Riemann Surfaces

Dr C. Teleman[1]

Lent Term 2003

[1]Originally LaTeXed by James Lingard — please send all comments and corrections to teleman@dpmms.cam.ac.uk

Lecture 1

What are Riemann surfaces?

1.1 Problem: Natural algebraic expressions have 'ambiguities' in their solutions; that is, they define multi-valued rather than single-valued functions.

In the real case, there is usually an obvious way to fix this ambiguity, by selecting one *branch* of the function. For example, consider $f(x) = \sqrt{x}$. For real x, this is only defined for $x \geq 0$, where we conventionally select the positive square root (Fig.1.1).

We get a continuous function $[0, \infty) \to \mathbb{R}$, analytic everywhere except at 0. Clearly, there is a problem at 0, where the function is not differentiable; so this is the best we can do.

In the complex story, we take '$w = \sqrt{z}$' to mean $w^2 = z$; but, to get a single-valued function of z, we must make a choice, and a *continuous* choice requires a *cut* in the domain.

A standard way to do that is to define '$\sqrt{z}$' $: \mathbb{C} \setminus \mathbb{R}^- \to \mathbb{C}$ to be the square root with positive real part. There is a unique such, for z away from the negative real axis. This function is continuous and in fact *complex-analytic*, or *holomorphic*, away from the *negative* real axis.

A different choice for $\sqrt{z}$ is the square root with positive imaginary part. This is uniquely defined away from the *positive* real axis, and determines a complex-analytic function on $\mathbb{C} \setminus \mathbb{R}^+$.

In formulae: $z = re^{i\theta} \Longrightarrow \sqrt{z} = \sqrt{r}e^{i\theta/2}$, but in the first case we take $-\pi < \theta < \pi$, and, in the second, $0 < \theta < 2\pi$.

Either way, there is no continuous extension of the function over the missing half-line: when z approaches a point on the half-line from opposite sides, the limits of the chosen values of $\sqrt{z}$ differ by a sign. A restatement of this familiar problem is: starting at a point $z_0 \neq 0$ in the plane, any choice of $\sqrt{z_0}$, followed continuously around the origin once, will lead to the opposite choice of $\sqrt{z_0}$ upon return; z_0 needs to travel around the origin twice, before $\sqrt{z_0}$ travels once.

Clearly, there is a problem at 0, but the problem along the real axis is our own — there is no discontinuity in the function, only in the choice of value. We could avoid this problem by allowing multi-valued functions; but another point of view has proved more profitable.

The idea is to replace the complex plane, as domain of the multi-valued function, by the graph of the function. In this picture, the function becomes projection to the w-axis, which is well-defined single-valued! (Fig. 1.2)

In the case of $w = \sqrt{z}$, the graph of the function is a closed subset in $\mathbb{C}^2$,

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

In this case, it is easy to see that the function $w = w(z)$,

$$S \to \mathbb{C}, \qquad (z,w) \mapsto w$$

defines a homeomorphism (*diffeomorphism*, in fact) of the graph S with the w-plane. This is exceptional; it will not happen with more complicated functions.

The graph S is a very simple example of a (concrete, non-singular) Riemann surface. Thus, the basic idea of Riemann surface theory is to replace the domain of a multi-valued function, e.g. a function defined by a polynomial equation

$$P(z,w) = w^n + p_{n-1}(z)w^{n-1} + \cdots + p_1(z)w + p_0(z)$$

by its graph

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

and to study the function w as a function on the 'Riemann surface' S, rather than as a multi-valued function of z.

This is all well, provided we understand

- what kind of objects Riemann surfaces are;
- how to do complex analysis on them (what are the analytic functions?)

The two questions are closely related, as will become clear when we start answering them properly, in the next lecture; for now, we just note the moral definitions.

1.2 Moral definition. An *abstract Riemann surface* is a surface (a real, 2-dimensional manifold) with a 'good' notion of complex-analytic functions.

The most important examples, and the first to arise, historically, were the graphs of multi-valued analytic functions:

1.3 Moral definition: A *(concrete) Riemann surface* in $\mathbb{C}^2$ is a locally closed subset which is locally — near each of its points (z_0, w_0) — the graph of a multi-valued complex-analytic function.

1.4 Remarks:

(i) *locally closed* means *closed in some open set*. The reason for 'locally closed' and not 'closed' is that the domain of an analytic function is often an open set in $\mathbb{C}$, and not all of $\mathbb{C}$. For instance, there is no sensible way to extend the definition of the function $z \mapsto exp(1/z)$ to $z = 0$; and its graph is not closed in $\mathbb{C}^2$.

(ii) Some of the literature uses a more restrictive definition of the term *multi-valued function*, not including things such as $\sqrt{z}$. But this need not concern us, as we shall not really be using multi-valued functions in the course.

The Riemann surface $S = \{(z, w) \in \mathbb{C}^2 \mid z = w^2\}$ is identified with the complex w-plane by projection. It is then clear what a holomorphic function on S should be: an analytic function of w, regarded as a function on S. We won't be so lucky in general, in the sense that Riemann surfaces will not be identifiable with their w- or z-projections. However, a class of greatest importance for us, that of *non-singular* Riemann surfaces, is defined by the following property:

1.5 Moral definition: A Riemann surface S in $\mathbb{C}^2$ is *non-singular* if each point (z_0, w_0) has the property that

- either the projection to the z-plane
- or the projection to the w-plane
- or both

can be used to identify a neighbourhood of (z_0, w_0) on S homeomorphically with a disc in the z-plane around z_0, or with a disc in the w-plane around w_0.

We can then use this identification to define what it means for a function on S to be holomorphic near (z_0, w_0).

1.6 Remark. We allowed concrete Riemann surfaces to be singular. In the literature, that is usually disallowed (and our singular Riemann surfaces are called *analytic sets*). We are mostly concerned with non-sigular surfaces, so this will not cause trouble.

An interesting example

Let us conclude the lecture with an example of a Riemann surface with an interesting shape, which cannot be identified by projection (or in any other way) with the z-plane or the w-plane.

Start with the function $w = \sqrt{(z^2-1)(z^2-k^2)}$ where $k \in \mathbb{C}$, $k \neq \pm 1$, whose graph is the Riemann surface

$$T = \{(z, w) \in \mathbb{C}^2 \mid w^2 = (z^2-1)(z^2-k^2)\}.$$

There are *two* values for w for every value of z, other than $z = \pm 1$ and $z = \pm k$, in which cases $w = 0$. A real snapshot of the graph (when $k \in \mathbb{R}$) is indicated in Fig. (1.3), where the dotted lines indicate that the values are imaginary.

Near $z = 1$, $z = 1 + \epsilon$ and the function is expressible as

$$w = \sqrt{\epsilon(2+\epsilon)(1+\epsilon+k)(1+\epsilon-k)} = \sqrt{\epsilon}\sqrt{2+\epsilon}\sqrt{(1+k)+\epsilon}\sqrt{(1-k)+\epsilon}.$$

A choice of sign for $\sqrt{2(1+k)(1-k)}$ leads to a holomorphic function $\sqrt{2+\epsilon}\sqrt{(1+k)+\epsilon}\sqrt{(1-k)+\epsilon}$ for small ϵ, so $w = \sqrt{\epsilon}\ \times$ (a holomorphic function of ϵ), and the qualitative behaviour of the function near $w = 1$ is like that of $\sqrt{\epsilon} = \sqrt{z-1}$.

Similarly, w behaves like the square root near -1, $\pm k$. The important thing is that there is no continuous single-valued choice of w near these points: any choice of w, followed continuously round any of the four points, leads to the opposite choice upon return.

Defining a continuous branch for the function necessitates some cuts. The simplest way is to remove the open line segments joining 1 with k and -1 with $-k$. On the complement of these segments, we can make a continuous choice of w, which gives an analytic function (for $z \neq \pm 1, \pm k$). The other 'branch' of the graph is obtained by a global change of sign.

Thus, ignoring the cut intervals for a moment, the graph of w breaks up into two pieces, each of which can be identified, via projection, with the z-plane minus two intervals (Fig. 1.4).

Now over the said intervals, the function also takes two values, except at the endpoints where those coincide. To understand how to assemble the two branches of the graph, recall that the value of w jumps to its negative as we cross the cuts. Thus, if we start on the upper sheet and travel that route, we find ourselves exiting on the lower sheet. Thus,

- the far edges of the cuts on the top sheet must be identified with the near edges of the cuts on the lower sheet;
- the near edges of the cuts on the top sheet must be identified with the far edges on the lower sheet;
- matching endpoints are identified;
- there are no other identifications.

A moment's thought will convince us that we cannot do all this in $\mathbb{R}^3$, with the sheets positioned as depicted, without introducing spurious crossings. To rescue something, we flip the bottom sheet about the real axis. The matching edges of the cuts are now aligned, and we can perform the identifications by stretching each of the surfaces around the cut to pull out a tube. We obtain the following picture, representing two planes (ignore the boundaries) joined by two tubes (Fig. 1.5.a).

For another look at this surface, recall that the function

$$z \mapsto R^2/z$$

identifies the exterior of the circle $|z| \leq R$ with the punctured disc $\{|z| < R \mid z \neq 0\}$. (This identification is even bi-holomorphic, but we don't care about this yet.) Using that, we can pull the exteriors of the discs, missing from the picture above, into the picture as punctured discs, and obtain a torus with two missing points as the definitive form of our Riemann surface (Fig. 1.5.b).

Lecture 2

The example considered at the end of the Lecture 1 raises the first serious questions for the course, which we plan to address once we define things properly: What shape can a Riemann surface have? And, how can we tell the topological shape of a Riemann surface, other than by creative cutting and pasting?

The answer to the first question (which will need some qualification) is that any orientable surface can be given the structure of a Riemann surface. One answer to the second question, at least for a large class of surfaces, will be the *Riemann-Hurwitz theorem* (Lecture 6).

2.1 Remark. Recall that a surface is *orientable* if there is a continuous choice of clockwise rotations on it. (A non-orientable surface is the Möbius strip; a compact example without boundary is the Klein bottle.) Orientability of Riemann surfaces will follow from our desire to do complex analysis on them; notice that the complex plane carries a natural orientation, in which multiplication by i is counter-clockwise rotation.

Concrete Riemann Surfaces

Historically, Riemann surfaces arose as graphs of analytic functions, with multiple values, defined over domains in $\mathbb{C}$. Inspired by this, we now give a precise definition of a concrete Riemann surface; but we need a preliminary notion.

2.2 Definition: A complex function $F(z,w)$ defined in an open set in $\mathbb{C}^2$ is called *holomorphic* if, near each point (z_0, w_0) in its domain, F has a convergent power series expansion

$$F(z,w) = \sum_{m,n\geq 0} F_{mn}(z-z_0)^m(w-w_0)^n.$$

The basic properties of 2-variable power series are assigned to Problem 1.4; in particular, F is differentiable in its region of convergence, and we can differentiate term by term.

2.3 Definition: A subset $S \subseteq \mathbb{C}^2$ is called a *(concrete, possibly singular) Riemann surface* if, for each point $s \in S$, there is a neighbourhood U of s and a holomorphic function F on U such that $S \cap U$ is the zero-set of F in U; moreover, we require that $\partial^n F/\partial w^n(s) \neq 0$ for some n.

In particular, the continuity of F implies that S is locally closed. The condition $\partial^n F/\partial w^n(s) \neq 0$ rules out vertical lines through s, which cannot reasonably be viewed as 'graphs'. (Indeed, we can see from the power series expansion that $S \cap U$ will contain a vertical line precisely when $F_{0n} = 0$ for all n.)

2.4 Definition: The Riemann surface is called *non-singular* at $s \in S$ if a function F can be found with the gradient vector $(\partial F/\partial z, \partial F/\partial w)$ non-zero at s.

2.5 Theorem (Local structure of non-singular Riemann surfaces):

(i) Assume $\partial F/\partial w(s) \neq 0$. Then, in some neighbourhood of s, S is the graph of a holomorphic function $w = w(z)$.

(ii) Assume $\partial F/\partial z(s) \neq 0$. Then, in some neighbourhood of s, S is the graph of a holomorphic function $z = z(w)$.

(iii) Assume both. Then, the two holomorphic functions above are inverse to each other.

2.6 Remark: The domains of these function will be small neighbourhoods of the components of s; the functions may extend to a larger region, or again they may not.

Proof:

(i) Writing all in real variables, we have $z = x + iy$, $w = u + iv$, $F = R + iM$. The Jacobian matrix of $F = (R, M)$ is

$$J = \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} & \frac{\partial R}{\partial u} & \frac{\partial R}{\partial v} \\ \frac{\partial M}{\partial x} & \frac{\partial M}{\partial y} & \frac{\partial M}{\partial u} & \frac{\partial M}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} & \frac{\partial R}{\partial u} & \frac{\partial R}{\partial v} \\ -\frac{\partial R}{\partial y} & \frac{\partial R}{\partial x} & -\frac{\partial R}{\partial v} & \frac{\partial R}{\partial u} \end{pmatrix}$$

where the second equality uses the Cauchy-Riemann equations. If $\partial F/\partial w(s) \neq 0$, then $(\partial R/\partial u, \partial M/\partial u)$ = $(\partial R/\partial u, -\partial R/\partial v) \neq (0,0)$, and then the matrix

$$\begin{pmatrix} & J & & \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

has full rank. This is, then, the Jacobian for a local change of coordinates from (x, y, u, v) to (x, y, R, M) near s. The inverse function theorem says that the smooth map $(x, y) \mapsto$

$(x, y, R = 0, M = 0)$ can be rewritten in u, v coordinates, $(x, y) \mapsto (x, y, u(x, y), v(x, y))$ and defines *smooth* functions $u(x, y)$ and $v(x, y)$, whose graph constitute the zero-set of F.

The proof that the function $(x, y) \mapsto (u(x, y), v(x, y))$ is actually holomorphic is assigned to Problem 1.5.

(ii) and (iii) are obvious consequences of Part (i).

Abstract Riemann surfaces

For most of the course, we shall consider Riemann surfaces from an abstract point of view. This suffices to establish their general properties, and dispenses with unnecessary embedding information. (Moreover, smoothness is built in, whereas in the embedded case it must be checked). However, the abstract definition is somewhat complicated and less intuitive. One way to motivate their introduction is by the following observation.

2.7 Proposition: Every Riemann surface in $\mathbb{C}^2$ is non-compact. (Proof in the next lecture).

This is clear for a Riemann surface defined as the zero-set of an algebraic equation $P(z, w(z)) = 0$; it projects surjectively to the complex z-plane, Because the image of a compact set under a continuous map is compact, it follows that the solution-set is not compact.

So, there is an obstacle to constructing *compact* Riemann surfaces, such as the torus *without* punctures, as graphs of multi-valued functions within $\mathbb{C}^2$. On the other hand, it's easy to produce compact topological surfaces with enough analytic structure to be worthy of Riemann's name. Here are two examples:

(2.8) *The Riemann sphere* $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1$ (Fig. 2.2).

The topological description of how $\mathbb{C} \cup \{\infty\}$ becomes a sphere is best illustrated by the stereographic projection, in which points going off to ∞ in the plane converge to the north pole in the sphere. (The south pole maps to 0.)

We can understand $\mathbb{P}^1$ as a Riemann surface is by regarding $z^{-1} = w$ as a *local coordinate* near ∞. We say that a function f defined in the neighbourhood of ∞ on $\mathbb{P}^1$ is holomorphic if the following function is holomorphic, in a neighbourhood of $w = 0$:

$$w \mapsto \begin{cases} f\left(w^{-1}\right), & \text{if } w \neq 0 \\ f(\infty), & \text{if } w = 0 \end{cases}$$

There is another descrition of $\mathbb{P}^1$ as a Riemann surface. Consider two copies of $\mathbb{C}$, with coordinates z and w. The map $w = z^{-1}$ identifies $\mathbb{C} \setminus \{0\}$ in the z-plane with $\mathbb{C} \setminus \{0\}$ in the w-plane, in analytic and invertible fashion. (We say that the map $z \mapsto w = z^{-1}$ from $\mathbb{C}^*$ to $\mathbb{C}^*$

is *bianalytic*, or *biholomorphic*.) Define a new topological space by gluing the two copies of $\mathbb{C}$ along this identification. Clearly, we get a topological sphere, but now there is an obvious notion of holomorphic function on it: we have $\mathbb{P}^1 = \mathbb{C}_{(z)} \cup \mathbb{C}_{(w)}$, and we declare a function f on $\mathbb{P}^1$ to be homomorphic precisely if its restrictions to the open sets $\mathbb{C} = \mathbb{P}^1 \setminus \{\infty\}$ and $\mathbb{C} = \mathbb{P}^1 \setminus \{0\}$ are holomorphic.

Because the identification map is holomorphic, and the composition of holomorphic maps is holomorphic, it follows that a function on $\mathbb{C} = \mathbb{P}^1 \setminus \{\infty\}$ is holomorphic in the new sense iff it was so in the old sense; so we have really enlarged our 'Riemann surface' $\mathbb{C}$ to form a sphere.

This procedure of obtaining Riemann surfaces by gluing is much more general (Problem 1.3c), and is one reason why abstract Riemann surfaces are more fun to play with than concrete ones.

(2.9) *Tori as Riemann surfaces*

Let A be the annulus $1 < |z| < R + \epsilon$. Identify the boundary strip $1 < |z| < 1 + \epsilon/R$ with the boundary strip $R < |z| < R + \epsilon$ via multiplication by a fixed complex number q, with $|q| = R$. Again, this identification is biholomorphic. Let T be the surface obtained by identifying the two boundary strips. Clearly T is a torus, and we have an open, surjective map $A \xrightarrow{\pi} T$. If $U \subseteq T$ is open, call a function $f : U \to \mathbb{C}$ holomorphic iff $f \circ \pi : \pi^{-1}(U) \to \mathbb{C}$ is holomorphic.

Every point $t \in T$ has a sufficiently small neighbourhood so that $\pi^{-1}(U)$ is either one or two disjoint open sets in A. In the second case, the two sets are identified analytically by the map $z \mapsto qz$. Hence, to check analyticity near t, it suffices to check it *near a singleone* of the inverse images of t. Thus we have shown:

2.10 Proposition: With these definitions, every point $t \in T$ has some neighbourhood U_t, identifiable via π with a disc in $\mathbb{C}$, in such a way that holomorphic functions go to holomorphic functions, in both directions.

These examples, I hope, serve to motivate the definition of an abstract Riemann surface, which we are finally ready to state. Recall first a topological notion.

2.11 Definition: A *topological surface* is a topological Hausdorff space in which every point has a neighbourhood homeomorphic to the open unit disc in $\mathbb{R}^2$.

For an open set U in a topological space, denote by $C^0(U)$ the algebra of complex-valued continuous functions on U.

2.12 Definition: A *(non-singular, abstract) Riemann surface* S is a topological surface with the following extra structure:

(i) For every open set U, there is given a subalgebra $\mathcal{O}(U) \subseteq C^0(U)$, called the *holomorphic functions* on U.

(ii) If $V \subseteq U$, the restriction $C^0(U) \to C^0(V)$ takes $\mathcal{O}(U)$ into $\mathcal{O}(V)$. (The restriction of a holomorphic function to on open subset is holomorphic.)

(iii) If $U = \bigcup_{\alpha \in A} U_\alpha$, with $U_\alpha \subseteq U$ open, and a function $f \in C^0(U)$ is holomorphic on each U_α, then it is holomorphic on U; that is, it lies in $\mathcal{O}(U)$. (Holomorphy is a local condition).

(iv) Every $s \in S$ has some neighbourhood U_s, which admits a homeomorphic identification $h_s : U_s \to \Delta$ with the unit disc Δ, with the property that composition with h_s identifies the holomorphic functions on Δ with $\mathcal{O}(U_s)$. (Riemann surfaces look locally like the unit disk).

It should be clear that the two examples above, $\mathbb{P}^1$ and T, satisfy this definition. Here is another one, important for a sanity check.

2.13 Example: An open subset $U \subseteq \mathbb{C}$ inherits the structure of an abstract Riemann surface, with the natural definition of holomorphic function. (We just use small disks sitting in U for condition (iv).)

*Appendix to Lecture 2: Singularities in Riemann surfaces

A1. Local structure of singularities

There is a basic result about the *local structure of singular Riemann surfaces* (they are more often called *analytic spaces* in the literature), which justifies the moral definition we gave in the first lecture. It also says that all local information near singular points can be captured algebraically.

2.14 Theorem (Weierstrass Preparation Theorem): Let $F(z, w)$ be holomorphic near $(0, 0)$ and assume that

$$F(0,0) = 0, \quad \frac{\partial F}{\partial w}(0,0) = 0, \quad \ldots, \quad \frac{\partial^{n-1} F}{\partial w^{n-1}}(0,0) = 0, \quad \text{but} \quad \frac{\partial^n F}{\partial w^n}(0,0) \neq 0.$$

Then

- (weak form) there exists a function Φ of the form

  $$\Phi(z, w) = w^n + f_{n-1}(z)w^{n-1} + \cdots + f_1(z)w + f_0(z)$$

  with $f_0, \ldots, f_{n-1}$ analytic near $z = 0$, whose zero-set agrees with that of F near 0.

- (strong form) there exists, additionally, a holomorphic function $u(z, w)$, non-zero near $(z, w) = (0, 0)$, such that

  $$F(z, w) = \Phi(z, w)\, u(z, w).$$

  Moreover, this factorization of F is *unique*.

Proof: See Problem 1.12 for the weak form; see Gunning and Rossi, Several Complex Variables, for the strong form.

2.15 Remark: For $n = 1$, the weak form recovers part (i) of theorem 2.5. For general n, it says that S represents the graph of an n-valued 'solution function' of a polynomial equation in w, with coefficients depending holomorphically on z.

A2. Removal of singularities

If we are not interested in the structure of a singularity, there arises the natural question how this singularity can be 'removed', or resolved (the official term). In algebraic geometry, this is done by a procedure called *normalization.* There is an analytic way to describe that; I shall do so informally, without attempting to define all terms or prove the statements.

There is no way to resolve the singularity of a surface S while keeping it in $\mathbb{C}^2$, so let us first clarify the question. First, it can be shown that the singularities of analytic sets are isolated. (This is plausible enough, as they are the zeroes of the gradient of F). The correct question is: can we find an abstract Riemann surface R (necessarily non-singular, in view of our definition), mapping holomorphically to S, so that the map is *bi-holomorphic* at the regular points of S? If so,we say that R resolves the singularities of S; in effect, we have replaced the singular points of S with smooth points (in R). The answer is, it can always be done, and R is unique up to isomorphism.

Note, first, that there are two kinds of singularities: topological ones and purely analytic ones. An example of a topological singularity is the solution set of $w^2 - z^2 = 0$ near the origin, whose neighbourhood is homeomorphic to the union of two disks meeting at their centre. Such singularities arise when the defining power series splits into distinct factors — in this case, $(z - w)(z + w)$. The first step in the resolution is then clear: we separate the disks by replacing their union with a disjoint union. There will now be two points in R maping to the singular point of S.

In general, if the power series F defining our surface near s can be factored into terms which are not units in the ring of holomorphic functions near s, we replace its zero-set by the disjoint union of the zero-sets of the factors.

The simplest example of a purely analytic singularity is $z^3 - w^2 = 0$, near the origin. As we shall see, the zero-set S is locally homeomorphic to the disc; however, neither coordinate can be used to define the structure of a non-singular surface on S. Instead, this can be done by the analytic map $u \mapsto (z, w) = (u^2, u^3)$, which defines a homeomorphism from $\mathbb{C}$ to S. Away from zero, the map is bi-holomorphic, because u can be recovered as w/z.

The following theorem shows that our example is no accident. Say that a power series $F(z, w)$ is *irreducible* near $(0, 0)$ if, for any factorization $F = F_1(z, w) \cdot F_2(z, w)$ into power series, near $(0, 0)$, one of the factors does not vanish at $(0, 0)$ (and hence is multiplicatively invertible there). This excludes the possibility of decomposing the zero-set, as in the previous example.

