $[0 : 1 : 0]$. This reflects that fact that for large x and y, the cubic looks like $y^2 = x^3$ which has a vertical asymptote.

A typical line meets the cubic curve V in 3 distinct points; thus the line at infinity is special, since it meets V in a single point with multiplicity 3. This shows that $\mathbb{P}^1_\infty$ is *tangent* to V, and that V has a flex at infinity. (Altogether, V has 9 flexes which correspond to $E[3]$.)

Any smooth cubic curve $C \subset \mathbb{P}^2$ can be put into the form above by applying a change of coordinates (an automorphism of $\mathbb{P}^2$). Later, will see that *every* cubic curve occurs for suitable choice of Λ.

Holomorphic functions and 1–forms on E. Since $E = \mathbb{C}/\Lambda$ is compact, the maximum principle implies that all holomorphic functions on E are constants; that is, $\mathcal{O}(E) = \mathbb{C}$.

In general, the vector space of all holomorphic 1–forms on a Riemann surface X is denoted $\Omega(X)$. We have seen that $\Omega(\widehat{\mathbb{C}}) = 0$.

Theorem 5.4 *For $E = \mathbb{C}/\Lambda$ a complex torus, the space $\Omega(E)$ is 1–dimensional; it is spanned by the form $\omega = dz$.*

Proof. Since the 1-form dz on $\mathbb{C}$ is translation invariant, it descends to give a nowhere zero holomorphic 1–form ω on E. Now for any other form $\eta \in \Omega(E)$, $f = \eta/\omega$ is a holomorphic function, and hence constant; so $\Omega(E) = \mathbb{C} \cdot \omega$. ∎

Compact Riemann surfaces. In general, $\dim \Omega(X) = g$ on a compact Riemann surface X of genus g. The following two properties of *meromorphic* functions are easily established, and apply to $\mathcal{M}(E)$.

Theorem 5.5 *If X is compact, then any nonconstant meromorphic function $f : X \to \widehat{\mathbb{C}}$ has the same number of zeros and poles.*

Proof. Apply the residue theorem to df/f, or use general properties of proper maps. ∎

Zeros and poles of elliptic functions. On $\widehat{\mathbb{C}}$, any collection of distinct points (a_i) and (p_i) can be realized as the zeros and poles of $f \in \mathcal{M}(\widehat{\mathbb{C}})$. Remarkably, on an elliptic curve, there is a constraint.

Theorem 5.6 *If f has poles $p_1, \ldots, p_d$ and zeros $a_1, \ldots, a_d$ in E, then $\sum p_i = \sum a_i$ in the group law on E.*

Proof 1. Let P be a period parallelogram, and assume f has no zeros or poles in P. Then we have

$$\sum a_i - \sum p_i = (2\pi i)^{-1} \int_{\partial P} \frac{z f'(z)\, dz}{f(z)}.$$

We wish to show that this quantity lies in Λ. The integrals over opposite edges cancel, up to a terms of the form $\lambda(2\pi i)^{-1} \int_e f'(z)/f(z)\, dz$ with $\lambda \in \Lambda$. Since the f is periodic, it has an integral winding number $N(e)$ on each edge, and these terms have the form $N(e)\lambda \in \Lambda$. ∎

Proof 2. Let $f : E \to \widehat{\mathbb{C}}$ be a nonconstant meromorphic function. Then $f_*(dz) = 0$, since $\Omega(\widehat{\mathbb{C}}) = 0$. Now choose a path $C \subset \widehat{\mathbb{C}}$ running from 0 to ∞ and avoiding the critical values of f. Then $\widetilde{C} = f^{-1}(C) \subset E$ gives a collection of arcs connecting (a_i) and (p_i) in pairs, which we can assume have the same indices. We then have

$$0 = \int_C f_* \omega = \int_{\widetilde{C}} dz = \sum p_i - a_i \mod \Lambda.$$

An elliptic function with given poles and zeros. We will now show that one can *construct* an elliptic function with given zeros and poles subject only to the constraint above.

It is natural to try to construct an elliptic function by forming the Weierstrass product for the lattice Λ:

$$\sigma(z) = z \prod_{\Lambda}{}' \left(1 - \frac{z}{\lambda}\right) \exp\left(\frac{z}{\lambda} + \frac{z^2}{2\lambda^2}\right).$$

Then $(\log \sigma)'' = -\wp(z)$, which is doubly–periodic, from which it follows that

$$\sigma(z + \lambda) = \sigma(z) \exp(a_\lambda + b_\lambda z)$$

for some $a_\lambda, b_\lambda \in \mathbb{C}$. (Note that Hadamard theory only insures the exponential of a quadratic polynomial, but linear is *required* for the argument that follows.)

From this transformation law, it is easy to see that

$$\frac{\sigma(z - a_1) \ldots \sigma(z - a_n)}{\sigma(z - p_1) \ldots \sigma(z - p_n)}$$

defines an elliptic function whenever $\sum a_i = \sum p_i$. This demonstrates:

Theorem 5.7 *A divisor $D = \sum a_i - \sum p_i$ on E is principal (it arises from a meromorphic function) iff $\deg D = 0$ and $\sum(a_i - p_i) = 0$ in the group law on E.*

The Abel–Jacobi condition. The pushforward argument works on any compact Riemann surface X of genus g, to show that if $(f) = \sum a_i - p_i$, and ω_j is a basis for $\Omega(X)$, then

$$\sum_1^d \int_{p_i}^{a_i} \omega_j = 0 \in \mathbb{C}^g/\Lambda,$$

where $\Lambda \cong H_1(X, \mathbb{Z})$ is the *period lattice* in $\mathbb{C}^g$. In fact this necessary condition is also sufficient for $\sum a_i - p_i$ to be a principal divisor.

The differential equation. We can now give a cubic equation relating $x = \wp(z)$ and $y = \wp'(z)$.

Since $\wp'(z)$ is an odd function of degree three, its zeros coincide with the points $E[2]$ of order two on E. Using our chosen basis $\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$, we can explicitly label representatives of these points:

$$E[2] = \{0, c_1, c_2, c_3\} = \left\{ 0, \frac{\alpha}{2}, \frac{\beta}{2}, \frac{\alpha + \beta}{2} \right\}.$$

Let $e_i = \wp(c_i)$.

We note that the critical values e_i are distinct; indeed, since $\wp'(z_i) = 0$, the function $\wp(z) - e_i$ has a double zero at z_i, so it cannot vanish anywhere else. This shows:

The zeros of $\wp'(z)$ coincide with the points of order two c_1, c_2, c_3.

(Morally there is also a critical point at $z = 0$.) Consequently:

The function $\wp(z) - e_i$ has a double zero at c_i and no other zeros.

(Note: it is not at all easy to say where the two zeros of $\wp(z)$ lie, except in the case of symmetric lattices.)

Theorem 5.8 *For all $z \in \mathbb{C}$, we have*

$$\wp'(z)^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3).$$

Proof. From what we have seen above, the two sides are doubly-periodic functions with the same zeros and poles. Thus they are multiples of one another. They are equal since near zero they are both asymptotic to $4/z^6$. ∎

Corollary 5.9 *The map $\wp : E \to \widehat{\mathbb{C}}$ presents E as a 2-sheeted covering space of $\widehat{\mathbb{C}}$, branched over e_1, e_2, e_3 and ∞. It gives the quotient of E by the involution $z \mapsto -z$.*

Proof. By construction $\wp$ is even and 2-to-1, and we have just identified its critical points and values. ∎

Corollary 5.10 *The map $\pi(z) = (\wp(z), \wp'(z))$ maps $E = \mathbb{C}/\Lambda$ bijectively to the smooth projective cubic curve E defined by*

$$y^2 = 4(x - e_1)(x - e_2)(x - e_3).$$

Proof. Since E is compact, the function $x = \wp(z)$ maps E onto $\widehat{\mathbb{C}}$; and for a given x, the two possible values of y solving the equation above are given by $y = \wp'(\pm z)$. ∎
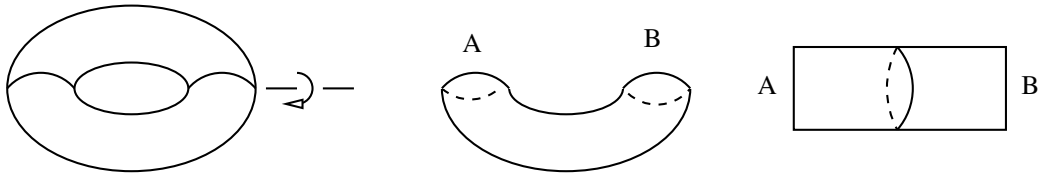

Figure 5.2. Visualizing the degree two map from E to $E/\eta \cong \widehat{\mathbb{C}}$, where $\eta(z) = -z$.

Power series expansion. Let $P = 0$ be the origin in E, and let $\mathcal{O}(d) = \mathcal{O}(d \cdot P) \subset \mathcal{M}(E)$ be the vector space of doubly–periodic functions $f(z)$ with a pole of order d at P and no other poles. Note that such a function is *uniquely determined* by its Laurent series at $z = 0$, up to the constant term:

$$f(z) = \frac{a_d}{z^d} + \cdots \frac{a_1}{z} + a_0 + O(z).$$

Theorem 5.11 *We have $\dim \mathcal{O}(d) = d$ for $d \geq 1$, and $\dim \mathcal{O}(0) = 1$.*

Proof. We have $a_1 = 0$ by applying the residue theorem to $\omega = f(z)\,dz$. Thus $\dim \mathcal{O}(d) \leq d$. Using $\wp(z)$ and $\zeta_k(z)$ we obtain the reverse inequality. ∎

The condition $\sum \mathrm{Res}_p(f\omega) = 0$ for every $f \in \mathcal{M}(X)$ and $\omega \in \Omega(X)$ holds on any compact Riemann surface X, and leads to the *Riemann–Roch theorem*.

Which element of $\mathcal{O}(2)$ is $\wp$? In fact, it is the unique element with $a_2 = 1$ and $a_0 = a_1 = 0$. This is evident from its definition, since each term in the

145

series for $\wp(z)$, apart from the initial term $1/z^2$, is normalized to vanish at the origin. The fact that $a_0 = 0$ gives rise to the fact that $\sum e_i = 0$, as we will see below.

