Riemann Surfaces

Dr C. Teleman[1]

Lent Term 2001

[1] LaTeXed by James Lingard — please send all comments and corrections to `james@lingard.com`

Lecture 1

What are Riemann surfaces and where do they come from?

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 $x \geq 0$, and then we conventionally select the positive square root.

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

In the complex story, we can take '$w = \sqrt{z}$' to mean $w^2 = z$; but then to get a single-valued function of z we must make a choice, and to make a *continuous* choice we need to make 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* ('holomorphic') away from the *negative* real axis.

A different choice could be to let $\sqrt{z}$ be the square root with positive imaginary part. This is uniquely defined, away from the positive real axis, and again determines a complex-analytic function away from the *positive* real axis.

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

In either case, there is no way to extend the function continuously to the missing half-line: when z approaches a point on the half-line in question from opposite sides, the values chosen for $\sqrt{z}$ approach values which differ by a sign.

A restatement of this familiar problem is: starting with 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 $\sqrt{z_0}$ on return. (z_0 needs to travel around the origin twice before $\sqrt{z_0}$ travels once.)

It is thus clear that there is a genuine problem at 0; but the problem along the real axis was our own doing — there is no discontinuity in the function until we insist on choosing a single value. We could avoid this problem by allowing multi-valued functions; but another point of view has proved more profitable.

The (simple) 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 and is thus well-defined and unambiguous!

In the case $w = \sqrt{z}$, the graph of the function is

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

a closed subset in $\mathbb{C}^2$.

In this case, it is easy to see that the function 'w',

$$\begin{aligned} S &\to \mathbb{C} \\ (z, w) &\mapsto w \end{aligned}$$

defines a homeomorphism (diffeomorphism if you know the term) of the graph S with the w-plane. But that is rather 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:

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

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

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 a 'Riemann surface' is;
- how to do complex analysis on them (what are analytic functions on them?).

The two questions are closely related; in a sense, a Riemann surface is a (real, 2-dimensional) surface with a notion of complex analytic functions on it. But for now, we just note a moral definition of a (concrete) Riemann surface.

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

Remark: '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 $\mathbb{C}$.

Remark: We are really abusing the term 'multi-valued' function, and including things such as $\sqrt{z}$, whereas the literature uses a more restrictive definition. But a proper definition of a Riemann surface will wait for next lecture and we shall not really be using multi-valued functions.

In the case of $S = \{(z, w) \in \mathbb{C}^2 \mid z = w^2\}$, we can identify S with the complex w-plane over projection. It is then clear what a holomorphic function on S should be — it would be an analytic function of w (regarded as a function on S). We won't be so lucky in general, as Riemann surfaces will not be identifiable with their w- or z-projections; however, the most interesting case of *non-singular* Riemann surfaces has the following property:

Moral definition: A *non-singular* Riemann surface S in $\mathbb{C}^2$ is a Riemann surface where 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

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

Before moving on to start the outline of the course and some important highlights, let me give an example of a Riemann surface with an interesting 'shape', which cannot be identified with either the z-plane or the w-plane.

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

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

A real snapshot of the graph is

where the dotted lines indicate that the values are imaginary. There are thus *two* values for w for every value of z, other than $z = \pm 1$ and $z = \pm k$.

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 $\sqrt{2}\sqrt{1+k}\sqrt{1-k}$ leads to a choice of 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, $\neq 0)$ 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, would lead 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 this, 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 for a moment what happens on the intervals, the graph of w breaks up into two pieces:

each of which can be identified, via projection, with the z-plane minus two intervals.

Now over the said intervals, the function also has two values, except at the endpoints where those are merged. 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 interval $(1, k)$. Thus, if we start on the upper sheet and travel that route, we find ourselves exiting on the lower sheet. Thus

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

A moment's thought will convince us that we cannot do all these identifications in $\mathbb{R}^3$, with the sheets positioned as above, without introducing spurious crossings. To rescue something, we flip the bottom sheet about the real axis. Now the corresponding edges of the cuts are aligned and we can perform the gluing by stretching each of the surfaces around the cut to pull out a tube, and get the following picture,

which represents two planes (ignore the boundaries) joined by two tubes.

Lecture 2

For a second look at the Riemann surface drawn above, recall that the function

$$z \mapsto \frac{R^2}{z}$$

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

This raises a first question for the course:

- What kind of shapes can Riemann surfaces have?
- How can I tell the topological shape (other than in an ad hoc fashion, as we did above)?

The answer to the first question is that any orientable surface can be endowed with the structure of a Riemann surface. (An answer to the second, for compact surfaces, will be the Riemann-Hurwitz theorem.)

Recall that a surface is *orientable* if there is a continuous choice of clockwise rotations on it A typical example of a non-orientable surface is the Möbius strip. Every *compact* surface you can draw in $\mathbb{R}^3$ is orientable; but there are non-orientable compact surfaces (e.g. the Klein bottle).

Remark: The reason a Riemann surface is orientable is that we require the ability to identify little pieces of the surface with the unit disc in $\mathbb{C}$; and the disc carries a natural orientation, in which multiplication by i is counter-clockwise rotation.

So, can we get compact Riemann surfaces — such as the torus without punctures, or the genus g surface

Well, almost. We'll soon be prove the following proposition:

Proposition: Every Riemann surface in $\mathbb{C}^2$ is non-compact.

(You can see that clealy for Riemann surfaces of algebraic equations, $P(z, w(z)) = 0$: they project surjectively to $\mathbb{C}$.)

So to get compact Riemann surfaces, we need the notion of an *abstract* Riemann surface.

Moral definition: An *abstract Riemann surface* is a topological surface, together with a local notion of holomorphic function. So we must be able to decide, for a function defined near a point, whether it is holomorphic or not.

Before the real definitions, some examples:

The Riemann sphere $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1$.

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.)

But the way we inderstand $\mathbb{P}^1$ as a Riemann surface is by regarding $z^{-1} = w$ as a *local coordinate* near ∞. That is, we say a function f defined in the neighbourhood of ∞ on $\mathbb{P}^1$ is holomorphic if the function defined in a neighbourhood of 0 by

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

is holomorphic there.

There is another way to describe $\mathbb{P}^1$ 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. (The map $z \mapsto w = z^{-1}$ from $\mathbb{C}^*$ to $\mathbb{C}^*$ is bianalytic / biholomorphic.)

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

Because the identification map is holomorphic and the composition of holomorphic maps is holomorphic, we see 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 3c), and is one reason why abstract Riemann surfaces are more fun to play with than concrete ones.

One other example of gluing: Let A be the annulus $1 < |z| < R+\epsilon$. Define an identification of the boundary strip $1 < |z| < 1 + \epsilon/R$ with the boundary strip $R < |z| < R + \epsilon$ via multiplication by R. Again this 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 $A \xrightarrow{\pi} T$. Define a function $f : U \to \mathbb{C}$, where $U \subseteq T$ is open, to be 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 sets in A; and in the latter case, the two sets are identified analytically by means of $z \mapsto Rz$. Then in the latter case, to check analyticity near t it suffices to check it on a single one of the inverse images of t. Thus we have shown:

Proposition: With these definitions, every point $t \in T$ has some neighbourhood O_t which is identified via π with a disc in $\mathbb{C}$; moreover the identification takes holomorphic functions to holomorphic functions, in both directions.

This is one definition of an abstract Riemann surface:

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

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

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

- 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 (so it lies in $\mathcal{O}(U)$).

- 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, such that h_s takes holomorphic functions to holomorphic functions in both directions.