2.16 Theorem. (Puiseux Expansion): Let the power series $F(z, w)$ be *irreducible* near $(0, 0)$. Assume that $F(z, w) \neq z$ (this rules out a vertical line). Then, there exist integers a, b and holomorphic functions $z(u) = u^a$, $w(u) = const \cdot u^b + O(u^{b+1})$ which give a homeomorphism from a small disc in the u-plane with a neighbourhood of the zero-set S of F near $(0, 0)$. Moreover, this homeomorphism is *bi-holomorphic* away from 0.

In other words, the 'solution function' of $F(z, w) = 0$ can be expanded as a power series in $u = z^{1/a}$, with starting term $u^b = z^{b/a}$.

2.17 Remark: A more precise formulation of the theorem gives a simple test for the irreducibility of F, a simple determination for the integers a, b, and a prescription for expressing w in terms of u. Let m and n be the smallest integers for which the coefficients F_{m0} and F_{0n} are non-zero. (If all F_{m0} vanish, then F is a multiple of z, and by irreducibility, $F = z$; similarly, if all F_{0n} are null, $F = w$.) Let a, b be the smallest pair of positive integers with $am - bn = 0$. The test for irreducibility of F is that F_{pq} should vanish for all positive p, q satisfying $p/b + q/a \leq m/b$ (all non-vanishing coefficients F_{pq} should lie above the line joining $(m, 0)$ with $(0, n)$).

Lecture 3

An alternative definition of a Riemann surface can be given using the notion of *holomorphic atlas*.

3.1 Definition: A *holomorphic atlas* on a topological surface is a collection of open sets U_α (the *coordinate charts*) which cover the surface, together with homeomorphic identifications $h_a : U_\alpha \to \Delta$ with the unit disc in $\mathbb{C}$, subject to the following condition: for all α and β, the transition functions $h_\alpha \circ h_\beta^{-1} : h_\beta(U_\alpha \cap U_\beta) \to h_\alpha(U_\alpha \cap U_\beta)$ should be holomorphic maps between open subsets of Δ.

3.2 Proposition. A holomorphic atlas on a topological surface determines a structure of an abstract Riemann surface. Conversely, on an abstract Riemann surface, the sets U_s and the maps h_s define a holomorphic atlas.

Proof. Given a holomorphic atlas and an open set U on the surface, we declare that $f \in \mathcal{O}(U)$ iff every $f \circ h_\alpha^{-1} : h_\alpha(U \cap U_\alpha) \to \mathbb{C}$ is a holomorphic function. It is clear that (i)–(iii) are satisfied. Moreover, the holomorphic patching condition ensures that $f \in \mathcal{O}(U_\alpha)$ iff $f \circ h_\alpha^{-1}$ is holomorphic on Δ: there is no need to check the other charts, because the function there is modified by a holomorphic change of variable. So condition (iv) follows as well. Conversely, the structural (U_s, h_s) on an abstract Riemann surface define a holomorphic atlas. (Which of the conditions implies the holomorphy of the transition functions?)

3.3 Example. A holomorphic atlas on $\mathbb{P}^1$ can be given using the charts $U_0 = \{z \mid |z| < 1\}$ and $U_\infty = \{z \mid |z| > 1/2\}$, with h_0 being the identity map and $h_1(z) = 1/2z$.The transition function $h_1 \circ h_0^{-1} : U_0 \cap U_\infty \to U_0 \cap U_\infty$ takes z to $1/2z$.

A large class of examples of abstract Riemann surfaces comes from the following result.

3.4 Proposition. Any concrete, non-singular Riemann surface in $\mathbb{C}^2$ carries a natural structure of abstract Riemann surface.

Proof. We have seen that every point has a neighbourhood identifiable with a disk in the z-plane, or in the w-plane, by projection. Any disk can be identified with Δ by a complex scaling and translation. We claim that all these disks define a holomorphic atlas. We need to check holomorphy of the transition functions. This is obvious were both coordinates are defined by the same projections (the transition function is linear). Where both $\partial F/\partial z$ and $\partial F/\partial w$ are non-zero, on the other hand, each coordinate is holomorphically expressible in terms of the other; so again the transition function is holomorphic. Alternatively (using the original definition), call a function on an open set of S holomorphic if it is so when transported to all the small z- and w- discs. Conditions (i)–(iii) are obvious. To check condition (iv), we must verify, near a point s where both projections can be used, that a function holomorphically expressible in z is also holomorphically expressible in w. (Else, our identification of little subsets with disks would not take holomorphic functions to holomorphic functions). But this is clear for the same reason, namely, z is a holomorphic function of w near s.

As a quick application, and before we bring in even more definitions, let us prove a theorem. Recall first that a topological space is *connected* if it cannot be decomposed as a disjoint union of two open non-empty subsets. Otherwise put, no *proper* subset is both closed and open.

3.5 Theorem: Every holomorphic function defined everywhere on a *compact* connected Riemann surface is constant.

We shall reduce this to a local property of holomorphic functions, which is an immediate consequence of Condition (iv) in the definition of an abstract Riemann surface. (We'll see more of this in the next lecture).

3.6 Theorem (Maximum Principle): Let f be a holomorphic function defined in a neighbourhood of a point s in a Riemann surface. If $|f|$ has a local maximum at s, then f is constant in a neighbourhood of s.

Proof of 3.6: Identify a neighbourhood U_s of s with the unit disc, as in condition (iv) of the definition, and apply the maximum principle in the disc.

Proof of 3.5: The function f is continuous, so $|f|$ achieves a maximum value M. Let z be any point on the surface with $|f(z)| = M$; then $|f|$ achieves a local maximum at z, hence is constant in a neighbourhood of z. Therefore, the set where $|f(z)| = M$ is open. As it is also closed (by continuity) it must be the entire surface (by connectedness).

3.7 Remark: If the surface has several connected components, f will be constant on every one of them.

Holomorphic maps between Riemann surfaces

We have generalized the notion of holomorphic function by allowing the domain to be an abstract Riemann surface. If we generalize the target space in the same way, we arrive at the notion of *holomorphic map* between Riemann surfaces. The following definition is a bit indirect, but it ties in directly with the abstract definition (2.12).

3.8 Definition: A continuous map $f : R \to S$ between Riemann surfaces is said to be *holomorphic* if it takes holomorphic functions to holomorphic functions: for every holomorphic $h : U \to \mathbb{C}$, with $U \subseteq S$ open, the function $h \circ f : f^{-1}(U) \to \mathbb{C}$ is holomorphic (i.e. is in $\mathcal{O}(f^{-1}(U))$).

3.9 Example: (Sanity check) Viewing $\mathbb{C}$ as a Riemann surface, a map $f : R \to \mathbb{C}$ is holomorphic as a map of Riemann surfaces iff it is a holomorphic function in the old sense. Indeed, let id denote the identity map $w \mapsto w$ from $\mathbb{C}$ to $\mathbb{C}$. If $f : R \to \mathbb{C}$ is holomorphic as a map, according to the definition above, then $\mathrm{id} \circ f = f$ must be a holomorphic function. Conversely if $f : R \to \mathbb{C}$ is a holomorphic function and $h : V \to \mathbb{C}$ is holomorphic, with V open in $\mathbb{C}$, then $h \circ f : f^{-1}(V) \to \mathbb{C}$ is a holomorphic function, since it is the composition of holomorphic (old style) functions. So f is a holomorphic map of Riemann surfaces.

The local form and valency of a holomorphic map

We now return to the study of holomorphic maps between abstract Riemann surfaces.

3.10 Theorem (The local form of a holomorphic map): Let $f : R \to S$ be holomorphic, with $r \in R$, $f(r) = s$, and f not constant near r.

Then, given any analytic identification $\psi : V_s \to \Delta$ of a sufficiently small neighbourhood of $s \in S$ with the unit disc Δ, sending s to 0, there exists an analytic identification $\phi : U_r \to \Delta$ of a suitable neighbourhood U_r of r with Δ such that $f(U_r) \subseteq V_s$ and the following diagram commutes:

$$\begin{array}{ccc} U_r & \xrightarrow{\;f\;} & V_s \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\psi} \\ \Delta & \longrightarrow & \Delta \\ z & \longmapsto & z^n \end{array}$$

That is, $(\psi \circ f)(x) = \phi(x)^n$ for all $x \in U_r$. In words, 'f looks locally like the map $z \mapsto z^n$'.

Proof: Choose a local chart $h : U_r \to \Delta$, centered at r. Then, $g := \psi \circ f \circ h^{-1}$ is defined and analytic near 0. Let n be the order of the zero, so $g(z) = g_n z^n + O(z^{n+1})$. An analytic n^{th} root $g^{1/n}$ of g is then defined near 0, of the form $g_n^{1/n} z + O(z^2)$. The derivative at 0 does not vanish, so $g^{1/n}$ is a local analytic isomorphism near 0. But then, $g^{1/n} \circ h$ is a local analytic isomorphism from a neighbourhood of r to a neighbourhood of 0, and its n^{th} power is clearly $(\psi \circ f)$. After shrinking V_s, if necessary, and rescaling ψ appropriately, we just take $\phi = g^{1/n}$.

3.11 Proposition: The number n above does not depend on the choice of neighbourhoods and is called the *valency* of f at r, $v_f(r)$.

Proof: Given the theorem, we see that $v_f(r)$ has the following nice description: it is the number of solutions to $f(x) = y$ which are contained in a very small neighbourhood U_r of r, as y approaches s (the number of solutions to $f(x) = y$ which converge to r as y converges to s.) Clearly this does not depend on any choices.

To use the theorem on the local form with more ease, observe the following.

3.12 Lemma: If a holomorphic map between Riemann surfaces is constant in a neighbourhood of a point, then it is constant on a connected component of that surface which contains the point.

Proof: The set U of points where the function f takes the value s in question is closed, by continuity. Let V be the interior of U, and let r be a point on the boundary of V; then, f is not constant near r, so the theorem on the local form applies. But there must be a sequence of points from U converging to r, and $f = s$ at those points, which contradicts the theorem. So the boundary of V is empty, hence V is open and closed, and is a union of connected components of the surface.

Local consequences of the theorem on the local form

We assume everywhere that R is connected, and $f : R \to S$ is holomorphic and non-constant. By Lemma (3.12), f has the local form (4.8) near each point of R; the following are immediate consequences.

- **Open mapping theorem:** The map f is *open*; that is, the image of any open set in R is open in S.
- **'Good behaviour' almost everywhere:**
  - (i) The set of points $r \in R$ with $f(r) = s$, for fixed $s \in S$, has no accumulation point in R.
  - (ii) The set of points $r \in R$ with $v_f(r) > 1$ has no accumulation point in R.

  (Recall that a point $x \in R$ is an *accumulation point* or *boundary point*) of a subset X if there exists a sequence x_n of points in $X \setminus \{x\}$ converging to x.)
- **Local test for local injectivity:** f is injective when restricted to a small neighbourhood of $r \iff v_f(r) = 1 \iff f$ gives an analytic isomorphism between a neighbourhood of $r \in R$ and a neighbourhood of $f(r) \in S$.
- **Inverse function theorem:**
  - (i) If f is holomorphic and bijective, then f is an analytic isomorphism; that is, the inverse mapping f^{-1} is analytic.
  - (ii) If f is injective then f gives an isomorphism of R with $f(R)$, an open subset of S.
- **Maximum Principle:** (See 3.6) Assume $S = (C)$; then, $|f|$ has no local maximum in R.

Lecture 4

Meromorphic functions and maps to $\mathbb{P}^1$

We now study the special case of holomorphic maps with target space $\mathbb{P}^1$. It turns out that we recover the more familiar notion of *meromorphic function.*

4.1 Definition: A function $f : U \to \mathbb{C} \cup \{\infty\}$ ($U \subseteq \mathbb{C}$ open) is called *meromorphic* if it is holomorphic at every point where it has a finite value, whereas, near every point z_0 with $f(z_0) = \infty$, $f(z) = \phi(z)/(z - z_0)^n$ for some holomorphic function ϕ, defined and non-zero around z_0. The positive number n is the *order of the pole* at z_0.

4.2 Remark: Equivalently, we ask that, locally, $f = \phi/\psi$ with ϕ and ψ holomorphic. We can always arrange that $\phi(z_0)$ or $\psi(z_0)$ are non-zero, by dividing out any $(z - z_0)$ power, and we define $a/0 = \infty$ for any $a \neq 0$.

4.3 Theorem: A meromorphic function on U is the same as a holomorphic map $U \to \mathbb{P}^1$ which is not identically ∞.

Proof: Let f be meromorphic. Clearly it defines a continuous map to $\mathbb{P}^1$, because $f(z) \to \infty$ near a pole. Clearly, also, it is holomorphic away from its poles. Holomorphicity near a pole z_0 means: for every function g, defined and holomorphic near $\infty \in \mathbb{P}^1$, $g \circ f$ is holomorphic near z_0. But g is holomorphic at ∞ iff the function h defined by

$$h(z) = \begin{cases} g(1/z) & \text{if } z \neq 0 \\ g(\infty) & \text{if } z = 0 \end{cases}$$

is holomorphic near 0. But then, $g \circ f = h(1/f) = h((z - z_0)^n/\phi(z))$ which is holomorphic, being the composition of holomorphic functions. (Recall $\phi(z) \neq 0$ near z_0.)

Conversely, let $f : U \to \mathbb{P}^1$ be a holomorphic map. By definition, using the function $w \mapsto w$ defined on $\mathbb{C} \subset \mathbb{P}^1$, the composite function $f : (f^{-1}(\mathbb{C}) = U \setminus f^{-1}(\infty)) \to \mathbb{C}$ is holomorphic; so we must only check the behaviour near the infinite value. For that, we use the function $w \mapsto 1/w$ holomorphic on $\mathbb{P}^1 \setminus \{0\}$ and conclude that $1/f$ is holomorphic on U, away from the zeroes of f. But then f is meromorphic.

We can now transfer this to a Riemann surface.

4.4 Definition: A function $f : S \to \mathbb{C} \cup \{\infty\}$ on a Riemann surface is meromorphic if it is expressible locally as a ratio of holomorphic functions, the denominator not being identically zero.

4.5 Proposition: A meromorphic function on a Riemann surface is the same as a holomorphic map to $\mathbb{P}^1$, not identicaly ∞.

Proof: Clear from the local case in $\mathbb{C}$, because every point on the surface has a neighbourhood identifiable with a disc, as far as holomorphic functions are concerned.

4.6 Definition: If the meromorphic function f on R has a pole at $r \in R$, the *order of the pole* is its valency at r, when viewed as a holomorphic map to $\mathbb{P}^1$.

Exercise: When $R \subset \mathbb{C}$, check that this definition agrees with (4.1). (Use the local coordinate $w = z^{-1} Son\mathbb{C}$.

Algebra with meromorphic functions

There is a slight difference between meromorphic functions and maps to $\mathbb{P}^1$; it stems from the condition that f should not be identically ∞, to be called meromorphic. This has a significant consequence, as far as algebra is concerned:

4.7 Proposition: The meromorphic functions on a connected Riemann surface form a field, called the *field of fractions* of the Riemann surface.

Recall that a field is a set with associative and commutative operations, addition and multiplication, such that multiplication is distributive for addition; and, moreover, the ratio a/b of any two elements, with b not equal two zero, is defined and has the familiar property $a/b * b = a$.

Proof: This is clear from the local definition (4.1).

Remark: From another point of view, this may seem curious. Recall from calculus that certain arithmetic operations involving ∞ and 0 cannot be consistently defined: $\infty - \infty$, ∞/∞, $0/0$ and $\infty \cdot 0$ cannot be assigned meanings consistent with the usual arithmetic laws. For meromorphic functions, we assign a meaning to this undefined expressions by taking the limit of the nearby values of the function; the local expression (4.1) of a meromorphic function ensures that the limit exists.

Unique Presentations of meromorphic functions on $\mathbb{P}^1$

We shall describe more closely the holomorphic maps from $\mathbb{P}^1$ to itself. By the results (3.14) of the previous lecture, these are the same as the meromorphic functions on $\mathbb{P}^1$, plus the constant map ∞.

Recall that a *rational function* $R(z)$ is one expressible as a ratio of two polynomials, $p(z)/q(z)$ (q not identically zero). Clearly, it is meromorphic. We may assume p and q to have no common factors, in which case we call $\max(\deg p, \deg q)$ the *degree* of $R(z)$.

4.8 Theorem: Every meromorphic function on $\mathbb{P}^1$ is rational.

We shall prove two stronger statements.

4.9 Theorem (Unique Presentation by principal parts): A meromorphic function on $\mathbb{P}^1$ is uniquely expressible as

$$p(z) + \sum_{i,j} \frac{c_{ij}}{(z-p_i)^j},$$

where $p(z)$ is a polynomial, the c_{ij} are constants and the sum is *finite*.

4.10 Remark: The p_i are the finite poles of the function.

Proof: Recall that, near a pole p, a meromorphic function has a convergent *Laurent expansion*:

$$a_n(z-p)^{-n} + a_{-n+1}(z-p)^{-n+1} + \cdots + a_{-1}(z-p)^{-1} + \sum_{k\geq 0} a_k(z-p)^k$$

and the negative powers form the *principal part* of the series.

Note, from the local expression of a meromorphic function, that the poles are isolated. By compactness of $\mathbb{P}^1$, this implies that f has finitely many poles. We start by subtracting the principal parts at all the finite poles. What is left is a meromorphic function with poles only at ∞. I claim this is a polynomial. Indeed, in the local coordinate $w = z^{-1}$, we can subtract the principal part of the function, which is a polynomial in $z = w^{-1}$; and are left with a meromorphic function with no poles, that is, a holomorphic function on $\mathbb{P}^1$. But that must be constant. (Use the Theorem in lecture 3, or, if you have come across it, Liouville's theorem that a bounded holomorphic function on $\mathbb{C}$ is constant.)

4.11 Theorem (Unique Presentation by zeroes and poles): A meromorphic function on $\mathbb{P}^1$ has a unique expression as

$$c \times \frac{\prod_{i=1}^{n}(z - z_i)}{\prod_{j=1}^{m}(z - p_j)},$$

where c is a constant, the z_i are the (finite) zeroes of the function (repeated as necessary) and the p_j the (finite) poles (repeated as necessary).

Proof: The ratio of f by the product above will be a meromorphic function on $\mathbb{P}^1$, having no zeroes or poles in $\mathbb{C}$. But, as it has no poles, it must be a polynomial (see the previous proof), and a polynomial without roots in $\mathbb{C}$ is *constant*.

If we count properly, there is in fact one constraint on the distribution of zeroes and poles of a meromorphic function on $\mathbb{P}^1$:

4.12 Proposition: A meromorphic function on $\mathbb{P}^1$ has just as many zeroes as poles, if multiplicities are counted.

Proof: Just note that the product in (4.11) has a pole of order $n - m$ at ∞ if $n > m$, or else a zero of order $m - n$ there, if $m > n$.

4.13 Discussion The importance of the two Unique Presentation Theorems goes beyond $\mathbb{P}^1$. They have analogues for arbitrary compact Riemann surfaces. Later in the course we shall discuss the torus in detail, but a few words about the general statement are in order now.

The part of the theorem which generalises easily is the 'uniqueness up to a constant', additive or multiplicative. That is,

(i) Two meromorphic functions on a compact Riemann surface having the same principal part at each of their poles must differ by a constant.

(ii) Two meromorphic functions having the same zeroes and poles (multiplicities included) agree up to a constant factor.

The argument is the same as for $\mathbb{P}^1$; to wit, the difference of the functions, in case (i), and the ratio, in case (ii), would be a global holomorphic function on the surface, and as such would be constant.

By contrast, the existence problem — functions with specified principal parts, or with specified zeroes and poles — is more subtle, and there are obstructions *coming from the topology of the surface*. As a general rule, there will be g conditions imposed on the principal parts, or on the locations of the zeroes and poles, on a surface of genus g. But, in genus 2 and higher, there will be exceptions to this rule.

4.14 Caution: The 'order of a zero' or 'order of a pole' of a meromorphic function on a Riemann surface is defined without ambiguity, via the notion of valency (3.11). However, defining a 'principal part at a pole' requires a *local coordinate* on the surface in question. Nonetheless, statement (i) above is sensible, because it compares the principal parts of two functions at the same point, and we can use the same local coordinate for both.

Lecture 5

Global consequences of the theorem on the local form

5.1 Theorem: Let $f : R \to S$ be a non-constant holomorphic map, with R connected and compact. Then f surjects onto a compact connected component of S.

5.2 Corollaries:

(i) A non-constant holomorphic map between compact connected Riemann surfaces is surjective.

(ii) A global holomorphic function on a compact Riemann surface is constant.

(iii) (Fundamental Theorem of Algebra) A non-constant complex polynomial has a least one root.

Proof of the theorem: f is open and continuous and R is compact, so $f(R)$ is open in S and compact, hence closed. As R is also connected, $f(R)$ is connected, so it is a connected component of S. ($S = f(R) \cup (S \setminus f(R))$ with $f(R)$ and $S \setminus f(R)$ both open.)

Proof of the corollaries:

(i) Clear from the theorem and connectedness of S.

(ii) A holomorphic function determines a map to $\mathbb{C}$, hence a holomorphic map to $\mathbb{P}^1$. By the previous corollary, the image of any non-constant map would be contain ∞; so the map must be constant.

(iii) A polynomial determines a holomorphic map $\mathbb{P}^1 \to \mathbb{P}^1$. If not constant, the image of this map must contain 0, so the polynomial must have a root.

There is, of course, a more elementary proof of the fundamental theorem of algebra using complex analysis, straight from the Cauchy integral formula. The point is that we just proved a far-reaching generalisation of the fundamental theorem.

In this context, note that the FTA has a stronger form, asserting that a polynomial f of degree d will have exactly d roots, if they are counted properly. We can generalise this, too.

The degree of a holomorphic map between compact Riemann surfaces

5.3 Theorem/Definition: Let $f : R \to S$ be a non-constant holomorphic map between compact, connected Riemann surfaces. Then the number

$$\deg(f) = \sum_{r \in f^{-1}(s)} v_f(r)$$

is independent of the choice of point $s \in S$, and is called the *degree* of the map f.

5.4 Note: If f is constant, we define $\deg(f) = 0$. Note that $\deg(f) > 0$ otherwise.

5.5 Proposition: For all but finitely many $s \in S$, $\deg(f) = |f^{-1}(s)|$, the number of solutions to $f(x) = s$. For any s, $|f^{-1}(s)| \leq \deg(f)$.

Proof: Clear from the theorem and the fact that the points r with $v_f(r) > 1$ are finite in number (see 'good behaviour' theorem).

For the proof of (5.3), we need the following lemma.

5.6 Lemma: Let $f : X \to Y$ be a continuous map of topological Hausdorff spaces, with X compact. Let $y \in Y$ and U be a neighbourhood of $f^{-1}(y)$. Then there exists some neighbourhood V of y with $f^{-1}(V) \subseteq U$.

Proof of the lemma: As V varies over all neighbourhoods of $y \in Y$, $\bigcap \bar{V} = \{y\}$, by the Hausdorff property. Then, $\bigcap f^{-1}(\bar{V}) = f^{-1}(y)$. But then, $\bigcap f^{-1}(\bar{V}) \cap (X \setminus U) = \emptyset$. Now the $f^{-1}(\bar{V})$ and $(X \setminus U)$ are closed sets, and by compactness of X, some finite intersection is already empty. So $X \setminus U \cap f^{-1}(\bar{V}_1) \cap \cdots \cap f^{-1}(\bar{V}_n) = \emptyset$, or $f^{-1}(\bar{V}_1 \cap \cdots \cap \bar{V}_n) \subseteq U$, in particular $f^{-1}(V_1 \cap \cdots \cap V_n) \subseteq U$. But $V_1 \cap \cdots \cap V_n$ is a neighbourhood of y in Y.