The cubic made explicit. More precisely, using the expansion

$$
\begin{aligned}
\frac{1}{(z-\lambda)^2} &= \frac{d}{dz}\frac{1}{\lambda - z} = \frac{1}{\lambda}\frac{d}{dz}\frac{1}{(1 - z/\lambda)} \\
&= \frac{1}{\lambda^2} + \frac{2z}{\lambda^3} + \frac{3z^2}{\lambda^4} + \frac{4z^3}{\lambda^5} + \cdots,
\end{aligned}
$$

we obtain:

Theorem 5.12 *The Weierstrass $\wp$-function is given near $z = 0$ by:*

$$
\wp(z) = \frac{1}{z^2} + 3G_2 z^2 + 5G_3 z^4 + \cdots = \frac{1}{z^2} + \sum_1^\infty (2n+1)z^{2n}G_{n+1}
$$

where

$$
G_n = G_n(\Lambda) = \sideset{}{'}\sum_\Lambda \frac{1}{\lambda^{2n}}.
$$

Corollary 5.13 *We have $\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3$, where $g_2 = 60G_2$ and $g_3 = 140G_3$.*

Proof. Neglecting terms of order $O(z^2)$, we have:

$$
\begin{aligned}
\wp(z)^3 &= \frac{1}{z^6} + \frac{9G_2}{z^2} + 15G_3 \quad \text{and} \\
\wp'(z)^2 &= \left(\frac{-2}{z^3} + 6G_2 z + 20G_3 z^3 + \cdots\right)^2 = \frac{4}{z^6} - \frac{24G_2}{z^2} - 80G_3.
\end{aligned}
$$

Thus $4\wp(z)^3 - \wp'(z)^2 = 60G_2/z^2 + 140G_3 + O(z^2) = g_2\wp(z) + g_3$.   ∎

Corollary 5.14 *We have $\sum e_i = 0$.*

A mnemonic. To remember the power series for $1/(z - \lambda)^2$, it is useful to note that every term has the form az^b/λ^c, where (a, b, c) are adjacent, different integers; and that the series starts with $1/\lambda^2 = 1 \cdot z^0/\lambda^2$.

Function fields. We now return to the algebraic geometry of elliptic curves.

Theorem 5.15 *The function field of $E = \mathbb{C}/\Lambda$ is generated by $x = \wp$ and $y = \wp'$; more precisely, we have*

$$\mathcal{M}(E) = \mathbb{C}(x)[y]/(y^2 - 4x^3 + g_2 x + g_3).$$

Proof. To see that $\wp$ and $\wp'$ generate $\mathcal{M}(E)$ is easy. Any even function $f : E \to \widehat{\mathbb{C}}$ factors through $\wp$: $f(z) = F(\wp(z))$, and so lies in $\mathbb{C}(\wp)$. Any odd function becomes even when multiplied by $\wp'$; and any function is a sum of one even and one odd.

To see that the field is exactly that given is also easy. It amounts to showing that $\mathcal{M}(E)$ is of degree exactly two over $\mathbb{C}(\wp)$, and $\wp$ is transcendental over $\mathbb{C}$. The first assertion is obvious (else $\wp$ would be constant), and if the second fails we would have $\mathcal{M}(E) = \mathbb{C}(\wp)$, which is impossible because $\wp$ is even and $\wp'$ is odd. ∎

Remark: other constructions of elliptic functions. To construct elliptic functions of degree two: if $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$, one can first sum over $\mathbb{Z}$ to get:

$$f_1(z) = \sum_{-\infty}^{\infty} \frac{1}{(z-n)^2} = \frac{\pi^2}{\sin^2(\pi z)};$$

then

$$f(z) = \sum_{-\infty}^{\infty} f_1(z - n\tau)$$

converges rapidly, and defines an elliptic function of degree two of the form $f(z) = \wp(z) + C$ for some constant C. Similarly, if $E = \mathbb{C}^*/\langle z \mapsto \lambda z\rangle$, with $|\lambda| \neq 0, 1$, and we let

$$f_1(z) = \frac{z}{(z-1)^2},$$

then

$$f(z) = \sum_{-\infty}^{\infty} f_1(\lambda^n z)$$

defines a function on $\mathbb{C}^*$ with a double pole at $z = 1$ satisfying $f(\lambda z) = f(z)$; thus $F(z) = f(e^z)$ is a degree two doubly-periodic function for the lattice $\Lambda = \mathbb{Z}2\pi i \oplus \mathbb{Z}\log \lambda$. (Note that $f_1(z)$ has a double pole at $z = 1$, the identity element of the group $\mathbb{C}^*$, and that $|f_1(z)|$ goes to zero like $1/|z|$ as $|z| \to \infty$, so the sum converges geometrically fast.)

147

These functions are not quite canonical; there is a choice of direction in the lattice to sum over first. As a consequence they agree with $\wp(z)$ only up to an additive constant. This constant is a multiple of the important *quasimodular form*

$$G_1(\tau) = \sum_n \left(\sum_{m\,:\,(m,n)\neq(0,0)} \frac{1}{(m+n\tau)^2} \right). \tag{5.1}$$

5.2 Cubic curves and the addition law

Let us now expand our study of $E = \mathbb{C}/\Lambda$ as a cubic curve, and various perspectives on the group structure on E.

The addition law on an elliptic curve. Consider the curve $V \subset \mathbb{P}^2$ defined by $y^2 = 4x^3 - g_2 x - g_3$. As we have seen, we have an isomorphism $P : E \to V$ given in (x, y) coordinates by

$$P(z) = (\wp(z), \wp'(z)).$$

Note that $P(0) = \infty$, the unique point at infinity for V; it corresponds to the vertical asymptote of $V \subset \mathbb{C}^2$.

Theorem 5.16 *The following are equivalent for $a, b, c \in E$:*

1. *We have $a + b + c = 0$ in the group law on E.*

2. *The points $P(a), P(b)$ and $P(c)$ lie on a line $L \subset \mathbb{P}^2$. (And hence they coincide with $V \cap L$).*

3. *There exists an $f \in \mathcal{M}(E)$ with a pole of order 3 at 0 and zeros at $\{a, b, c\}$.*

4. *There exist A, B, C such that $\{a, b, c\}$ is the zero set of $f(z) = A + B\wp(z) + C\wp'(z)$.*

5. *We have*
$$\begin{vmatrix} 1 & 1 & 1 \\ \wp(a) & \wp(b) & \wp(c) \\ \wp'(a) & \wp'(b) & \wp'(c) \end{vmatrix} = 0.$$

Proof. The equivalence uses two facts: (i) the space $\mathcal{O}(3)$ of functions $f \in \mathcal{M}(E)$ with a pole of order ≤ 3 at $z = 0$, and otherwise holomorphic, is 3–dimensional, spanned by $(1, \wp, \wp')$; and (ii) $\{a, b, c\}$ is the zero set of some $f \in \mathcal{O}(3) \iff a + b + c = 0$ in the group law on E. $\blacksquare$

Corollary 5.17 *The map $z \mapsto -z$ on E is given by $(x, y) \mapsto (x, -y)$ on V.*

Proof. Then the line passes through ∞ which is the origin of E, consistent with the equation $p + (-p) + 0 = 0$. (Alternatively, observe that $x = \wp(z)$ is even and $y = \wp(z)$ is odd.) $\blacksquare$

Corollary 5.18 *Given $a, b \in V$, the point $c = a + b$ is constructed geometrically by drawing the line L through a and b, finding its third point of intersection (x, y) on V, and then reflecting through the x–axis to get $c = (x, -y)$.*

Corollary 5.19 *The point $2a$ can be constructed by taking the line tangent to E at a, finding its other point of intersection with E, and then negating its y coordinate.*

The addition law without coordinates. We remark that any smooth cubic $V \subset \mathbb{P}^2$ has an addition law with the property that $a + b + c = 0$ iff a, b, c are collinear. The law is uniquely determined by this condition and the choice of one of the 9 cusps to play the role of the origin. One can put V into Weierstrass form $(y^2 = P_3(x))$ by normalizing so that the origin is on the line at infinity, at its intersection with $x = 0$, and that the line through the 3 nonzero points of order 2 in V is given by $y = 0$.

A formula for addition. Using the geometric picture of addition, one can show:

$$\wp(a + b) + \wp(a) + \wp(b) = \frac{1}{4} \left(\frac{\wp'(a) - \wp'(b)}{\wp(a) - \wp(b)} \right)^2 \tag{5.2}$$

Note that both $\wp$ and $\wp'$ are needed to compute $\wp(a + b)$, just as both sin and cos appear in the formula for $\sin(a + b)$.

Proof. Let (x_i, y_i), $i = 1, 2, 3$, denote the coordinates of the points on the cubic curve V corresponding to $(a, b, -a - b)$ on E. Since these three points are collinear, they satisfy both the cubic

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

149

and a linear equation of the form

$$y = mx + c.$$

Just using the first two points, we find

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Then, plugging in $mx + c$ for y, we find that (x_1, x_2, x_3) satisfy a cubic equation of the form

$$4x^3 - m^2 x^2 + Ax + B = 0.$$

Thus shows that

$$x_1 + x_2 + x_3 = \frac{m^2}{4} = \left(\frac{y_2 - y_1}{x_2 - x_1} \right)^2,$$

which is equivalent to the formula above. ∎

Flexes. A point p on a smooth cubic curve $V \subset \mathbb{P}^2$ is a *flex* if there exists a line L such that $L \cap V = \{p\}$. By the additional law on an elliptic curve, this is equivalent to the condition that $3p = 0$ in E. In other words, we have:

Corollary 5.20 *The cubic curve V has 9 flexes. Under P, they corresponds to the 9 points $E[3] \subset E$.*

The Hessian. Any cubic curve V in $\mathbb{P}^2$, defined by a homogeneous polynomial equation $F(z_1, z_2, z_3) = 0$, determines a second cubic, HV, its *Hessian*, defined by

$$HF(x, y, z) = \det \left(\frac{d^2 F}{dz_i \, dz_j} \right) = 0. \tag{5.3}$$

The 9 points of $HV \cap V$ coincide with the 9 flexes of V. For much more on the geometry of cubic curves, see [Sal].

Maps to $\mathbb{P}^1$. For a general smooth cubic $V \subset \mathbb{P}^2$, we note that the pencil of lines through any point $p \in \mathbb{P}^2$ defines a $\pi_p : V \to \mathbb{P}^1$. If $p \notin V$, this map has degree 3, and its 6 critical points are the intersection of V with its *polar conic* C_p. If $p \in V$, this map has degree 2, and it is equivalent to $\wp$; and $\pi_p | V$ has a critical point at p iff p is a flex.

Remark: Dynamics of rational maps. There is a rational function $f(z)$ such that

$$\wp(2z) = f(\wp(z)).$$

The preimages of the critical points of f lie along the images of $E[2^n]$, so they give a way of visualizing the $\wp$ function; see Figure 5.3 and, for more details, [Mil2].
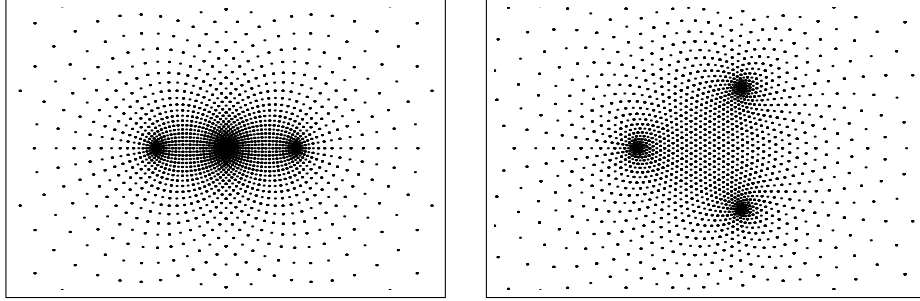

Figure 5.3. Image of $2^{-5}\Lambda$ under $\wp$, for $\Lambda = \mathbb{Z}[i]$ and $\mathbb{Z}[\rho]$.