A corollary of the last condition is that all *local* properties of holomorphic functions in $\mathbb{C}$ carry over to Riemann surfaces. For example,

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

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

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, $h \circ f : f^{-1}(U) \to \mathbb{C}$ is holomorphic (i.e. is in $\mathcal{O}(f^{-1}(U))$).

Proposition:

(i) An open subset $U \subseteq \mathbb{C}$ inherits the structure of an abstract Riemann surface, with the natural definition of holomorphic function.

(ii) A map $f : U \to \mathbb{C}$, where $U \subseteq \mathbb{C}$ is open, is holomorphic in the new sense iff it is a holomorphic function in the old sense.

Proof:

(i) Obvious, check that the holomorphic functions satisfy the condition in the definition of Riemann surfaces.

(ii) Let id denote the identity map $w \mapsto w$ from $\mathbb{C}$ to $\mathbb{C}$. If $f : U \to \mathbb{C}$ is holomorphic as a map, according to the definition above, $\mathrm{id} \circ f$ is a holomorphic function, thus f is holomorphic in the old sense. Conversely if f is a holomorphic function (old style) 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 holomorphic, since it is the composition of holomorphic (old style) functions.

As a quick application, let us prove a theorem.

Theorem: Every holomorphic map defined everywhere on a *compact* Riemann surface is locally constant.

Remark: Thus if the domain is connected, the function is constant. (Recall that a topological space is *connected* if it cannot be decomposed as a disjoint union of two open subsets.)

Proof: The map f is continuous, so $|f|$ achieves a maximum value M. Let z be a point on the surface with $|f(z)| = M$; then $|f|$ achieves a local maximum at z, so it is constant in a neighbourhood of z. Thus, the set of points where $|f(z)| = M$ is *open*. But by continuity of $|f|$, it is also *closed* in the surface. Thus f is constant on the connected component of the surface which contains z. Repeating this for all components shows that f is constant on each connected component.

Side remark: We have swept under the rug a topological fact about surfaces which is not altogether obvious; namely,

Proposition: A topological surface is the disjoint union of its connected components.

This is false for general topological spaces (e.g. $\mathbb{Q}$).

Proof: Define an equivalence relation on points in the surface by relating two points that can be joined by a continuous path. It is easy to check now that the equivalence classes are open and connected subsets of the surface.

Lecture 3

Concrete Riemann surfaces (in $\mathbb{C}^2$)

For most of the course, we shall regard Riemann surfaces from an abstract point of view. Historically, however, they arose as graphs of analytic functions over domains in $\mathbb{C}$, with multiple values; and we shall still need to study them from that angle and understand their properties (later in the course).

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 for Problem 4; in particular, P is differentiable in its region of convergence and we can differentiate term by term.

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 with $S \cap U =$ zero-set of F in U; and, moreover, $\partial^n F/\partial w^n(s) \neq 0$ for some n. (In particular, we see that S is *locally closed*, by continuity of F.)

The conclusion $\partial^n F/\partial w^n(s) \neq 0$ is imposed to rule out vertical lines through s. ($S \cap U$ would contain a vertical line iff $F_{0n} = 0$ for each n, as can be seen from the power series expansion.)

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

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 holomorphic functions above are inverse to each other.

Remark: In all cases, we can only assume the domain of the function to be a small neighbourhood of the components of s.

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

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

using 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. But this is then the Jacobian for a 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 constitutes 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 5.

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

Finally, as a cultural fact (which we shall not use in this course), let me mention the basic result about the *local structure of singular Riemann surfaces*. It says that all local information about a singularity is captured algebraically.

Theorem (Weierstrass Preparation Theorem in 2 dimensions): Let $F(z, w)$ be holomorphic near $(0, 0)$ and

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

Then

- (weak form) there exists a function 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 $z = 0$, such that the zero-set of F agrees with the zero-set of Φ, in a neighbourhood of 0.

- (strong form)

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

with $\Phi(z, w)$ as above and $u(z, w)$ holomorphic and non-zero near $(z, w) = (0, 0)$. Moreover, the factorization is *unique*.

Proof: See Problem 7 for the weak form; see Gunning and Ross, Several Complex Variables, for the strong form.

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

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

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 number n is the *order of the pole* at z_0.

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 $a \neq 0$.

Theorem: A meromorphic function on U is the same as a holomorphic map $U \to \mathbb{P}^1$, 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 move to a Riemann surface.

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.

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

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

Remark: There is also a notion of the 'order of a pole'. It is, however, a special case of the notion of 'valency of a map at a point', which will wait until next lecture.

Corollary: Meromorphic functions on $\mathbb{P}^1$ are the holomorphic maps from $\mathbb{P}^1$ to itself, not identically ∞.

We shall describe these more closely. 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 $R(z)$.

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

We shall prove two stronger statements.

Theorem (Unique Presentation Theorem 1): 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*.

Remark: Clearly the p_i are the poles of the function.

Proof: 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.

We now subtract from our meromorphic function f all the principal parts at the finite poles. We note first that there are finitely many poles:

Lemma: For a non-constant holomorphic map $f : R \to S$ between Riemann surfaces, the inverse images $f^{-1}(s)$ of any point $s \in S$ are *isolated* in R.

Proof: Let r be such that $f(r) = s$, U_s a neighbourhood of s in S analytically identifiable with the unit disc, and V_r a finite neighbourhood of r in R such that $f(V_r) \subseteq U_s$ (possible by continuity of f). Then f is a holomorphic function on the disc V_r (taking values in a disc but that does not matter) and we are reduced a familiar statement that the zeroes of a holomorphic function are isolated.

To continue, this shows that f has finitely many poles on $\mathbb{P}^1$, and we can subtract all the principal parts at the finite points. What is left is a meromorphic function with poles only at ∞. But this is knows to be a polynomial (Liouville's Theorem, for instance); we can also argue directly, using $w = z^{-1}$ as local coordinates near ∞. We can then subtract the principal part of the function, which is a polynomial in z, and are left with a meromorphic function with no poles, that is, a holomorphic function on $\mathbb{P}^1$. But that must be constant (Theorem in lecture 2).

Theorem (Unique Presentation Theorem 2): 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*.

Lecture 4

Meromorphic functions on Riemann surfaces
Meromorphic functions on $\mathbb{P}^1$ and the Unique Presentation Theorems

These are both discussed in the notes for lecture 3, so I only make two comments:

Firstly, 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:

Proposition: The meromorphic functions on a Riemann surface form a field (sometimes 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 properties. That is:

> we can add, subtract and multiply meromorphic functions, and we can divide by any function that is not identically zero; and the usual rules of arithmetic hold.

This is not as naïve as we might first think; recall from calculus that some arithmetic operations involving ∞ and 0 cannot be consistently defined; e.g. $\infty - \infty$, ∞/∞, $0/0$ and $\infty \cdot 0$ cannot be assigned meanings consistent with the usual arithmetic laws. Nonetheless, for meromorphic functions ϕ and ψ, the values of $(\phi + \psi)(z)$, $(\phi \cdot \psi)(z)$ and $(\phi/\psi)(z)$ can be defined for any z in the domain, even if $\phi(z)$ and $\psi(z)$ are both ∞, or both 0, or one of the forbidden combinations, at the same point $z = z_0$. The reason is the existence of convergent Laurent expansions near each point, and arithmetic on Laurent expansions with finite principal parts is easily seen to work, with the advertised properties.

