Riemann Surfaces

S. K. Donaldson

December 2, 2004

Contents

Part I

Preliminaries

Chapter 1

Holomorphic functions

1.1 Simple examples; algebraic functions

This is an introductory Chapter in which we recall some examples of holomorphic functions in complex analysis. We emphasise the idea of "analytic continuation" which is a fundamental motivation for Riemann surface theory.

One naturally encounters holomorphic functions in various ways. One way is through power series, say $f(z) = \sum a_n z^n$. It often happens that a function which is initially defined on some open set $U \subset \mathbf{C}$ turns out to have natural extensions to larger open sets. for example the power series

$$f(z) = 1 + z + z^2 + z^3 + \dots$$

converges only for $|z| < 1$, but writing $f(z) = 1/(1-z)$ we see that the function actually extends to $\mathbf{C} \setminus \{1\}$. A more subtle example is the Gamma function. For $\Re(z) > 0$ we write

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$$

The integral is convergent and defines a holomorphic function of z on this half-plane. Integration by parts shows that

$$\Gamma(n) = (n-1)!$$

if n is a positive integer. It is clear that $\Gamma(z)$ tends to infinity as z tends to 0 but let us examine this more carefully by writing

$$\Gamma(z) = \int_0^1 t^{z-1} e^{-t} dt + \int_1^\infty t^{z-1} e^{-t} dt.$$

The second integral is defined for all z, and holomorphic in z. We write the first integral as
$$\int_0^1 t^{z-1}(e^t - 1)dt + \int_0^1 t^{z-1}dt.$$
Now the term
$$\int_0^1 t^{z-1}(e^t - 1)dt$$
is defined, and holomorphic in z, for $\Re(z) > -1$. The other integral we can evaluate explicitly:
$$\int_0^1 t^{z-1}dt = \frac{1}{z}.$$
So we conclude that, for $\Re(z) > 0$,
$$\Gamma(z) = \frac{1}{z} + \Gamma_1(z)$$
say, where Γ_1 extends to a holomorphic function on the larger half-plane $\{z : \Re(z) > -1\}$. So Γ has a meromorphic extension to the larger half-plane. Repeating the procedure, by considering
$$e^{-t} - (1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \ldots + \frac{(-t)^k}{k!}),$$
we get a meromorphic extension to $\Re(z) > -(k+1)$, and thence to the whole of $\mathbf{C}$.

Exercise. Show that Γ has a simple pole at the point $z = -k$ for positive integers k, with residue $(-1)^k/k!$.

It often happens that when extending a function one encounters "multiple valued functions". For example
$$f(z) = 1 + \frac{z}{2} - \frac{z^2}{2^2 2!} + \frac{3z^3}{2^3 3!} - \frac{5.3z^5}{2^4 4!} + \ldots,$$
is a perfectly good holomorphic function on the disc $|z| < 1$ which we recognise as $\sqrt{1+z}$. This cannot be extended holomorphicaly to $z = -1$ but, more, if we try to extend the function to $\mathbf{C} \setminus \{-1\}$ we find that going once around the origin the function switches to the other branch of the square root. Particularly important examples of this phenomena occur for "algebraic functions".

Let $P(z,w)$ be a polynomial in two complex variables. We want to think of the equation $P(z,w)=0$ as defining w "implicitly" as a function of z. For example if $P(z,w)=w^2-(1+z)$ then we would get the function $w=\sqrt{1+z}$ above. Or if $P(z,w)=z^3+w^2-1$ we would get the function $w=\sqrt{1-z^3}$. Of course this does not make sense precisely because of the problem of multiple values. The next theorem, which will be fundamental later, expresses precisely the way in which such functions are defined.

First, some notation. Let $X\subset \mathbf{C}^2$ be the set of points (z,w) with $P(z,w)=0$. Second, we define the partial derivatives

$$\frac{\partial P}{\partial z},\frac{\partial P}{\partial w},$$

in the obvious way. They are again polynomial functions of the two variables z,w.