The Hesse pencil. The Hesse family of cubics $V_t \subset \mathbb{P}^2$ is defined by

$$X^3 + Y^3 + Z^3 - 3tXYZ = 0. \tag{5.4}$$

The curve A_0 is the Fermat cubic, while A_∞ is the triangle defined by $XYZ = 0$. The base locus $B = V_0 \cap V_\infty$ of this pencil coincides with the flex points of every member of the family. This family is convenient for many reasons; for example, the Hessian cubic HV_t is also in the Hesse pencil.

Rational points. When V is defined $\mathbb{Q}$, the group law make $V(\mathbb{Q})$ into a finitely generated Abelian group. This means that one can often find infinitely many rational points on V, provided a few exist.

As an example, consider the curve V defined by

$$y^2 = x^3 + 8.$$

It clearly contains $P = (3, 1)$. Now on this curve in general, we have $dy/dx = m = 3x^2/2y$, and thus

$$x(2P) = \frac{9x^4}{4(x^3 + 8)} - 2x(P)$$

151

by (a limiting form) of equation (5.2). (There is no factor of $1/4$ because our cubic is monic.) Iterating this operation starting with $x_0 = 1$ we find

$$x_1 = \frac{-7}{4}, \ x_2 = \frac{31073}{2704}, \ x_3 = \frac{892933489418780033}{326211856441766464}, \ \text{etc.}$$

It is almost miraculous that $\sqrt{x_n^2 + 8}$ is rational for all these values.

5.3 The rectangular case

Suppose $\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$ with $\alpha \in \mathbb{R}_+$ and $\beta \in i\mathbb{R}_+$. Then the critical points $(0, c_1, c_2, c_3)$ of $\wp(z)$ bound a rectangle $S \subset \mathbb{C}$ (see Figure 5.4).

Theorem 5.21 *The value $\wp(z)$ is real if and only if z lies on one of the vertical or horizontal lines through $(1/2)\Lambda$.*

Proof. Since Λ is invariant under both negation and complex conjugation, we have $\wp(z) = \wp(\bar{z})$; since $\wp(z)$ is even and periodic, more generally we have

$$\overline{\wp(z)} = \wp(\lambda \pm \bar{z})$$

for any $\lambda \in \Lambda$. Thus $\wp(z)$ is real whenever

$$z \pm \bar{z} \in \Lambda.$$

The solutions to this equation are just the vertical and horizontal lines through $(1/2)\Lambda$. In particular, $\wp$ is real along the real and imaginary axes.

Now as z traverses ∂S, it moves monotonically along $\widehat{\mathbb{R}}$ passing through ∞ just once. Thus $\wp$ maps S bijectively to $\widehat{\mathbb{R}}$, and the same for $-S$. Since $\wp$ has degree two, there are no other preimages of $\widehat{\mathbb{R}}$. ∎

Corollary 5.22 *The map $\wp|S$ gives the unique conformal map from S to $-\mathbb{H}$ such that $\wp(c_i) = e_i$ for $i = 1, 2, 3$ and $\wp(0) = \infty$.*

Proof. The region S is a component of the complement of $\wp^{-1}(\widehat{\mathbb{R}})$, containing no critical points of $\wp$, so it maps homeomorphically to $\mathbb{H}$ or $-\mathbb{H}$. In fact the image is $-\mathbb{H}$ because $\arg \wp(z) \sim \arg(1/z^2) = -2\arg(z)$ for small z in S. ∎
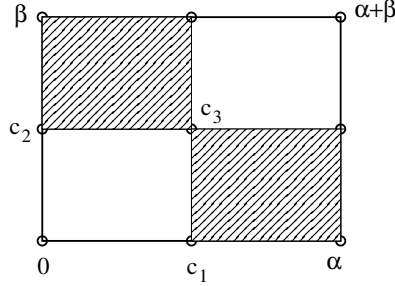
Figure 5.4. The shaded rectangles map under $\wp$ to $\mathbb{H}$; the unshaded ones, to $-\mathbb{H}$.

Theorem 5.23 *Any branch* $f : (\pm\mathbb{H}) \to \mathbb{C}$ *of* $\wp^{-1}$ *satisfies*

$$f(z) = \int \frac{d\zeta}{\sqrt{4\zeta^3 - g_2\zeta - g_3}}.$$

Proof. If $z = f(\zeta)$ then $\wp(z) = \zeta$, and hence $\wp'(z)f'(\zeta) = 1$. Consequently

$$f'(\zeta) = \frac{1}{\wp'(z)} = \frac{1}{\sqrt{4\wp(z)^3 - g_2\wp(z) - g_3}} = \frac{1}{\sqrt{4\zeta^3 - g_2\zeta - g_3}}.$$

$\blacksquare$

Corollary 5.24 *The inverse of* $\wp(z)$ *gives an explicit solution to the Schwarz–Christoffel formula for the map from* $\mathbb{H}$ *to a rectangle.*

From the Riemann mapping theorem to double periodicity. Conversely, starting from a *rectangle* $S = [0, \alpha/2] + [0, \beta/2] \subset \mathbb{C}$, we can construct a doubly–periodic function $f(z)$ by first choosing a Riemann map $f : S \to \mathbb{H}$, and then developing f by Schwarz reflection to obtain a degree 2 map $f : E = \mathbb{C}/\Lambda \to \widehat{\mathbb{C}}$, where $\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}i\beta$. If we arrange that $f(0) = \infty$, then we will have $f(z) = \alpha\wp(z) + \beta$ for some constants α, β.

Location of roots. Note that the argument shows the roots of $4x^3 - g_2x - g_3 = 0$ are real, and satisfy $e_2 < e_3 < e_1$. We also note that $\wp'(z) \in \mathbb{R}$ along the horizontal lines forming $\wp^{-1}(\mathbb{R})$, and $\wp'(z) \in i\mathbb{R}$ along the vertical lines, since $\wp$ most rotate the latter by 90° to make them real. Thus only the horizontal lines map to the real points of the cubic $y^2 = 4x^3 - g_2x - g_3$.
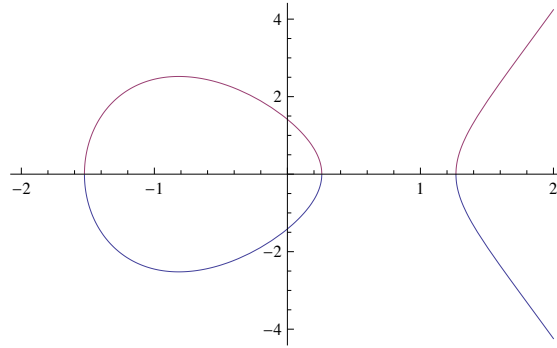
153

Figure 5.5. The elliptic curve $y^2 = 4x^3 - 8x + 2$.

Loops. In particular, the preimage under $\wp(z)$ of the real points of V is a *pair of disjoint loops* on E, corresponding to the horizontal lines through 0 and $\beta/2$.

Periods. This picture makes clear the close relationship between the generators α, β of Λ and the cubic polynomial

$$4x^3 - g_2 x - g_3 = 4(x - e_1)(x - e_2)(x - e_3);$$

namely, we have

$$\frac{\alpha}{2} = \int_{e_1}^{\infty} (4x^3 - g_2 x - g_3)^{-1/2} \, dx = \int_{e_2}^{e_3} (4x^3 - g_2 x - g_3)^{-1/2} \, dx,$$

and

$$\frac{\beta}{2i} = \int_{-\infty}^{e_2} (g_3 + g_2 x - 4x^3)^{-1/2} \, dx = \int_{e_3}^{e_1} (g_3 + g_2 x - 4x^3)^{-1/2} \, dx.$$

In all 4 integrals we take the positive square-root; thus, these are simply integrals of $|(\wp')^{-1}(x)|$.

In each case, the fact that the two integrals agree follows from Cauchy's theorem (applied to two homotopic loops in $\mathbb{C} - \{e_1, e_2, e_3\}$), *or* via a change of variables coming from a Möbius transformation that swaps one interval for the other.

More generally, we have:

154

Theorem 5.25 *The rectangular lattice Λ with invariants g_2 and g_3 is generated by the values of*

$$\pm \int_\gamma \frac{dz}{\sqrt{4z^3 - g_2 z - g_3}},$$

where γ ranges over all oriented loops in $\mathbb{C}$ which enclose exactly two roots of the cubic in the denominator.

The condition on the roots insures that the integrand can be defined continuously on a neighborhood of γ.

Evaluation of g_2 and g_3. The preceding formula sometimes permits the evaluation of g_2 and/or g_3. For example, we have

$$140 \sum_{\mathbb{Z}[\rho]}' \lambda^{-6} = 4 \left(\int_1^\infty \frac{dx}{\sqrt{x^3 - 1}} \right)^6 = \frac{256\pi^3 \Gamma[7/6]^6}{\Gamma[2/3]^6}.$$

This follows from the fact that for $\Lambda = \mathbb{Z}[\rho]$, we have $g_2 = 0$ and $g_3 > 0$ satisfies

$$\frac{1}{2} = \int_{e_1}^\infty \frac{dx}{\sqrt{4x^3 - g_3}},$$

where $e_1 = (g_3/4)^{1/3}$. Similarly, we have

$$60 \sum_{\lambda \in \mathbb{Z}[i]}' \lambda^{-4} = 4 \left(\int_1^\infty \frac{dx}{\sqrt{x^3 - x}} \right)^4 = \frac{64\pi^2 \Gamma[5/4]^4}{\Gamma[3/4]^4}.$$

5.4 Aside: Conics and singly-periodic functions

As a point of reference, we describe the parallel theory for conics and rank one lattices in $\mathbb{C}$.

Let $\Lambda \subset \mathbb{C}$ be a discrete subgroup isomorphic to $\mathbb{Z}$. Then we can rescale so $\Lambda = \mathbb{Z}$, and form, for $k \geq 2$, the singly-periodic functions

$$Z_k(z) = \sum_{-\infty}^\infty \frac{1}{(z-n)^k}$$

and the function

$$P(z) = \frac{1}{z} + \sum_{-\infty}^\infty{}' \frac{1}{(z-n)} - \frac{1}{n}.$$

Note that we have the Laurent series (with $\zeta(0) = -1/2$),

$$P(z) = \frac{1}{z} - z \sum' \frac{1}{n^2} - z^3 \sum' \frac{1}{n^4} - \cdots = \frac{2}{z} \sum_0^\infty \zeta(2k) z^{2k},$$

i.e. the coefficients of $P(z)$ carry interesting invariants of the lattice $\Lambda = \mathbb{Z}$.
We then find:

Theorem 5.26 *The map $P : \mathbb{C}/\mathbb{Z} \to \widehat{\mathbb{C}} - (\pm \pi i)$ is an isomorphism, sending $[0, 1]$ to $[\infty, -\infty]$.*

Proof. As we have seen,

$$P(z) = \pi \cot \pi z = \pi i \frac{e^{2\pi i z} + 1}{e^{2\pi i z} - 1} = A(e^{2\pi i z}),$$

where $A(z) = \pi i(z + 1)/(z - 1)$ sends the omitted values 0 and ∞ for $e^{2\pi i z}$ to $\pm i$. ∎

Just as we did for elliptic functions, we next note that the power series

$$P(z) = \frac{1}{z} - \frac{\pi^2 z}{3} + \cdots$$

gives

$$-P'(z) = \frac{1}{z^2} + \frac{\pi^2}{3} + \cdots \quad \text{and} \quad P(z)^2 = \frac{1}{z^2} - \frac{2\pi^2}{3} + \cdots$$

yielding the differential equation

$$-P'(z) = P(z)^2 + \pi^2.$$

Here we have used the fact that both $P(z)$ and $P'(z)$ are bounded as $|\operatorname{Im} z| \to \infty$. Put differently, we have:

Theorem 5.27 *The map $\pi : \mathbb{C} \to \mathbb{P}^2$ given by*

$$\pi(z) = (x, y) = (P(z), -P'(z))$$

gives an isomorphism between the $\mathbb{C}/\mathbb{Z} \cong \mathbb{C}^*$ and the smooth projective conic defined by $y = x^2 + \pi^2$ with two points removed.