5.7 Remark: A map $f : X \to Y$ between locally compact Hausdorff spaces is called *proper* if f^{-1}(any compact set) is compact. (For instance, the inclusion of a closed subset is proper, the inclusion of an open set is not.) The proof above works for proper maps, without assuming compactness of X.

Proof of theorem (5.3): Using the fact that $f^{-1}(s)$ is finite, we can now find a neighbourhood V of y such that $f^{-1}(V)$ is a union of neighbourhoods U_i of the $r_i \in f^{-1}(V)$ in which f has the local form (3.10). The result now follows from the obvious fact that the map $z \mapsto z^n$ has n solutions near zero: so the sum of valencies does not change in any of the U_i.

5.8 Remark: The theorem, with the same proof, holds for any *proper* holomorphic map of Riemann surfaces.

As an example, let us prove:

5.9 Proposition: The degree of a non-constant holomorphic map $f : \mathbb{P}^1 \to \mathbb{P}^1$ equals its degree as a rational function.

Proof: Let $f(z) = p(z)/q(z)$ and assume $\deg p \geq \deg q$ (else. replace f by $f+1$). The solutions to $f(z) = 0$ are precisely the solutions to $p(z) = 0$, assuming the expression p/q to be reduced. Moreover, the valency of f at a solution will equal the root multiplicity of p, because, if $q(z) \neq 0$, as we assume, f and p will have the same number of consecutive vanishing derivatives at 0. So the sum of root multiplicities is $\deg p = \deg f$.

5.10 Remark: If you try this with $f(z)$ of the form p/q with $\deg p < \deg q$, you must also count the root ∞, with valency $\deg q - \deg p$.

5.11 Corollary: For all but finitely many w, the equation

$$\frac{p(z)}{q(z)} = w$$

has exactly $\max(\deg p, \deg q)$ solutions.

(This can also be proved algebraically, but is a bit tedious.)

5.12 Example: Degree of a concrete Riemann surface over the z-plane.

Consider the Riemann surface of the equation $P(z, w) = 0$ for a polynomial

$$P(z, w) = w^n + p_{n-1}(z) w^{n-1} + \cdots + p_1(z) w + p_0(z).$$

5.13 Proposition: The projection π of the zero-set R of P to the z-plane is a proper map (cf. Remark 5.7).

Proof: By continuity of π, the inverse image of any closed set is closed; so we only need to see that the inverse image of a bounded set is bounded. But if z ranges over a bounded set, then all $p_k(z)$ range over some bounded set, and the roots of a polynomial with leading term 1 can be bounded in terms of the coefficients (e.g. by the argument principle).

Hence, whenever the concrete Riemann surface is non-singular, the notion of degree, and Thm. (5.3) are applicable. Assume now that, for one z, $P(z,w)$ has *no multiple roots* (hence, exactly n roots). Then the surface S maps properly to $\mathbb{C}_{(z)}$, and the degree of the map is n.

5.14 Comment: We shall see in th next lecture that the assumption that $P(z,w)$ has no multiple roots, for general z, is equivalent to the condition that the irreducible factorization of $P(z,w)$, as a polynomial of two variables, has no repeated factors.

The Riemann-Hurwitz formula

For a holomorphic map f between compact connected Riemann surfaces R and S, the theorem gives a formula relating

- the degree of the map,
- the topologies of R and S,
- the (local) valencies of the map.

As a preliminary, we need the following result from topology.

5.15 Classification of compact orientable surfaces: Any compact, connected, orientable surface without boundary is homeomorphic to one of the following:

g is the *genus* and counts the 'holes'. There is another depiction of these surfaces as 'spheres with handles',

and then g counts the number of handles. The *Euler characteristic* of these surfaces is (provisionally) defined as $2-2g$.

5.16 Definition: Let $f : R \to S$ be a non-constant holomorphic map between compact connected Riemann surfaces. The *total branching index* b of f is

$$\sum_{s\in S}\sum_{r\in f^{-1}(s)}(v_f(r)-1)=\sum_{s\in S}\left(\deg(f)-|f^{-1}(s)|\right).$$

Note that this sum is finite. It counts the total number of 'missing' solutions to $f(x) = s$.

5.17 Theorem (Riemann-Hurwitz formula): With f as above,

$$\chi(R) = \deg(f)\chi(S) - b.$$

Equivalently, in terms of the genus,

$$g(R) - 1 = (\deg f)(g(S) - 1) + \tfrac{1}{2}b,$$

where $g(X)$ denotes the genus of X.

(In particular, b must be even.)

5.18 Complement: Compactness of R and S is not essential in (5.17). As will emerge from the proof, what is necessary is *properness* of the map f and finiteness of $\chi(S)$; finiteness of $\chi(R)$ and the formula follow. In practice, we will study situations, as in (5.13), where $S = \mathbb{C}$; however, the preferred method will be to compactify R and S, and understand the branching over ∞.

5.19 Example: Recall the torus T given by the equation

$$w^2 = (z^2 - 1)(z^2 - k^2) \qquad k \neq \pm 1.$$

The projection to the z-axis has degree 2, and there are four branch points with total index 4. Now, T can be compactified by the addition of 2 points at ∞. The valencies must then be 1, because the degree was 2; so there is no branching there. We get

$$g(T) - 1 = 2 \cdot (0 - 1) + \tfrac{1}{2} \cdot 4 = 0,$$

or $g(T) = 1$, as expected.

Lecture 6

The Riemann-Hurwitz formula — Applications

6.1 Example: $f : \mathbb{P}^1 \to \mathbb{P}^1$ a polynomial of degree d. We have $g(R) = g(S) = 0$, so Riemann-Hurwitz gives

$$-1 = -d + \tfrac{1}{2}b,$$

or $b = 2(d - 1)$.

To see why that is, we pull the following theorem out of our algebraic hat.

6.2 Theorem: Let $f : \mathbb{C} \to \mathbb{C}$ be a polynomial mapping of degree d; then the total branching index over the *finite* branch points is exactly $(d - 1)$.

This means that the equation $f(x) = y$ has multiple roots for precisely $(d - 1)$ values of y, counting y k times when the equation $f(x) - y$ has only $d - k$ distinct roots.

6.3 Remark: This is quite clear in examples such as $f(x) = x^n$, or for any polynomial of degree 2. In general, f has multiple roots iff a certain expression — the discriminant of f — vanishes; and $\mathrm{disc}(f(x)-\alpha)$ is a polynomial in α of degree $(d-1)$. For generic f, $f-\alpha$ will have a double root for precisely $(d-1)$ values of α. Of course, reversing our arguments here will deduce thm. (6.2) from the Riemann-Hurwitz formula.

But we need another $(d-1)$ to agree with Riemann-Hurwitz. This of course comes from the point at ∞.

6.4 Lemma: $v_f(\infty) = \deg f$ for a polynomial f.

Proof: Let
$$f(z) = z^n + f_{n-1}z^{n-1} + \cdots + f_1 z + f_0.$$
Changing coordinates to $w = 1/z$, we have
$$\frac{1}{f(\frac{1}{w})} \to 0 \quad \text{as} \quad w \to 0$$
and
$$\frac{1}{f(\frac{1}{w})} = \frac{w^n}{1 + f_{n-1}w + \cdots + f_1 w^{n-1} + f_0 w^n},$$
and this has a zero of order n at 0.

As an application of the Riemann-Hurwitz formula, we shall now determine the topological type of certain concrete Riemann surfaces in $\mathbb{C}^2$.

We shall consider *algebraic* Riemann surfaces, which are solution sets of polynomial equations of the form $P(z,w) = 0$. We confine ourselves to polynomials P of the special form
$$P(z,w) = w^n + p_{n-1}(z)w^{n-1} + \cdots + p_1(z)w + p_0(z),$$
with the $p_k(z)$ polynomial functions of z. In practice, we shall study examples where P is simple enough, and the algebra is manageable.

We saw in Prop. (5.13) that the degree of $\pi : R \to \mathbb{C}_{(z)}$ is well-defined. Concerning the question whether R really is an (abstract) Riemann surface and π is holomorphic, recall the following from Props. (2.5) and (3.4):

6.5 Proposition: If the vector $(\partial P/\partial z, \partial P/\partial w)$ does not vanish anywhere on R, then R has a natural structure of an abstract Riemann surface; π is holomorphic for that structure, and its valency is 1 at each point where $\partial P/\partial w \neq 0$.

6.6 Remark: There are some obvious bad choices of P, for which he gradient condition in (6.5) fails; for instance, if $P(z,w) = f^2(z,w) \cdot g(z,w)$, its gradient will vanish identically on the zero-set of f. Because of that, we assume that P has *no repeated factors* in its factorization, as a polynomial in w (with coefficients in the ring $\mathbb{C}[z]$).

6.7 Proposition: The condition $(\partial P/\partial z, \partial P/\partial w) \neq 0$ on R guarantees that R is well-behaved in the following two ways:

(i) $\partial P/\partial w \neq 0$ at all but finitely many points on R.

(ii) For all but finitely many z, there are precisely n solutions to $P(z,w) = 0$. In particular, the degree of π is exactly n.

6.8 Remark: This no longer follows from general arguments, because R is not compact.

The complete proof is algebraic and takes us a bit off target now; so we only outline it here.

Idea of proof: View P as a polynomial in w, with coefficients in the unique factorization domain $\mathbb{C}[z]$ of polynomials in z. For part (i), if $\partial P/\partial w$ vanishes at a point (z_0, w_0) of R, it follows that the w-polynomials P and $\partial P/\partial w$ have a common root when $z = z_0$; but that happens iff the discriminant of P (6.3), as polynomials in w, vanishes when $z = z_0$. Now, the discriminant is a polynomial in z; if it has infinitely many roots, it vanishes identically. But then, it follows that P and $\partial P/\partial w$ have a common factor, as polynomials with coefficients in $\mathbb{C}[z]$, in violation of the multiplicity-free assumption (6.6).

For part (ii), the condition $P(z_0, w)$ has multiple roots, for fixed z_0, is equivalent to the vanishing of $\partial P/\partial w$ at one of the roots.

6.9 Terminology: A point $z_0 \in \mathbb{C}$ is a *branch point* for R if its inverse image contains points of valency > 1.

Although the Euler characteristic version can apply in the non-compact case, we get the most out of the Riemann-Hurwitz theorem when dealing with compact surfaces. The z-plane $\mathbb{C}_{(z)}$ is compactified easily, yielding the Riemann sphere $\mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$; we must now compactify R to R^{cpt}, by adding some points over ∞, in such a way that the extended map $\pi : R^{\text{cpt}} \to \mathbb{P}^1$ is holomorphic. Assuming this can be done, we obtain the following result, where N denotes the number of points at ∞ in R^{cpt}.

6.10 Proposition: R is homeomorphic to a compact surface of genus g with N points removed, where

$$\begin{aligned} g - 1 &= n(-1) + \tfrac{1}{2}b_{\text{finite}} + \tfrac{1}{2}(n - N) \\ &= \tfrac{1}{2}b_{\text{finite}} - \tfrac{1}{2}n - \tfrac{1}{2}N, \end{aligned}$$

b_{finite} being the total branching index over the finite branch points of π.

Proof: We are simply asserting that the total branching index over $\infty \in \mathbb{P}^1$ is $n - N$; but this is clear from the fact that the sum of the valencies over ∞ is the degree n of π.

The fact that R has a 'well-behaved' compactification follows from the following fact:

6.10 Proposition: Let $D^\times$ be the *outside* of some large disc in $\mathbb{C}$, which contains all branch points. Then $\pi^{-1}(D^\times)$ is analytically isomorphic to a union of N punctured discs, each mapping to $D^\times$ via a 'power map' $u \mapsto z = u^k$.

Restated slightly differently:

6.12 Proposition: Let $f : S \to \Delta^\times$ be a proper holomorphic map from a connected Riemann surface to the punctured unit disk. Assume that $v_f(s) = 1$ everywhere on S. Then S is isomorphic to $\Delta^\times$, in such a way that the map f becomes the d^{th} power map, where $d = \deg(f)$.

6.13 Remark: This result is really topological in nature. The proof is simple enough: we observe that the d^{th} power map from $\Delta^\times$ to itself can be 'lifted' to S, and that the lifting is a bijection. (Try to spell out the argument, and see exactly where connectedness comes in). Since all maps were local analytic isomorphisms, it follows that the new map is also a global analytic isomorphism.

From (6.12), it is clear that R can be compactified to a Riemann surface R^{cpt} by the addition of N points at ∞ — one for each disc — in such a way that $\pi : R^{\text{cpt}} \to \mathbb{P}^1$ is holomorphic.

The number N of points over ∞ will be the number of 'discs at ∞', and this is the number of connected components of $\pi^{-1}(D^\times)$, where $D^\times$ is the outside of a very large disc.

6.12 Example: $w^3 = z^3 - z$.

So we have

$$P(z,w) = w^3 - (z^3 - z), \qquad \frac{\partial P}{\partial z} = -3z^2 + 1, \qquad \frac{\partial P}{\partial w} = 3w^2.$$

So $P(z,w) = 0$ and $\partial P/\partial w = 0$ imply $w = 0$ which implies $z = 0$ or ± 1, and so $\partial P/\partial z \neq 0$. So R is a non-singular Riemann surface.

The branch points are the roots of $z^3 - z$, that is, $z = 0$ and $z = \pm 1$. Indeed, everywhere else there are three solutions for w. The valency of the projection at these points in 3, whence $b_{\text{finite}} = 2 + 2 + 2 = 6$.

So how many points are there over ∞? For $|z| > 1$, we can write $z^3 - z = z^3(1 - 1/z^2)$, and $\sqrt[3]{1 - 1/z^2}$ has the following convergent expansion:

$$1 - \binom{1/3}{1}\frac{1}{z^2} + \binom{1/3}{2}\frac{1}{z^4} - \binom{1/3}{3}\frac{1}{z^6} + \cdots$$

with

$$\binom{\alpha}{p} = \frac{\alpha(\alpha-1)\cdots(\alpha-p+1)}{p!}.$$

So $w^3 = z^3 - z$ has the three holomorphic solution functions

$$w = (3^{\text{rd}} \text{ root of } 1) \times z \times \sqrt[3]{1 - 1/z^2}$$

if $|z| > 1$, which describe three components of $\pi^{-1}(D^\times)$, if $D^\times$ is the outside of the unit disc. So $N = 3$. Riemann-Hurwitz gives

$$g(R^{\text{cpt}}) - 1 = -3 + \tfrac{1}{2} \cdot 6 = 0,$$

so $g = 1$ and R is a torus minus three points.

6.14 Example: $w^3 - 3w - z^2 = 0$.

6.15 Remark: Clearly this is best handelled by projecting to the w-axis, but we shall ignore this clever fact and proceed as before.

The branch points of π are the zeroes of $\partial P/\partial w = 3w^2 - 3$, so $w = \pm 1$, so $z^2 = w^3 - 3w = \mp 2$.

So there are four branch points, $z = \pm\sqrt{2}$ and $\pm i\sqrt{2}$. Now at those points, the polynomial factors as

$$w^3 - 3w - 2 = (w+1)^2(w-2)$$
$$w^3 - 3w + 2 = (w-1)^2(w+2),$$

so over each branch point we have a point of valency 1 and one of valency 2. So $b_{\text{finite}} = 1 + 1 + 1 + 1 = 4$.

Now for ∞: waving my arms a little, when z is very large, w must be large, too, for $w^3 - 3w = z^2$ does not have small solutions; so the leading term on the left is w^3, and the equation is roughly the same as

$$w^3 = z^2.$$

Notice in this case that the three 'sheets' of the solution $w(z)$ can be connected by letting z wind around zero, one or two times. ($3w$, being small, cannot 'alter' the number of times w winds around 0 when following z.) So in this case there is a single punctured disc going out to ∞, and $N = 1$. So

$$g - 1 = -3 + \tfrac{1}{2} \cdot 4 + \tfrac{1}{2} \cdot 2 = 0,$$

so $g = 1$ again, but $N = 1$ so R is a torus with a single point removed.

Lecture 7

Proof of Riemann-Hurwitz

To prove the Riemann-Hurwitz formula, we need to introduce a new notion — that of a triangulation of a surface — and give a rigorous definition of the Euler characteristic. To be completely rigorous, though, a mild digression on topological technology is required.

7.1 Definition: A topological space is *second countable*, or has a *countable base*, if it contains a countable family of open subsets U_n, such that every open set is a union of some of the U_n.

7.2 Example: A countable base for $\mathbb{R}$ is the collection of open intervals with rational endpoints. The exact same argument for this also proves the following.

7.3 Proposition: A metric space has a countable base iff it contains a dense countable subset.

7.4 Remark: Such metric spaces are often called *separable*.

Another easy observation is:

7.5 Proposition: A toplogical surface has a countable base iff it can be covered by countably many discs.

In particular, compact surfaces have a countable base. Every connected surface you can easily imagine has a countable base, but Prüfer has given an example of a connected surface admitting none. Such examples are necessarily quite pathological; it is common to exclude them by building the 'countable base' requirement into the definition of a surface.

7.6 Remark: It turns out that one does not exclude any interesting Riemann surfaces by insisting on the countable base condition. That is, it can be proved that every *connected* Riemann surface has a countable base, even if the condition was not included in the definition to begin with. (The proof is not obvious; see, for example, Springer, *Introduction to Riemann Surfaces*; you'll also find Prüfer's example there.)

The relevance of this topological techno-digression is the following theorem; 'triangulable' means pretty much what you'd think, but is defined precisely below.

7.7 Proposition: A connected surface is triangulable iff it admits a countable base. In particular, every Riemann surface is triangulable. (This was given a direct proof by Radò (1925).)

7.8 Definition: (see also 7.9) A *triangulation* of a surface S is the following collection of data:

(i) A set of isolated points on S, called *vertices*.

(ii) A set of continuous paths, called *edges*, each joining a pair of vertices.

The paths are required to be homeomorphic images of the closed interval, with endpoints at vertices; and two paths may not intersect, except at a vertex. Finally, only finitely many paths may meet at a given vertex.

(iii) The connected components of the complement of the edges are called *faces*. The closure of every face is required to be homeomorphic to a triangle.

One can show that the conditions imply the following:

- every point on an edge, which is not an endpoint, has a neighbourhood homeomorphic to Fig. 7.1.a, the edge being the diameter of the disc.
- Every endpoint of an edge is a vertex, and has a neighbourhood homeomorphic to Fig. 7.1.b, the (finitely many) spokes being edges.
- Every face is homeomorphic to the interior of a triange, with the homeomorphism extending continuously to the boundary of the triangle, taking edges homeomorphically to edges and vertices to vertices, as in Fig. 7.1.c.

7.9 Remark: In the literature one often imposes two further conditions:

- no two vertices are joined by more than one edge;
- two triangles sharing a pair of vertices share the corresponding edge.

This disallows things like Fig. 7.2.a. (Note that loops are already disallowed, because an edge must join a (disjoint) pair of vertices.) These extra restrictions are not material for our purposes; in fact, one could even be more generous and allows loops, and get funny triangles like the one in Fig. 7.2.b,

and he definition (7.11) below for the Euler characteristic still applies. A more useful generalization is the notion of a *polygonal decomposition*, where the faces are required only to be homeomorphic to convex polygons, rather than triangles. (Again, there is a strict version, where one requires two polygons to share no more than two vertices, and if so, they must share an edge, etc.)

7.10 Definition: A triangulation is called *finite* if it has finitely many faces.

Necessarily, then, it has finitely many edges and vertices. Note that any triangulation of a compact surface must be finite.

7.11 Definition: The *Euler characteristic* of a finitely triangulated surface is

$$\chi = V - E + F = \#\text{ vertices} - \#\text{ edges} + \#\text{ faces}.$$

7.12 Theorem: The Euler characteristic of a compact surface is a topological invariant — it does not depend on the triangulation. It is even computed correctly by any polygonal decomposition.

7.13 Proposition: The Euler characteristic of the orientable surface of genus g is $2 - 2g$.

7.14 Note: The propositions must be proved in the said order! We never proved the genus was a topological invariant of a surface, so the proposition really provides the first honest definition.

Sketch of proof of the theorem:

- One first checks that χ is unchanged when a polygon is subdivided into more polygons; that is, the Euler characteristic of the polygon is 1, no matter how it is polygonally decomposed. (This is essentially Problem 2, Sheet 2.)

  This shows that polygonal decompositions are as good as triangulations, because we can always decompose each polygon into triangles.

- One then shows that any two polygonal decompositions have a common refinement, possibly after perturbing one of them a bit.

  Perturbation may be necessary because edges may intersect badly, such as Fig. 7.3.a, which we must perturb as in Fig. 7.3.b:

  Following this, we add the 'nice' crossing point as a new vertex of the polygonal decomposition. Showing that such deformation is always possible, while intuitively obvious, is the one slight technical difficulty in the argument.

Proof of the Riemann-Hurwitz formula

Let $f : R \to S$ be a map of compact orientable surfaces, assumed to satisfy the condition described in the 'local form of holomorphic maps'; that is, we assume that near each $r \in R$,

$s = f(r) \in S$, there are neighbourhoods U_r and V_s homeomorphic to the unit disc, such that the map becomes $z \mapsto z^n$:

$$\begin{array}{ccc} U_r & \xrightarrow{f} & V_s \\ \phi \downarrow & & \downarrow \psi \\ \Delta & \xrightarrow{z^n} & \Delta \end{array}$$

Under these circumstances, the degree of f, valency at a point, and branching index are defined and satisfy the usual properties, and we have:

Theorem (Riemann-Hurwitz for χ):

$$\chi(R) = \deg(f)\chi(S) - b,$$

b being the total branching index $\sum_{r\in R}(v_f(r) - 1)$.

Proof: Start with a triangulation of S making sure that the branch points are included among the vertices. (Subdivide if necessary.) Also make sure that all edges and faces lie in small open sets, over which f has the required local form. It is then clear that the inverse images in R of the edges on S form the edges of a triangulation of R, and that there are $\deg(f)F$ faces and $\deg(f)E$ edges upstairs, if there were f faces and E edges downstairs. However, there are only $\deg(f)V - b$ vertices on R, if V was the number for S : the missing b vertices are the missing points over the branch points.

7.15 Remark: It is not necessary to subdivide, in order to ensure that the inverse image of a triangulation is also a triangulation and that the count comes out correctly; but that requires an extra argument (see 'lifting of simply connected spaces' in the next but last lecture).

Lecture 8

Elliptic functions

We now turn to the study of meromorphic functions on Riemann surfaces of genus 1.

The only Riemann surface of genus 0 is the Riemann sphere $\mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$. (This is not obvious: we are saying that any abstract Riemann surface structure on the 2-sphere ends up being isomorphic to the standard one. If you recall hat Riemann surface structures can be defined by gluing, you see why this is not a simple consequence of any definition). On $\mathbb{P}^1$, the meromorphic functions are rational, and those we understand quite explicitly; so it is natural to study tori next.

The tori we shall study are of the form $\mathbb{C}/L$, where $L \subset \mathbb{C}$ is a lattice — a free abelian subgroup for which the quotient is a topological torus. A less tautological definition is, viewing $\mathbb{C}$ as $\mathbb{R}^2$, that L should be generated over $\mathbb{Z}$ by two vectors which are not parallel. Calling them ω_1 and ω_2, the conditions are

$$\omega_1, \omega_2 \neq 0 \qquad \text{and} \qquad \frac{\omega_1}{\omega_2} \notin \mathbb{R}.$$

8.1 Exercise: Show, if $\omega_1/\omega_2 \in \mathbb{R}$, that $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \subset \mathbb{C}$ is either generated over $\mathbb{Z}$ by a single vector, or else its points are dense on a line. (The two cases correspond to $\omega_1/\omega_2 \in \mathbb{Q}$ and $\omega_1/\omega_2 \in \mathbb{R} \setminus \mathbb{Q}$.)