Theorem 1 *Suppose (z_0,w_0) is a point in X and $\frac{\partial P}{\partial w}$ does not vanish at (z_0,w_0). Then there is a disc D_1 centred at z_0 in $\mathbf{C}$ and a disc D_2 centred at w_0 and a holomorphic map $\phi : D_1\to D_2$ with $\phi(z_0)=w_0$ such that*

$$X\cap(D_1\times D_2)=\{(z,\phi(z)) : z\in D_1\}.$$

The reader will recognise the set in the statement as the "graph" of the map ϕ. Essentially the theorem says that $w=\phi(z)$ gives the unique local solution to the equation $P(z,w)=0$, where local means close to (z_0,w_0).

To prove the Theorem, recall that if f is a holomorphic function on an open set containing the closure of a disc D which does not vanish on the boundary ∂D then the number of solutions of the equation $f(w)=0$ in D, counted with multiplicity, is given by the contour integral

$$\frac{1}{2\pi i}\int_{\partial D}\frac{f'(w)}{f(w)}dw.$$

If there is only one solution, w_1, it is given by another contour integral

$$w_1=\frac{1}{2\pi i}\int_{\partial D}\frac{wf'(w)}{f(w)}dw.$$

We apply these formulae to the family of functions of the variable w: $f_z(w)=P(z,w)$, where we regard z as a parameter. First take $z=z_0$. Then the hypothesis that $\frac{\partial P}{\partial w}\neq 0$ means that f'_{z_0} does not vanish at w_0. Thus we can find a small disc D_2 centred at w_0 so that f_{z_0} has no other zeros in the closure of D_2. Since the boundary of D_2 is compact there is some $\delta > 0$ such that $|f_{z_0}| > 2\delta$ on ∂D_2. By continuity, this means that if z is sufficiently close to z_0 we still have $|f_z| > \delta$, say, on ∂D_2. Thus we can apply the formula above for the number of roots of the equation $f_z(w) = 0$. When $z = z_0$ this must be 1 so, by continuity, the same is true for z close to z_0. Then we define $\phi(z)$ to be this unique root. The second formula shows that

$$\phi(z) = \int_{\partial D_2} \frac{w}{P}\frac{\partial P}{\partial w} dw,$$

which is clearly holomorphic in the variable z. This completes the proof.

1.2 Analytic continuation; differential equations

Next we want to give a precise meaning to possibly many-valued extensions of a holomorphic function.

Definition 1 *Let ϕ be a holomorphic function defined on a neighbourhood of a point $z_0 \in \mathbf{C}$. Let $\gamma : [0,1] \to \mathbf{C}$ be a continuous map with $\gamma(0) = z_0$. An analytic continuation of ϕ along γ consists of a family of holomorphic functions ϕ_t, for $t \in [0,1]$, where ϕ_t is defined on a neighbourhood U_t of $\gamma(t)$ such that*

- *$\phi_0 = \phi$ on some neighbourhood of z_0;*
- *for each $t_0 \in [0,1]$ there is a $\delta_{t_0} > 0$ such that if $|t - t_0| < \delta_{t_0}$ the functions ϕ_t and ϕ_{t_0} are equal on their their common domain of definition $U_t \cap U_{t_0}$.*

For example suppose $z_0 = 0$ and ϕ is the function defined by the power series above, giving a branch of $\sqrt{1+z}$. Let γ be the path which traces out the circle centred at -1

$$\gamma(t) = -1 + e^{2\pi i t}.$$

The the reader will see how to construct an analytic continuation of ϕ along this path with ϕ_1 equal to the other branch of the square root,

$$\phi_1 = -\phi,$$

on a suitable neighbourhood of 0.

Alongside the algebraic equations discussed above, another very important way in which holomorphic functions arise is as solutions to *differential equations*. Of course the simplest example here is the differential equation

$$\frac{du}{dz} = g,$$

where g is a given function, whose solution is the indefinite integral of g and is given by a contour integral. We know that we may again encounter many valued functions, for example when g is the function $1/z$. In our current language, the contour integral along a (smooth) path γ furnishes an analytic continuation of a local solution along γ.