156

The omitted values on the parabola $y = x^2 + \pi^2$ are $y = 0$, where $x = \pm i\pi$.

We could also verify the final identity using the fact that

$$-P'(z) = \frac{\pi^2}{\sin^2 \pi z}.$$

The fundamental period of Λ can now be expressed as

$$\int_{-\infty}^{\infty} \frac{dx}{y} = \int_{-\infty}^{\infty} \frac{dx}{\pi^2 + x^2} = \int_0^1 dz = 1.$$

I.e. the change of variables $x = P(z)$ transforms the integrand into the standard form dz on $\mathbb{C}/\mathbb{Z}$.

Remark. Any smooth conic in $\mathbb{P}^2$ is equivalent to the parabola above, so we have uniformized all conics. The familiar conics $x^2 + y^2 = 1$, $x^2 - y^2 = 1$ and $xy = 1$ are isomorphic to $\mathbb{C}/2\pi\mathbb{Z}$ or $\mathbb{C}/2\pi i\mathbb{Z}$, and are uniformized by $(\cos(t), \sin(t))$, $(\cosh(t), \sinh(t))$ and (e^t, e^{-t}) respectively. Note that all 3 curves have, over $\mathbb{C}$, two asymptotes, corresponding to the ends of $\mathbb{C}^*$.

5.5 Moduli spaces of elliptic curves

We now turn to the meta–theory of elliptic curves; that is, the study of space of *all* complex tori and their invariants. We will explain the meaning of the objects and maps in the statement below, and prove:

Theorem 5.28 *In the commutative diagram of maps between moduli spaces below:*

$$\mathbb{H}/\operatorname{SL}_2(\mathbb{Z}) \;\cong\; \mathcal{M}_1 \xleftarrow{\ \pi\ } \widetilde{\mathcal{M}}_1 \;\cong\; \mathbb{H}/\Gamma(2)$$
$$\Big\downarrow{\scriptstyle J} \qquad\qquad \Big\downarrow{\scriptstyle \lambda}$$
$$\mathbb{C} \;\cong\; \mathcal{M}_{0,4} \xleftarrow{\ p\ } \widetilde{\mathcal{M}}_{0,4} \;\cong\; \widehat{\mathbb{C}} - \{0, 1, \infty\},$$

the maps J and λ is isomorphisms, and p and π are degree 6 covering maps (of orbifolds).

Unwinding the definitions we will find:

Corollary 5.29 *Every smooth cubic curve*

$$V \subset \mathbb{P}^2,$$

157

and every degree two cover

$$V \to \mathbb{P}^1$$

branched oveer 4 points, is isomorphic to a complex torus $\mathbb{C}/\Lambda$.

Corollary 5.30 *For any $g_2, g_3 \in \mathbb{C}$ such that the discriminant*

$$\Delta = g_2^3 - 27 g_3^2 \neq 0,$$

there exists a lattice $\Lambda \subset \mathbb{C}$ with these invariants.

Corollary 5.31 *Let $E_i = \mathbb{C}/\mathbb{Z} \oplus \tau_i \mathbb{Z}$, $i = 1, 2$, $\tau_i \in \mathbb{H}$. Then $E_1 \cong E_2$ iff $J(\tau_1) = J(\tau_2)$.*

Moduli spaces. We begin by discussing $\mathcal{M}_1$, the *moduli space of lattices*. As a set, we let

$$\mathcal{M}_1 = \{\text{all lattices } \Lambda \subset \mathbb{C}\}/(\Lambda \sim \alpha\Lambda)$$

denote the space of lattices up to similarity. By associating to Λ the complex torus $E = \mathbb{C}/\Lambda$, we find

$$\mathcal{M}_1 = \{\text{all Riemann surfaces of genus one}\}/(\text{isomorphism}).$$

This is because any isomorphism $E_1 \cong E_2$ lifts to an automorphism $f(z) = \alpha z + \beta$ of the universal cover $\mathbb{C}$.

There is a natural map $\pi : \mathbb{H} \to \mathcal{M}_1$, given by

$$\pi(\tau) = [\mathbb{Z} \oplus \mathbb{Z}\tau] \in \mathcal{M}_1.$$

This map is surjective, since $\mathbb{Z}\alpha \oplus \mathbb{Z}\beta \sim \mathbb{Z} \oplus \mathbb{Z}(\beta/\alpha)$. We can think of a point $\tau \in \mathbb{H}$ as providing both a lattice $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$ and a marking isomorphism,

$$\phi : \mathbb{Z}^2 \to \Lambda,$$

coming from the basis $(\tau, 1)$. All other bases for Λ with the same orientation as this one are given by $(a\tau + b, c\tau + d)$, where $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. The marked lattice with this new basis is equivalent to $(g(\tau), 1)$ where

$$g(\tau) = \frac{a\tau + b}{c\tau + d}.$$

Thus forgetting the marking altogether is the same as taking the quotient of $\mathbb{H}$ by the action of all $g \in \mathrm{SL}_2(\mathbb{Z})$, and hence:
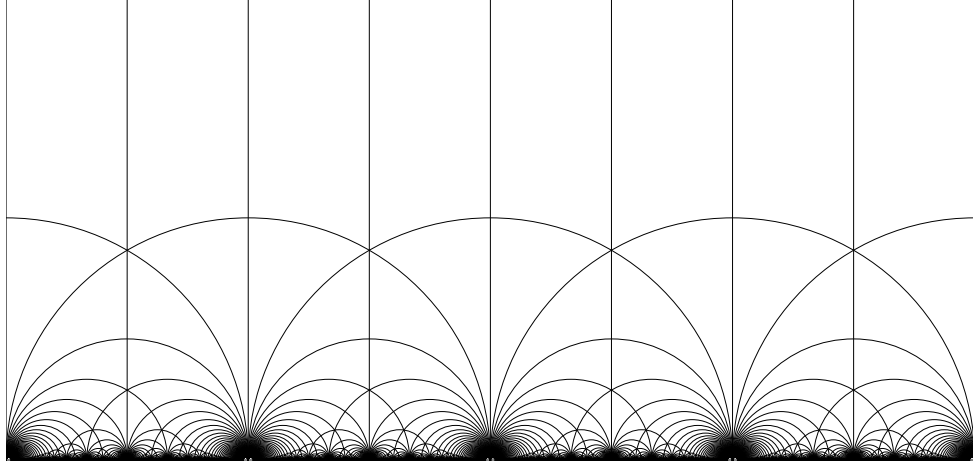
Figure 5.6. Tiling of $\mathbb{H}$ for $\mathrm{SL}_2(\mathbb{Z})$.

Theorem 5.32 *The moduli space $\mathcal{M}_1$ is naturally isomorphic to the complex orbifold $\mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$.*

Here $\mathrm{SL}_2(\mathbb{Z})$ acts by Möbius transformations.

Fundamental domains. We can always normalize a lattice Λ by $\mathbb{C}^*$ so that its shortest nonzero vector is $z = 1$ and the shortest vector in $\Lambda - \mathbb{Z}$ is $\tau \in \mathbb{H}$. Then $|\tau| \geq 1$, and $\mathbb{Z} + \tau \subset \Lambda$ so $|\mathrm{Re}\,\tau| \leq 1/2$.

Moreover $\mathbb{Z} \oplus \mathbb{Z}\tau = \Lambda$; otherwise there would be a vector $v \in \Lambda - \mathbb{Z}$ of the form $v = a + b\tau$ with $a, b \in [0, 1/2]$; but then

$$|v| < 1/2 + |\tau|/2 \leq |\tau|,$$

contrary to our assumption that the shortest vector in $\Lambda - \mathbb{Z}$ has length $|\tau|$. (This argument is special to lattices in $\mathbb{R}^2$!)

Conversely, if τ lies in this fundamental domain then it is the 'next shortest' vector in $\mathbb{Z} \oplus \tau\mathbb{Z}$. This shows:

Theorem 5.33 *The region $|\mathrm{Re}\,\tau| \leq 1/2$, $|\tau| > 1$ in $\mathbb{H}$ is a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$.*

The quotient $\mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$ is an *orbifold* because the action is not quite free, e.g. $\tau \mapsto -1/\tau$ fixes $\tau = i$. In fact $\mathcal{M}_1$ is the $(2, 3, \infty)$ orbifold.

159

The subgroup $\Gamma(2)$. We now turn to the degree 6 covering space $\widetilde{\mathcal{M}}_1 \to \mathcal{M}_1$ whose objects are complex tori with a *level structure*, that with a choice of an *ordering* or *labelling* of the points of order two:

$$E[2] = \{0, c_1, c_2, c_3\}.$$

For a given value of τ, we can set

$$E[2] = (0, c_1, c_2, c_3) = (0, 1, \tau, 1 + \tau)/2 \subset E.$$

It turns out that $\mathrm{SL}_2(\mathbb{Z})$ does not preserve this ordering, in fact it permutes it arbitrarily; but the subgroup

$$\Gamma(2) = \{A \in \mathrm{SL}_2(\mathbb{Z}) \ : \ A = I \bmod 2\}$$

preserves the ordering of $E[2]$, in fact we have an exact sequence

$$0 \to \Gamma(2) \to \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/2) \to 0,$$

and $\mathrm{SL}_2(\mathbb{Z}/2) \cong S_3$ acts naturally on $E[2] \cong (\mathbb{Z}/2)^2$, realizing every possible permutation of its nonzero elements.