Secondly, the Unique Presentation Theorems have analogues for arbitrary compact Riemann surfaces. Later in the course we shall discuss the torus, but a few words abour 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.

Cautionning comment: The 'order of a zero' or 'order of a pole' of a meromorphic function on a Riemann surface is defined without ambiguity (see 'valency of a map at a point', below). However, the notion of 'principal part at a pole' requires a choice of a local coordinate on the surface in question. Nonetheless, the first of the two statements above is unambiguous, because it compares the principal parts of two functions at the same point. (Just use the same local coordinate for both.)

By contrast, the existence problem — functions with specified principal parts, or with specified zeroes and poles — is more subtle, and there are obstructions to that 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.

In this context, notice that even on $\mathbb{P}^1$, there is a restriction on the number of zeroes and poles, namely:

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

Proof: Just note that the function at the end of lecture 3 has a pole of order $n-m$ at ∞ if $n>m$, and a zero of order $m-n$ there if $m>n$.

Local properties of holomorphic maps and their consequences

The main theorem is the following:

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

Then, given an analytic identification $\psi : V_s \to \Delta$ of a small neighbourhood of $s \in S$ with the unit disc Δ, 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:

$$\begin{tikzcd} U_r \arrow[r, "f"] \arrow[d, "\phi"'] & V_s \arrow[d, "\psi"] \\ \Delta \arrow[r] & \Delta \\ z \arrow[r, mapsto] & z^n \end{tikzcd}$$

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: Postponed.

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 a nice description as 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.

Consequences of the theorem on the local form:

- **Open mapping theorem:** If $f : R \to S$ is non-constant and R is connected then f is *open*; that is, the image of any open set in R is open 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.

- **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 $f(r) \in S$.
- **Maximum modulus theorem:** If $|f|$ has a local maximum then f is locally constant. (Clear from the open mapping theorem.)
- **'Good behaviour almost everywhere' theorem:**
  (i) If $f : R \to S$ is holomorphic and nowhere locally constant then the set of points $r \in R$ with $v_f(r) > 1$ has no accumulation point in R.
  (ii) With the same assumptions, the set of points with $f(r) = s$, for any fixed $s \in S$, has no accumulation point in R.

  Proof: In the neighbourhood of such a point, the theorem on the local form of an analytic map leads to a contradiction.

  Remark: 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.

Lecture 5

Global consequences of the theorem on the local form

We have seen that

- non-constant holomorphic maps are open;
- injective maps are local analytic isomorphisms;
- f is injective near $r \in R \iff v_f(r) = 1$.

Here is now a global consequence.

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.

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, and so $f(R)$ is open in S and compact. 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.

Remark: There is, of course, a much more elementary proof of the fundamental theorem of algebra using complex analysis (straight from the Cauchy integral formula). The point is that the theorem just proved is 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.

Theorem/Definition (Degree of a map): Let $f : R \to S$ be a non-constant holomorphic map between compact connected Riemann surfaces. Let $s \in S$. Then the number

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

is independent of the choice of point s and is called the *degree* of the map f.

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

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 end of previous lecture).

Proof of the theorem: We need the following lemma.

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 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 of them 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.

Remark: A map $f : X \to Y$ between locally compact Hausdorff spaces is *proper* if f^{-1}(any compact set) is compact, e.g. the inclusion of a closed set is proper, the inclusion of an open set is not proper. The proof above can be adapted to proper maps, without requiring compactness of X.

Proof of the theorem continued: 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 $r_i \in f^{-1}(V)$ to which the theorem on the local form of a holomorphic map applies. The result now follows from the obvious fact that the map $z \mapsto z^n$ has n solutions near zero.

Remark: The theorem does apply to any proper holomorphic map of Riemann surfaces.

As an example, let us prove:

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 $\deg p \geq \deg q$ (if not, 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 f will have a zero of the same order as p, if $q(z) \neq 0$. So the sum of root multiplicities is $\deg p = \deg f$.

Remark: If you try the argument with $f(z)$ of the form p/q with $\deg p < \deg q$, you get the wrong answer, unless you also count the root ∞, with valency $\deg q - \deg p$.

Corollary: For all but finitely many w, the equation

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

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

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

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

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

Assume that, for general z, $P(z, w)$ has no multiple roots. Then the surface S maps properly to $\mathbb{C}_{(z)}$, and the degree of the map is n.

Comment: The assumption that $P(z, w)$ has no multiple roots, for general z, can be shown to be 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

The theorem gives a formula relating

- the degree of a holomorphic map between compact connected Riemann surfaces R and S
- the topologies of R and S
- the valencies of the map.

As a preliminary, we need:

Theorem (Classification of compact orientable surfaces): Any compact orientable surface is homeomorphic to one of the following:

g is called the *genus* and counts the 'doughnut holes'. There is another description of these surfaces as 'spheres with handles':

and then g counts the number of handles.

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 $f(x) = s$. (Not really missing, just multiple.)

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

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

where $g(X)$ denotes the genus of X. (In particular, b must be even.)

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. We have seen that T can be compactified by the addition of 2 points at ∞. The valencies must then be 1 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.

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.

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

That is, $f(x) = \alpha$ has multiple roots for exactly $(d-1)$ values of y, counting y k times if $f(x) - \alpha$ has excess multiplicity k — that is, the number of distinct roots is $d - k$.

Example: This is clear in examples such as $f(x) = x^n$, or for any polynomial of degree 2. In general, you must use the fact that 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 $(d-1)$. So, for generic f, $f - \alpha$ will have a double root for precisely $(d-1)$ values of α.

But we need another $(d-1)$ to make the theorem work. This of course comes from the point at ∞.

Lemma: $v_f(\infty) = \deg f - 1$ for a polynomial f.

Proof: Let

$$f(z) = z^n + f_{n-1}z^{n-1} + \cdots + f_1 z + f_0.$$

Then if $w = 1/z$, 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 exactly n at 0.

The Riemann-Hurwitz formula — two examples

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

The surfaces we shall consider are all *algebraic*, in the sense that they are the solution sets of polynomial equations of the form $P(z,w) = 0$. We shall restrict ourselves to polynomials P of the special form

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

with the $p_k(z)$ polynomial functions of z. A general procedure could be described, which involves algebraic computations, but we shall limit ourselves to the case when P is simple enough, and the algebra is quite manageable.

To explain the restriction $(*)$, we note first:

Proposition: The projection π of the zero-set R of $(*)$ to the z-plane is a proper map; that is, the inverse image of a compact set is compact.

Proof: By continuity of π, the inverse image of any closed set is closed; so we 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 a 'variation of the argument' principle).

By our discussion in lecture 4, this implies that the degree of $\pi : R \to \mathbb{C}_{(z)}$ can be defined as the sum of the valencies of π over any point. Actually, we are jumping a bit too far — we must first check that R is an (abstract) Riemann surface and π is holomorphic. But in lecture 2, we saw that the following held:

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; and π is holomorphic for that structure.

(At the time, we had called R 'non-singular'.) Indeed, we saw that the projection to the z-axis, or the projection to the w-axis, or both, could be used to define a little holomorphic disc near each non-singular point and hence define a local notion of holomorphic function.

It follows in particular that the valency of π is 1 at each point where $\partial P/\partial w \neq 0$.

Remark: When P is as in $(*)$, 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.)

The proof takes us a bit out of the bounds of the course, but briefly, to see (i), if $\partial P/\partial w$ were to vanish everywhere on R, it would follow that all irreducible factors of $\partial P/\partial w$ are also factors of P; whence one can show that P itself factors only into squares or higher powers; whence $\partial P/\partial z$ would also have to vanish identically on R. (Here, we are using P as a polynomial in w, with coefficients in the unique factorization domain $\mathbb{C}[z]$ of polynomials in z.)