We will consider second order linear, homogeneous equations of the form;

$$u'' + Pu' + Qu = 0,$$

where P and Q are given functions of z and u is to be found. Suppose first that P and Q are holomorphic near $z = 0$. So they have power series expansions

$$P(z) = \sum p_n z^n \quad , \quad Q(z) = \sum q_n z^n,$$

valid in some common region $|z| < R$. We seek a solution to the equation in the form $u(z) = \sum u_n z^n$. Equating terms we get, for each $n \geq 0$,

$$(n+2)(n+1)u_{n+2} + \sum_{i \geq 0}(n+1-i)p_i u_{n+1-i} + \sum_{j \geq 0} q_j u_{n-j} = 0.$$

Both the sums are finite and only contain terms u_i for $i < n+2$, so this gives recursion formula. Given any choice of u_0, u_1 there is a unique way to define all the u_i satisfying the equations.

Exercise Show that the power series $\sum u_n z^n$ converges for $|z| < R$.

We conclude that the solutions of our equation on the disc $|z| < R$ form a 2-dimensional complex vector space.

Now suppose that P, Q are holomorphic on some open set $\Omega \subset \mathbf{C}$ and let $\gamma : [0,1] \to \Omega$ be a path in Ω.

Proposition 1 *If u is a solution to the equation (*) on a neighbourhood of $\gamma(0)$ then u has an analytic continuation along γ, through solutions of the equation*

We leave the proof as an exercise. (In fact one can generalise the result to the case when P, Q are themselves defined initially in a neighbourhood of $\gamma(0)$ and have analytic continuations along γ.)

This leads to the notion of the "monodromy" of solutions to a differential equation. Suppose γ is a loop in Ω, with $\gamma(0) = \gamma(1)$. Let u_1, u_2 be a basis for the solutions of the differential equation on a small neighbourhood of $\gamma(0)$. Analytic continuation of u_1, u_2 along γ yields another pair of solutions$\tilde{u}_1, \tilde{u}_2$ say. these are linear combinations of the original pair

$$\tilde{u}_1 = au_1 + bu_2, \tilde{u}_2 = cu_1 + du_2.$$

In more invariant language, let V be the two dimensional vector space of local solutions, then analytic continuation along γ gives a linear map

$$M_\gamma : V \to V.$$

Interesting examples arise when the complement of Ω is a discrete subset of $\mathbf{C}$ and P, Q are meromorphic.

Definition 2 *A point $z_0 \in \mathbf{C}$ is a regular singular point of the equation $u'' + Pu' + Qu = 0$ if P has at worst a simple pole at z_0 and Q has at worst a double pole.*

Consider a model case,

$$u'' + \frac{A}{z}u' + \frac{B}{z^2}u = 0,$$

where A, B are complex constants. We try a solution $u = z^\alpha$ defined on a cut plane, say. This satisfies the equation if

$$\alpha(\alpha - 1) + A\alpha + B = 0.$$

If this quadratic equation (the *indicial equation*) has two distinct roots α_1, α_2 we get two solutions to our equation $z^{\alpha_1}, z^{\alpha_2}$ in the cut plane. If there is a double root α, the second solution is $z^\alpha \log z$. The general case is similar

Proposition 2 *If $P(z) = \frac{A}{z} + P_0(z)$ and $Q(z) = \frac{B}{z^2} + \frac{C}{z} + Q_0(z)$ where P_0, Q_0 are holomorphic near 0 and if the indicial equation has roots α_1, α_2 with $\alpha_1 - \alpha_2 \notin \mathbf{Z}$ then there are solutions*

$$u_1(z) = z^{\alpha_1} w_1(z) \quad , \quad u_2(z) = z^{\alpha_2} w_2(z),$$

to the equation $u'' + Pu' + Qu = 0$ where w_1, w_2 are holomorphic in a neighbourhood of 0.