By definition, a function f is *holomorphic* on an open subset $U \subseteq \mathbb{C}/L$ iff $f \circ \pi$ is holomorphic on $\pi^{-1}(U) \subseteq \mathbb{C}$, where $\pi : \mathbb{C} \to \mathbb{C}/L$ is the projection.

Note that a 'fundamental domain' for the action of L on $\mathbb{C}$ is the 'period parallelogram'

Strictly speaking, to represent each point only once, we should take the interior of the parallelogram, two open edges and a single vertex; but it is more sensible to view $\mathbb{C}/L$ as arising from the closed parallelogram by identifying opposite sides. The notion of holomorphicity is pictorially clear now, even at a boundary point P — we require matching functions on the two half-neighbourhoods of P.

8.2 Remark: Division by ω_1 turns the period parallelogram into the form depicted in Fig. 8.2, with $\tau = \omega_2/\omega_1 \notin \mathbb{R}$.

Another presentation of the Riemann surface $T = \mathbb{C}/L$ is then visibly as $\mathbb{C}^*/\mathbb{Z}$, where the abelian group $\mathbb{Z}$ is identified with the multiplicative subgroup of $\mathbb{C}^*$ generated by $q = e^{\pi i \tau}$. We have a map $\exp : \mathbb{C} \to \mathbb{C}^*$ which descends to an isomorphism of Riemann surfaces, between $\mathbb{C}/L$ and $\mathbb{C}^*/\{q^{\mathbb{Z}}\}$.

Returning to the $\mathbb{C}/L$ description, we see that functions on T correspond to doubly periodic functions on $\mathbb{C}$, that is, functions satisfying

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

for all $u \in \mathbb{C}$. For starters, we note the following:

8.3 Proposition: Any doubly periodic holomorphic function on $\mathbb{C}$ is constant.

First Proof: Global holomorphic functions on $\mathbb{C}/L$ are constant.

Second Proof: By Liouville's theorem, bounded holomorphic functions on $\mathbb{C}$ are constant.

To get any interesting functions, we must allow poles.

8.4 Definition: An *elliptic function* is a doubly periodic meromorphic function on $\mathbb{C}$.

Elliptic functions are thus meromorphic functions on a torus $\mathbb{C}/L$. The reason for the name is lost in the dawn of time. (Really, elliptic functions can be used to express the arc-length on the ellipse.)

Constructing the first example of an elliptic function takes some work. We shall in fact describe them all; but we must start with some generalities.

8.5 Theorem: Let $z_1, \ldots, z_n$ and $p_1, \ldots, p_m$ denote the zeroes and poles of a non-constant elliptic function f in the period parallelogram, repeated according to multiplicity. Then:

(i) $m = n$,

(ii) $\sum_{k=1}^{m} \mathrm{Res}_{p_k}(f) = 0$,

(iii) $\sum_{k=1}^{n} z_k = \sum_{k=1}^{m} p_k \pmod{L}$.

8.6 Remark: Zeroes and poles that are on the boundary should be counted only on a single edge, or at a single vertex. In fact, we can easily avoid zeroes and poles on the boundary by shifting our parallelogram by a small complex number λ; the relations (i)–(iii) are unchanged.

8.7 Remark: Compared to rational functions, relation (i) is familiar, but (ii) and (iii) are new. They place some constraints on the existence part of the Unique Presentation Theorems for elliptic functions. We shall later see that those are the only constraints, that is, we shall prove:

8.8 Theorem (Unique Presentation by principal parts): An elliptic function is specified uniquely, up to an additive constant, by prescribing its principal parts at all poles in the period parallelogram. The prescription is subject only to condition (ii).

8.9 Theorem (Unique Presentation by zeroes and poles): An elliptic function is specified uniquely, up to a multiplicative constant, by prescribing the location of its zeroes and poles in the period parallelogram, with multiplicities. The prescription is subject to conditions (i) and (iii).

We'll prove these results next time. Meanwhile, we have yet to construct a single elliptic function!

Proof of theorem 8.5: Assume as before that no zeroes or poles are on the boundary of the period parallelogram.

(i) The 'variation of the argument' principle for meromorphic functions says

$$\oint_C \frac{f'(u)}{f(u)} : du = \oint_C d\log(f) = 2\pi i \times (\text{number of zeroes} - \text{number of poles}),$$

the number referring to the number enclosed by the contour. Taking C to be the boundary of the parallelogram, the integrands on opposite sides cancel, by periodicity; and this gives (i).

(ii) The Cauchy formula says

$$\frac{1}{2\pi i} \oint_C f(u)\, du = \sum_p \mathrm{Res}_p(f),$$

the sum going over all poles in the contour. Again we get zero by cancellation of opposite sides, thus concluding (ii).

(iii) Consider now

$$\frac{1}{2\pi i}\oint_C u\frac{f'(u)}{f(u)}\,du.$$

By another application of contour integrals over the boundary of the parallelogram, this gives

$$\sum_{k=1}^{n} z_k - \sum_{k=1}^{m} p_k.$$

This time we don't get zero because the opposite sides no longer cancel. Instead, comparing opposite sides:

$$\begin{aligned}\int_0^{\omega_1} u\frac{f'(u)}{f(u)}\,du + \int_{\omega_1+\omega_2}^{\omega_2} u\frac{f'(u)}{f(u)}\,du &= \int_0^{\omega_1} u\frac{f'(u)}{f(u)}\,du - \int_0^{\omega_1}(u+\omega_2)\frac{f'(u)}{f(u)}\,du \\ &= -\omega_2\int_0^{\omega_1}\frac{f'(u)}{f(u)}\,du \\ &= -\omega_2(\log f(\omega_1) - \log f(0))\end{aligned}$$

Now $f(\omega_1) = f(0)$, but the reason the expression fails to be 0 is the multi-valuedness of log. Indeed, $\log f(\omega_1) - \log f(0)$ can be any integer multiple of $2\pi i$.

All in all, we conclude that the value of our integral is equal to a lattice element.

Lecture 9

The Weierstraß $\wp$-function

We assume a lattice $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \subset \mathbb{C}$ has been chosen, with $\omega_1, \omega_2 \neq 0$, $\omega_1/\omega_2 \notin \mathbb{R}$.

We know that an elliptic function cannot have a *single, simple* pole in the period paralellogram. So the simplest assignment of principal parts is a double pole at $u = 0$. This leads to the *$\wp$-function of Weierstraß.*

9.1 Theorem/Definition: The *Weierstraß $\wp$-function* is the sum of the series

$$\wp(u) = \frac{1}{u^2} + \sum_{\omega\in L^*}\left[\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2}\right]$$

which converges to an elliptic function; the convergence is uniform on any compact subset $K \subset \mathbb{C}$, once the terms with poles in K are set aside.

Proof of convergence: For a compact $K \subset \mathbb{C}$, only finitely many terms will have poles in K. If the others converge uniformly, as we claim, holomorphy of their sum, and hence meromorphy of the entire series is a consequence of Morera's Theorem. Now, the individual terms are estimated by

$$\left|\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2}\right| < \frac{|u|^2 + 2|u||\omega|}{|\omega|^2|u-\omega|^2} = \frac{|u|^2}{|\omega|^2|u-\omega|^2} + 2\frac{|u|}{|\omega||u-\omega|^2}$$

and we have estimates $|u-\omega| > a^{-1}|\omega|$, $|u| < b$ for $u \in K$ and $\omega \in L \setminus K$; so

$$\left|\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2}\right| < \frac{a^2b^2}{|\omega|^4} + \frac{2a^2b}{|\omega|^3}$$

and the series on the right converges. (Proof: estimate by comparing with $\iint (x^2+y^2)^{-k} : dx\,dy$, with $k = \frac{3}{2}$ and $k = 2$. Alternatively, sum the series on parallelograms centered at the origin, and show that the sum over the boundary of each parallelogram decays at least quadratically with respect to the size of the parallelogram.)

9.2 Corollary: The series for $\wp(u)$ can be differentiated term by term and gives

$$\wp'(u) = -2 \sum_{\omega \in L} \frac{1}{(u-\omega)^3},$$

an elliptic function with a triple pole at 0.

9.3 Proposition: $\wp$ is even, that is, $\wp(u) = \wp(-u)$; $\wp'$ is odd, that is, $\wp'(u) = -\wp'(-u)$.

Proof: Clear from the series expansion.

First proof of periodicity of $\wp$: The derivative $\wp'(u)$ is evidently periodic, as a period shift in u can be reabsorbed by the lattice. But then,

$$\frac{d}{du}\left(\wp\left(u\right) - \wp\left(u+\omega_i\right)\right) = \wp'(u) - \wp'(u+\omega_i) = 0,$$

so $\wp(u) - \wp\left(u+\omega_i\right)$ is constant. Setting $u = -\omega_i/2$ and using the parity of $\wp$, we see that the constant is zero.

Second proof of periodicity: As a consequence of convergence,

$$\begin{aligned}
\wp(u+\omega_1) &= \frac{1}{(u+\omega_1)^2} + \sum_{\omega \in L^*} \left[\frac{1}{(u+\omega_1-\omega)^2} - \frac{1}{\omega^2}\right] \\
&= \frac{1}{(u+\omega_1)^2} + \left[\frac{1}{u^2} - \frac{1}{{\omega_1}^2}\right] + \sum_{\substack{\omega \in L^* \\ \omega \neq \omega_1}} \left[\frac{1}{(u+\omega_1-\omega)^2} - \frac{1}{\omega^2}\right] \\
&= \frac{1}{u^2} + \frac{1}{(u+\omega_1)^2} - \frac{1}{{\omega_1}^2} + \sum_{\substack{\omega \in L^* \\ \omega \neq -\omega_1}} \left[\frac{1}{(u-\omega)^2} - \frac{1}{(\omega+\omega_1)^2}\right] \\
&= \frac{1}{u^2} + \left[\frac{1}{(u+\omega_1)^2} - \frac{1}{{\omega_1}^2}\right] + \sum_{\substack{\omega \in L^* \\ \omega \neq -\omega_1}} \left[\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2}\right] + \sum_{\substack{\omega \in L^* \\ \omega \neq -\omega_1}} \left[\frac{1}{\omega^2} - \frac{1}{(\omega+\omega_1)^2}\right] \\
&= \wp(u) + \sum_{\substack{\omega \in L^* \\ \omega \neq -\omega_1}} \left[\frac{1}{\omega^2} - \frac{1}{(\omega+\omega_1)^2}\right].
\end{aligned}$$

But the last term vanishes due to an obvious symmetry, which cancels matching terms, namely

$$m\omega_1 + n\omega_2 \longrightarrow -(m+1)\omega_1 - n\omega_2.$$

9.4 Remark: We cannot quite break up the last series into a difference of two cancelling sums, because the separate series diverge. However, a clever order of summation can be used to show the vanishing, without exploiting the symmetry used in the proof (try to find that).

9.5 Proposition: $\wp'(\omega_1/2) = \wp'(\omega_2/2) = \wp'((\omega_1+\omega_2)/2) = 0$; these are all simple zeroes, and there are no others, modulo L.

Proof: $\wp'(\omega_1/2) = \wp'(\omega_1/2 - \omega_1) = \wp'(-\omega_1) = -\wp'(\omega_1)$ by periodicity and parity; same for the other half-lattice points. To see that there are no other zeroes, note the following important observation:

9.6 Proposition: $\wp$ defines a holomorphic map of degree 2 from $\mathbb{C}/L$ to $\mathbb{P}^1$; $\wp^{prime}$ defines a map of degree 3.

Proof: From the order of the pole and the definition of degree.

Let now e_1, e_2, e_3 be the values of $\wp$ at the half-lattice points $\omega_1/2, \omega_2/2, (\omega_1 + \omega_2)/2$.

9.7 Proposition:

(i) The e_i are all distinct.

(ii) For any $a \in \mathbb{C}$, $a \neq e_1, e_2, e_3$, the equation $\wp(u) = a$ has two simple roots in the period parallelogram; for those three exceptional values of a, it has a single double root.

Proof:

(ii) General theory ensures that we have either two simple roots or a double root. Since a double root is a zero of the derivative, (ii) follows. Note that the two solutions will always differ by a sign (mod L), by parity of $\wp$.

(i) If, say, $e_1 = e_2$, then $\wp(u) = e_1$ would have too many roots (a double root at ω_1 and another double root at ω_2.

9.8 Remark: The result about the number of roots can also be given a more elementary proof, using a contour integral argument.

Let us rephrase some of the last results in the following:

9.9 Proposition: $\wp : \mathbb{C}/L \to \mathbb{P}^1$ is a degree 2 holomorphic map with branch points over e_1, e_2, e_3, ∞.

Those of us who solved Example Sheet 1, Question 2, have seen the same picture of branching for the Riemann surface of the cubic equation

$$w^2 = (z - e_1)(z - e_2)(z - e_3);$$

in Lecture 10, we shall establish a deep connection between the two.

We will use the $\wp$-function to prove the Unique Presentation by principal parts. Uniqueness being clear on general grounds (cf. Lecture 4), we merely need to prove the existence statement; and this will emerge from the proof of the first theorem below. Remarkably, this will also allow us to describe the field of meromorphic functions over $\mathbb{C}/L$.

9.10 Theorem: Every elliptic function is a rational function of $\wp$ and $\wp'$. Specifically, every *even* elliptic function is a rational function of $\wp$, every *odd* elliptic function is $\wp' \times$(a rational function of $\wp$); and every elliptic function can be expressed uniquely as

$$f(u) = R_0(\wp(u)) + \wp'(u) \cdot R_1(\wp(u)),$$

with R_0, R_1 rational functions, where the two terms are the even and odd parts of f.

A remarkable consequence is that the function $\wp'(u)^2$, being elliptic and even, is expressible in terms of $\wp$. Explicitly, we have the following.

9.11 Theorem (Differential equation for $\wp$):

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

where $g_2 = 60G_4$, $g_3 = 140G_6$, and

$$G_r = G_r(L) = \sum_{\omega \in L^*} \omega^{-r}.$$

We postpone the proofs until next lecture. The two theorems immediadely lead to a description of the field of meromorphic functions on $\mathbb{C}/L$.

9.12 Corollary: The field of meromorphic functions on $\mathbb{C}/L$ is isomorphic to

$$\mathbb{C}(z)[w]/(w^2 - 4z^3 + g_2 z + g_3),$$

the degree 2 extension of the field of rational functions $\mathbb{C}(z)$ obtained by adjoining the solutions w to the equation $w^2 = 4z^3 - g_2 z - g_3$.

Lecture 10

Proof of theorems 8.8 and 9.10

First Proof of (9.10): It suffices to prove the statement for even elliptic functions; division by $\wp'$ reduces odd ones to even ones. Recall (Prop. 9.9) that $\wp : \mathbb{C}/L \to \mathbb{P}^1$ is a degree 2 holomorphic map. This map realizes $\mathbb{P}^1$ as the quotient space of the torus $\mathbb{C}/L$ under the identification of u with $-u$. (Because $\wp$ is even, these two points have the same image; u and $-u$ agree in $\mathbb{C}/L$ when u is a half-lattice point, but that is precisely where the fibre of $\wp$ contains a single point.) Hence, any even *continuous* map from $\mathbb{C}/L \to \mathbb{P}^1$ has the form $R \circ \wp$, for some continuous map $R : \mathbb{P}^1 \to \mathbb{P}^1$. Moreover, $\wp$ is a local analytic isomorphism away from the four branch points, which implies that R is holomorphic there, if $R \circ \wp$ was so. The following result rules out any possible trouble at those points and completes the proof.

10.1 Proposition: Let $f : S \to R$ be a continuous map between Riemann surfaces, known to be holomorphic except at isolated points. Then f is holomorphic everywhere.

Proof: Choosing coordinate neighbourhoods near the questionable points and their images we are reduced to the statement that a continuous function on Δ which is holomorphic on $\Delta^\times$ is, in fact, holomorphic at 0 as well. See now Problem 1.9.

Proof of (8.8): This follows from (9.10) and the UPT1 for rational functions.We'll be sketchy here, because we give a more constructive proof below. An even assignment of principal parts on $\mathbb{C}/L$ is pulled back,via $\wp$ from $\mathbb{P}^1$; and we just pull back the corresponding rational function. For odd assignments, division by $\wp'$ produces an even assignment. There is a fine point to check at 0, where $\wp'$ is singular, but we shall do so in the proof below.

Second Proof of (8.8) and (9.10): This is more computational, but also more concrete. We first show that we can realize any even assignment of principal parts on $\mathbb{C}/L$ using a suitable rational function $R(\wp(u))$. Such assignment involves finitely many points $\lambda \in \mathbb{C}/L$ and principal parts

$$\sum_{k=1}^{n_\lambda} a_k^{(\lambda)} (u - \lambda)^{-k},$$

with the properties that

- if $2\lambda \notin L$, then $(-\lambda)$ also appears, with assignment
$$\sum_{k=1}^{n_\lambda} (-1)^k a_k^{(\lambda)} (u+\lambda)^{-k},$$
i.e. $a_k^{(-\lambda)} = (-1)^k a_k^{(\lambda)}$;
- if $2\lambda \in L$ then only even powers of $(u-\lambda)^{-1}$ are present.

Now if $2\lambda \notin L$, $(\wp(u) - \wp(\lambda))^{-1}$ has a simple pole at $u = \lambda$ and we can create any principal part there as a sum of $(\wp(u) - \wp(\lambda))^{-k}$. Evenness of $\wp$ takes care of the symmetry. If $2\lambda \in L$ then we can use either powers of $\wp$, if $\lambda \in L$, or powers of $(\wp(u) - e_{1,2,3})^{-1}$, which have double poles with no residue.

Now, onto the odd functions. Odd assignments of principal parts are of the form
$$\sum_{k=1}^{n_\lambda} a_k^{(\lambda)} (u-\lambda)^{-k},$$
with a matching term
$$-\sum_{k=1}^{n_\lambda} (-1)^k a_k^{(\lambda)} (u+\lambda)^{-k}$$
at $-\lambda$ (i.e. $a_k^{(-\lambda)} = (-1)^{k+1} a_k^{(\lambda)}$), or else with vanishing $a_k^{(\lambda)}$ (k even) if $2\lambda \in L$. The principal parts
$$\left(\frac{P_\lambda}{\wp'(u)}, \frac{P_{-\lambda}}{\wp'(u)}\right)$$
can be realized by a sum of powers of $(\wp(u) - \wp(\lambda))^{-1}$. If $2\lambda \in L$ but $\lambda \notin L$ (not 0), then $P_\lambda^{(u)}/\wp'(u)$ is also a well-defined even principal part, expressible via $(\wp(u) - \wp(\lambda))^{-1}$. Same goes for $P_0^{(u)}/\wp'(u)$. So there exists a function of the form $R_1(\wp(u))$ whose principal parts agree with the $P_\lambda(u)/\wp'(u)$ everywhere. The principal parts of $R_1(\wp(u))\,\wp'(u)$ agree with the P_λ, except possibly at $\lambda = 0$, where the cubic pole of $\wp'$ could introduce unwanted or incorrect u^{-3} and u^{-1} terms. We can adjust the u^{-3} term by shifting R by a constant. We have no control over the u^{-1} term, but that is determined from the condition $\sum \mathrm{Res} = 0$; which indeed must be met if a function with prescribed principal parts is to exist.

Verification of the differential equation:

We start with the following.

10.2 Lemma (Laurent expansion of the Weierstrass function):

$$\wp(u) = u^{-2} + 3G_4(L)u^2 + 5G_6(L)u^4 + \cdots$$
$$\wp'(u) = -2u^{-3} + 6G_4(L)u + 20G_6(L)u^3 + \cdots$$

Proof:

$$(u-\omega)^{-k} = \frac{(-1)^k}{\omega^k}\left[1 + k\frac{u}{\omega} + \frac{k(k+1)}{2!}\frac{u^2}{\omega^2} + \frac{k(k+1)(k+2)}{3!}\frac{u^3}{\omega^3} + \cdots\right],$$

convergent for $|u| < |\omega|$.

Expanding each term in the series expansion for the $\wp$-function as above, and leaving the justification of convergence of the double series, for small values of u, as an enjoyable exercise (cf. Lecture 8), we notice that the odd powers of u cancel, and obtain

$$\wp(u) = u^{-2} + \sum_{m=1}^{\infty} \binom{-2}{2m} G_{2m+2}(L) u^{2m}$$

Similarly, for $\wp'(u)$ we get

$$\wp'(u) = -2u^{-3} + \sum_{m=0}^{\infty} -2\binom{-3}{2m+1} G_{2m+4}(L) u^{2m+1}.$$

Using the lemma, we compare the first few terms in the Laurent expansion of $(\wp')^2$ and $4\wp^3 - g_2\wp - g_3$ at $u = 0$ agree, so their difference is an elliptic function with no poles, vanishing at $u = 0$. Thus the two functions agree.

Geometric interpretation for $\wp$

Recall the Differential equation for the Weierstrass function,

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

10.3 Lemma: $g_2^3 \neq 27g_3^2$ and e_1, e_2, e_3 are the roots of the equation

$$4z^3 - g_2 z - g_3 = 0.$$

Proof: $\wp'$ vanishes at the half-lattice points, while $\wp$ takes the values e_1, e_2, e_3 there.

10.4 Theorem (Geometric interpretation): The map $\mathbb{C}/L \setminus \{0\} \to \mathbb{C}^2$ given by

$$u \mapsto (z(u), w(u)) = (\wp(u), \wp'(u))$$

gives an analytic isomorphism between the Riemann surface $\mathbb{C}/L \setminus \{0\}$ and the (concrete) Riemann surface R of the function $w^2 = 4z^3 - g_2 z - g_3$ in $\mathbb{C}^2$.

Proof: We have

$$\begin{array}{ccc} \mathbb{C}/L \setminus \{0\} & \xrightarrow{(\wp,\wp')} & R \\ & \searrow^{\wp} & \downarrow \pi \\ & & \mathbb{C} \end{array}$$

and we know that

- π is proper and 2–to–1 except at the branch points e_1, e_2, e_3, which are the roots of $4z^3 - g_2 z - g_3$.
- $\wp$ is proper and 2–to–1 except at the half-period points $\omega_1/2, \omega_2/2, \omega_1/2 + \omega_2/2$, which map to the roots e_1, e_2, e_3.
- $\wp(u) = \wp(-u)$ and $\wp'(u) = -\wp'(-u)$: this means that, unless u is a half-period, $\wp'$ takes both values $\pm w = \pm\wp'(u)$ at two points $\pm u$ mapping to the same point $z = \wp(u)$ of $\mathbb{C}$.

Together, these three properties show that the map we just constructed is bijective. Note, further, that at no point $u \in \mathbb{C}/L \setminus \{0\}$ is $\wp'(u) = \wp''(u) = 0$, because $\wp'$ has simple zeroes only (because there are three of them); this means that, for every $u \in \mathbb{C}/L \setminus \{0\}$, either the map $\wp$ or the map $\wp'$ gives an analytic isomorphism of a neighbourhood of u with a small disc in the z-plane or in the w-plane.

Since the Riemann surface structure on the (concrete, non-singular) Riemann surface R was defined by using the projections to the z- and w-planes, appropriately, we conclude that $(\wp, \wp')$ gives an analytic isomorphism $\mathbb{C}/L \to R$.

10.5 Remark: Clearly this map extends to a continuous map $\mathbb{C}/L \to R^{\text{cpt}} = R \cup \{\infty\}$. It is easy to show, proceeding as in lecture 6, that $R \cup \{\infty\}$ can be given a structure of abstract Riemann surface, so that the projection π extends to a holomorphic map $R^{\text{cpt}} \to \mathbb{P}^1$. It is then possible to show directly that the compactified Riemann surface R^{cpt} is 'the same' as $\mathbb{C}/L$, but in fact this follows from a much more general fact which we now state.