((ii) is dealt with in the next remark.)

A point $z_0 \in \mathbb{C}$ is a *branch point* for R if its inverse image contains points of valency > 1. (Equivalently, there are fewer that n points in the fibre; or again, $P(z_0, w)$ has multiple roots.

Remark: The condition $P(z_0, w)$ has multiple roots can be rewritten as: the discriminant of $P(z_0, w)$, viewed as a polynomial in w, vanishes. But this discriminant is a polynomial in the $p_k(z_0)$, hence a polynomial in z_0; hence *there are only finitely many branch points for R.*

To apply the Riemann-Hurwitz theorem, we must deal with compact surfaces. Now 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}.

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:

Proposition: Let $D^\times$ be the *outside* of some very large disc in $\mathbb{C}$, large enough to contain 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' $u \mapsto z = u^k$.

Remark: It's the wrong place to prove this result, which is really of topological nature. Let me merely restate it in slightly different form:

Proposition: Let $f : S \to \Delta^\times$ be a proper holomorphic map from a connected Riemann surface to the punctured unit disc. 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 $d = \deg(f)$.

From the proposition, it is now 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.

Example 1:

$$w^3 = z^3 - z$$

So

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

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.

Example 2:

$$w^3 - 3w - z^2 = 0$$

Remark: Clearly this is best handelled by projecting to the w-axis, but we shall be oblivious to 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 6

Proof of Riemann-Hurwitz

To prove the Riemann-Hurwitz formula (we shall really only sketch the proof), we need to introduce new notions — that of a triangulation of a surface, and that of Euler characteristic. They require first a mild digression into some topological technology.

Definition: A topological space 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.

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

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

Remark: Such metric spaces are called *separable*.

Another easy observation is:

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. Pretty much 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 requirement on the countable base into the definition of a surface. So, in many texts, every surface has a countable base.

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 described 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.

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).)

Definition: (see also the remark below) 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 pairs 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 compact, and bounded by exactly three edges.

It takes a bit of work that the conditions imply the following:

- every point on an edge, which is not an endpoint, has a neighbourhood homeomorphic to:

  the edge being the diameter of the disc.

- Every endpoint of an edge is a vertex and has a neighbourhood homeomorphic to:

  the (finitely many) spokes being edges.

- Every face is the homeomorphic image of 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.

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

(Note that loops are already disallowed because an edge is required to 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:

and the definition of the Euler characteristic below still holds. But a more useful generalization is the notion of a *polygonal decomposition*, where the faces are required only to be homeomorphic to arbitrary 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.)

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.

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

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

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

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

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

  which we must perturb to

  following which 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.

The Riemann-Hurwitz formula for χ

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 $z \mapsto z^n$:

$$\begin{tikzcd} U_r \arrow[r, "f"] \arrow[d, "\phi"'] & V_s \arrow[d, "\psi"] \\ \Delta \arrow[r, "z^n"] & \Delta \end{tikzcd}$$

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

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 (sketch): Start with a triangulation of S making sure that the branch points are included among the vertices. (Subdivide if necessary.) The inverse images in R of the edges on S form the edges of a triangulation of R. (Check the definition and the properties of the local form of the function f carefully.)

Now, if there were V vertices, E edges and F faces downstairs, there will be $\deg(f)F$ faces, $\deg(f)E$ edges but only $\deg(f)V - b$ vertices upstairs: the missing b vertices are the missing points over the branch points.

Corollary: Riemann-Hurwitz in terms of the genus (lecture 5) holds.

Proof: Use the above theorem and the equality $\chi(R) = 2 - 2g(R)$.

Lecture 7

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 fact is far from obvious — we are saying that any Riemann surface structure on the 2-sphere ends up being isomorphic to the 'standard' one; if you recall that Riemann surface structures can be defined by gluing, you see why this result 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 turn to 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 so that 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 linearly independent vectors. Calling them $2\omega_1$ and $2\omega_2$, the conditions are

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

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 obtained 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.

Remark: Division by $2\omega_1$ turns the period parallelogram into the form

with $\tau = \omega_2/\omega_1 \notin \mathbb{R}$. Another way to construct 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 $q = e^{2\pi i\tau}$. 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

$$f(z + 2\omega_1) = f(z + 2\omega_2) = f(z)$$

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

Proposition: Any doubly periodic holomorphic function on $\mathbb{C}$ is constant. (We assume $\tau \notin \mathbb{R}$.)

Proofs:

(i) Global holomorphic functions on $\mathbb{C}/L$ are constant.

(ii) Use Liouville's theorem, that bounded holomorphic functions on $\mathbb{C}$ are constant.

So to get anything interesting, we must allow poles.

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.

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

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 function f by a small complex number λ; the relations (i)–(iii) are unchanged.

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:

Theorem (Unique Presentation Theorem 1): 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).

Theorem (Unique Presentation Theorem 2): 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 this next time. Meanwhile, we have yet to construct a single elliptic function!

Proof of the theorem: 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

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

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

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

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

(iii) Consider now

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

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:

$$\begin{aligned}\int_0^{2\omega_1} z\frac{f'(z)}{f(z)}\,dz + \int_{2(\omega_1+\omega_2)}^{2\omega_2} z\frac{f'(z)}{f(z)}\,dz &= \int_0^{2\omega_1} z\frac{f'(z)}{f(z)}\,dz - \int_0^{2\omega_1} (z+2\omega_2)\frac{f'(z)}{f(z)}\,dz \\ &= -2\omega_2\int_0^{2\omega_1} \frac{f'(z)}{f(z)}\,dz \\ &= -2\omega_2(\log f(2\omega_1) - \log f(0))\end{aligned}$$

Now $f(2\omega_1) = f(0)$, but the reason the expression fails to be 0 is the multi-valuedness of log. Indeed, $\log f(2\omega_1) - \log f(0)$ 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 8

The Weierstrass functions

We assume a lattice $L = 2\mathbb{Z}\omega_1 + 2\mathbb{Z}\omega_2 \subset \mathbb{C}$ has been chosen, with $\omega_1, \omega_2 \neq 0$, $\omega_1/\omega_2 \notin \mathbb{R}$.

Because of the properties we established for elliptic functions, we know that the simplest possible one cannot have a single simple pole in the period paralellogram. So the simplest assignment of principal parts is a double pole at $z = 0$. This leads to the so-called $\wp$-function of Weierstrass.

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

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

which converges, uniformly on compact subsets $K \subset \mathbb{C}$, once the terms with poles are set aside, to an elliptic function.

Proof: For a compact $K \subset \mathbb{C}$, finitely many terms will have poles in K. If the others converge uniformly, as we claim, meromorphicity of the limit is a consequence of Morera's Theorem. Now the individual terms are easily estimated by

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

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

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

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 $k = 2$.) Now periodicity is a consequence of convergence essentially:

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

But the last term vanishes due to an obvious symmetry, namely

$$2m\omega_1 + 2n\omega_2 \longrightarrow -2(m+1)\omega_1 - 2n\omega_2$$

will cancel matching terms. (The argument relies of course on convergence of the series — otherwise the last two steps in the chain would be inadmissible.)

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

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

an elliptic function with a triple pole at 0.

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

Proof: Clear from the series expansion.