The proof goes by power series expansion, as before. If $\alpha_1 - \alpha_2 \in \mathbf{Z}$ the second solution may involve log terms. Expressed in terms of monodromy, if z_0 is a regular single point of the equation $u'' + Pu' + Q = 0$ and γ is a loop around z_0 then in the case when $\alpha_1 - \alpha_2 \notin \mathbf{Z}$ the monodromy around γ is (in a suitable basis) the diagonal matrix with entries $e^{2\pi i \alpha_1}, e^{2\pi i \alpha_2}$. In the other case we may get a nontrivial Jordan form.

A very important example, which we will return to later, is the *hypergeometric equation*:

$$z(1-z)u'' + (c - (a+b+1)z)u' - abu = 0,$$

which has regular singular points at $0, 1$. (It also has a regular single point "at infinity" as we will explain later.) Here a, b, c are fixed parameters.

Exercise Show that the indicial equation at $z = 0$ has roots $0, c$ and that the solution corresponding to the root 0 is the hypergeometric function

$$F(z) = 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{c(c+1)2!}z^2 + \frac{a(a+1)a+2)b(b+1)(b+2)}{c(c+1)(c+2)3!}z^3 + \dots,$$

in $|z| < 1$.

Chapter 2

Surface Topology

2.1 Classification of surfaces

In this second introductory chapter we change direction completely. We discuss the topological classification of surfaces, and outline one approach to a proof. Our treatment here is almost entirely informal; we do not even define precisely what we mean by a "surface". (Definitions will be found in the following Chapter.) However it is not in fact too hard to turn our informal account into a precise proof. The reasons for including this material here are, first, that it gives a counterweight to the previous chapter: the two together illustrating two themes—complex analysis and topology—which run through the subject of Riemann surfaces; and, second, that we are able to introduce some more advanced ideas that will be taken up later in the book.

The statement of the clasification of closed surfaces is probably well-known to many readers. We write down two families of surfaces Σ_g, Ξ_h for integers $g \geq 0, h \geq 1$.

The surface Σ_0 is the 2-sphere S^2. The surface Σ_1 is the 2-torus T^2. For $g \geq 2$ we define the surface Σ_g by taking the "connected sum" of g copies of the torus. In general if X and Y are (connected) surfaces the connected sum $X\sharp Y$ is a surface constructed as follows. We choose small discs D_X in X and D_Y in Y and cut them out to get a pair of "surfaces-with-boundaries", coresponding to the circle boundaries of D_X and D_Y. Then we glue these boundary circles together to form $X\sharp Y$. [DIAGRAM 1]

One can show that this resulting surface is (up to natural equivalence) independent of the choices of discs. Also the operation $\sharp$ is commutative and associative, up to natural equivalence. Now we define inductively, for $g \geq 2$

$$\Sigma_g = \Sigma_{g-1} \sharp \Sigma_1 = \Sigma_{g-1} \sharp T^2,$$

which we can write as

$$\Sigma_g = T^2 \sharp \ldots \sharp T^2.$$

[DIAGRAM 2]

There are many other representations of these surfaces, topologically equivalent. For example we can think of Σ_g as being obtained by deleting $2g$ discs from the 2-sphere and adding g cylinders to form g "handles". Or we start with a disc and add g ribbons in the manner shown:

[DIAGRAM 3]

The boundary of the resulting surface-with-boundary is a circle and we form Σ_g by attaching a disc to this boundary to get a closed surface (i.e. a surface with no boundary).

The surface Ξ_1 is the real projective plane $\mathbf{R}P^2$. We can form it by starting with a Mobius band and attaching a disc to the boundary circle. We cannot do this within ordinary 3-dimensional space without introducing self-intersections: more formally $\mathbf{R}P^2$ cannot be embedded in $\mathbf{R}^3$. But we can still perform the construction to make a topological space and if we like we can think of embedding this in some $\mathbf{R}^n$ for larger n. Again there are many other models possible. Notice that we can think of our Mobius band as a disc with a twisted ribbon attached [DIAGRAM 4]. Then the construction falls into the same pattern as our third representation of Σ_g.