The space $\widetilde{\mathcal{M}}_1$. Thus the space of complex tori E with labelled points of order 2 is isomorphic to

$$\widetilde{\mathcal{M}}_1 = \mathbb{H}/\Gamma(2).$$

We have a natural degree 6 covering map of orbifolds,

$$\pi : \widetilde{\mathcal{M}}_1 \to \mathcal{M}_1.$$

It can be shown that $\Gamma(2)/(\pm I)$ is a free group, and in fact

$$\Gamma(2) = \langle \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \rangle,$$

with a fundamental domain given the ideal quadrilateral

$$Q = \mathrm{hull}(\infty, -1, 0, 1).$$

As we have seen before, using the Riemann mapping theorem, we have:

Theorem 5.34 *The moduli space $\widetilde{\mathcal{M}}_1$ is isomorphic to $\widehat{\mathbb{C}} - \{0, 1, \infty\}$.*
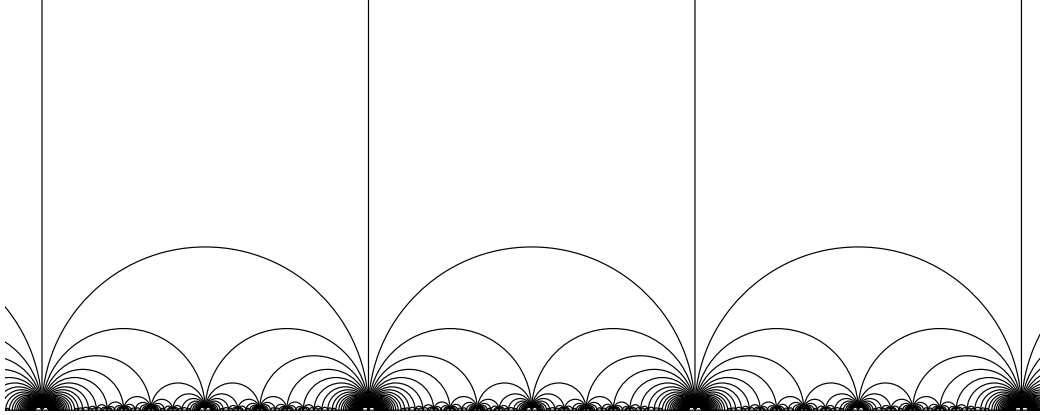
Figure 5.7. Tiling for $\Gamma(2)$.

Cross-ratios and $\widetilde{\mathcal{M}}_{0,4}$. Next we turn to the more elementary consideration of the space $\widetilde{\mathcal{M}}_{0,4}$ of *ordered* quadruples of *distinct points* (z_0, z_1, z_2, z_3) in $\widehat{\mathbb{C}}$, *up to* the action of $\operatorname{Aut}\widehat{\mathbb{C}}$. The latter group allows us to normalize uniquely so the first three points are $(0, 1, \infty)$, and then the last point,

$$\lambda = [z_1, z_2, z_3, z_4]$$

is the *cross ratio* of the (z_i). Any value distinct from $0, 1, \infty$ can arise, and thus we have

$$\widetilde{\mathcal{M}}_{0,4} \cong \widehat{\mathbb{C}} - \{0, 1, \infty\}.$$

If we reorder the quadruple, the cross-ratio changes, ranging among the six values

$$\lambda, \frac{1}{\lambda}, 1 - \lambda, \frac{1}{1 - \lambda}, \frac{\lambda}{\lambda - 1}, \frac{\lambda - 1}{\lambda}.$$

(There is a natural action of S_4, but the Klein 4-group $\mathbb{Z}/2 \times \mathbb{Z}/2$ acts trivially.) There are two cross–ratios with order 3 symmetries: namely ρ and $\bar{\rho}$, the zeros of $\lambda^2 - \lambda + 1$. There are also three special values with order 2 symmetry: $-1, 1/2$ and 2.

Unordered 4–tuples. The automorphism group of $\widetilde{\mathcal{M}}_{0,4}$ as a Riemann surface is S_3; its action amounts to forgetting the ordering of the 4 points, and we have a degree 6 quotient map

$$\pi : \widetilde{\mathcal{M}}_{0,4} \to \mathcal{M}_{0,4} \cong \mathbb{C}$$

161

given by

$$\pi(\lambda) = \frac{4}{27} \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(1 - \lambda)^2}.$$

This map is normalized so that

$$\pi(\rho) = 0 \quad \text{and} \quad \pi(2) = 1.$$

In these coordinates, $\mathcal{M}_{0,4}$ is the $(2, 3, \infty)$ orbifold.

The modular function $J(\tau)$. We now define a remarkable map

$$\lambda : \widetilde{\mathcal{M}}_1 \to \widetilde{\mathcal{M}}_{0,4}$$

by sending E, with $E[2]^* = \{c_1, c_2, c_3\}$, to the cross–ratio

$$\lambda = [\infty, e_1, e_2, e_3],$$

where $e_i = \wp(c_i)$. To extract an invariant of E which does not depend on this choice of ordering, we define

$$J(\tau) = p(\lambda(\tau)).$$

One can check that

$$J(\tau) = \frac{g_2^3}{g_2^3 - 27g_3^2} = \frac{g_2^3}{\Delta}.$$

Note that $J(i) = 1$ and $J(\rho) = 0$, since in these two cases $g_3 = 0$ and $g_2 = 0$ respectively.

This completes the definition of the maps in Theorem 5.28.

Isomorphism. We can now give a few different proofs of:

Theorem 5.35 *The map* $\lambda : \widetilde{\mathcal{M}}_1 \to \widetilde{\mathcal{M}}_{0,4}$ *is a bijection.*

Proof 1. For a modern geometric proof, one constructs an inverse to λ by simplying taking V to be the curve defined by

$$y^2 = x(x - 1)(x - \lambda)$$

in $\mathbb{P}^2$, with the nowhere–vanishing holomorphic 1–form $\omega = dx/y$, and oberving that $V \cong \mathbb{C}/\Lambda$ for some lattice Λ. In fact, Λ is simply the lattice of periods of the form ω, or even more simply, the span of the values of the integrals

$$\pm \int_\gamma \frac{dx}{\sqrt{x(x - 1)(x - \lambda)}},$$

162

where γ ranges over all loops in $\mathbb{C}$ enclosing exactly 2 of the points $(0, 1, \lambda)$.

An isomorphism between the universal cover $\widetilde{V}$ of V and $\mathbb{C}$ is constructed by integrating the nowhere–vanishing form ω to get an immersion $\iota : \widetilde{V} \to \mathbb{C}$, and the observing that ι is proper, because $(V, |\omega|)$ is geodesically complete. ∎

Proof 2. We first check that λ is a *proper map*. For this it suffices to show that $J(\tau) \to \infty$ as $\tau = iy \to \infty$. In fact in the limit we have

$$g_2 = 120\zeta(4) = (4/3)\pi^4 \quad \text{and} \quad g_3 = 280\zeta(6) = (8/27)\pi^6,$$

which gives $\Delta = g_2^3 - 27g_3^2 = \pi^{12}(4^3 - 8^2)/27 = 0$, and hence $J(\tau) \to \infty$.

Since both the domain and range of λ are triply–punctured spheres, it now follows by Euler characteristic considerations that λ has degree one. ∎

Proof 3. (Classical.) By directly computing e_1, e_2 and e_3 as functions of $\tau = iy$, one shows that λ tends to zero as $y \to \infty$, and indeed λ has a simple zero in the local parameter $z = \exp(\pi i \tau)$ near infinity. This gives properness and degree 1 at the same time.

In more detail, we can estimate

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

for $\tau = 2iy$ using the fact that

$$\wp(z) = C + p(z), \quad \text{where} \quad p(z) = \sum_n \frac{\pi^2}{\sin^2(\pi(z + n\tau))}.$$

The constant C is irrelevant since we are taking differences. We have $(c_1, c_2, c_3) = (1/2, iy, iy+1/2)$. Since $1/\sin(z)$ decays rapidly as $\mathrm{Im}(z) \to \infty$, we find (setting $e_i = p(c_i)$, ignoring C):

$$e_1 = p(1/2) \to \pi^2, \; e_2 \to 0 \quad \text{and} \quad e_3 \to 0.$$

More precisely, we have

$$\sin(\pi iy) = i\sinh(\pi y) \quad \text{and} \quad \sin(\pi(iy + 1/2) = \cosh(\pi y),$$

so (taking the terms for $p(z)$ with $n = 0$ and -1):

$$e_2 \sim -8\pi^2 \exp(-2\pi y) \quad \text{and} \quad e_3 \sim +8\pi^2 \exp(-2\pi y).$$

This gives
$$\lambda \sim 16\exp(-2\pi y) = 16\exp(\pi i\tau) = 16q^{1/2},$$
where $q^{1/2} = \exp(\pi i\tau)$ is the natural local parameter on $\mathbb{H}/(2\mathbb{Z})$ corresponding to the cusp at infinity for $\Gamma(2)$. This shows λ has a zero of order exactly one at this cusp; similarly, it assumes to values 1 and ∞ with multiplicity one at the others, and thus λ has degree one. ∎

Proof 4. It suffices to show that the modular *form* Δ has a simple zero at $\tau = \infty$. But when regarded as a form, Δ gives a nowhere vanishing, symmetric, holomorphic cubic differential on $\widetilde{\mathcal{M}_1} \cong \widehat{\mathbb{C}} - \{0, 1, \infty\}$. Such a differential must have 6 more poles than zeros, so it has double poles at each puncture. This translates into a simple zero for Δ. (See the discussion of modular forms which follows.) ∎

Completion of the proof of Theorem 5.28. It only remains to show that J is a bijection. But this follows from the fact that λ is a bijection and that the maps π and p are S_3–equivariant. ∎

5.6 Interlude

We collect two brief remarks.

The $\boldsymbol{\lambda}$ function. It is also possible to show directly, using the explicit formula for λ and $\wp(z)$:

Corollary 5.36 *The map λ gives an explicit Riemann map from the ideal triangle spanned by 0, 1 and ∞ to $\mathbb{H}$, sending $(0, 1, \infty)$ to $(1, \infty, 0)$.*

Thus the Schwarzian $S\lambda^{-1}$ is given by the quadratic differential $Q(0, 0, 0)$ discussed earlier (see equation (4.2)), and λ^{-1} itself can be given as a ratio of solutions to a linear differential equation of order two.

We then also have a one line proof of the Little Picard Theorem: 'consider $\lambda^{-1} \circ f : \mathbb{C} \to \mathbb{H}$.'

> *The question recently arose in conversation whether a dissertation of 2 lines could deserve and get a Fellowship... in mathematics the answer is yes....*

(Theorem.) An integral function never 0 or 1 is a constant. (Proof.) $\exp\{i\Omega(f(z))\}$ is a bounded integral function.

...But I can imagine a referee's report: 'Exceedingly striking and a most original idea. But, brilliant as it undoubtedly is, it seems more odd than important; an isolated result, unrelated to anything else, and not likely to lead anywhere.'

<div align="right">—J. E. Littlewood.</div>

Cf. [Bol, p.39–40]: Here Ω is our λ^{-1}.

Lattices in $\mathbb{R}^n$ One can similarly discuss the moduli space of lattices in $\mathbb{R}^n$.

Indeed, we can specify a marked lattice $\Lambda \subset \mathbb{R}^n$, normalized so $\mathbb{R}^n/\Lambda$ has volume one, by a linear map

$$T : \mathbb{Z}^n \to \Lambda \subset \mathbb{R}^n$$

with $T \in \mathrm{SL}_n(\mathbb{R})$. Two such lattices are similar iff they differ by a rotation, i.e. $T_1 = R \circ T_2$ where $R \in \mathrm{SO}_n(\mathbb{R})$. Thus the Teichmüller space of lattices in $\mathbb{R}^n$ is the homogeneous space

$$\mathcal{H} = \mathrm{SO}_n(\mathbb{R})\backslash \mathrm{SL}_n(\mathbb{R}).$$

For $n = 2$ we have $\mathcal{H} \cong \mathbb{H}$ because $\mathrm{SO}_2(\mathbb{R})$ is the stabilizer of $\tau = i$ for the usual action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$.