Proposition: $\wp'(\omega_1) = \wp'(\omega_2) = \wp'(\omega_1 + \omega_2) = 0$; these are all simple zeroes and there are no other zeroes, mod L.

Proof: $\wp'(\omega_1) = \wp'(\omega_1 - 2\omega_2) = \wp'(-\omega_1) = -\wp'(\omega_1)$ by periodicity and parity; same for the other half-lattice points. To see that there are no other zeroes, note the following important observation:

Proposition: $\wp$ defines a holomorphic map of degree 2 from $\mathbb{C}/L$ to $\mathbb{P}^1$; $\wp$ 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 $\omega_1, \omega_2, \omega_1 + \omega_2$.

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 $\wp(z) = a$ 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) Note that $\wp(z) = e_i$ would have too many roots, if two of the e_is agreed.

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:

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

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

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

indeed it is now our purpose to establish a connection between the two. From now on, we change the name of the argument of $\wp$ to u.

Theorem (Differential equation for the Weierstrass function):

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

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

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

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

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

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 $\mathbb{C}^2$.

Lecture 9

For the proof of the differential equation theorem we require a lemma:

Lemma (Laurant 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 amusing exercise, we notice that the odd powers of u cancel, and we obtain

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

and similarly for $\wp'(u)$ we get

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

Proof of the differential equation theorem: Using the lemma, we check directly that the first few terms in the Laurant expansion of $f(u) = (\wp')^2 - 4\wp^3 - g_2\wp - g_3$ at $u = 0$ are

$$0 \cdot u^{-6} + 0 \cdot u^{-4} + 0 \cdot u^{-2} + 0 + O(u^2),$$

so $f(u)$ is an elliptic function with no poles, vanishing at $u = 0$. Thus f is identically zero.

Proof of the geometric interpretation theorem: We have

$$\begin{tikzcd} \mathbb{C}/L \setminus \{0\} \arrow[r, "(\wp,\wp')"] \arrow[dr, "\wp"'] & R \arrow[d, "\pi"] \\ & \mathbb{C} \end{tikzcd}$$

and we know that

- π is proper and 2–to–1 except at the branch points e_1, e_2, e_3, roots of $4z^3 - g_2 z - g_3$.
- $\wp$ is proper and 2–to–1 except at the half-period points $\omega_1, \omega_2, \omega_1 + \omega_2$, which map to 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 $\mathbb{C}$.

These three properties show that the map we just constructed is bijective. It is also continuous, and it would be easy to show that it was a homeomorphism, but we can and must do better; which we do by noticing that at no point $u \in \mathbb{C}/L \setminus \{0\}$ is $\wp'(u) = \wp''(u) = 0$, which 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 $\mathbb{C}/L \to R$.

Remark: Clearly this map extends to a continuous map $\mathbb{C}/L \to R^{\text{cpt}} = R \cup \{\infty\}$. It is easy to show that, proceeding as in lecture 5, R can be compactified by the addition of a single point at ∞, so that π extends to a holomorphic map $R^{\text{cpt}} \to \mathbb{P}^1$. It is possible to show directly that the compactified Riemann surface R^{cpt} is 'the same' as $\mathbb{C}/L$, but this follows from a much more general fact which we now state.

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

For the proof, one chooses coordinate neighbourhoods near the questionable points and their images, and is then 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 Problem 9, Sheet 1.)

Concluding remarks: 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 next show that, choosing ω_1 and ω_2 suitably, one can produce any preassigned g_2 and g_3, subject only to the condition $g_2^3 - 27g_3^2 \neq 0$. ('Non-examinable material', below; see also Cohn, *Conformal Mappings on Riemann Surfaces*, for more details on the topic.) Thus, the Riemann surface of *any* equation $w^2 = p(z)$, with p a cubic polynomial with simple roots, is isomorphic to $\mathbb{C}/L \setminus \{0\}$. (Note that we can get rid of the quadratic term in p by shifting the z variable.)

Much more difficult is the following result:

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'.

More about elliptic functions (non-examinable material)

Let us for now 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 $\alpha \in \mathbb{C}$; 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.

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. Note that on $\mathbb{C}/L$, we have a global (i.e. everywhere holomorphic) 1-form, namely, du.

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 $\mathbb{C}/L$).

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:

$(2\omega_1, 2\omega_2) = (\int_{C_1} \phi, \int_{C_2} \phi)$ 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.

Proposition: The global holomorphic diferential du is dz/w.

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

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

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 $\wp(\omega_1), \wp(\omega_2), \wp(\omega_1 + \omega_2)$. So, changing variables to u, we get the obvious statements
$$\int_{\omega_2}^{\omega_1+\omega_2} du = \omega_1, \qquad\qquad \int_{\omega_1}^{\omega_1+\omega_2} du = \omega_2.$$

Remark: We can give an integral formula for the inverse $u(z) = \wp^{-1}(z)$ of the $\wp$-function as
$$u(z) = \int_{e_1}^{z} \frac{dz}{\sqrt{4z^3 - g_2 z - g_3}} + \omega_1 \left(= \int_{\omega_1}^{u(z)} du + \omega_1 \right).$$

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.

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

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:

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:

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.

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 $T \to \mathbb{C}/L$.

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 $v = 1/\sqrt{z}$.

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:

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

Proposition: { Lattices in $\mathbb{C}$, up to scaling } are in bijection with the orbits of $PSL(2, \mathbb{Z})$ on the upper half-plane $\mathscr{H} = \{z \mid \mathrm{Im}(z) > 0\}$, 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 $\mathrm{Im}(\tau) > 0$.

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 $\mathrm{Im}(\tau) > 0$ cuts down our group to $PSL(2, \mathbb{Z})$.

Fact: A *fundamental domain* for the action of $SL(2, \mathbb{Z})$ on $\mathscr{H}$ consists of the portion of the strip $-\frac{1}{2} \leq \mathrm{Re}(z) \leq \frac{1}{2}$ 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.

Definition: The *elliptic modular function* $J : \mathscr{H} \to \mathbb{C}$ sends $\tau \in \mathscr{H}$ 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:

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

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

Lecture 10

Today we shall describe the field of meromorphic functions over $\mathbb{C}/L$, and prove the Unique Presentation Theorems of lecture 7. Finally we introduce a new class of functions with interesting periodicity properties.

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 as

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

with R_0, R_1 rational functions.

Note: The two terms are the even and odd parts of f.

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

Proof of the theorem: We shall 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.

Corollary of proof: Unique Presentation Theorem by principal parts.

Two new functions

To prove the Unique Presentation Theorem 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.

Definition: The *Weierstrass ζ-function* $\zeta(u)$ is the unique odd antiderivative of $\wp$; that is,

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

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

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

Note: More explicitly, we can write

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

Proposition:

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

Proof: Once uniform convergence is established (this is left as an exercise), the relation $\zeta'(u) = \wp(u)$ follows by differentiating term by term.

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.

Proposition (Periodicity of ζ):

$$\zeta(u + 2\omega_i) = \zeta(u) + 2\eta_i,$$

where the η_i are given by

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

on any path of integration which avoids the poles.

Proof: The difference $\zeta(u+2\omega_i)-\zeta(u)$ has vanishing derivative, so it must be constant. Inserting $u=\omega_i$ 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

Proposition (Legendre Identity):

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

Proof: We consider the integral of $\xi\,\wp(\xi)$ along a period parallelogram, shifted by $-(\omega_1+\omega_2)$. 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 vertical sides of the parallelogram gives, using periodicity of $\wp$,

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

while the horizontal sides give similarly

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

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

Definition:

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

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

Note: $\int\zeta(\xi)\ d\xi$ is well-defined up to $2\pi i\mathbb{Z}$ and has only logarithmic singularities, so its exponential is well-defined and holomorphic.