10.6 Summary: Starting with a lattice L, we have produced two elliptic functions $\wp, \wp'$ and realised the Riemann surface $\mathbb{C}/L\setminus\{0\}$ as the Riemann surface of the equation $w^2 = 4z^3 - g_2 z - g_3$. The coefficients g_2 and g_3 are determined from the lattice, and satisfy $g_2^3 - 27g_3^2 \neq 0$, which is the condition that the polynomial $4z^3 - g_2 z - g_3$ should have simple roots only.

We shall explain in the next lecture why the converse also holds, in the sense that a suitable choice of ω_1 and ω_2 suitably, will produce any preassigned g_2 and g_3, subject only to the condition $g_2^3 - 27g_3^2 \neq 0$. Thus, the Riemann surface of *any* equation $w^2 = p(z)$, with p a cubic polynomial with simple roots, is isomorphic to $\mathbb{C}/L \setminus \{0\}$. (We can get rid of the quadratic term in p by shifting z.)

Much more difficult is the following result, whose proof requires some serious analysis of the Laplace operator:

10.7 Theorem: Any Riemann surface homeomorphic to a torus is analytically isomorphic to $\mathbb{C}/L$, for a suitable lattice L.

Assuming that, we shall see that the lattice L is uniquely determined, save for an overall scale factor. In particular, *there is a continuous family of complex analytic structures on the torus.*

Lecture 11

Recovering the period lattice of a Riemann surface of genus 1

Let us first take for granted the fact that every Riemann surface of an equation

$$4z^3 - g_2 z - g_3 = w^2$$

with $g_2^3 \neq 27g_3^2$, 'comes from' the $\wp$-function of a suitable lattice; how could we recover the lattice, knowing only the equation?

Let us rephrase the question: say we are given an abstract Riemann surface T, and we are told it is of the form $\mathbb{C}/L$ for a certain lattice L. How can we recover L?

Note first that we can only recover L up to a scale factor $\alpha \in \mathbb{C}^\star$; indeed the Riemann surface $\mathbb{C}/\alpha L$ is isomorphic to T, by sending $u \in \mathbb{C}/\alpha L$ to $u/\alpha \in \mathbb{C}/L$. To get L up to scale requires an idea borrowed from calculus.

11.1 Definition: A *holomorphic differential 1-form* on an abstract Riemann surface is a 1-form which, in any local coordinate z, can be expressed as $f(z)\,dz$ with $f(z)$ a holomorphic function. Here, $dz = dx + i \cdot dy$. Note that on $\mathbb{C}/L$, we have a global holomorphic 1-form, namely, du.

11.2 Proposition: Every global holomorphic 1-form on T is a constant multiple of du.

Proof: The form must be expressible as $f(u)\,du$, with f a function on T that is holomorphic everywhere. So f must be constant.

Let now ϕ be any non-zero holomorphic differential on T. We know ϕ is a constant multiple of du (but of course we do not know the constant, if we do not know the presentation $\mathbb{C}/L$).

11.3 Proposition: Up to a constant, the lattice L is the set of values

$$\left\{ \int_C \phi \mid C \text{ is a closed curve in } T \right\}.$$

Proof: Closed curves in T correspond to curves in $\mathbb{C}$ whose endpoint differs from the origin by a lattice point. The integral of du along such a path is precisely the lattice element in question. We also see that we can obtain a basis of L by integrating ϕ along a meridian and along a parallel on the torus:

$(\omega_1, \omega_2) = (\int_{C_1} \phi, \int_{C_2} \phi)$ is a basis of (the scalar multiple of) L.

Note that the latter step involves a choice of cross-cuts on the torus, such choices corresponding precisely to the choice of basis of the lattice. So describing the lattice as the set of all integrals is a bit more 'canonical' that describing the basis.

Let us now identify the differential form and the cross-cuts on the surface R.

11.4 Proposition: The global holomorphic differential du is dz/w.

Proof:

$$\frac{dz}{w} = \frac{d\wp(u)}{\wp'(u)} = \frac{\wp'(u)\,du}{\wp'(u)} = du.$$

11.5 Proposition:

$$\omega_1 = 2 * \int_{e_2}^{e_3} \frac{dz}{w(z)}, \qquad \omega_2 = 2 * \int_{e_1}^{e_3} \frac{dz}{w(z)}.$$

11.6 Remark: We could write $dz/\sqrt{4z^3 - g_2 z - g_3}$ instead of $dz/w(z)$ to make the integrals look more sensible, but of course dz/w has the advantage that it is manifestly well-defined on R.

Proof: Recall that e_1, e_2, e_3 were the branch points of R, roots of $4z^3 - g_2 z - g_3$, but also they were the values $\wp(\omega_1/2), \wp(\omega_2/2), \wp(\omega_1/2 + \omega_2/2)$. So, changing variables to u, we get the obvious statements

$$\int_{\omega_2}^{\omega_1+\omega_2} du = \omega_1, \qquad \int_{\omega_1}^{\omega_1+\omega_2} du = \omega_2.$$

11.7 Remark: We can give an integral formula for the inverse $u(z) = \wp^{-1}(z)$ of the $\wp$-function as

$$u(z) = \int_{e_1}^{z} \frac{dz}{\sqrt{4z^3 - g_2 z - g_3}} + \omega_1/2 \left(= \int_{\omega_1/2}^{u(z)} du + \omega_1/2 \right).$$

Of course, u is determined only up to sign, and up to translation by L, which is reflected in a choice of paths of integration from e_1 to z and sign of $\sqrt{}$ along it.

A better explanation for these formulae lies in identifying the cross-cuts in R. Note that there are two choices of the sign of w on the segment from e_2 to e_3 in the z-plane, and the integral represents half of the contour integral that travels from e_2 to e_3 on one sheet of R and from e_3 to e_2 on the other sheet. It would also equal the contour integral on any simple loop C that surrounds e_2 and e_3, by an application of Cauchy's formula to the region between C and the degenerate contour represented by the segment travelled back and forth, as in Fig. 11.2:

$$\int_C \frac{dz}{w} = \int_{e_2}^{e_3} \frac{dz}{w} + \int_{e_3}^{e_2} \frac{dz}{w} = 2\int_{e_2}^{e_3} \frac{dz}{w} \qquad \text{by symmetry.}$$

Let us now identify the contour on R, and the corresponding contour linking e_1 with e_3:

Clearly in the identification of R^{cpt} with the torus, the two contours become a pair of cross-cuts (Fig. 11.3.b):

11.8 Caution: There is a choice hidden in $\int_{e_2}^{e_3} dz/w$, in the sign of the square root in the expression of w in terms of z. More seriously, if e_1, e_2, e_3 happen to be colinear, and, say, e_1 lies between the other two, there is even more choice, as we can choose the signs of w independently on the two subintervals. Combined with the fact that the roots of a cubic are not naturally ordered, this clearly shows that there is no canonical determination of ω_1 and ω_2 from the polynomial $4z^3 - g_2 z - g_3$; only the lattice L is canonical.

We'd like to summarize our results in the following:

11.9 Theorem: There is a bijective correspondence between lattices $L \subset \mathbb{C}$ and polynomials $f(z) = 4z^3 - g_2 z - g_3$ with *simple* roots, so that the compactified Riemann surface of $w^2 = f(z)$ is $\mathbb{C}/L$. The gs are expressed in terms of the Eisenstein series of L, while the lattice elements are the values of the integral $\int dz/\sqrt{f(z)}$ along closed loops on R.

Unfortunately, we have not quite proved this yet, because we have been assuming, in going from R to L, that R was already parametrized by some $\wp$-function. If that was not the case, our computation of ω_1, ω_2 and $\wp^{-1}$ would be meaningless. The missing steps are supplied by the following propositions.

11.10 Proposition: Let T be a Riemann surface of genus 1. If T has a non-zero global holomorphic differential ϕ, then $T = \mathbb{C}/L$, with L being the 'lattice of periods' $\int_C \phi$ for closed curves $C \subset T$.

Proof: Simply put, pick a base point $P \in T$ and note that the map $\theta \mapsto \int_P^\theta \phi$ defines a bijective analytic map $T \to \mathbb{C}/L$.

11.11 Proposition: If $f(z) = 4z^3 - g_2 z - g_3$ has simple roots, then dz/w is a global holomorphic differential on R^{cpt}, the compactified Riemann surface of $w^2 = f(z)$.

Proof: Clear when $w \neq 0, \infty$; near $w = 0$, note that w, not z, is a local coordinate on R, and that $(z - e_1) = O(w^2)$; whence dz/w is holomorphic. A similar argument works at ∞ using a local coordinate $v = 1/\sqrt{z}$.

*Appendix: The elliptic modular function J

It turns out (but this is much more difficult) that any Riemann surface of genus 1 carries non-zero holomorphic differentials. Combining that with out knowledge, we get:

11.12 Theorem: Isomorphism classes of Riemann surfaces of genus 1 are in bijection with lattices in $\mathbb{C}$, modulo scaling.

11.13 Proposition: { Lattices in $\mathbb{C}$, up to scaling } are in bijection with the orbits of $PSL(2, \mathbb{Z})$ on the upper half-plane $\mathfrak{H} = \{z \mid \text{Im}(z) > 0\}$, acting by Möbius transformations. The bijection takes a basis $(2\omega_1, 2\omega_2)$ of L to the number $\tau = \pm\omega_2/\omega_1$, the sign being adjusted so that $\text{Im}(\tau) > 0$.

Proof: Clearly, we can rescale the lattice so that one period is 1, so all the information is in τ. But this involves a choice of basis of L, and another basis differs from ω_1, ω_2 by the action of $PGL(2, \mathbb{Z})$. This would send τ to a suitable Möbius transform of τ. The condition that $\text{Im}(\tau) > 0$ cuts down our group to $PSL(2, \mathbb{Z})$.

11.14 Fact: A *fundamental domain* for the action of $SL(2, \mathbb{Z})$ on $\mathfrak{H}$ consists of the portion of the strip $-\frac{1}{2} \leq \text{Re}(z) \leq \frac{1}{2}$ which lies outside the unit circle. The map $\tau \mapsto \tau + 1$ identifies the two sides of the strip, while $\tau \mapsto -1/\tau$ identifies the two curved arcs on the unit circle by reflection. The result of these identifications is a topological space that can be seen to be homeomorphic to $\mathbb{C}$. Actually, much more is true.

11.15 Definition: The *elliptic modular function* $J : \mathfrak{H} \to \mathbb{C}$ sends $\tau \in \mathfrak{H}$ to $g_2^3/(g_2^3 - 27g_3^2) \in \mathbb{C}$, where g_2, g_3 are the coefficients of the cubic function associated to the lattice represented by τ. (We can use $L = \mathbb{Z} + \mathbb{Z}\tau$.) From the bijection between lattices and pairs (g_1, g_2) we now get:

11.16 Theorem: J is holomorphic, invariant under the action of $PSL(2,\mathbb{Z})$, and maps the orbits of $SL(2,\mathbb{Z})$ on $\mathfrak{H}$ bijectively onto $\mathbb{C}$.

In other words, the quotient $\mathfrak{H}/SL(2,\mathbb{Z})$ inherits the structure of an abstract Riemann surface; and J establishes an *analytic isomorphism* between this surface and $\mathbb{C}$.

Lecture 12

Two new Weierstraß functions

To prove the Unique Presentation by zeroes and poles we'd like a function like 'z' — an elliptic function with a single zero and no poles. That of course does not exist, but if we weaken the condition of double periodicity, a suitable function will emerge.

12.1 Definition: The *Weierstraß ζ-function* $\zeta(u)$ is the negative of the unique odd antiderivative of $\wp$,

$$\zeta(u) = -\int_{u_0}^{u} \wp(\xi)\, d\xi + c,$$

where $u_0 \notin L$ and c is chosen so that $\zeta(u) = -\zeta(-u)$.

Note that the integral is *path-independent* by Cauchy's residue formula, because $\wp$ has no residue at any of its poles.

12.2 Note: Explicitly, we can write

$$\zeta(u) = -\frac{1}{2}\int_{-u}^{u} \wp(\xi)\, d\xi.$$

12.3 Proposition:

$$\zeta(u) = \frac{1}{u} + \sum_{\omega \in L^*} \left[\frac{1}{u-\omega} + \frac{1}{\omega} + \frac{u}{\omega^2}\right].$$

Proof: Having established uniform convergence is established (this is left as an exercise), the relation $\zeta'(u) = \wp(u)$ follows by differentiating term by term; and the sum of the series is an odd function of u.

12.4 Proposition: $\zeta(u)$ has simple poles at lattice points, with residue 1.

Proof: This is clear from the uniform convergence of the series; away from a lattice point, all terms are holomorphic, while the unique singular term at a given lattice point has a simple pole with residue 1.

12.5 Proposition (Periodicity of ζ):

$$\zeta(u + \omega_i) = \zeta(u) + \eta_i,$$

where the η_i are given by

$$\eta_i = -\int_{-\omega_i/2}^{\omega_i/2} \wp(u)\, du,$$

on any path of integration which avoids the poles.

Proof: The difference $\zeta(u+\omega_i)-\zeta(u)$ has vanishing derivative, so it must be constant. Inserting $u=\omega_i/2$ in the definition of ζ gives the values.

There is no algebraic expression of the ηs in terms of the periods; but we have the

12.6 Proposition (Legendre Identity):

$$\eta_1\omega_2 - \eta_2\omega_1 = 2\pi i.$$

Proof: We consider the integral of $\xi\,\wp(\xi)$ along a period parallelogram, shifted by $-(\omega_1+\omega_2)/2$. It encloses a single pole (at 0), and, by the residue formula, the answer must be $2\pi i$. On the other hand, collecting the two horizontal sides of the parallelogram gives, using periodicity of $\wp$,

$$\int_{(-\omega_1-\omega_2)/2}^{(\omega_1-\omega_2)/2} \xi\,\wp(\xi)\,d\xi - \int_{(-\omega_1+\omega_2)/2}^{(\omega_1+\omega_2)/2} \xi\,\wp(\xi)\,d\xi = -\int_{(-\omega_1+\omega_2)/2}^{(\omega_1+\omega_2)/2} \omega_2\,\wp(\xi)\,d\xi = \omega_2\eta_1,$$

while the vertical sides give similarly

$$\int_{(\omega_1-\omega_2)/2}^{(\omega_1+\omega_2)/2} \xi\,\wp(\xi)\,d\xi - \int_{(-\omega_1-\omega_2)/2}^{(-\omega_1+\omega_2)/2} \xi\,\wp(\xi)\,d\xi = \int_{\omega_1-\omega_2}^{\omega_1+\omega_2} \omega_1\,\wp(\xi)\,d\xi = -\omega_1\eta_2.$$

The second Weierstrass function we shall consider is the *exponential antiderivative* of ζ.

12.7 Definition:

$$\sigma(u) = \exp\int^u \zeta(\xi)\,d\xi,$$

with the (multiplicative) constant ambiguity is adjusted such that $\sigma'(0)=1$.

The ambiguities in the antiderivative $\int \zeta(\xi)\,d\xi$ are are resolved by a choice of path of integration; two paths from u_0 to u will give answers differing by $2\pi i\times$(the number of lattice points enclosed), so the exponential is well-defined. Moreover, all singularities of the integral are logarithmic, at the lattice points of L, and convert into simple zeroes upon exponentiation; indeedn, the Laurent expansion of ζ at $u=0$, $\zeta(u)=1/u+O(u^3)$ gives the antiderivative $\log u + O(u^4) + const$, and the behaviour at the other lattice points is similar, by periodicity. The overall multiplicative factor depends on a lower bound of integration, which has not been specified. We adjust it by requiring that $\sigma'(0)=1$ (this amounts to using the antiderivative with no constant term at $u=0$.)

Integrating the series expansion of ζ and exponentiating leads to the expression

$$\sigma(u) = u\prod_{\omega\in L^*}\left[\left(1-\frac{u}{\omega}\right)\exp\left(\frac{u}{\omega}+\frac{u^2}{2\omega^2}\right)\right].$$

12.8 Proposition: The infinite product converges uniformly on compact subsets. Moreover, its logarithm converges uniformly on compact subsets of $\mathbb{C}$, once the singular terms are set aside, and a the principal branch is chosen for the remaining logarithms.

Proof: Convergence follows from the estimate

$$\left|\log\left(1-\frac{u}{\omega}\right)+\frac{u}{\omega}+\frac{u^2}{\omega^2}\right| < \frac{C}{|\omega|^3}$$

for suitable C, if u ranges over a compact set K and $\omega\in L\setminus K$.

12.9 Remark: In connection with infinite products, recall the following from calculus: the product $\prod(1+a_n)$ and its inverse $\prod(1+a_n)^{-1}$ converge as soon as $\sum |a_n| < \infty$ (and $a_n \neq -1$). In particular, uniform convergence of $\prod(1+a_n(z))$ and $\prod(1+a_n(z))^{-1}$ follows from a uniform convergence of $\sum a_n(z)$; in particular the limit functions are then holomorphic if the $a_n(z)$ were so (except of course we'll get poles for $\prod(1+a_n(z))^{-1}$ wherever some $a_n(z) = -1$).

12.10 Proposition (Periodicity of σ):

$$\sigma(u+\omega_i) = -\sigma(u)\exp(2\eta_i(u+\omega_i)).$$

Proof: Taking log derivatives on both sides shows the ratio of the two to be constant. Evaluation at $u = -\omega_i$ plus the relation $\sigma(-u) = -\sigma(u)$ shows that we got the factor right.

Proposition: Let $z_1, \ldots, z_n$ and $p_1, \ldots, p_n \in \mathbb{C}$ be such that $z_1 + \cdots + z_n = p_1 + \cdots + p_n$. Then

$$f(u) = \prod_{i=1}^{n} \frac{\sigma(u-z_i)}{\sigma(u-p_i)}$$

is an elliptic function with a simple zero at each z_i and a simple pole at each p_i. (Obviously, repeated values lead to multiple zeroes or poles.)

Proof: The periodicity factors

$$\frac{f(u+2\omega_i)}{f(u)}$$

of the product are

$$\exp\left(\sum_{j=1}^{n} 2\eta_i(u+\omega_i-z_j) - \sum_{j=1}^{n} 2\eta_i(u+\omega_i-p_j)\right) = 1.$$

12.11 Corollary: Unique Presentation by zeroes and poles (Lecture 8).

We shall now make the σ-function nicer by making it as periodic as possible. Clearly we cannot remove both periodicity factors because we would get an elliptic holomorphic function which would have to be constant. But we can remove one of the periodicity factors by multiplying by a quadratic exponential. This leads to:

Lecture 13

The Jacobi θ-functions

Let $f(u) = \sigma(u)\exp(-\eta_1 u^2/2\omega_1)$, and let $\tau = \omega_2/\omega_1$.

Proposition:

$$f(u+2\omega_1) = -f(u)$$
$$f(u+2\omega_2) = -f(u)\cdot\exp\left(-\frac{\pi i u}{\omega_1} - \pi i\tau\right).$$

Proof: Follows by direct computation, using the Legendre identity $\eta_1\omega_2 - \eta_2\omega_1 = \pi i/2$ of lecture 10.

From now on we assume $\omega_1 = 1$, whence $\omega_2 = \tau$, which can always be accomplished by rescaling the lattice (and u). Recall the condition $\mathrm{Im}(\tau) > 0$; we can always meet these requirements by rescaling the lattice (and u) and swapping ω_1 and ω_2, if needed.

Let $p = e^{\pi i u}$, $q = e^{\pi i \tau}$. The relation $f(u+z) = f(u)$ shows that f depends on u only via $p \in \mathbb{C}^*$, and has a convergent Laurent expansion (Sheet 2, Ex. ?)

$$f(u) = \sum_{n\in\mathbb{Z}} f_n(\tau)\, p^n.$$

The periodicity relations become now

$$\begin{aligned} f(u+1) &= -f(u) \\ f(u+\tau) &= -f(u)\cdot p^{-2}q^{-1}, \end{aligned}$$

or

$$\begin{aligned} \sum f_n(\tau)\,(-p)^n &= -\sum f_n(\tau)\, p^n \\ \sum f_n(\tau)\,(pq)^n &= -\sum f_n(\tau)\, p^n p^{-2} q^{-1}. \end{aligned}$$

Proposition: Up to a multiplicative constant, $f(u)$ is the *first Jacobi θ-function*

$$\theta_1(u) = \theta_1(u \mid \tau) = -i \sum_{n\in\mathbb{Z}} (-1)^n p^{2n+1} q^{(n+1/2)^2}.$$

Proof: Exercise (Problem 11); the $f_n(\tau)$ are determined from the periodicity relations, once one of them is known. Note also, since $p^k - p^{-k} = 2i\sin(k\pi u)$, that we can rewrite

$$\theta_1(u) = 2\sum_{n\geq 0} (-1)^n q^{(n+1/2)^2} \sin((2n+1)\pi u).$$

By analogy with the trigonometric functions, we can define new θ-functions by translating θ_1 by half-periods. We thus get

$$\begin{aligned} \theta_2(u \mid \tau) &= \theta_1(u + \tfrac{1}{2} \mid \tau) \\ \theta_3(u \mid \tau) &= pq^{1/4}\theta_1(u + \tfrac{1}{2} + \tfrac{\tau}{2} \mid \tau) \\ \theta_4(u \mid \tau) &= ipq^{1/4}\theta_1(u + \tfrac{\tau}{2} \mid \tau). \end{aligned}$$

The neatest of the series expansions is no doubt

$$\theta_3(\tau) = \sum p^{2n} q^{n^2}$$

while the others are

$$\begin{aligned} \theta_2(\tau) &= \sum p^{2n+1} q^{(n+1/2)^2}, \\ \theta_4(\tau) &= \sum (-1)^n p^{2n} q^{n^2}. \end{aligned}$$

Proposition: Each of the θ-functions is entire holomorphic in u, and has a simple zero in the fundamental parallelogram. θ_1 is odd, while the others are even.

Remark: The limits $\tau \to +i\infty$, so $q \to 0$ are enlightening: $q^{-1/4}\theta_1 \to 2\sin(\pi u)$, $q^{-1/4}\theta_2 \to 2\cos(\pi u)$, while $\theta_3, \theta_4 \to 1$.

The half-period transformation properties of the $\theta's$ are summarized in the following table, the top row indicating the shift in u; the periodicity factors are $a = p^{-1}q^{-1/4}$ and $b = p^{-2}q^{-1}$:

	$+1/2$	$+\tau/2$	$+(1+\tau)/2$	$+1$	$+\tau$
θ_1	θ_2	$ia\theta_4$	$a\theta_3$	$-\theta_1$	$-b\theta_1$
θ_2	$-\theta_1$	$a\theta_3$	$-ia\theta_4$	$-\theta_2$	$b\theta_2$
θ_3	θ_4	$a\theta_2$	$ia\theta_1$	θ_3	$b\theta_3$
θ_4	θ_3	$ia\theta 1$	$a\theta_2$	$\theta 4$	$-b\theta_4$

Like the trigonometric functions, the θ-functions satisfy a remarkable collection of identities, some of which have remarkable combinatorial and number-theoretic applications. The identities come from their relation to elliptic functions (the ratio of any two θ-functions is elliptic); the applications come from the presence of an 'n^2' in the exponent of q, in the power series expansion. We start with some simple ones.

Proposition.

$$\theta_2^2(0)\theta_2^2(u) + \theta_4^2(0)\theta_4^2(u) = \theta_3^2(0)\theta_3^2(u)$$

Proof: The ratio

$$\frac{\theta_3^2(0)\theta_3^2(u) - \theta_2^2(0)\theta_2^2(u)}{\theta_4^2(0)\theta_4^2(u)}$$

is periodic for 1 and τ and has poles of degree 2 at lattice translates of $\tau/2$. But $\theta_3(\tau/2) = a\theta_2(0)$ and $\theta_2(\tau/2) = a\theta_3(0)$, so the numerator vanishes and the pole has order 1 at most. But this means the function is constant. Setting $u = 1/2$ reveals the constant to be 1.