Finally $T_1(\mathbb{Z}^n) = T_2(\mathbb{Z}^n)$ iff $T_1 \circ T_2^{-1}$ gives an isomorphism of $\mathbb{Z}^n$ to itself, which shows

Theorem 5.37 *The moduli space of lattices in $\mathbb{R}^n$ is isomorphic to*

$$\mathcal{L}(\mathbb{R}^n) = \mathrm{SO}_n(\mathbb{R})\backslash \mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z}).$$

5.7 Modular forms

In this section we will find an explicit formula for $J(\tau)$, and give a complete analysis of the holomorphic forms and functions on $\mathbb{H}$ (with controlled growth) that are invariant under $\mathrm{SL}_2(\mathbb{Z})$.

Modular forms: analytic perspective. A holomorphic function $f : \mathbb{H} \to \mathbb{C}$ is said to be a *modular form* of weight $2k$ for $\mathrm{SL}_2(\mathbb{Z})$ if it has the following properties:

1. $f(\tau + 1) = f(\tau)$;

2. $f(-1/\tau) = \tau^{2k} f(\tau)$; and

3. $\sup_{\text{Im } \tau > 1} |f(\tau)| < \infty$.

The vector space of all such forms will be denoted by M_k. The product forms of weight $2k$ and 2ℓ has weight $2(k+\ell)$, so $\oplus M_k$ forms a graded ring.

The first two properties imply that

$$f(g(\tau)) = (c\tau + d)^{2k} f(\tau)$$

for all $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \text{SL}_2(\mathbb{Z})$. The first and third properties imply that we can write

$$f(\tau) = \sum_0^\infty a_n q^n,$$

where $q = \exp(2\pi i \tau)$. In other words, $f(\tau)$ descends to a holomorphic function on $\Delta^* = \mathbb{H}/\mathbb{Z}$ with a removable singularity at $q = 0$. In particular, $f(\tau) \to a_0$ as $\text{Im } \tau \to \infty$. One sometimes writes $f(i\infty) = a_0$.

If $f(i\infty) = 0$, one says that $f(\tau)$ is *cusp form*. The cusp forms (*Spitzenformen*) give a natural subspace $S_k \subset M_k$.

Examples.

1. Any holomorphic invariant $F(\Lambda)$ of lattices that is homogeneous of degree $-2k$ and satisfies the right growth conditions defines a modular form of weight $2k$. That is, if $F(t\Lambda) = t^{-2k}\Lambda$, then the function

$$f(\tau) = F(\mathbb{Z} \oplus \tau\mathbb{Z})$$

satisfies $f(\tau + 1) = f(\tau)$ and

$$f(-1/\tau) = F(\mathbb{Z} \oplus \tau^{-1}\mathbb{Z}) = F(\tau^{-1}(\mathbb{Z} \oplus \tau\mathbb{Z})) = \tau^{2k} f(\tau).$$

2. In particular, for $k \geq 2$ the *Eisenstein series*

$$G_k(\tau) = \sideset{}{'}\sum (n + m\tau)^{-2k}$$

are modular forms of weight $2k$. Evidentally G_k converges to $2\zeta(2k)$ as $\text{Im } \tau \to \infty$, so these are holomorphic at infinity but not cusp forms.

3. Similarly for the normalized functions $g_2 = 60G_2$, $g_3 = 140G_3$.

Modular forms: algebraic perspective. If we associate to $f(\tau)$ the form

$$\omega = f(\tau) \, d\tau^k,$$

then the first two conditions above just say that $g^*\omega = \omega$ for all $g \in \mathrm{SL}_2(\mathbb{Z})$. (Recall that $g'(z) = (cz + d)^{-2}$.) Since $dz = (2\pi i)^{-1} dq/q$, the third condition says that ω descends to a form

$$\omega = (2\pi i)^{-k} f(q) \left(\frac{dq}{q} \right)^k$$

on $\Delta^* = \mathbb{H}/\Gamma$ with at worst a pole of order k at $q = 0$.

Thus ω also descends to an S_3-invariant holomorphic k-form on

$$\widetilde{\mathcal{M}}_{0,4} = \widehat{\mathbb{C}} - \{0, 1, \infty\} = \mathbb{H}/\Gamma(2)$$

with poles of order $\leq k$ at $0, 1$ and ∞. The converse is also true. This shows:

Theorem 5.38 *The space M_k is naturally isomorphic to the space of S_3-invariant rational forms $\omega(z) \, dz^k$ on $\widehat{\mathbb{C}}$ with poles of order $\leq k$ at $0, 1, \infty$ and no other poles.*

This provides a purely algebraic perspective on the initially transcendental–looking theory of automorphic forms.

Forms of weight $2k$. We now wish to find all the rational holomorphic k-forms $\omega \in M_k$. Here are 2 key properties of any nonzero $\omega \in M_k$:

(i) The zeros $Z(\omega)$ are invariant under S_3 and satisfy $|Z(\omega)| \leq k$, when counted with multiplicity, and determine ω up to constant multiple.

(ii) The poles of ω at $0, 1, \infty$ are all of the same order, which can be k, $k - 2$, $k - 4$, etc. We have $\omega \in S_k$ iff the order of pole is $k - 2$ or less.

Part (i) comes from the fact that ω has at most $3k$ poles, hence at most $3k - 2k = k$ zeros. The parity constraint in (ii) comes from the fact that $g(z) = 1 - z$ leaves ω invariant, and $g'(\infty) = -1$.

Residues. We remark that a modular form $\omega \in M_k$ has just the right order of pole to allow for a nonzero *residue* at $(0, 1, \infty)$. In general the residue of such a form is defined by

$$\mathrm{Res}_0 \left((a_k/z^k + a_{k-1}/z^{k-1} + \cdots) \, dz^k \right) = a_k.$$

The quantity is invariant under changes of coordinates.

For a modular form, the residues at the three points $0, 1, \infty$ must be equal, and for a cusp form they must vanish.

Examples.

1. We have $M_0 = \mathbb{C}$; it consists of the constant functions.

2. We have $\dim M_1 = 0$; the group S_3 has no fixed point, so there is no candidate for $Z(\omega)$. Alternatively, the residues must all be zero, by the residue theorem, forcing ω to be holomorphic; but the only holomorphic 1–form on $\widehat{\mathbb{C}}$ is $\omega = 0$.

3. We have $\dim M_2 = 1$. The only possibility is $Z(\omega) = \{\rho, \bar\rho\}$; thus the quadratic differential

$$F_2 = \frac{(z^2 - z + 1)\, dz^2}{z^2(z-1)^2}$$

spans M_2. (We have met this differential before in the study of Schwarz triangle functions.)

4. Similarly, $\dim M_3 = 1$; it is spanned by the cubic differential

$$F_3 = \frac{(z-2)(z-1/2)(z+1)\, dz^3}{z^3(z-1)^3},$$

which has zeros at $-1, 1/2, 2$.

5. The products F_2^2 and $F_3 F_2$ span M_4 and M_5. This is because S_3 has unique invariant sets with $|Z| = 4$ and 5.

6. We have $\dim M_6 = 2$; it is spanned by F_2^3 and F_3^2. These two forms are linearly independent because they have different zero sets.

7. *The discriminant*

$$D_6 = \frac{4}{27}(F_2^3 - F_3^2) = \frac{dz^6}{z^4(z-1)^4},$$

is the first nontrivial cusp form; it has poles of order 4 at $(0, 1, \infty)$ and no zeros. It spans S_6.

8. Ratios of forms of the same weight give all S_3-invariant rational functions. Indeed, as we have seen, an isomorphism $\widehat{\mathbb{C}}/S_3 \cong \widehat{\mathbb{C}}$ sending $(\infty, \rho, 2)$ to $(\infty, 0, 1)$,is given by

$$F(z) = \frac{F_2^3(z)}{F_2(z)^3 - F_3(z)^2} = \frac{4}{27} \frac{(z^2 - z + 1)^3}{z^2(z - 1)^2}.$$

More generally, any S_3-invariant rational function of degree d can be expressed as a ratio of modular forms of degree d.

9. If desired, we can regard F_2 and F_3 as forms on $\widehat{\mathbb{C}}/S_3$ by substituting $w = J(z)$. They then come:

$$F_2 = \frac{dw^2}{4w(w-1)} \quad \text{and} \quad F_3 = \frac{dw^3}{8w^2(w-1)}.$$

We also find $D_6 = (1/432)(w^{-4}(w-1)^{-3}) \, dw^6$.

Cusp forms. We let $S_k \subset M_k$ denote the space of *cusp forms* with poles of order $\leq k - 1$ at $0, 1, \infty$.

Theorem 5.39 *The map $\omega \mapsto D_6\omega$ gives an isomorphism from M_k to S_{k+6}.*

Proof. Any $F \in M_k$ has poles of order $\leq k$, so FD_6 has poles of order $\leq k + 4$ and hence lies in S_{k+6}. Conversely, if $G \in S_{k+6}$ then $F = G/D_6$ has poles of order at most k at $0, 1, \infty$, and it is otherwise holomorphic since D_6 has no zeros. ∎

Every space M_k, $k \geq 2$ contains a form of the type $F_2^i F_3^j$ which is *not* a cusp form; thus $M_k \cong \mathbb{C} \oplus S_k$. By inspection, dim $S_k = 0$ for $k \leq 5$. On the other hand, any cusp form is divisible by D_6. This shows:

Corollary 5.40 *We have dim $M_{6n+1} = n$, and dim $M_{6n+i} = n + 1$ for $i = 0, 2, 3, 4, 5$.*

Corollary 5.41 *The forms F_2 and F_3 generate the ring $M = \oplus M_k$.*

Corollary 5.42 *The forms $F_2^i F_3^j$ with $2i + 3j = k$ form a basis for M_k.*

Proof. By the preceding Corollary these forms span M_k, and the number of them agrees with dim M_k as computed above. ∎

Corollary 5.43 *The ring of modular forms M is isomorphic to the polynomial ring $\mathbb{C}[F_2, F_3]$.*

Proof. The preceding results shows the natural map from the graded ring $\mathbb{C}[F_2, F_3]$ to M is bijective on each graded piece. ∎

We have already seen that, under the isomorphism $\mathcal{M}_1 \cong \mathcal{M}_{0,4}$, an analytic modular form is the same as an algebraic modular form. Thus we also have:

Corollary 5.44 *The ring of modular forms is isomorphic to $\mathbb{C}[g_2, g_3]$.*

Corollary 5.45 *For every $k \geq 2$, the quantity $\sum' \lambda^{-2k}$ can be expressed as a polynomial in $\sum' \lambda^{-4}$ and $\sum' \lambda^{-6}$.*

Values of g_2 and g_3. We now note that for $\tau = i$, the zeros of $4\wp^3 - g_2\wp - g_3$ must look like $(-1, 0, 1)$ and thus $g_3(i) = 0$. Similarly for $\tau = \rho$ the zeros are arrayed like the cube roots of unity and hence $g_3(\rho) = 0$.