Lecture 11

The Weierstrass σ-function and the Jacobi θ-functions

Recall that we defined

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

This is well-defined and holomorphic: the ambiguities in the integral are additive multiples of $2\pi i$, and its only singularities are logarithmic. There is an overall multiplicative factor, depending on the (unspecified) lower bound of integration. We adjust it by requiring that $\sigma'(0)=1$; using the Laurent expansion of ζ at $u=0$, $\zeta(u)=1/u+O(u^3)$ (Problem 8, Sheet 2), this amounts to using the antiderivative $\log u + O(u^4)$, with no constant term.

Note: The antiderivative $\int\zeta(\xi)\ d\xi$ is multi-valued, but the ambiguities in a definite integral 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 by the two paths).

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

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

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 choice of branch for the remaining logarithms is made.

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 $\omega \in L \setminus K$.

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 $a_n(z) = -1$).

Proposition (Periodicity of σ):

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

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

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

Corollary: Unique Presentation Theorem by zeroes and poles (see lecture 7).

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:

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:

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

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 $\omega_1 = 1/2$, whence $\omega_2 = \tau/2$. 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 Laurent expansion

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

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

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

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 θ-functions satisfy a dizzying collection of identities, with remarkable combinatorial and number-theoretic applications. The identities come from their relation to elliptic functions — ratios of θ-functions are always elliptic; the applications come from the presence of an 'n^2' in the exponent of q, in the power series expansion.

As an illustration, we shall prove the following 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

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

and comparing coefficients will give the result.

The series in $(*)$ is four times

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

(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. ($p = e^{\pi i u}$ 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

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

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

$$\theta_1'(0) = \pi\, \theta_2(0)\, \theta_3(0)\, \theta_4(0). \tag{$***$}$$

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

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

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

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

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

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

and substituting 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 12

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 sort of know 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 arieses 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 some serious analysis — finding solutions of the Laplace equation in various surface domains, with prescribed singularities. ('Green's functions'.)

Let $\mathbb{C}(R)$ be the field of meromorphic functions on the connected Riemann surface R. (Note: $\mathbb{C}(R)$ is only a ring, and not a field, if R is disconnected — why?) A non-constant meromorphic function z, whose existence is guaranteed by the theorem, 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 $\le n$ with coefficients in $\mathbb{C}(z)$,

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

(iii) Let now f be a meromorphic function on R with the following property: there exists some point $z_0 \in \mathbb{P}^1$ such that f takes n distinct values at the points of R over z_0. 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 as 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:

- General algebraic arguments imply, from (i), that any two elements $f, g \in \mathbb{C}(R)$, not constant, are *algebraically related* over $\mathbb{C}$; that is, there is an equation

$$\sum_{p,q=0}^{m,n} a_{pq} f^p g^q \equiv 0.$$

  In particular, (i) $\Rightarrow$ (ii) by 'standard' arguments.

- 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. Actually, 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!

- Needless to say, a function f as in (iii) exists; but we shall not prove that.

We shall sketch the proof of some of these statements next time. The proof of (iii) is assigned Problem 7, Sheet 3. Meanwhile, let us pursue the theoretical developments of the subject.

Theorem: There is a bijection between

$$\left\{ \begin{array}{c} \text{Isomorphism classes of field} \\ \text{extensions of } \mathbb{C}(z) \text{ of degree } n \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Isomorphism classes of compact} \\ \text{Riemann surfaces with a map } \pi \\ \text{of degree } n \text{ to } \mathbb{P}^1 \end{array} \right\}.$$

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

Theorem: There is a bijection between

$$\left\{ \begin{array}{c} \text{Isomorphism classes of fields} \\ \text{which can be realized as finite} \\ \text{extensions of } \mathbb{C}(z) \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{Isomorphism classes of compact} \\ \text{connected Riemann surfaces} \end{array} \right\}.$$

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.

To establish the correspondence, one assigns, in the 'easy' direction, 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 $\mathbb{C}^2$.

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 considered in detail, in two respects. Firstly, the poles of the a_i lead to 'branches running off to ∞', as in the picture:

Secondly, nothing in our assumptions guarantees that the surface is non-singular everywhere. There can indeed by self-crossings such as, for $w^2 = z^2 - z^3$,

or even 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 then turns out 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 $w \mapsto -w$.

Lecture 13

Today:

- Proofs of some easy statements from last time.
- A new tool: holomorphic differentials.
- Application to hyperelliptic Riemann surfaces.

We shall prove first, in connection with the four-part theorem from last time:

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

(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,

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

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 $\mathbb{P}^1$', as in part (iii). 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 7 on Sheet 3 (save for the existence of such an f, which is more difficult).

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 great enough 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 $(z, w) = (z(P), f(P))$. (Actually this only defines a map away from finitely many 'problem points'; but, once the Riemann surface of $P(z, w)$ has been properly constructed and conpactified, we can use the theorem which asserts tht 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

$$\phi(z) = \psi(u(z))\, u'(z). \tag{$*$}$$

(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 14

Analytic methods

Question: We investigated elliptic Riemann surfaces (tori) and elliptic functions very 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 shall see that a related question has a much better asnwer. The reason we cannot do very much 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: Obvious from the description of automorphisms of $\mathbb{C}$ (see Problem 11b, Sheet 1).

Corollary: The only quotient Riemann surfaces of $\mathbb{C}$ under a group which acts *freely* are: $\mathbb{C}$, $\mathbb{C}^*$ and eliptic Riemann surfaces. The groups are: $\{0\}$, $\mathbb{Z}\omega$ and $L = 2(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$, $(\bar{\omega}_1\omega_2 \notin \mathbb{R})$, acting by translations.

Proof: Problem 8, Sheet 3.

Remark:

(i) Freedom of the action is related to the condition that the map $\mathbb{C} \to$ 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 different ways, e.g. $\mathbb{C} = \mathbb{C}^*/\{\pm 1\}$, but that is not interesting.

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

Theorem (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.

Note: In case (i), $R = \mathbb{P}^1/\{1\}$; in case (ii), $R = \mathbb{C}/\Gamma$, with $\Gamma = \{0\}, \Gamma \cong \mathbb{Z}$ or $\Gamma \cong \mathbb{Z}^2$; so 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:

(i) One can (and must) strengthen that a bit: Γ acts *properly discontinuously*; see 'covering spaces' below.

(ii) The simply connected surface in question is the *universal covering surface* of 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 isomorphic to U via p.

Remarks:

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

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

Proposition: If p is *proper* (meaning $p^{-1}(\text{compact}) = \text{compact}$) 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 $p : \tilde{U} \to p(\tilde{U})$ is a homeomorphism. (In general the covering condition is stronger.)

Proof: We proved '$\Leftarrow$' in lecture 3 (local form of holomorphic maps), while '$\Rightarrow$', without any properness condition, is obvious.

Example: $S^1 \to S^1$, $z \mapsto z^n$ 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 $\mathbb{C} \to \mathbb{C}^*$.) However, the inclusion of an open subset in a surface is *not* a covering map.

Here are three basic properties of covering surfaces.

Theorem: 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 $\tilde{f} : X \to \tilde{R}$ with $\tilde{f}(x_0) = \tilde{r}$ (and $p \circ \tilde{f} = f$, the 'lifting' condition).