Proposition.

$$\wp(u) = e_1 + \left[\frac{\theta_1'(0)}{\theta_1(u)} \cdot \frac{\theta_2(u)}{\theta_2(0)}\right]^2$$

Proof: Exercise (check periodicity on the right and compare poles and principal parts; finally, evaluate at $u = 1/2$.

A remarkable source of combinatorial identities is the *product expansion* of the theta-functions. The formulas are

$$\begin{aligned}
i\theta_1(u) &= Cq^{1/4}(p - p^{-1}) \prod_{n=1}^{\infty} \left(1 - q^{2n}p^2\right)\left(1 - q^{2n}p^{-2}\right) \\
\theta_2(u) &= Cq^{1/4}\left(p + p^{-1}\right) \prod_{n=1}^{\infty} \left(1 + q^{2n}p^2\right)\left(1 + q^{2n}p^{-2}\right) \\
\theta_3(u) &= C \prod_{n=1}^{\infty} \left(1 + q^{2n-1}p^2\right)\left(1 + q^{2n-1}p^{-2}\right) \\
\theta_4(u) &= C \prod_{n=1}^{\infty} \left(1 - q^{2n-1}p^2\right)\left(1 - q^{2n-1}p^{-2}\right),
\end{aligned}$$

with the 'constant' C independent of p, but depending on q, as described below.

Sample proof for θ_3: The zeroes are simple and in the right places, so θ_3 over the product is holomorphic. Periodicity under $u \mapsto u+1$ ($p \mapsto -p$) is clear, while $u \mapsto u+\tau$ has the effect $p \mapsto pq$ under which the product picks up an overall factor of b, just as θ_3 does; so the ratio is elliptic and hence constant.

Half-period translation produces the other three relations, with the *same* constant C. To determine it, one enters the product expansions , evaluated at $p=1$, into Jacobi's null value identity discussed below; C appears on one side and C^3 on the other, which determines C^2. The sign of C is found in the limit $q \to 0$, and the final answer is the attractive

$$C = \prod_{n=1}^{\infty} \left(1 - q^{2n}\right).$$

Equating the sum and product expansion for θ_3 gives a Jacobi's famous *triple product identity*:

$$\sum_{n\in\mathbb{Z}} p^{2n} q^{n^2} = \prod_{n=1}^{\infty} \left(1-q^{2n}\right)\left(1+q^{2n-1}p^2\right)\left(1+q^{2n-1}p^{-2}\right)$$

Finally, let us use theta-functions to prove the a serious theorem.

Theorem: The number of ways of writing a number n as a sum of two squares equals four times the difference between the number of its divisors of the form $(4k+1)$ and the number of its divisors of the form $(4k+3)$.

In particular, primes of the form $(4k+3)$ cannot be written as sums of two squares; primes of the form $(4k+1)$ can, in a unique way.

Remark: For a prime of the form $4k+1$, the number in question is 8; but $p = a^2 + b^2$ leads to seven other obvious expressions by changing the signs and the order of a and b.

Proof: The proof, due to Jacobi, starts with the expression of $\theta_3(0)^2$ as $\sum_{m\geq 0} r_2(m)\, q^m$, where $r_2(m)$ is the number of ways of writing m as a sum of two squares (see Problem 12).

We shall next show that we also have

$$\theta_3(0)^2 = 1 + 4\sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\left[q^{n(4l+1)} - q^{n(4l+3)}\right],$$

and comparing coefficients will give the result.

The series in $(*)$ is four times

$$\begin{aligned}\sum_{n=0}^{\infty}(q^n - q^{3n})\sum_{l=0}^{\infty} q^{4ln} &= \sum_{n=1}^{\infty}(q^n - q^{3n})\frac{1}{1-q^{4n}} \\ &= \sum_{n=1}^{\infty}\frac{q^n - q^{3n}}{1-q^{4n}} \\ &= \sum_{n=1}^{\infty}\frac{q^n}{1+q^{2n}}\end{aligned}$$

(because the expression is symmetric under $n \longleftrightarrow (-n)$)

$$= \frac{1}{2}\sum_{n\in\mathbb{Z}} \frac{q^n}{1+q^{2n}} - \frac{1}{4},$$

and so

$$(*) = 2\sum_{n\in\mathbb{Z}} \frac{q^n}{1+q^{2n}}.$$

But this is $2i$ times the value at $p = i$ of the series

$$\sum_{n\in\mathbb{Z}} \frac{q^n p^{-1}}{1-q^{2n}p^{-2}}.$$

(To see convergence of the latter, break it up as

$$\sum_{n=0}^{\infty} \frac{q^n p^{-1}}{1-q^{2n}p^{-2}} - \sum_{n=0}^{\infty} \frac{q^n p}{1-q^{2n}p^{2}},$$

and use $|q| = |e^{\pi i\tau}| < 1$.) Now the latter is an elliptic function of u, with period lattice spanned by 2 and τ. ($p = e^{\pi iu}$ so 2 is clear, while $u \mapsto u+\tau$ has the effect $p \mapsto pq$, and this clearly preserves the series.) It has poles at $u = 0$ and $u = 1$ ($p = \pm 1$) and of course all points obtained from these by translation by $\mathbb{Z}$ and $\mathbb{Z}\tau$, but nowhere else; and the residues are easily computed to be

$$\mathrm{Res}_{u=0} \frac{p^{-1}}{1-p^{-2}} = \mathrm{Res}_{u=0} \frac{1}{p-p^{-1}} = \frac{1}{2i}\,\mathrm{Res}_{u=0} \frac{1}{\sin(\pi u)} = \frac{1}{2\pi i},$$

and similarly, at $u = 1$, the residue is $-1/(2\pi i)$.

Now we shall write a ratio of θ-functions with the same periods, poles and residues. (This is not surprising, because θ-functions are closely related to σ and ratios of translates of σ are elliptic functions.) The correct combination is

$$\frac{1}{2\pi i}\cdot\frac{\theta_1'(0)}{\theta_1(u)}\cdot\frac{\theta_4(u)}{\theta_4(0)},$$

the values at 0 being there to normalize the residue correctly. Indeed, $\theta_1(u+1) = -\theta_1(u)$ so $\theta_1(u+2) = \theta_1(u)$; and

$$\theta_4(u+1) = -ipq^{1/4}\theta_1(u+1+\tfrac{\tau}{2}) = ipq^{1/4}\theta_1(u+\tfrac{\tau}{2}) = \theta_4(u),$$

while $\theta_1(u+\tau) = -p^{-2}q^{-1}\theta_1(u)$ and

$$\begin{aligned}\theta_4(u+\tau) = ipq^{5/4}\theta_1(u+\tau+\tfrac{\tau}{2}) &= (-p^{-2}q^{-2})ipq^{5/4}\theta_1(u+\tfrac{\tau}{2})\\ &= (-p^{-2}q^{-1})ipq^{1/4}\theta_1(u+\tfrac{\tau}{2}) = (-p^{-2}q^{-1})\,\theta_4(u)\end{aligned}$$

so the ratio $\theta_4(u)/\theta_1(u)$ is periodic for τ and for 2; the residues are left as an exercise.

Remark: We can also now identify the functions in $(**)$ as $\sqrt{\wp(u)-e_3}$, $\wp$ for the original lattice $(1,\tau)$; we see the equality by comparing zeroes and poles. (θ_4 has a zero at lattice translates of $(1+\tau)/2$.)

Finally, combining $(*)$ and $(**)$, the desired identity is

$$\theta_3(0)^2 = \frac{1}{\pi}\cdot\frac{\theta_1'(0)\,\theta_4(\frac{1}{2})}{\theta_1(\frac{1}{2})\,\theta_4(0)} = \frac{1}{\pi}\cdot\frac{\theta_1'(0)\,\theta_3(0)}{\theta_2(0)\,\theta_4(0)}$$

(note that $p = i$ corresponds to $u = 1/2$), which is

$$\theta_1'(0) = \pi\, \theta_2(0)\, \theta_3(0)\, \theta_4(0).$$

This important 'null values' identity was established by Jacobi. No 'direct' proof from the definitions is known; in terms of power series, the identity reads

$$\sum_{n\in\mathbb{Z}} (-1)^n (2n+1)\, q^{n^2+n} = \sum_{(k,l,m)\in\mathbb{Z}^3} (-1)^m q^{k^2-k+l^2+m^2},$$

which is far from obvious!

In what follows, to conclude the chapter with an honest proof, I reproduce Whittaker and Watson's proof of the identity. It involves more θ-function work.

Proof of $(***)$**:** We start with *Jacobi's duplication formula*

$$\theta_1(2u) = 2\, \frac{\theta_1(u)\, \theta_2(u)\, \theta_3(u)\, \theta_4(u)}{\theta_2(0)\, \theta_3(0)\, \theta_4(0)}$$

by checking by hand that the ratio is periodic for $(1, \tau)$ (you do the checking, using the definitions) and has no poles (both sides have simple zeroes at all the half-lattice points). So the ratio is constant, and the constant is 1 as can be seen from the limit as $u \to 0$.

Log differentiation gives

$$2\, \frac{\theta_1'}{\theta_1}(2u) = \sum_{k=1}^{4} \frac{\theta_k'}{\theta_k}(u)$$

and differentiating again gives

$$2 \frac{d}{du}\left(\frac{\theta_1'}{\theta_1}(2u)\right) - \frac{d}{du}\left(\frac{\theta_1'}{\theta_1}(u)\right) = \sum_{k=2}^{4} \frac{\theta_k''}{\theta_k}(u) - \sum_{k=2}^{4} \left[\frac{\theta_k'}{\theta_k}(u)\right]^2.$$

Note now that θ_1 is odd while the other θs are even; so $\theta_1(0) = \theta_1''(0) = 0 = \theta_2'(0) = \theta_3'(0) = \theta_4'(0)$. Evaluating at 0 leads to

$$\text{(left-hand side as } u \to 0\text{)} \;= \sum_{k=2}^{4} \frac{\theta_k''}{\theta_k}(u).$$

A calculation using the leading terms $\theta_1(u) = \theta_1'(0)u + (u^3/6)\theta_1'''(0) + O(u^5)$ shows that the left-hand side gives $\theta_1'''(0)/\theta_1'(0)$ as $u \to 0$, so

$$\frac{\theta_1'''(0)}{\theta_1'(0)} = \frac{\theta_2''(0)}{\theta_2(0)} + \frac{\theta_3''(0)}{\theta_3(0)} + \frac{\theta_4''(0)}{\theta_4(0)}.$$

The final ingredient needed to exploit the relation is the *heat equation* identity

$$\frac{\partial \theta_k}{\partial \tau} = \frac{1}{4\pi i} \cdot \frac{\partial^2 \theta_k}{\partial u^2},$$

easily derived from the series expansion, and substituting it in $(* * **)$ shows that the log derivative with respect to τ of $\theta_2(0)\theta_3(0)\theta_4(0)/\theta_1'(0)$ vanishes:

$$\frac{\partial}{\partial \tau} \log\left(\frac{\theta_2(0)\, \theta_3(0)\, \theta_4(0)}{\theta_1'(0)}\right) = 0,$$

so the ratio $(\theta_2\, \theta_3\, \theta_4/\theta_1')(0)$ is a constant, which is found by letting $q \to 0$: $\theta_3, \theta_4 \to 1$ while $\theta_2 \to 2q^{1/4}\cos(\pi u)$ and $\theta_1 \to 2q^{1/4}\sin(\pi u)$, so the ratio is $1/\pi$.

This concludes the proof of the theorem.

Lecture 14

Algebraic methods in the study of compact Riemann surfaces

The fundamental result of the theory, conjectured by Riemann circa 1850, and proved over the next five decades, is:

Theorem: Every compact Riemann surface is algebraic.

We have an idea what this means, because we have considered Riemann surfaces defined by polynomial equations

$$P(z, w) = w^n + a_{n-1}(z)w^{n-1} + \cdots + a_1(z)w + a_0(z) = 0,$$

and we have seen how to compactify these; and indeed, the result does imply that every compact Riemann surface arises in such manner. But we would like now to do more that just explain the meaning of the theorem, and survey the basic algebraic tools available for the study of compact Riemann surfaces.

The truly hard part of the theorem is to get started. Nothing in the definition of an abstract Riemann surface implies in any obvious way the existence of the basic algebraic objects of study, the meromorphic functions.

Theorem: Every compact Riemann surface carries a non-constant meromorphic function.

Equivalently, every compact Riemann surface can be made into a branched cover of $\mathbb{P}^1$.

Remarks: This is the difficult part of the theorem; once we have a branched cover of $\mathbb{P}^1$, we can start studying it by algebraic methods. The proof involves serious analysis, specifically finding solutions of the Laplace equation in various surface domains, with prescribed singularities. ('Green's functions'.)

Contained in Riemann's theorem, there is a second result which we shall use without proof.

Proposition: Let $\pi : R \to \mathbb{P}^1$ be a holomorphic map of degree $n > 0$. There exists, then, an additional meromorphic function f on R which *separates the sheets of R over $\mathbb{P}^1$*, in the following sense: there exists a point $z_0 \in \mathbb{P}^1$ such that f takes n distinct values at the points of R over z_0.

Exercise: Show that such an f must then take n distinct values over all but finitely many points of $\mathbb{P}^1$. (Consider a limit point of a sequence z_k over which k takes fewer values and use the fact that the zeroes of a non-constant analytic functions are isolated. The case when the limit point is a branch point will require extra care.)

Assuming now that the Riemann surface R is connected, let $\mathbb{C}(R)$ be its field of meromorphic functions. A non-constant meromorphic function z defines an inclusion of fields

$$\mathbb{C}(z) \subset \mathbb{C}(R).$$

In algebra, this is commonly called a *field extension* rather that 'field inclusion'. The *degree* of the field extension, denoted $[\mathbb{C}(R) : \mathbb{C}(z)]$, is the dimension of $\mathbb{C}(R)$, as a vector space over $\mathbb{C}(z)$. Let $\pi : R \to \mathbb{P}^1$ denote the holomorphic map associated to the meromorphic function z.

Theorem:

(i) $[\mathbb{C}(R) : \mathbb{C}(z)] = \deg \pi \ (= n)$.

(ii) Any $f \in \mathbb{C}(R)$ satisfies a polynomial equation of degree $\leq n$ with coefficients in $\mathbb{C}(z)$,

$$f^n + a_{n-1}(z)f^{n-1} + \cdots + a_0(z) \equiv 0.$$

(iii) Let f be a meromorphic function on R which separates the sheets of R over $\mathbb{P}^1$. Then , $\mathbb{C}(R)$ is generated by f over $\mathbb{C}(z)$:

$$\mathbb{C}(R) = \mathbb{C}(z)[f].$$

(iv) Let now $f^n + a_{n-1}(z)f^{n-1} + \cdots + a_0(z) \equiv 0$ be the equation satisfied by the f in (iii). Then R is isomorphic to the non-singular, compactified Riemann surface of the equation

$$w^n + a_{n-1}(z)w^{n-1} + \cdots + a_1(z)w + a_0(z) = 0.$$

Additional remarks:

- It can be shown that quite easily that any two non-constant functions $f, g \in \mathbb{C}(R)$ are *algebraically related* over $\mathbb{C}$, that is, there is an equation, for some N,

$$\sum_{p,q=0}^{N} a_{pq} f^p g^q \equiv 0;$$

  See Problem Sheet 2. But statement (ii) is much more precise.

- By continuity, the function f in (iii) will take n distinct values over the points in $\pi^{-1}(z)$ for all z near z_0. An algebraic argument shows that the number of points $z \in \mathbb{P}^1$ over which f takes fewer than n values is finite. (A polynomial $(*)$ will either have multiple roots for every value of z, or else it will only have multiple roots for finitely many values of z.)

- $\mathbb{C}(R) \cong \mathbb{C}(z)[w]/(w^n + a_{n-1}(z)w^{n-1} + \cdots + a_0(z))$, in algebraic terms, by sending w to f. The point is that f cannot satisfy an equation of degree $< n$, because at $z = z_0$, the polynomial $(*)$ must have n roots!

We shall sketch the proof of some of these statements next time. The proof of (iii) is assigned Problem 3.7. Meanwhile, let us pursue the theory.

Theorem: There is a bijection between isomorphism classes of field extensions of $\mathbb{C}(z)$ on one hand, and isomorphism classes of compact Riemann surfaces, together with a degree n map to $\mathbb{P}^1$.

Forgetting the map to $\mathbb{P}^1$, we have:

Theorem: There is a bijection between isomorphism classes of fields which can be realized as finite extensions of $\mathbb{C}(z)$, on one hand, and isomorphism classes of compact Riemann surfaces, on the other.

The theorem follows essentially from part (iv) of the previous result; the only missing ingredient, which rounds up the correspondence between Riemann surfaces and their fields of functions, is:

Theorem: Homomorphisms from $\mathbb{C}(S)$ to $\mathbb{C}(R)$ are in bijection with holomorphic maps from R to S.

In the easy direction of the correspondence, one assigns to each Riemann surface its field of meromorphic functions. In the other direction, starting with a finite extension of $\mathbb{C}(z)$, a theorem from algebra (of the *primitive element*) asserts that the field extension is generated by a single element w. This w must satisfy an equation of degree n, with coefficients in $\mathbb{C}(z)$ (where n is the degree of the field extension); and the desired Riemann surface is obtained from the concrete Riemann surface defined by that very same equation in $\mathbb{C}^2$ by removing the singular points and compactifying.

Remark on Riemann surfaces in $\mathbb{C}^2$: The Riemann surface of

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

with the $a_i(z) \in \mathbb{C}(z)$, is more general than the ones we met before, in two respects. Firstly, the poles of the a_i lead to 'branches running off to ∞', as in Fig. 14.1.a. Secondly, our assumptions do not guarantee that the surface is non-singular everywhere. There can indeed by self-crossings such as for $w^2 = z^2 - z^3$, illustrated in Fig. 14.1.b, or worse singularities.

Both of these problems are handled roughly in the way we compactified the nicer kind of surfaces: we remove any problem points first; it can then be shown that all non-compact 'ends' of the resulting surface are analytically isomorphic to a punctured disc; we compactify the surface by filling in the disc. For example, removing the origin in the second picture leads to *two* punctured discs, each of which is compactified by adding a point.

Finally, the correspondence between homomorphisms of fields and maps between Riemann surfaces leads to an attractive geometric interpretation of the basic definitions of Galois theory.

Recall that a finite field extension $k \subset K$ is called *Galois*, with group Γ, if Γ acts by automorphisms of K and k is precisely the set of elements fixed by Γ.

Proposition: The automorphisms of a Riemann surface R are in bijection with those of its field of meromorphic functions $\mathbb{C}(R)$.

Let now $\pi : R \to S$ be holomorphic; it gives a field extension $\mathbb{C}(S) \subset \mathbb{C}(R)$.

Proposition: The automorphisms of R that commute with π are precisely the automorphisms of $\mathbb{C}(R)$ which fix $\mathbb{C}(S)$.

Corollary: A map $\pi : R \to S$ defines a Galois extension on the fields of meromorphic functions iff there exists a group Γ of automorphisms of R, commuting with π, and acting *simply transitively* on the fibres $\pi^{-1}(s)$, for a general $s \in S$.

Proof: Note first that any automorphism of R, commuting with π, which fixes a point of valency 1 must be the identity. Indeed, by continuity, it will fix an open neighbourhood of the point in question, and unique continuation property of analytic maps shows it to be the identity. Now, if $\mathbb{C}(R)$ is Galois over $\mathbb{C}(S)$, the order of the group of automorphisms is $[\mathbb{C}(R) : \mathbb{C}(S)]$. So the automorphism group must act simply transitively on the fibres which do not contain branch points. Conversely, an automorphism group acting simply transitively on even one fibre with no branch points must have order $\deg \pi$. But since that is $[\mathbb{C}(R) : \mathbb{C}(S)]$ it follows that the extension is Galois.

Remark: Such a map is called a 'Galois cover with group Γ'.

Remark: Note that $R/\Gamma = S$, set theoretically. Topology tells us that the Γ-invariant continuous functions on R are precisely the continuous functions on S. We have just shown the same for the meromorphic functions.

Examples of Galois covers:

(i)

$$\begin{aligned} \mathbb{P}^1 &\longrightarrow \mathbb{P}^1 \\ w &\longmapsto z = w^3 \end{aligned}$$

The automorphisms are $z \mapsto \zeta z$, where ζ is any cube root of 1.

(ii)

$$\begin{aligned} \mathbb{C}/L &\longrightarrow \mathbb{P}^1 \\ u &\longmapsto \wp(u) \end{aligned}$$

The automorphism is $u \mapsto (-u)$.

Rewriting it, the surface $w^2 = 4z^3 - g_2 z - g_3$ is a Galois cover of the z-plane, with Galois group $\mathbb{Z}/2$ and automorphism $w \mapsto -w$.

*Some Proofs

We now prove the algebraic theorem from the last lecture. In addition, we discuss holomorphic differentials and give an example of their use in the study of Riemann surfaces.

We show first, in connection with statement (i):

$$[\mathbb{C}(R) : \mathbb{C}(z)] \leq \deg \pi \ (= n).$$

(This is the easy inequality.) We must show, given $f_1, \ldots, f_{n+1} \in \mathbb{C}(R)$ that we can find $a_1(z), \ldots, a_{n+1}(z) \in \mathbb{C}(R)$ with

$$f_i a_1(z) + \cdots + f_{n+1} a_{n+1}(z) \equiv 0$$

on R.

Let $U \subset \mathbb{P}^1$ be such that $\pi^{-1}(U)$ is a disjoint union $\coprod_{j=1}^n U_j$ of open sets isomorphic to U. (Every point which is not a branch point has such a neighbourhood.) Let f_{ji} be $f_i|_{U_j}$. The linear system, with coefficients f_{ij} in the meromorphic functions on U,

$$\sum_{j=1}^{n+1} f_{ij} a_j = 0, \qquad 1 \leq i \leq n$$

has non-zero solutions, a_j, meromorphic on U. Moreover, we can produce a *canonical* solution, by row-reducing the matrix f_{ij}, setting the first free variable a_j to 1, the subsequent ones (if any) to zero and solving for the leading variables. The solution is canonical in that it does not depnd on the ordering of the sets U_j. (It would of course depend on the ordering of the f_i, but that we can fix once for all.)

If V is now another open set with $\pi^{-1}(V) = \coprod_{j=1}^n V_j$ then applying the same procedure will give a solution $b_1(z), \ldots, b_{n+1}(z)$ which agrees with $a_1(z), \ldots, a_{n+1}(z)$ on the overlap $U \cap V$.

All in all we have produced a solution to $(*)$ which is meromorphic on $\mathbb{P}^1$, away from the branch points. We could refine the argument to account for those, but there is an easier way. Note that the $a_i(z)$ are expressible algebraically (rationally, in fact) in terms of the $f_i(z)$. Now the f_i have at most polynomial singularities $(u - u_0)^{-k}$ at all points $u_0 \in R$, including the branch points. The local form of the holomorphic map says that $(z - z_0)$ is related polynomially to a local coordinate $(u - u_0)$ on R (where $\pi(u_0) = z_0$). So the $a_i(z)$ have at most polynomial growth $(z - z_0)^{-k}$ at the branch points. But then they have at most pole singularities and they must be meromorphic on all of $\mathbb{P}^1$.

The opposite inequality $[\mathbb{C}(R) : \mathbb{C}(z)] \geq n$ follows from the existence of a function separating the sheets of R over $\mathbb{P}^1$. Such a function cannot satisfy a polynomial equation

$$\sum_{k=0}^{d} a_k(z) f^k \equiv 0$$

over R of degree *less* that d, because near the point z_0 in question, the equation must have n distinct solutions. However, we now prove:

(ii) Every $f \in \mathbb{C}(R)$ satisfies an equation $\sum_{k=0}^{d} a_k(z) f^k \equiv 0$ with $a_k(z) \in \mathbb{C}(R)$ not all zero and $d = [\mathbb{C}(R) : \mathbb{C}(z)]$.