To determine the values at infinity, we observe that as $\tau \to \infty$ we have

$$G_2(\tau) \to 2\zeta(4) = \pi^4/45$$

and

$$G_3(\tau) \to 2\zeta(6) = 4\pi^6/945.$$

This gives the values $g_2(\infty) = 60G_2(\infty) = (4/3)\pi^4$ and $g_3(\infty) = 140G_3(\infty) = (8/27)\pi^6$, and thus

$$(g_2^3/g_3^2)(i\infty) = 27.$$

The cusp form Δ and the modular function J. By the preceding calculation, the *discriminant*

$$\Delta(\tau) = g_2^3(\tau) - 27g_3(\tau)^2$$

is a cusp form of weight 12. We have seen such a form is unique up to a scalar multiple, and is nonvanishing everywhere except for a simple zero at infinity.

This form has a natural meaning: up to a multiple, it is the discriminant of the cubic polynomial $4x^3 - g_2 x - g_3$. In fact, since the sum of the roots of this polynomial is zero, it has a multiple root iff it has the form

$$4(x - a)^2(x + 2a) = 4x^2 - 12a^2 x + 8a^3,$$

which happens iff

$$g_2^3 - 27g_3^2 = (12^3 - 3^3 \cdot 8^2)a^3 = 0.$$

As we have seen, this polynomial has distinct roots whenever g_2 and g_3 come from a lattice; this explains why $\Delta(\tau)$ has no zeros in $\mathbb{H}$.

Theorem 5.46 *We have $J(\tau) = g_2^3(\tau)/\Delta(\tau)$.*

Proof. First note that this is a ratio of forms of weight 12 and hence a modular function, i.e. it is invariant under $\mathrm{SL}_2(\mathbb{Z})$. Since $g_2(\infty) \neq 0$, it has a pole at infinity. Since the forms have weight 12, it defines a rational function of degree 6 on $\widetilde{\mathcal{M}}_1$ and hence of degree one on $\mathcal{M}_1$. Finally we have $J(\rho) = 0$ since $g_2(\rho) = 0$, and $J(i) = 1$ since $g_3(i) = 0$. ∎

Remarks. One can use residues to determine the exact relationship between F_2 and g_2. Namely, $F_2 \sim (1/4)\,dw^2/w^2$, while for $q = \exp(2\pi i \tau)$ we have $dq = 2\pi i q\, d\tau$, and hence

$$g_2 = 60 G_2(\tau)\, d\tau^2 \sim \frac{4\pi^4}{3} \frac{dq^2}{(2\pi i q)^2} = -\frac{\pi^2}{3} \frac{dq^2}{q^2}.$$

Thus $g_2 = -(4\pi^2/3)F_2$. One can similarly calculate g_3/F_3.

If we let $j(\tau) = 1728 J(\tau)$, then

$$j(\tau) = \frac{1}{q} + 744 + \sum_1^\infty a_n q^n$$

with $a_n \in \mathbb{Z}$.

Connections with additive and multiplicative number theory. Incredibly, we have

$$\Delta(q) = (2\pi i)^{12} q \prod_1^\infty (1 - q^n)^{24}.$$

This gives a close connection between the theory of modular forms and the partition function $p(n)$, since

$$\prod(1 - q^n)^{-1} = \sum p(n)q^n.$$

The coefficients power series for $G_k(\tau)$ involves the function $\sigma_k(n) = \sum_{d|n} d^k$.

The Riemann function $\zeta(s)$ arises as the Mellin transform of a theta function, and then modularity translates into the functional equation.

5.8 Exercises.

1. Let $\Lambda \subset \mathbb{R}^n$ be a discrete group. Prove that $\Lambda \cong \mathbb{Z}^a$ for some $a \leq n$, and $a = n$ if and only if $\mathbb{R}^n/\Lambda$ is compact.

2. Let $\Lambda \subset \mathbb{R}^n$ be a lattice, i.e. a discrete subgroup isomorphic to $\mathbb{Z}^n$. Choose a sequence of vectors $a_1, a_2, \ldots, a_n \in \Lambda$ such that a_1 is a shortest nonzero vector, and (for $i > 1$) a_i is a shortest vector linearly independent from $(a_1, a_2, \ldots, a_{i-1})$.

   Is it always the case then that $\Lambda = \mathbb{Z}a_1 \oplus \cdots \oplus \mathbb{Z}a_n$?

3. Let $\Lambda \subset \mathbb{C}$ be a lattice, let $X = \mathbb{C}/\Lambda$ and let $\text{End}(\Lambda) = \{\alpha \in \mathbb{C} : \alpha\Lambda \subset \Lambda\}$. Show that for each $\alpha \in \text{End}(\Lambda)$, the formula $[f(z)] = [\alpha z]$ defines an analytic covering map $f : X \to X$ of degree $|\alpha|^2$.

   For what values of $\alpha \in \mathbb{C}$ does there exist a lattice with $\alpha \in \text{End}(\Lambda)$?

   Conclude that $\text{End}(\mathbb{Z} \oplus \mathbb{Z}\tau) = \mathbb{Z}$ for almost all values of τ.

4. Where are the zeros of the $\wp$-function for the lattice $\mathbb{Z} \oplus \mathbb{Z}i$? For the lattice $\mathbb{Z} \oplus \mathbb{Z}\omega$? (Here $\omega = (1 + \sqrt{-3})/2$.)

5. Let $\omega = \exp(2\pi i/3)$, and let $\Lambda = \mathbb{Z} \oplus \mathbb{Z}i$ and $L = \mathbb{Z} \oplus \mathbb{Z}\omega$. Prove that $\sum'_\Lambda \lambda^{-6} = \sum'_L \lambda^{-4} = 0$. (The prime means $\lambda = 0$ is omitted.)

6. Prove that there exists a pair of nonconstant meromorphic functions on $\mathbb{C}$ such that $f(z)^3 + g(z)^3 = 1$. (Hint: show the equation $x^3 + y^3 = 1$ defines the same elliptic curve in $\mathbb{P}^2$ as $y^2 = 4x^3 - 1$.)

7. Prove the identity:

$$60 \sideset{}{'}\sum_{\lambda \in \mathbb{Z}[i]} \lambda^{-4} = 4 \left(\int_1^\infty \frac{dx}{\sqrt{x^3 - x}} \right)^4.$$

8. Given $\tau \in \mathbb{H}$ let $\wp(z)$ be the Weierstrass $\wp$–function for the lattice $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$, and let

$$f(z) = \sum_{n \in \mathbb{Z}} \frac{\pi^2}{\sin^2(\pi(z - n\tau))}.$$

Prove that $f(z) = \wp(z) + C$, and express the constant C in terms of the function $G_1(\tau)$ defined by equation (5.1).

9. (i) What is the fundamental group of the real projective plane. $\mathbb{RP}^2$?

(ii) Let $V \subset \mathbb{RP}^2$ be the closed cubic curve defined by $y^2 = x(x-1)(x+1)$ (including its unique point at infinity). Show there is no topological disk D^2 such that
$$V \subset D^2 \subset \mathbb{RP}^2.$$

Conclude that there is no automorphism g of $\mathbb{RP}^2$ such that $g(V) \subset \mathbb{R}^2 \subset \mathbb{RP}^2$.

10. Where are the 9 flexes of the cubic curve $V \subset \mathbb{CP}^2$ defined by $x^3 + y^3 = 1$? How many of these are real?

11. Let $V \subset \mathbb{P}^2$ be a smooth cubic curve, and let $F \subset V$ be its 9 flexes.

(i) Show that any line that passes through 2 points of F passes through a third.

(ii) Show that there are no 9 points in $\mathbb{R}^2$ with this property, unless they all lie on the same line.

12. Let $\wp(z)$ be the Weierstrass $\wp$–function for $E = \mathbb{C}/\Lambda$, and consider the meromorphic 1–form
$$\omega = \frac{\wp(z)\wp''(z)}{\wp'(z)} \, dz$$
on E. (i) Prove that $\mathrm{Res}_0(\omega) = 0$. (ii) Using the residue theorem, prove that $e_1 + e_2 + e_3 = 0$.

Here $e_i = \wp(c_i)$ are the images of the nonzero points of order 2 in the group E.

173

13. State and prove a 'double angle' formula for the Weierstrass $\wp$-function. That is, find a rational function $f(z)$ (that may depend on (g_2, g_3)) such that $\wp(2z) = f(\wp(z))$.

14. (Continuation.) Let $X = \mathbb{C}/\Lambda$ be a complex torus, and define a map $F : X \to X$ by $[F(z)] = [2z]$. Show that F has a dense orbit on X, i.e. that there exists a $p \in X$ such that $\overline{\{F^n(p) : n > 0\}} = X$, where $F^n(p) = F(F(\cdots F(p)))$.

    Then prove the rational function $f(z)$ of the double-angle formula has a dense orbit on $\widehat{\mathbb{C}}$.

15. Suppose $\sum'\lambda^{-4}$ and $\sum'\lambda^{-6}$ are the same for two lattices $\Lambda_1, \Lambda_2 \subset \mathbb{C}$. Does it follow that $\Lambda_1 = \Lambda_2$?

16. Suppose $(x, y) = (\wp(z), \wp'(z))$ satisfies $y^2 = 4x^3 + ax + b$ with $a, b \in \mathbb{R}$, and the polynomial $4x^3 + ax + b = 0$ has only one real root. What can you say about the shape of the lattice Λ used to define $\wp(z)$?

17. Let $\Lambda = \mathbb{Z} \oplus iy\mathbb{Z}$. Compute the limit as $y \to \infty$ of $\wp(z)$, $G_2(\Lambda)$ and $G_3(\Lambda)$, and verify directly that in the limit we still have
$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3.$$
(Note: the limit of $\wp(z)$ will be a trigonometric function.)

18. Prove that in the Hesse pencil of cubics V_t defined by equation (5.4), we have $HV_t = V_s$ for some s, and compute s as a function of t. Here HV_t is defined using equation (5.3).

19. Let $T \subset \mathbb{C}$ be the region bounded by the triangle with vertices $(0, 1, 1 + i)$. Give an explicit formula for a conformal mapping $f : T \to \mathbb{H}$ in terms of the Weierstrass $\wp$-function for a suitable lattice.

20. Let $\Lambda = \mathbb{Z}\lambda_1 \oplus \mathbb{Z}\lambda_2$ be a lattice in $\mathbb{C}$, with associated Weierstrass $\wp$-function $\wp(z)$. Let
$$\sigma(z) = z \prod_\Lambda{}'(1 - z/\lambda) \exp(z/\lambda + (1/2)(z/\lambda)^2) \qquad (5.5)$$
be the canonical product with zeros at the points of Λ.

    Prove there is a unique odd meromorphic function $\zeta(z)$ on $\mathbb{C}$ such that $\zeta'(z) = -\wp(z)$, and relate $\zeta(z)$ to $\sigma(z)$.