(iii) (Non-existence of interesting covering surfaces 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) You can 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, and uniqueness is equally clear; so concatenating these little liftings as you travel along ω produces your $\tilde{\omega}$.

(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 $\tilde{\omega}$, by (i), and we 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 path independent. By simple connectivity of X, any path γ' can be continuously deformed to γ. It's fairly easy to see (from the description 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.

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

$$\begin{tikzcd} & \tilde{U} \arrow[d, "p"] \\ U \arrow[r, "\mathrm{id}"'] \arrow[ur, dashed, "s"] & U \end{tikzcd}$$

Now $p \circ s = \mathrm{id}$, 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 said to be *properly discontinuous* if every $x \in X$ has some neighbourhood U all of 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.

Main theorem: Let X be a connected, locally path-connected and locally simply connected topological space (e.g. a connected surface). Then $X \cong \tilde{X}/\Gamma$, with $\tilde{X}$ simply connected and Γ acting properly discontinuously. Moreover, $\tilde{X}$ and Γ are unique up to isomorphism. Γ is called the *fundamental group* of X.

Remark: Connoisseurs will know that an ambiguity is hidden in this definition (We have not chosen a base point on X, so Γ 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 $\tilde{X}', \tilde{x}_0', \Gamma'$, we have *unique* isomorphisms

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

Remark: Γ is the group of automorphisms of $\tilde{X}$ that commute with the projection to X (see lecture 15, Appendix).

Remark: The proof of the theorem is rather easy but we only give the idea. Uniqueness follows from the lifting theorems, so we only need to construct $\tilde{X}$ and Γ. We 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', while keeping the endpoints fixed. Γ is the set of continuous deformation classes of loops on X based at x_0. It can be shown to form a group, if multiplication of loops is defined as concatenation, and a loop γ acts on $\tilde{X}$ by sending the pair (ω, x) to $(\gamma \cdot \omega, x)$. The map to $\tilde{X}$ to X sends (ω, x) to x. This ensures that any loop in X which lifts to a *loop* (rather than path) in $\tilde{X}$ is *contractible* in X. From here, using the lifting theorems, one concludes that every loop in $\tilde{X}$ is contractible.

Application: The Little Picard Theorem

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 missed two values a and b. By translation and rescaling we can assume that $a = 0$ and $b = 1$. So we have a holomorphic map $f : \mathbb{C} \to \mathbb{C} \setminus \{0, 1\}$. Now, 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 $\Delta \to \mathbb{C} \setminus \{0, 1\}$.

Lecture 15

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 $\mathscr{H}$ 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 $\mathscr{H}$, namely $PSL(2,\mathbb{Z}) \subset PSL(2,\mathbb{R})$, acting via Möbius transformations.

Proposition: The orbits of $PSL(2,\mathbb{Z})$ on $\mathscr{H}$ 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 $\tau \in \mathscr{H}$ 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 $\omega_2/\omega_1 \in \mathscr{H}$ (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 $\mathscr{H}$ is the set $\{z \in \mathscr{H} \mid |\operatorname{Re} z| \leq \frac{1}{2}, |z| \geq 1\}$. 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: $\mathscr{H}/SL(2,\mathbb{Z}) \cong \mathbb{C}$, 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 $\mathscr{H}$.

(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 $\mathscr{H}$.

Proof: Problem 10, Sheet 3.

Proposition: A fundamental domain for the action of $\Gamma(2)$ on $\mathscr{H}$ is as depicted below; it incldues 6 fundamental domains for $PSL(2,\mathbb{Z})$, labelled 1–6.

Proposition: $\mathscr{H}/\Gamma(2) \cong \mathbb{C} \setminus \{0,1\}$ 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

$$\begin{tikzcd} \lambda \, : \, \mathscr{H} \arrow[r] \arrow[rr, bend right, "J"'] & \mathbb{C} \setminus \{0,1\} \arrow[r, "\phi"] & \mathbb{C} \end{tikzcd}$$

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 $J : \mathscr{H}/\Gamma(2) \to \mathbb{C}$ 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 Poincarré, 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.)

Appendix: Fundamental groups, covering spaces and Galois groups

Classification of covering spaces

Every (connected, locally path-connected and locally simply connected) X is a quotient $\tilde{X}/\Gamma$ with $\tilde{X}$ simply connected and Γ isomorphic to the fundamental group of X.

Now 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:

$$\begin{tikzcd} & Y \arrow[d, "p"] \\ \tilde{X} \arrow[ur, dashed, "\tilde{\pi}"] \arrow[r, "\pi"'] & X \end{tikzcd}$$

It easily follows that $\tilde{\pi} : \tilde{X} \to Y$ is also a covering map. (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 $V \cong U$.) We make the following observation.

Proposition: Γ is the group of automorphisms of $\tilde{X}$ which commute with $\pi : \tilde{X} \to X$.

Proof: Indeed this follows from the uniqueness of the lifting in the lifting theorem. If we fix $x \in X$, and two liftings $\tilde{x}_1$ and $\tilde{x}_2$ for it, there will be a unique map $\gamma : \tilde{X} \to \tilde{X}$ commuting with π and taking $\tilde{x}_1$ to $\tilde{x}_2$.

$$\begin{tikzcd} & \tilde{X} \arrow[d, "\pi"] \\ \tilde{X} \arrow[ur, dashed, "\gamma"] \arrow[r, "\pi"'] & X \end{tikzcd}$$

This must be then the unique $\gamma \in \Gamma$ taking $\tilde{x}_1$ to $\tilde{x}_2$ (which map to the same $x \in X$, so must be in the same Γ-orbit).

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 is contained in Γ. We have established:

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 $\tilde{X}/H$.

Remark: We are still being a bit careless, and to be precise we should choose a base point $x \in X$, a point $y \in Y$ lifting above it, and by 'isomorphism' between (Y, y, π) and (Y', y', π') we mean an isomorphism $f : Y \to Y'$ taking y to y' such that $\pi' \circ f = \pi$.

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

Definition: A covering space $Y \xrightarrow{p} X$ is *normal* 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 $Y/(\Gamma/H) = X$.

Analogy with Galois groups

There is a formal analogy between this picture 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. Then, $K = (\bar{K})^\Gamma$, the Γ-invariants. Any algebraic field extension $K \subseteq L$ can be placed inside $\bar{K}$

and $\bar{K}$ is a Galois extension of L with Galois group H, a subgroup of Γ. If it happens that H is normal in Γ, then L is a Galois extension of K, that is, Γ/H acts on L and the fixed point field is K. Clearly, $\bar{K}$ is analogous to the 'universal cover' of K.

In 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

the shaded areas representing normal coverings, on the topological side, and Galois extensions, on the algebraic side. (These notions agree on the overlap.) The intersection consists of the maps of finite degree; on the algebraic side, these are called 'unramified field extensions'.

Lecture 16

Analytic continuation and the Riemann surface of an analytic function

Here 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 we took in lecture 1, and, historically, this was one of the major motivations 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:

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 z_0, then $f = g$ on U.

Say f is defined in a small disc around z_0. 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: How could we recover g from U, assuming it exists?

There is a procedure, in principal (rarely used in practice) based on Taylor expansion. It is based on:

Proposition: Let f be holomorphic in a disc centered at a. Then the Taylor expansion of f about a converges (uniformly on compact subsets) under the disc.

This shows that we can immediately extend f to the largest disc that 'does not hit a singularity', if we know the Taylor expansion at a point.