This is clear because $1, f, \ldots, f^d$ will be linearly dependent over $\mathbb{C}(z)$.

Part (iii) is Problem 3.7.

Sketch of proof of part (iv): We cannot give complete details because we have not described the process of removing singular points on a Riemann surface in $\mathbb{C}^2$ in complete detail (but this is the only difficulty, the sketch below is accurate aside from that).

We define a map ϕ from R to the Riemann surface R' of

$$P(z, w) = \sum a_k(z) w^k$$

by sending $P \in R$ to $(z, w) = (z(P), f(P))$. This only defines a map away from finitely many 'problem points'; but, once the compactified Riemann surface of $P(z, w)$ has been shown to exist, we can use the fact that a continuous map between Riemann surfaces which is holomorphic away from finitely many points is in fact holomorphic everywhere.

Let us indicate why the map ϕ is a bijection (and thus an analytic isomorphism). Because f generates $\mathbb{C}(R)$ over $\mathbb{C}(z)$, the pull-back of functions from R' to R, $\phi^* : \mathbb{C}(R') \to \mathbb{C}(R)$, is surjective. But R' is an n-sheeted covering of the z-plane, so $\dim \mathbb{C}(R')$ over $\mathbb{C}(z)$ is n by the argument in part (i). But then ϕ^* is an isomorphism. (This implies in particular that $\mathbb{C}(R')$ is a field, so that R' is connected.) Now ϕ is non-constant, so it is surjective, as the target is connected. Finally, f takes n distinct valaues over some $z_0 \in \mathbb{C}$; R' has never more than n sheets over any $z \in \mathbb{C}$ (because its equation has degree n); so there is a point in R' with no more that one inverse image in R. That means that $\deg \phi = 1$ and ϕ is an isomorphism.

***Holomorphic and meromorphic differentials**

This is the final algebraic tool I would like to introduce in the course, with an application to hyperelliptic Riemann surfaces. We have already used holomorphic differentials when recovering the lattice of a Riemann surface $\mathbb{C}/L$, as the 'periods' of the integral dz/w. We'll now see how to determine holomorphic differentials over more general Riemann surfaces.

Definition: A differential 1-form on a Riemann surface is called *holomorphic* if, in any local analytic coordinate, it has an expression $\phi(z)\,dz = \phi(z)(dx + i\,dy)$, with ϕ holomorphic.

For those of you unfamiliar with the notion of differential forms on a suface, there is a hands-on (but dirty) definition:

Definition: A *holomorphic differential* on a Riemann surface R is a quantity which takes the form $\phi(z)\,dz$ in a local coordinate z, and on the overlap region with another coordinate u, where it has the form $\psi(u)\,du$, it satisfies the gluing law

$$\phi(z) = \psi(u(z))\,u'(z).$$

(Formally, $du = u'(z)\,dz$.)

Proposition: If f is a holomorphic function on R, then df represents a holomorphic differential. In a local coordinate z, df is expressed as

$$df = f'(z)\,dz.$$

The transition formula $(*)$ for differentials is then a consequence of the chain rule.

Remark: We are trying to talk about derivatives of functions on a Riemann surface. However, the derivative of a function does not transform like a function under a change of coordinates, because of the chain rule $(df/dz) = (df/du)(du/dz)$. Differentials are quantities which transform like derivatives of functions.

Proposition: If ϕ is a holomorphic differential and f is a holomorphic function, then $f \cdot \phi$ is a holomorphic differential.

If ϕ and ψ are two holomorphic differentials, then ϕ/ψ is a *meromorphic* function. If is holomorphic iff the zeroes of ψ are 'dominated' by the zeroes of ϕ, that is, in local coordinate z when $\phi = \phi(z)\,dz$ and $\psi = \psi(z)\,dz$, the order of the zeroes of ψ is $\leq$ the order of the zeroes of ϕ.

Remark: There is an obvious notion of a meromorphic differential and there are analogous properties to the above.

Examples of computation of holomorphic differentials:

(i) Holomorphic differentials on $\mathbb{P}^1$ are zero.

Indeed, over the usual chart $\mathbb{C}$, the differential must take the form $f(z)\,dz$ with f holomorphic. Near ∞, $w = 1/z$ is a coordinate, and the differential becomes $f(1/w)\,d(1/w) = -f(1/w)\,dw/w^2$. So we need $f(1/w)/w^2$ to be holomorphic at $w = 0$, so f should extend holomorphically at ∞ and have a double zero there. But then f must be zero.

(ii) Holomorphic differentials on the Riemann surface $w^4 + z^4 = 1$.

The branch points of the projection to the z-plane are at $z = \pm 1, \pm i$; $w = 0$ at all of them. The map has degree 4 and branching index 3 at each of the points. At ∞, we have four separate sheets defined by $w = \sqrt[4]{1-z^4}$ which has four convergent expansions in $1/z$, as soon as $|z| > 1$. So Riemann–Hurwitz gives

$$g(R) - 1 = -4 + \tfrac{1}{2} \cdot 12 = 2, \qquad\qquad g(R) = 3.$$

R is a genus 3 surface with 4 points at ∞.

Now dz defines a meromorphic differential on R^{cpt}, because z is a meromorphic function there. At ∞, on R^{cpt}, $u = z^{-1}$ is a local homomorphic coordinate, and $dz = -u^{-2}du$ has a double pole.

On the other hand, I claim that dz has a triple zero at each of the branch points. Indeed, by the theorem on the local form of an analytic map, there is a local coordinate v with $z - 1 = v^4$. So $dz = d(v^4) = 4v^3dv$ has a triple zero over $z = 1$, and similarly over the other branch points.

So dz/w^2, dz/w^3 are still holomorphic at the branch points (and everywhere else when $z \neq \infty$, because $w \neq 0$). At $z = \infty$, w has a simple pole on R^{cpt} and we see that $w^{-2}dz$ and $w^{-3}dz$ (and higher powers) are non-singular there. Moreover, we can even afford to add $z\,dz/w^3$ to our list, and we have produced three holomorphic differentials on R^{cpt}.

Remark: It is easy to see that the three are linearly independent. It takes more work to show that any holomorphic differential is a linear combination of these three.

At any rate, we observe the following:

Proposition: The ratios of holomorphic differentials on R^{cpt} generate the field of meromorphic functions.

Proof: $(dz/w^2)\,/\,(dz/w^3) = w$, $(z\,dz/w^3)\,/\,(dz/w^3) = z$ and z and w generate the field of meromorphic functions, by our theorem from last time.

Application to hyperelliptic Riemann surfaces

Definition: A compact Riemann surface is called *hyperelliptic* if it carries a meromorphic function of degree 2. Equivalently, it can be presented as a double (branched) cover of $\mathbb{P}^1$.

Proposition: Any hyperelliptic Riemann surface is siomorphic to the compactification of the Riemann surface of

$$w^2 = f(z)$$

when f is a polynomial over $\mathbb{C}$ with simple roots only.

Proof: The degree 2 map $\pi : R \to \mathbb{P}^1$ realises $\mathbb{C}(R)$ as a degree 2 field extension of $\mathbb{C}(z)$. Let $u \in \mathbb{C}(R) \setminus \mathbb{C}(z)$; then u generates $\mathbb{C}(R)$ and satisfies a degree 2 equation

$$u^2 + a(z)u + b(z) = 0 \qquad\qquad a(z), b(z) \in \mathbb{C}(z).$$

Completing the square leads to $(u + a(z)/2)^2 + b(z) - a^2(z)/4 = 0$ or $v^2 + c(z) = 0$. Multiplying out by the square of the denominator of $c(z)$ gives $w^2 = f(z)$ with $f(z)$ a polynomial. Any repreated factors of $f(z)$ can be divided out and incorporated in w, leading to a square-free $f(z)$.

Theorem: Not every Riemann surface of genus 3 is hyperelliptic.

Proof: On Problem Sheet 3 you prove that the ratios of holomorphic differentials generate the proper subfield $\mathbb{C}(z)$ of $\mathbb{C}(R)$. But we saw that in the case of $w^4 + z^4 = 1$, we get the entire field of functions.

Remark: It can be shown that any Riemann surface of genus 2 is hyperelliptic. In higher genus, the hyperelliptic surfaces are quite 'rare', that is, non-generic. That shows that the problem of existence of meromorphic functions with prescribed poles does not have as neat a solutions in higher genus, as it does in genera 0 and 1: some surfaces carry a degree two meromorphic function, while some do not.

Lecture 15

Analytic methods and covering surfaces

Question: We investigated elliptic Riemann surfaces (tori) and elliptic functions successfully by presenting the surface as $\mathbb{C}/L$. Could we study other surfaces by viewing them as quotients of $\mathbb{C}$, and lifting the meromorphic functions to $\mathbb{C}$?

Answer: We can do this for very few surfaces, but a closely related question has a much more satisfactory better asnwer. The reason we cannot build many other examples from $\mathbb{C}$ is that the automorphism group of $\mathbb{C}$, $\{z \mapsto az + b \mid a, b \in \mathbb{C}\}$ is too small to lead to any interesting quotients. In fact:

Proposition: Any automorphism of $\mathbb{C}$ acting *freely* is a translation.

Proof: Clear from the complete description of automorphisms of $\mathbb{C}$ (Prob. 11b, Sheet 1).

Corollary: The only quotient Riemann surfaces of $\mathbb{C}$ under a group which acts *freely* are: $\mathbb{C}$, $\mathbb{C}^*$ and genus 1 Riemann surfaces. The groups are: $\{0\}$, $\mathbb{Z}\omega$ and $L = (\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$, $(\bar{\omega}_1\omega_2 \notin \mathbb{R})$, acting by translations.

Proof: Problem 8, Sheet 3.

Remark:

(i) Freedom of the action is related to the condition that the map from $\mathbb{C}$ to the quotient has valency 1 everywhere; it is quite useful, and we shall insist on it.

(ii) We would not gain much by dropping the condition that the group acts freely; the only quotient to add to the list would be $\mathbb{P}^1$. (Hint: $\mathbb{P}^1 \cong (\mathbb{C}/L)/\{\pm 1\}$.) We can obtain $\mathbb{C}$ and $\mathbb{P}^1$ in many other ways, e.g. $\mathbb{C} = \mathbb{C}^*/\{\pm 1\}$, but that is not so interesting.

In modified form, this idea of quotients is still a winner, as the following (major) theorem shows.

Uniformization Theorem: Every connected Riemann surface R is isomorphic to one of the following:

(i) $\mathbb{P}^1$,

(ii) $\mathbb{C}$, $\mathbb{C}^*$ or $\mathbb{C}/L$,

(iii) Δ/Γ, where Δ is the open unit disc and $\Gamma \subset PSU(1,1)$ is a *discrete* group of automorphisms acting freely.

Notes:

(i) The action of the group in (iii) is not just free, but *properly discontinuous*; see the discussion of covering spaces below.

(ii) In case (i), $R = \mathbb{P}^1/\{1\}$; in case (ii), $R = \mathbb{C}/\Gamma$, with $\Gamma = \{0\}, \Gamma \cong \mathbb{Z}$ or $\Gamma \cong \mathbb{Z}^2$. Thus, the Uniformization Theorem breaks up into two pieces:

Theorem (hard): Every simply connected Riemann surface is isomorphic to to $\mathbb{P}^1$, $\mathbb{C}$ or Δ.

Theorem (fairly easy): Every connected Riemann surface R is a quotient of a simply connected one by a group Γ of automorphisms acting freely.

Note: The simply connected surface in question is called the *universal covering surface* of R. We also say that it *uniformizes* R.

The hard part of the theorem involves a great deal of analysis. As before, the problem is to construct a global holomorphic function (meromorphic, in case of $\mathbb{P}^1$) with special properties. The easy part turns out to be purely topological; it is related to the notion of covering spaces and fundamental group, which we now review.

Definition: A surface $\tilde{R}$ endowed with a map $p : \tilde{R} \to R$ is a *covering surface* of a surface R if every point $r \in R$ has a neighbourhood U for which $p^{-1}(U)$ is a disjoint union of open sets homeomorphic to U via p.

Examples: $S^1 \to S^1$, $z \mapsto z^n$ is a covering map; so is $\mathbb{R} \to S^1$, $x \mapsto e^{2\pi i x}$. If you prefer surfaces, use the maps $\mathbb{C}^* \to \mathbb{C}^*$ and $\exp : \mathbb{C} \to \mathbb{C}^*$. Note, however, that the inclusion of an open subset in a surface is *not* a covering map (check out the points on the boundary of the subset).

Remarks:

(i) The definition makes sense for a map of topological spaces $p : \tilde{X} \to X$, and $\tilde{X}$ is then called a *covering space* of X.

(ii) The definition models the properties of a quotient under a free and 'nice' group action, such as $\mathbb{C}/L$.

Proposition: If p is *proper*, then p is a covering map iff it is a local homeomorphism, i.e. every point $\tilde{r} \in \tilde{R}$ has a neighbourhood $\tilde{U}$ such that $p : \tilde{U} \to p(\tilde{U})$ s a homeomorphism.

Remark:

(i) Recall that p is *proper* iff p^{-1} of a compact set is compact.

(ii) In the absence of properness, the covering condition is stronger than the local homeomorphism condition.

Proof: We proved '$\Leftarrow$' in lecture 4 (when discussing the degree of a map), while '$\Rightarrow$', without any properness condition, is obvious.

Here are three basic properties of covering surfaces.

Theorem (Lifting properties): Let $p : \tilde{R} \to R$ be a covering surface.

(i) (Path lifting property.) Pick $r \in R$, $\tilde{r}$ mapping to r by p and a continuous path $\omega : [0,1] \to R$ with $\omega_0 = r$. Then there is a unique continuous *lifting* $\tilde{\omega} : [0,1] \to \tilde{R}$ with $p \circ \tilde{\omega} = \omega$, and $\tilde{\omega}_0 = \tilde{r}$.

(ii) (Lifting of simply connected spaces.) Let $f : X \to R$ be a continuous map from a path-connected, simply connected space to R, with $f(x_0) = r$ for some $x_0 \in X$. Then there is a unique *lifting* $\tilde{f} : X \to \tilde{R}$ satisfying $\tilde{f}(x_0) = \tilde{r}$. (A lifting is a map $\tilde{f}$ satisfying $p \circ \tilde{f} = f$).

(iii) (Non-existence of interesting covers of simply connected surfaces.) Assume that $\tilde{R}$ is connected and R is simply connected. Then $p : \tilde{R} \to R$ is a homeomorphism ('$\tilde{R} = R$').

Proof:

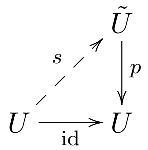
(i) Cover the path ω by finitely many good neighbourhoods U (the kind whose liftings to $\tilde{R}$ are disjoint unions of copies of U). In such a neighbourhood, it is clear how to lift the path; so concatenating these little liftings as you travel along ω produces $\tilde{\omega}$.

(ii) For any $x \in X$, choose a path γ from x_0 to x, and let $\omega = f \circ \gamma : [0,1] \to R$. ω lifts to $\tilde{\omega}$ on $\widetilde{R}$, by (i). Define

$$\tilde{f}(x) = \tilde{\omega}(1).$$

Clearly, $p \circ \tilde{f}(x) = \omega(1) = f(x)$; what is not clear is that the answer is independent of γ. By simple connectivity of X, any path γ' can be continuously deformed to γ. It is easy to see, from the construction of the lifting, that $\tilde{\omega}'$ gets deformed continuously to $\tilde{\omega}$ in the process. However, $\tilde{\omega}'(1)$ is restricted to the *discrete set* $p^{-1}(r)$, so in fact it must be constant throughout; so $\tilde{f}(x)$ is independent of the path, and the map $\tilde{f}$ is indeed well-defined.

(iii) We construct an inverse s to p be means of part (ii) of the theorem, applied to the diagram



Now $p \circ s = \text{id}$, so s is injective; further, given $\tilde{u} \in \tilde{U}$ and a path $\tilde{\omega}$ from $\tilde{u}$ to $\tilde{v}$, setting $\omega = p\tilde{\omega}$, $s\omega$ is a lift of ω to $\tilde{U}$ starting at $\tilde{u}$, so it must agree with $\tilde{\omega}$. Then $\tilde{v} = s(\omega(1)) \in s(U)$, so s is surjective as well.

Definition: Let a group Γ act on a topological space by homeomorphisms. The action is *properly discontinuous* if every $x \in X$ has some neighbourhood U whose translates $\gamma \cdot U$, as γ ranges over Γ, are mutually disjoint.

The connection with covering spaces is contained in the following statement, which is immediate from the definition.

Proposition: If Γ acts properly discontinuously on X, then the quotient map $X \to X/\Gamma$ is a covering space.

We are interested in surfaces only, but, as you shall see from the proof, the main theorem below, on the existence of the universal cover applies to a large class of topological spaces (they need to be connected, locally path-connected and locally simply connected).

Main theorem: Let X be a connected surface. Then $X \cong \tilde{X}/\Gamma$, for some simply connected $\tilde{X}$ and a group Γ acting properly discontinuously. Moreover, $\tilde{X}$ and Γ are unique up to isomorphism.

***Remark:** Γ is isomorphic to the so-called *fundamental group* of X. Connoisseurs will notice that we have hidden an ambiguity in this definition; we have not chosen a base point on X, so the identification of Γ with the fundamental group is only defined up to conjugacy. Choosing $x_0 \in X$ and $\tilde{x}_0 \in \tilde{X}$, we get a stronger uniqueness property. Namely, for any other $\tilde{X}', \tilde{x}'_0, \Gamma'$, there is a *unique* isomorphism

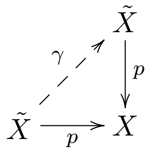
$$\begin{matrix} f : \tilde{X} \to \tilde{X}' \\ a : \Gamma \to \Gamma' \end{matrix} \text{ with } \begin{matrix} f(\tilde{x}_0) = \tilde{x}'_0 \\ f(\gamma\tilde{x}) = a(\gamma)f(\tilde{x}). \end{matrix}$$

The following Lemma, which we need for the proor of the main theorem, gives a meaning to the group Γ.

Lemma: Let $p : \widetilde{X} \to X$ be a connected and simply connected covering surface of X, and let Γ be the group of self-homeomorphisms of $\tilde{X}$ which commute with p; then, Γ acts properly discontinuously, and $X \cong \tilde{X}/\Gamma$.

Remark: Such automorphisms of $\widetilde{X}$ are called *deck transformations* of the covering $\widetilde{X} \to X$.

Proof of the lemma: For any $x \in X$ and any $a, b \in p^{-1}(x)$, the lifting theorem (ii) ensures the existence of a unique $\gamma : \widetilde{X} \to \widetilde{X}$ taking a to b and a γ' taking b to a:



Then, $\gamma'\gamma$ fixes a and $\gamma\gamma'$ fixes b, so these must be the identity maps on $\widetilde{X}$, by uniqueness of the lifting. So γ is a *homeomorphism* commuting with p, and we have just established that Γ acts *simply transitively* on $p^{-1}(x)$. As x was arbitrary, this implies that $X \cong \tilde{X}/\Gamma$. Choosing a small connected neighbourhood U of x for which $p^{-1}(U)$ is a disjoint union of copies of U, it follows that Γ acts simply transitively on these copies, hence it acs properly discontinuously on $\widetilde{X}$.

***Proof of the main theorem (sketch):** Uniqueness of $\tilde{X}$ follows from the lifting theorem (ii): for two such $\widetilde{X}$ and $\widetilde{X'}$, let p and p' be the quotient maps to X. There exists, then, liftings $\widetilde{p}$ of p to $\widetilde{X'}$ and $\widetilde{p'}$ of p' to $\widetilde{X}$. Moreover, if $\widetilde{x_0}$ and $\widetilde{x_0'}$ are liftings of x_0 in $\widetilde{X}$ and $\widetilde{X'}$, we can arrange that they map to each other under these liftings. Then, I claim that $\widetilde{p'} \circ \widetilde{p}$ is the identity on $\widetilde{X}$. Indeed, it is a lifting of the identity map on X, but then it must agree with the 'other' lifting, which is the identity on $\widetilde{X}$. So $\widetilde{p}$ is a homeomorphism. Now, as in the lemma, Γ is the group of deck transformations of the cover, so we must also have a compatible isomorphism $\Gamma \cong \Gamma'$.

To prove existence, it suffices, by our lemma, to find one simply connected covering space of X. We fix an $x_0 \in X$ and define $\tilde{X}$ to be the space of pairs (ω, x), with ω being a path in X from x_0 to x, modulo declaring two such pairs (ω, x), (ω', x') to be equivalent iff ω can be continuously deformed to ω', with the endpoints kept fixed. The covering map $p : \widetilde{X} \to X$ sends (ω, x) to x. Any simply connected neighbourhood U of x in X can be identified bijectively with a subset ωU in $\widetilde{X}$, containing (ω, x), as follows: for any $y \in U$, we choose a path λ from x to y in U and send y to $(\omega\lambda, y)$, where $\omega\lambda$ means the path ω, followed by λ. Simple connectedness of U shows that this is well-defined. We declare ωU to be a neighbourhood of (ω, x) in $\widetilde{X}$; it is easy to see that, in this definition, p becomes a covering map; morever, for any simply connected $U \subset X$, $p^{-1}(U)$ will be a disjoint union of copies of itself.

Any loop γ in X which lifts to a *loop* (rather than a path) in $\tilde{X}$ is *contractible* in X. (This is because a point in $\tilde{X}$ 'carries' as part of the information a path ω from x to it; following that around γ returns us to the concatenation of ω with γ. If this represents the original point in $\tilde{X}$, it must be that γ is contractible in X.) From here, using the lifting theorems, one concludes that every loop in $\tilde{X}$ is contractible.

Remark: There is a similar explicit description of Γ, as the set of continuous deformation classes of loops on X based at x_0. Define a 'multiplication' on Γ by concatenating the loops, and an 'action' of Γ on $\tilde{X}$ by sending a loop $\gamma \in \Gamma$ and an element $(\omega, x) \in \widetilde{X}$ to $(\gamma \cdot \omega, x)$. Clearly, these operations are well-defined, that is, they pass to continuous deformation classes. The identity element in Γ is the *constant* loop at x_0. The inverse of a loop γ can be represented by the same loop, travelled in the opposite direction: indeed, the constant loop can be continuously deformed to the circuit which travels first along γ, then backtracks along the same route, by progessively increasing the amount of travel along γ (for each parameter $s \in [0,1]$, we travel the corresponding fraction of the loop γ and return).

Application: The Little Picard Theorem

Little Picard Theorem: A non-constant entire holomorphic function on $\mathbb{C}$ misses at most a single value; that is, $f(z) = a$ can be solved for all but perhaps a single complex number a.

Remark: Clearly this is the best possible result, because the exponential map misses 0.

Proof: Assume f misses two values a and b. By translation and rescaling we can assume that $a = 0$ and $b = 1$. So we have a holomorphic map $f : \mathbb{C} \to \mathbb{C} \setminus \{0,1\}$. Because $\mathbb{C}$ is simply connected, f lifts to a holomorphic map $\tilde{f} : \mathbb{C} \to S$, where S is the universal covering surface of $\mathbb{C} \setminus \{0,1\}$. From the Uniformization Theorem we see that the only possibility for S is the unit disc. But then $\tilde{f}$ is a bounded holomorphic function on $\mathbb{C}$, so it is constant by Liouville's theorem.

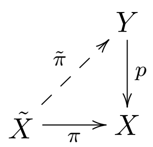
Remark: In the following lecture we shall prove directly that S is the unit disc, by constructing a covering map $\Delta \to \mathbb{C} \setminus \{0,1\}$.