21. (Continuation.) Show that $\zeta(z + \lambda_i) = \zeta(z) + \eta_i$ for suitable $\eta_i \in \mathbb{C}$. Show these 'dual periods' satisfy

$$\det \begin{pmatrix} \eta_1 & \eta_2 \\ \lambda_1 & \lambda_2 \end{pmatrix} = 2\pi i.$$

(Hint: translate the period parallelogram so it is centered at $z = 0$, where $\zeta(z)$ has a pole, and integrate around its boundary.)

22. (Continuation.) The canonical product satisfies $\sigma(z + \lambda_i) = \exp(a_i + b_i z)\sigma(z)$. Express a_i and b_i in terms of λ_i and η_i.

23. Let $[z_i]$ be a finite sequence of distinct points in $E = \mathbb{C}/\Lambda$. Let

$$F_i(z) = \sum_{j=N_i}^{1} a_{i,j}(z - z_i)^{-j}$$

be a Laurent tail centered at $z_i \in \mathbb{C}$. Prove there exists a meromorphic function $f : E \to \widehat{\mathbb{C}}$ with poles at z_i and these Laurent tails if and only if $\sum_i a_{i,1} = 0$. (Hint: use the residue theorem in one direction, and translates of the function

$$\zeta(z) = \frac{1}{z} + {\sum_{\Lambda}}' \frac{1}{z - \lambda} + \frac{1}{\lambda}$$

in the other.)

24. Prove that $\wp'(z) = -\sigma(2z)/\sigma(z)^4$, where $\sigma(z)$ is defined by (5.5).

25. Let $\Lambda = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$ be a rectangular lattice, with $\alpha \in \mathbb{R}_+$ and $\beta \in i\mathbb{R}_+$, let $V(\mathbb{R}) = \mathbb{R}^2 \cap V$ be the real points of the corresponding cubic curve, and let $P(z) = (\wp(z), \wp'(z))$.

(i) Show that $\gamma_1(t) = P(t)$, $t \in \mathbb{R}$ parameterizes one component of $V(\mathbb{R})$, and $\gamma_2(t) = P(t + \beta/2)$ parameterizes the other.

(ii) Draw a picture of $V(\mathbb{R})$, and put arrows on its two components to indicate the directions of the parameterizations above. Justify your answer.

26. Prove that $J(z) = g_2(z)^3/(g_2(z)^3 - 27g_3(z)^2)$ directly from the equation

$$J(z) = \frac{4}{27} \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(1 - \lambda)^2},$$

using the fact that $\lambda(z)$ is the cross-ratio of the roots of the cubic equation $4x^3 - g_2(z)x - g_3(z) = 0$, together with ∞.

27. For any $\lambda \neq 0, 1, \infty$ let $S_\lambda = \widehat{\mathbb{C}} - \{0, 1, \infty, \lambda\}$. Let $\mathrm{Aut}(S_\lambda)$ be the group of holomorphic bijections $f : S_\lambda \to S_\lambda$.

   (a) Prove that every $f \in \mathrm{Aut}(S_\lambda)$ is a Möbius transformation.

   (b) Prove that for all λ, $\mathrm{Aut}(S_\lambda)$ contains a subgroup isomorphic to $\mathbb{Z}/2 \times \mathbb{Z}/2$. How does this group permute the points $\{0, 1, \infty, \lambda\}$?

   (c) Find all values of λ such that $|\mathrm{Aut}(S_\lambda)| > 4$ and identify the group in each case.

28. Let $\rho(z)\,|dz|$ be the hyperbolic metric on $X = \widehat{\mathbb{C}} - \{0, 1, \infty\}$; that is, the unique metric such that $\lambda : \mathbb{H} \to X$ is a local isometry. Show there are constants $A, B > 0$ such that for small values of z,

$$\frac{A}{|z \log |z||} \leq \rho(z) \leq \frac{B}{|z \log |z||}.$$

   (Hint. Recall $|dz|/|z \log |z||$ is the hyperbolic metric on $\Delta^* - \Delta - \{0\}$. Using the fact that $\lambda(z + 2) = \lambda(z)$, write $\lambda(z) = f(\exp(\pi i z))$ where $f : \Delta^* \to X$ is a covering map. Observe that $f(0) = 0$ and $f'(0) \neq 0$ to complete the proof.)

29. Let L be length in the hyperbolic metric of the closed geodesic γ on $X = \widehat{\mathbb{C}} - \{0, 1, \infty\}$ that makes a figure 8 around 0 and 1. Show that $L = \log(17 + 12\sqrt{2})$. (Hint: show that γ corresponds to a matrix of trace 6 in $\pi_1(X) \cong \Gamma(2)$.)

30. Recall for any lattice we defined $G_n(\Lambda) = \sum' \lambda^{-2n}$. It can be shown (using the fact that there is a unique modular form of weight 8) that there is a universal constant A such that $G_2(\Lambda)^2 = AG_4(\Lambda)$ for all Λ. What is the value of A? (Hint: $\sum' 1/n^8 = \pi^8/4725$.)

176

31. What are the orbits of $\mathrm{SL}_2(\mathbb{Z})$ acting on $\mathbb{Q} \cup \{\infty\} \subset \partial\mathbb{H}$? What are the orbits of $\Gamma(2)$?

32. For $k \geq 0$, let M_k denote the space of rational forms $f(z)\, dz^k$ on $\widehat{\mathbb{C}}$ with poles of order at most k at $0, 1, \infty$ and no other poles. (i) Determine $\dim M_k$ and give a basis for this space. (ii) Give a finite set of generators for $\oplus M_k$ as a graded ring. (This is the ring of modular functions for $\Gamma(2)$.)

33. Prove that $\lambda(i/2) = 12\sqrt{2} - 16$.

    (Hint: Letting $X_\tau = \mathbb{C}/\mathbb{Z} \oplus \tau\mathbb{Z}$, $\lambda(\tau)$ is the cross-ratio of the critical values (suitably ordered) of any degree two map $f_\tau : X_\tau \to \widehat{\mathbb{C}}$. For the square torus, $\tau = i$, choose f_i so its critical values are the roots of $z^4 + 1 = 0$. Then one can choose $f_{i/2}(z) = (f_i(z) + f_i(z)^{-1})/2$, and find that its critical values are $\{-1, -\sqrt{1/2}, \sqrt{1/2}, 1\}$.)

34. Prove that
$$60 \sideset{}{'}\sum_{\lambda \in \mathbb{Z}[i]} \lambda^{-4} = \frac{64\pi^2\Gamma[5/4]^4}{\Gamma[3/4]^4}.$$

35. (Continuation.) Relate the equation above to the conformal radius of the unit square (with sides of length two), previously shown to be $4\Gamma[3/4]^2/\pi^{3/2} = 1.078705\ldots$.

36. Prove that
$$\int_1^\infty dx/\sqrt{x^3 - 1} = (1/3)B(1/2, 1/6),$$
where B is the Beta function, and use this result to show that
$$140 \sideset{}{'}\sum_{\mathbb{Z}[\rho]} \lambda^{-6} = \frac{256\pi^3\Gamma[7/6]^6}{\Gamma[2/3]^6}.$$

37. (i) Prove that $\mathrm{Isom}(\mathbb{H})$ is generated by reflections through geodesics.

    (ii) Prove that $\mathrm{Isom}(\mathbb{H}^3)$ is generated by reflections through (two–dimensional) planes.

6 Epilogue

To conclude we reflect on the disparate strands of complex analysis that are intertwined in the theory of elliptic curves:

- *Mittag–Leffler's theorem*, which underlies the construction of $\wp(z)$ as a function with prescribed poles at points of a lattice Λ;

- *Rouché's theorem*, which shows the divisor of a meromorphic function on $E = \mathbb{C}/\Lambda$ is *balanced*: it satisfies $\sum z_i = \sum p_i$;

- *Hadamard's factorization theorem*, which produces the simplest entire function $\sigma(z)$ with zeros on Λ, and in turn leads to the *existence* of a meromorphic function with a given balanced divisor;

- The *Riemann mapping theorem*, which can be used to construction $\wp(z)$ geometrically, starting with the conformal map from a rectangle to $\mathbb{H}$;

- The *Schwarz–Christoffel formula*, which connects the cubic relation between $\wp'(z)$ and $\wp(z)$ to the integral formula, going backwards, for a map from $\mathbb{H}$ to a rectangle;

- The Γ *function*, which figures in the explicit evaluation of g_2 and g_3 for the most symmetric lattices, $\mathbb{Z}[i]$ and $\mathbb{Z}[\rho]$; and

- *Picard's Theorem* — which can be proved using the modular invariant λ as an explicit covering map from $\mathbb{H}$ to the triply punctured sphere.

The next chapter in the story is the general theory of compact Riemann surfaces.

> *Slowly crossing the deck from the scuttle, Ahab leaned over the side and watched how his shadow in the water sank and sank to his gaze, the more and the more that he strove to pierce the profundity.* — Melville, Chapter 132 (The Symphony)

References

[Ad]   W. A. Adkins. Local algebraicity of some analytic hypersurfaces. *Proc. Amer. Math. Soc.* **49** (1980), 546–548.

[BD] F. Berteloot and J. Duval. Sur l'hyperbolicité de certain complèmentaires. *Enseign. Math.* **47** (2001), 253–267.

[Bol] B. Bollabas, editor. *Littlewood's Miscellany.* Cambridge University Press, 1986.

[CJ] L. Carleson and P. Jones. On coefficient problems for univalent functions. *Duke Math. J.* **66** (1992), 169–206.

[Gam] T. W. Gamelin. *Uniform Algebras.* Prentice–Hall, 1969.

[Gol] G. M. Goluzin. *Geometric Theory of Functions of a Complex Variable.* Amer. Math. Soc, 1969.

[GM] J. D. Gray and S. A. Morris. When is a function that satisfies the Cauchy-Riemann equations analytic? *The American Mathematical Monthly* **85** (1978), 246–256.

[Har] G. H. Hardy. *Divergent Series.* Chelsea, 1991.

[Ko] S. V. Kolesnikov. On the sets of nonexistence of radial limits of bounded analytic functions. *Russian Acad. Sci. Sb. Math.* **81** (1995), 477–485.

[MH] J. E. Marsden and M. J. Hoffman. *Basic Complex Analysis.* W. H. Freeman, 1999.

[Mc] C. McMullen. Moduli spaces in genus zero and inversion of power series. *Enseign. Math.* **60** (2014), 25–30.

[Mil1] J. Milnor. *Singular Points of Complex Hypersurfaces*, volume 61 of *Annals of Math. Studies.* Princeton University Press, 1968.

[Mil2] J. Milnor. On Lattès maps. In P. G. Hjorth and C. L. Petersen, editors, *Dynamics on the Riemann sphere*, pages 9–44. European Math. Soc., 2006.

[Mon] M. G. Monzingo. Why are 8:18 and 10:09 such pleasant times? *Fibonacci. Quart.* **2** (1983), 107–110.

[Re] R. Remmert. *Classical Topics in Complex Function Theory.* Springer, 1998.

[Sa]   R. Salem. *Algebraic Numbers and Fourier Analysis.* Wadsworth, 1983.

[Sal]   G. Salmon. *Higher Plane Curves.* Hodges, Foster and Figgis, 1879.