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 $z - a = -a$ $(z = 0)$ without 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. It is clear that repeating this process will recover the function everywhere (expect at $z = 0$ where it is singular). We'll need infinitely many steps to get all of f, but any point in the plane can be reached in finitely many steps.

This process can by systematized via something called 'analytic continuation along paths'. Say f is a holomorphic function $f : U \to \mathbb{C}$ but we only know it near $a \in U$. To recover the function near $b \in U$, choose a path γ from a to b and keep expanding f via a Taylor series at points along the path. Eventually you will reach b.

Say f has singularities on ∂U; still we can expand it as a Taylor series on discs contained in U.

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 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 can recover its maximal domain $\mathbb{C}^*$.

Unfortunately, the conjecture is *false*, for the reason discussed in 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 try to continue it along various paths, but we find that the function we get 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 always be an easy relation between the different continuations.

The reason a natural conjecture such as $(*)$ can be false is that we asked the question of a maximal domain — slightly wrong. It turns out that a maximal domain exists, and is *unique up to isomorphism*, but not in the class of open subsets of $\mathbb{C}$, but rather in the class of Riemann surfaces.

Let us then consider objects (R, r, π, F), where:

- R is a connected Riemann surface,
- $r \in R$,
- $\pi : R \to \mathbb{C}$ is a holomorphic map, with $\pi(r) = a$ (to keep with convention we also require that π is a local isomorphism — it has valency 1 everywhere),
- $F : R \to \mathbb{C}$ is a holomorphic function whose restriction to a small neighbourhood V of r agrees with $f \circ \pi$.

$$\begin{tikzcd}
V \arrow[d, "\cong"] & \subset & R \arrow[d, "\pi"] \arrow[dr, "F"] & \\
 & & & \mathbb{C} \\
U & \subset & \mathbb{C} &
\end{tikzcd}$$

$$U \xrightarrow{f} \mathbb{C}$$

Thus, 'F is an analytic continuation of f', if f is transported to R via the identificaton of V with U.

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* defined by the analytic function f. It only depends on knowing f in a tiny neighbourhood of a.

'Strict final object' means: any other quadruple (R, r, π, F) maps to it holomorphically, while preserving the entire structure;

$$\begin{tikzcd}
r \arrow[r, mapsto, "\phi"] & r_{\max} & \\
R \arrow[r, "\phi"] \arrow[d, "\pi"] & R_{\max} \arrow[d, "\pi_{\max}"] \arrow[dr, "F_{\max}"] & \\
 & & \mathbb{C} \\
\mathbb{C} \arrow[r, equal] & \mathbb{C} &
\end{tikzcd}$$

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 $F(w) = w$.

$$\begin{tikzcd}
w \arrow[dd, mapsto] & \mathbb{C}^* \arrow[dd] \arrow[dr, "F(w)=w"] & \\
 & & \mathbb{C} \\
z = w^2 & \mathbb{C}^* \arrow[ur, "f(z)=\sqrt{z}"'] &
\end{tikzcd}$$

Note that there is a 'bigger' cover of $\mathbb{C}^*_{(z)}$ we could use in place of w, namely

$$\begin{tikzcd} u \arrow[d, mapsto] & \mathbb{C}_{(u)} \arrow[d] \arrow[dr, "F(w)=e^{(u/2)}"] & \\ z = e^u & \mathbb{C}^*_{(z)} \arrow[r, "f(z)=\sqrt{z}"'] & \mathbb{C} \end{tikzcd}$$

but we can map $\mathbb{C}_{(u)}$ to $\mathbb{C}^*_{(w)}$ by $w = \exp(u/2)$

$$\begin{tikzcd} \mathbb{C}_{(u)} \arrow[r, "e^{(u/2)}"] \arrow[d, "z=e^u"'] & \mathbb{C}^*_{(w)} \arrow[d, "z=w^2"'] \arrow[dr, "w"] & \\ \mathbb{C}^*_{(z)} \arrow[r] & \mathbb{C}^*_{(z)} \arrow[r, "\sqrt{z}"'] & \mathbb{C} \end{tikzcd}$$

so going to this much bigger cover is unnecessary.

As a consequence of the existence of R, we shall prove

Theorem (The Monodromy Theorem): Let U be a simply connected domain, f a holomorphic function defined near $a \in U$. Assume that f can be analytically continued 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 redefinition of 'analytic continuation along a path' and on a topological lemma:

Proposition: f can be analytically continued along a path γ iff there exists a Riemann surface S, mapping locally isomorphically to $\mathbb{C}$, to which f extends and γ lifts.

Proof: This is really the cleanest definition of analytic continuation along a path. The Riemann surface arises by gluing the discs of successive power series expansion in the natural order. We avoid unnecessary gluings arising from self-crossings of γ:

Lemma: Let U be simply connected, $\tilde{U}$ connected and $\pi : \tilde{U} \to U$ a local isomorphism with the *path lifting property*: every path γ, starting at a prescribed $u \in U$, lifts to a path on $\tilde{U}$ starting at a prescribed lift $\tilde{u}$ of U. Then, $\tilde{U}$ is isomorphic to U.

Proof: This is really part (iii) of the theorem on lifting properties — the proof only uses the path lifting property of π; an inverse of π is constructed by sending a point $u \in U$ to the endpoint of a path in $\tilde{U}$ lifting any chosen path from a to u in U.

Remark: The topologically savvy reader will see how to modify the argument to show that, if U is locally simply connected, it follows that $\tilde{U}$ is a *covering space* of U. (Study simply connected neighbourhoods of any given point.)

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. Then, $\pi : \tilde{U} \to U$ has the path lifting property, for if there was a path γ that did not lift there would be a Riemann surface S, carrying an analytic continuation of f, which could not be mapped to R (γ lifts to S, but not to R). The Lemma ensures that $\tilde{U} = U$, and then $F/\tilde{U}$ is a single-valued analytic continuation of f.

Proof of the existence of the maximal domain (sketch): One starts by invoking Zorn's Lemma to conclude the existence of maximal domains in our category: these are quadruples $Q = (R, r, \pi, F)$ which do not map to any that are not isomorphic to them. (This is not completely obvious, because the maps in our category are not inclusions but only local isomorphisms, so the construction of an upper bound for an ascending chain takes an argument; this will come up again in a second.)

Now, having such a maximal quadruple Q, we must show that every other one maps to it. Given one $Q' = (S, s, \rho, G)$ that does not, we construct a 'bigger' quadruple $Q'' = (T, t, \sigma, H)$ as follows: T is the union of R and S, modulo identifying two points $r' \in R$ and $s' \in S$ if they lie over the same point in $\mathbb{C}$ *and* if F, in a neighbourhood of r', agrees with G in a neighbourhood of s'. (For instance we identify r with s.)

Continuity of F and G is used to prove that the resulting space is *Hausdorff* (this is the part that must also be checked in constructing the 'upper bound' in Zorn's Lemma); the fact that it is locally isomorphic to $\mathbb{C}$, hence a Riemann surface, is immediate. It is also clear by construction that F and G assemble to a well-defined holomorphic map on T, and that Q and Q' map to Q''. Maximality of Q would be contradicted, unless $Q = Q''$; but then Q' maps to $Q = Q''$, as desired. Uniqueness of that map follows because s must map to r, π and ρ must be compatible, and S is connected.

Historical remark: The construction of the Riemann surface $R_{\max}$ of an analytic function is essentially due to Wererstrass in full generality. 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 $R_{\max}$ than the one suggested above (and is the one given in most textbooks); 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 never performed by Taylor expansion, so that is not a practical way to construct $R_{\max}$.)