*Appendix: Fundamental groups and Galois groups

1. Classification of covering spaces

We have seen that every 'reasonable' (connected, locally path-connected and locally simply connected) X is a quotient $\tilde{X}/\Gamma$, where $\tilde{X}$ is simply connected and Γ is (isomorphic to) the fundamental group of X.

For every connected covering space $Y \xrightarrow{p} X$, the lifting theorem (ii) ensures that the map $\pi : \tilde{X} \to X$ lifts to Y:



It follows that $\tilde{\pi} : \tilde{X} \to Y$ is also a covering map; indeed, if V is a neighbourhood of $y \in Y$ isomorphic to $U := p(V)$, then $\tilde{\pi}^{-1}(V)$ is a union of components of $\pi^{-1}(U) = \tilde{\pi}^{-1}(p^{-1}(U))$ so is a disjoint union of isomorphic copies of $V \cong U$. Recall now that Γ is the group of *deck transformations* of $\tilde{X}$ over X, that is, the group of homeomorphisms of $\tilde{X}$ which preserve π.

Now as $\tilde{X}$ is a covering of Y and is simply connected, it is also the universal cover of Y, and then $Y = \tilde{X}/H$ for the group H of automorphisms of $\tilde{X}$ commuting with $\tilde{\pi}$. But then, H certainly commutes with π, so it is contained in Γ. This proves the following:

Theorem: There is a bijection between connected covering spaces Y of X, up to isomorphism, and subgroups H of Γ, the fundamental group of X. H is the fundamental group of Y and Y is $\tilde{X}/H$.

Clarification: To be precise, we must choose a base point $x \in X$ and a point $y \in Y$ above x; an isomorphism between (Y, y, π) and (Y', y', π') is then a homeomorphism $f : Y \to Y'$ taking y to y' and such that $\pi' \circ f = \pi$.

If we ignore base points, the isomorphism classes of Y correspond to subgroups of Γ *up to conjugacy.*

Definition: A covering space $Y \xrightarrow{p} X$ is *normal*, or *Galois*, if the corresponding subgroup H is normal in Γ.

Proposition: If $Y \xrightarrow{p} X$ is normal, then Γ/H acts freely on Y, commuting with p; and $Y/(\Gamma/H) = X$.

2. Analogy with Galois groups.

There is a formal analogy between the theory of covering spaces and the Galois theory of algebraic extensions. Let K be a (perfect) field (e.g. in characteristic 0), $\bar{K}$ its algebraic closure, with Galois group Γ. This means that Γ acts on $\bar{K}$ by automorphisms, and $K = (\bar{K})^\Gamma$, the subfield of Γ-invariants. Galois theory tells us that any algebraic field extension $K \subseteq L$ can be placed inside $\bar{K}$, and $\bar{K}$ is also Galois extension of L with Galois group H, the subgroup of Γ which leaves every element of L invariant. If it happens that H isa normal subgroup of Γ, then L is a Galois extension of K, with Galois group Γ/H: indeed, Γ/H acts naturally on L, and the fixed point field is K. In this respect, $\bar{K}$ is analogous to the universal cover of K.

For Riemann surfaces, this can be made more precise, via the correspondence between compact Riemann surfaces and their fields of meromorphic functions: the field of functions on R/Γ is the field of Γ-invariants in $\mathbb{C}(R)$. The analogy is imperfect, because a covering $p : R \to S$ between compact Riemann surfaces is necessarily finite, so algebraic covering surfaces R of S lead only to *finite index* subgroups of the fundamental group of S. On the other hand, covers coming from algebraic field extensions are usually branched over some points, and this is not allowed of topological covering maps. So the theories overlap like this:

the shaded areas representing normal coverings, on the topological side, and Galois extensions, on the algebraic side. The notions agree on the overlap. The intersection comprises, on the topological side, the covering maps of finite degree; on the algebraic side, it contains the so-called *unramified field extensions.*

*Appendix: The universal covering space of $\mathbb{C} \setminus \{0,1\}$, and its relation to genus 1 Riemann surfaces

We said last time that the universal cover of $\mathbb{C} \setminus \{0,1\}$ was Δ, so $\mathbb{C} \setminus \{0,1\} = \Delta/\Gamma$, with $\Gamma \subset PSU(1,1)$ a discrete subgroup acting freely. This followed from the Uniformization Theorem — the alternative cases $\mathbb{C}/\Gamma$ and $\mathbb{P}/\Gamma$ being easily ruled out — but we'd like to realize the covering of $\mathbb{C} \setminus \{0,1\}$ by Δ concretely.

We'll actually use the upper-half plane $\mathfrak{H}$ instead of Δ (the identification can be made via $z \mapsto (z-i)/(1-iz)$, which takes 0 to $-i$, ∞ to i and fixes 1 and -1). The reason is, we have an obvious discrete group acting on $\mathfrak{H}$, namely $PSL(2,\mathbb{Z}) \subset PSL(2,\mathbb{R})$, acting via Möbius transformations.

Proposition: The orbits of $PSL(2,\mathbb{Z})$ on $\mathfrak{H}$ are in bijection with the set of lattices in $\mathbb{C}$, modulo scaling, and also with the set of isomorphism classes of genus 1 Riemann surfaces $\mathbb{C}/L$.

See the end of lecture 9 for a refresher (and some info we did not discuss in class). In one direction, a point $\tau \in \mathfrak{H}$ leads to the lattice $\mathbb{Z} + \mathbb{Z}\tau \subset \mathbb{C}$; going back, a lattice $L \cong 2(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$ leads to $\tau = \omega_2/\omega_1$, if $\omega_2/\omega_1 \in \mathfrak{H}$ (or else we use ω_1/ω_2). However, τ depends on a choice of basis and a different basis, related to the first by an $SL(2,\mathbb{Z})$ transformation, will lead to a Möbius transform of τ.

Proposition: A fundamental domain for the action of $PSL(2,\mathbb{Z})$ on $\mathfrak{H}$ is the set $\{z \in \mathfrak{H} | |\operatorname{Re} z| \leq \frac{1}{2}, |z| \geq 1\}$. The only identifications are $\tau \to \tau+1$, between the two vertical edges, and $\tau \to -1/\tau$, between the two arcs on the unit circle.

Remark: It can also be shown that $\tau \to \tau + 1$ and $\tau \to -1/\tau$ (rather, the matrices $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$) generate $SL(2,\mathbb{Z})$.

Remark: The shaded region shows another fundamental domain.

Remark: To show that every orbit of $SL(2,\mathbb{Z})$ meets the domain above, one starts with any τ, brings it into the strip $|\operatorname{Re}\tau| \leq \frac{1}{2}$ by subtracting an integer, and applies $\tau \to -1/\tau$ if $|\tau| < 1$. If the result is outside the strip, one repeats the procedure. A geometric argument shows that the procedure terminates ($\operatorname{Im}\tau$ gets increased each time).

Proposition: $\mathfrak{H}/SL(2,\mathbb{Z}) \cong \mathbb{C}$, and the bijection is implemented by the (holomorphic) *elliptic modular function* J,

$$J(\tau) = \frac{{g_2}^3}{{g_2}^3 - 27{g_3}^2},$$

where $4z^3 - g_2 z - g_3 = w^2$ is the equation of the Riemann surface $\mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$.

Remarks:

(i) Recall that the g_i are expressed in terms of *Eisenstein series*

$$\sum_{m,n} (m + n\tau)^{-k};$$

so there is no 'obvious' expression for the map J.

(ii) Scaling the lattice scales g_2 by the fourth and g_3 by the sixth power of the scale, so the only scale-invariant quantity is ${g_2}^3/{g_3}^2$. The latter, however, has a pole, because one of the branch points could be 0. The quantity that is not allowed to vanish is the discriminant ${g_2}^3 - 27{g_3}^2$; so $J(\tau)$ is the simplest $SL(2,\mathbb{Z})$-invariant holomorphic function on $\mathfrak{H}$.

(iii) J is a bijection because the Riemann surface is determined up to isomorphism by the branch points e_1, e_2, e_3 up to scale, hence by the combination ${g_2}^3/{g_3}^2$; and every lattice does arise from such a Riemann surface.

Now J cannot be a covering map, because $\mathbb{C}$ is simply connected and does not have any interesting covering surfaces. So $PSL(2,\mathbb{Z})$ cannot act freely, and indeed we can spot two points:

- i with stabilizer $\mathbb{Z}/2$, generated by $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$,
- ω with stabilizer $\mathbb{Z}/3$, generated by $\left[\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. ($\omega = e^{\pi i/3}$.)

A bit of work shows these are the only problematic orbits: the action is free everywhere else.

We have a chance of getting a free action by considering subgroups of $PSL(2,\mathbb{Z})$ which miss the stabilizers.

Proposition: $\Gamma(2) \subset PSL(2,\mathbb{Z})$, represented by matrices $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ with a and d odd and b and c even, is a subgroup of index 6 and acts freely on $\mathfrak{H}$.

Proof: Problem 10, Sheet 3.

Proposition: A fundamental domain for the action of $\Gamma(2)$ on $\mathfrak{H}$ is as depicted below; it includes 6 fundamental domains for $PSL(2,\mathbb{Z})$, labelled 1–6.

Proposition: $\mathfrak{H}/\Gamma(2) \cong \mathbb{C} \setminus \{0,1\}$ as a Riemann surface. In fact, the isomorphism is realized by $\tau \mapsto \lambda(\tau)$,

$$\lambda(\tau) = \frac{e_1 - e_2}{e_3 - e_2},$$

where e_1, e_2, e_3 are the values of $\wp$ at the half-periods $1/2, \tau/2, (\tau+1)/2$.

Idea of proof: One shows the invariance of λ under $\Gamma(2)$ by noting that the half-lattice points are preserved mod L, by the action of $\Gamma(2)$: $A\mu \equiv \mu \pmod{L}$ if $\mu \in L/2$.

Then one checks that we have

$$\lambda : \mathfrak{H} \longrightarrow \mathbb{C} \setminus \{0,1\} \xrightarrow{\ \phi\ } \mathbb{C}, \qquad \underbrace{\mathfrak{H} \longrightarrow \mathbb{C}}_{J}$$

so $J = \phi \circ \lambda$, where

$$\phi(\lambda) = \frac{4(\lambda^2 - \lambda + 1)^3}{27\lambda^2(\lambda - 1)^2}.$$

Finally, ϕ has degree 6, while $J : \mathfrak{H}/\Gamma(2) \to \mathbb{C}$ is generically 6–to–1. If λ had valency > 1 at a point, or if two points had the same λ-value, then $\phi \circ \lambda$ would be more than 6–to–1 in a neighbourhood of the point in question — contradiction.

Remark: The 6 points with the same J-value correspond to the 6 possible orderings of the half-lattice points mod L (or of the branch points e_1, e_2, e_3 of $\wp$).

Remark: Identifying via the map $z \mapsto (z - i)/(1 - iz)$ realises the fundamental domain of $\Gamma(2)$ as the infinite hyperbolic square depicted below:

This is a very special illustration of a theorem of Poincaré, which describes the hyperbolic polygons that are fundamental domains of free discrete group actions on Δ. Qualifying polygons which are completely contained in Δ lead to compact Riemann surfaces; those with some vertices on $\partial\Delta$ lead to compact Riemann surfaces with points removed, while those with edges on $\partial\Delta$ lead to Riemann surfaces with discs removed. For instance, using slightly smaller circles gives the fundamental domain for the action of a group $G \subset PSU(1,1)$ with $\Delta/G \cong$ (the disc minus two small discs).

The missing points at 0, 1, ∞ have been replaced with true holes.

Lecture 16

The Riemann surface of an analytic function

We shall see how the notion of a Riemann surface provides a solution to a classical question of function theory. This is closely related to the point of view adopted in lecture 1, where we resolved ambiguities of multi-valued functions by changing their domains. Historically, this was the main motivation for introducing the notion of an abstract Riemann surface.

Recall that the zeroes of a holomorphic function which is not identically zero are isolated. This implies the following.

Proposition: Let f and g be holomorphic functions in a connected region $U \subset \mathbb{C}$. If $f = g$ in the neighbourhood of some point a, then $f = g$ on U.

Say now that f is defined in a small disc around a. If there exists an extension g of f to U, then g must be unique. It is called an *analytic continuation* of f to U.

Question: Assuming it exists, how could we recover g on all of U from f?

There is a procedure, in principle (although rarely used in practice), based on Taylor expansions. It relies on the following reslut from complex function theory.

Proposition: Let f be holomorphic in a disc centered at a. Then the Taylor expansion of f about a converges (uniformly on compact subsets) inside the disc.

This shows that the Taylor expansion extends f to the largest disc that does not contain a singularity.

For example, the Talor series for $f(z) = 1/z$ near $z = a$ is

$$\sum_{n=0}^{\infty} (-1)^n \frac{(z-a)^n}{a^n}$$

and converges for $|z - a| < |a|$. At $z - a = -a$ $(z = 0)$, we encounter a singularity.

However, we can now take the expansion of f at a point close to $2a$. This will have radius of convergence nearly $2a$, and extend this domain even further. Repeating this process will recover the function everywhere on the plane, except of course at $z = 0$, where it is singular. We'll need infinitely many steps to get all of f, but any given point in the plane can be reached after finitely many steps.

This process can by systematized, and leads to the *analytic continuation along paths.* Say f is a holomorphic function $f : U \to \mathbb{C}$, only known near a in the (connected) domain U. To recover the function in the neighbourhood of $b \in U$, choose a path γ from a to b in U, and keep expanding f into Taylor series at points along the path. As γ is compact, it is a non-zero distance away from the boundary of U, so the radii of convergence are bounded below by that distance; this ensures that, eventually, you reach b.

So, given f in a neighbourhood of a, and the domain U, we know how to recover it on U.

Question: What if we do *not* know U: what is the largest set on which f extends (can be continued) analytically?

It is natural to conjecture that such a largest domain exists; in other words:

Conjecture (∗): Every analytic function f, defined near some $a \in \mathbb{C}$, can be continued analytically to a uniquely defined maximal domain $U \subset \mathbb{C}$.

For example, for $f(z) = 1/z$, we recover its maximal domain $\mathbb{C}^*$.

Unfortunately, the conjecture is *false*, for a familiar reason (see lecture 1). For instance, start with $f(z) = \sqrt{z}$ near 1 (with $f(1) = 1$). Its Taylor expansion has radius of convergence 1. We can continue it along any path avoiding the origin, but the resulting function is path-dependent:

The continuation of $\sqrt{z}$ to -1 along the upper and lower half-cirles differ by a sign. For more complicated algebraic functions, there will not be a simple relation between the different continuations.

The reason a natural conjecture such as $(*)$ can be false is that we asked the question slightly incorrectly. It turns out that a maximal domain does exist, and is *unique up to isomorphism*, but not in the class of open subsets of $\mathbb{C}$, but rather in a class of Riemann surfaces.

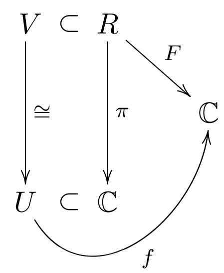
Here is the correct question. Consider quadruples (R, r, π, F), with the following properties:

(i) R is a connected Riemann surface,

(ii) $r \in R$,

(iii) $\pi : R \to \mathbb{C}$ is a holomorphic map, with $\pi(r) = a$ and with valency 1 everywhere (it is locally bi-holomorphic)

(iv) $F : R \to \mathbb{C}$ is a holomorphic function whose restriction to a small neighbourhood V of r agrees with $f \circ \pi$.

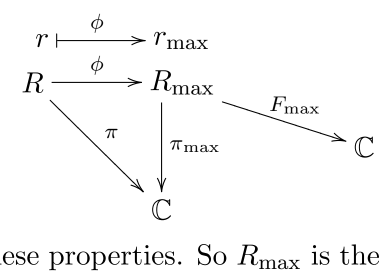


Thus, F *is an analytic continuation of* f, where f is transported to R via the identificaton of V with U.

A *morphism* between two such quadruples will be a holomorphic map preserving the entire structure. Note, by projecting to $\mathbb{C}$, and by condition (iii), that a morphism is necessarily locally bi-holomorphic. Further, a morphism between two objects is unique, if it exists, because it is determined in a neighbourhood of r.

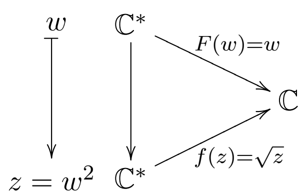
Theorem: The category of such quadruples has a *strict final object* $(R_{\max}, r_{\max}, \pi_{\max}, F_{\max})$. This is called the *Riemann surface of the analytic function* f. It is uniquely determined, as soon as f is known in a neighbourhood of a.

Strict final object means that any other quadruple (R, r, π, F) maps to it holomorphically, while preserving the entire structure,

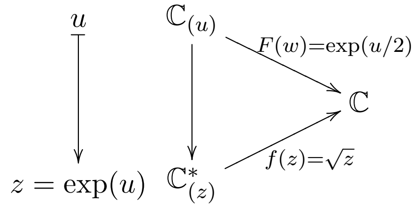


and the map ϕ is unique with these properties. So $R_{\max}$ is the largest domain of f, in the sense that every other possible domain maps to it.

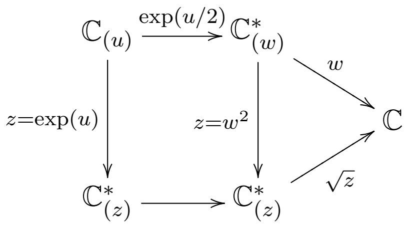
For example, $F(z) = \sqrt{z}$; $R_{\max} = \mathbb{C}^*$, with coordinate w, and $\pi : w \to z = w^2$, while $F(w) = w$.



Note that there is a 'bigger' cover of $\mathbb{C}^*_{(z)}$ we could use in place of w, namely



but we can map $\mathbb{C}_{(u)}$ to $\mathbb{C}^*_{(w)}$ by $w = \exp(u/2)$



so going to this much bigger cover is unnecessary. The Riemann surface of our function is 'maximal' relative to the direction of the arrows, but other surfaces carrying a continuation of f can have 'more points'. It is the 'most economical' surface that will carry the function.

As a consequence of the existence of R, we shall prove

The Monodromy Theorem: Let U be a simply connected domain, f a holomorphic function defined near $a \in U$. Assume that f can be continued analytically along any path in U, without encountering singularities. Then, this continuation defines a single-valued function on U.

Remark: This is false without the simply connectivity, e.g. take $U = \mathbb{C}^*$ and $f(z) = \sqrt{z}$.

The proof relies on a re-definition of analytic continuation along a path, and on a topological lemma:

Proposition: f can be analytically continued along a path γ, starting at a, iff γ can be lifted to the Riemann surface of f, starting at r.

Lemma: Let U be simply connected, $\tilde{U}$ connected and $\pi : \tilde{U} \to U$ a local homeomorphism with the *path lifting property*: every path γ, starting at some $u \in U$, lifts to a path on $\tilde{U}$ starting at any prescribed lift $\tilde{u}$ of u. Then, $\tilde{U}$ is homeomorphic to U.

Proof of the Monodromy Theorem: Let (R, r, π, F) be the Riemann surface of f, $\pi(r) = a$. Let $\tilde{U}$ be the component of $\pi^{-1}(U)$ containing r. We just asserted that $\pi : \tilde{U} \to U$ has the path lifting property; the Lemma ensures that $\tilde{U} = U$, and then $F/\tilde{U}$ is a single-valued analytic continuation of f.

Proof of the proposition: The 'if' part is easy; the lifting $\widetilde{\gamma}$ of γ can be covered by a finite sequence of small discs, mapping isomorphically to a discs centered on γ in $\mathbb{C}$. In each disc, the analytic continuation F admits convergent Taylor expansions, and these match on the successive overlaps. This provides an analytic contiuation of f along γ in $\mathbb{C}$. For the converse, we construct a quadruple (R, r, π, F) verifying (i)–(iv), together with a lifting of γ to R, from an analytic continuation of f along γ; the universal property of $R_{\max}$ provides our lifting $\widetilde{\gamma}$. The desired R arises by gluing together the discs of successive power series expansion along any overlaps *in which the values of the extended function agree.* We may have to avoid overlaps stemming from self-crossings of γ, as depicted, if the extended function takes different values in the different discs:

Proof of the lemma: We have really seen this in part (iii) of the theorem on lifting properties. An inverse of π is constructed as follows: for a point $u \in U$, choose a continuous path γ from a to u, lift it to $\widetilde{\gamma}$ on $\widetilde{U}$, and assign to u the endpoint $\widetilde{u}$ of $\widetilde{\gamma}$. This is independent of the choice of path, because any path can be continuously deformed into any other, and the set of possible $\widetilde{u}$ is discrete (from the local isomorphism theorem).

Remark: The topologically savvy reader will see how to modify the argument to show that, in general, $\tilde{U}$ is a *covering space* of U. (Assume that U and $\widetilde{U}$ are reasonable, e.g. Hausdorff and locally path connected).

Proof of the existence of the Riemann surface of an analytic function. We define (R, r, π, F) as follows. Let R be the disjoint union of the Riemann surfaces R_α in all our quadruples $(R_\alpha, r_\alpha, \pi_\alpha, F_\alpha)$, modulo the following equivalence relation: two points $x_\alpha \in R_\alpha$ and $x_\beta \in R_\beta$ are equivalent iff they map to the same point in $\mathbb{C}$ — that is, $\pi_\alpha(x_\alpha) = \pi_\beta(x_\beta)$ — *and* if F_α and F_β agree in neighbourhoods of x_α, x_β. (In light of the first condition, we can identify small neighbourhoods of the points via their projections to $\mathbb{C}$.) For instance, all the r_α will represent the same point $r \in R$.

The set R inherits two maps π, F to $\mathbb{C}$, from all the π_α and F_α, because these are compatible with our identifications; and the quadruple (R, r, π, F) accepts maps from all quadruples $(R_\alpha, r_\alpha, \pi_\alpha, F_\alpha)$. Any $x \in R$ comes from an $x_\alpha \in R_\alpha$, for some α . A small subset U_α of R_α is identified, via π_α, with a disc in $\mathbb{C}$. We declare it to be a neighbourhood of x in R, thus turning the latter into a topological space. Note that π and F are continuous in this topology. Moreover, no two points within U_α can be equivalent (they have different images under π), so π is a local homeomorphism from R to $\mathbb{C}$. I claim that R is Hausdorff.

First, two points $x, y \in R$ with distinct π-images in $\mathbb{C}$ must have disjoint neighbourhoods, by continuity of π. So let $x_{\alpha,\beta} \in R_{\alpha,\beta}$ and choose $U_{\alpha,\beta}$ identified with the same disc U in $\mathbb{C}$ under $\pi_{\alpha,\beta}$. If two points $u_{\alpha,\beta} \in U_{\alpha,\beta}$ are equivalent, then F_α and F_β agree in some open set in U. But then, they agree over all of U, and then x_α and x_β are equivalent, and represent the same point in R.

Thus, R is a topological surface. We use the same discs U_α, identified with their projections in $\mathbb{C}$, to define the Riemann surface structure; π is then locally bi-holomorphic, and F is holomorphic. Finally, every point of R can be connected by a path to r, because this holds in every R_α, so R is connected. This completes the proof.

Historical Remark: The construction of the Riemann surface of an analytic function, in full generality, is essentially due to Weierstraß. Of course, the concept of a Riemann surface did not exist at the time; Weierstraß was conceiving the Riemann surface as the totality of the analytic continuations of f along all possible paths. That idea gives an alternative, more concrete construction of $R_{\max}$, similar to the construction of the universal covering surface; but the construction is less improtant conceptually than the defining property of $R_{\max}$, and the fact of its existence. In practice, analytic continuation is rarely if ever performed by successive Taylor expansions, so that is not a practical way to construct the Riemann surface.