Now we make the family of surfaces Ξ_h by taking connected sums of copies of $\Xi_1 = \mathbf{R}P^2$:

$$\Xi_h = \mathbf{R}P^2 \sharp \ldots \sharp \mathbf{R}P^2.$$

Now let S be any closed, connected, surface. (More precisely we mean compact and without boundary, so for example $\mathbf{R}^2$ would not count as closed.) We say S is orientable if it does not contain any Mobius band, non-orientableif it does.

Classification Theorem

If S is orientable it is equivalent to one (and precisely one) of the Σ_g.

If S is not orientable it is equivalent to one (and precisely one) of the Ξ_h.

(We emphasise again that this statement has quite a different status from the other "Theorems" in this book, since we have not even defined the terms precisely.)

To see an example of this, consider the "Klein bottle" K. This can be pictured in $\mathbf{R}^3$ as shown in [DIAGRAM 5], except that there is a circle of self-intersection. So we should think of pushing the handle of the surface into a fourth dimension where it passes through the "side", just as we can take a curve in the plane as pictured in [DIAGRAM 6] and remove the intersection point by lifting one branch into three dimensions.

Now it is true but perhaps not immediately obvious, that K is equivalent to Ξ_2. To see this cut the picture in Diagram 5 down the vertical plane of symmetry. Then you can see that K is formed by gluing two Mobius bands along their boundaries. By definition of the connected sum this shows that $K = \mathbf{R}P^2 \sharp \mathbf{R}P^2$, since the complement of a disc in $\mathbf{R}P^2$ is a Mobius band.

Now we will outline a proof of the classification theorem. The proof uses ideas that (when developed in a rigorous way of course), go under the name of "Morse Theory". A detailed technical account is given in the book *Differential Topology* by Hirsch. The idea is that, given our closed surface S, we choose a typical real valued function on S. Here "typical" means, more precisely, that f is what is called a Morse function. What this requires is that if we introduce a choice of gradient vector field of f on S then there are only a finite number of points P_i, called critical points, in S where v vanishes, and near any of one of these points P_i we can parametrise the surface by two real numbers u, v such that the function is given by one of three local models:

- $u^2 + v^2 + constant$.
- $u^2 - v^2 + constant$.
- $-u^2 - v^2 + constant$.

The critical point P_i is said to have *index* $0, 1$ or 2 respectively in these three cases.

For example if S is a typical surface in $\mathbf{R}^3$ we can take the function f to be the restriction of the z co-ordinate (say): the "height" function on S. The vector field v is the projection of the unit vertical vector to the tangent spaces of S and the critical points are the points where the tangent space is horizontal. The points of index 0 and 2 are local minima and maxima respectively and the critical points of index 1 are "saddle points". [DIAGRAM 8]

Now for each $t \in \mathbf{R}$ we can consider the subset $S_t = \{x \in S : f(x) \leq t\}$. There are a finite number of special cases, the "critical values" $f(P_i)$ where P_i is a critical point. If f is a sufficiently typical function then the values $f(P_i)$ will all be different, so to each critical value we can associate just one critical point. If t is not a critical value then S_t is a surface with boundary. Morover if t varies over an interval not containing any critical values then the surfaces-with boundaries S_t are equivalent for different parameters t in the interval. To see this, if $t_1 < t_2$ and $[t_1, t_2]$ does not contain any critical value then one deforms S_{t_2} into S_{t_1} by pushing down the gradient vector field v. [DIAGRAM 9]

Now consider the exceptional case when t is a critical value, t_0 say. The set S_{t_0} is no longer a surface. However we can analyse the difference between $S_{t+\epsilon}$ and $S_{t-\epsilon}$ for small positive ϵ. The crucial thing is that this analysis is concentrated around the corresponding critical point, and the change is the surface follws one of three standard local models, depending on the index.

Index 0 Near the critical point the $S_{t_0 \pm \epsilon}$ corresponds to $\{u^2 + v^2 \leq \pm\epsilon\}$ which is empty in one case and a disc in the other case. In other words, $S_{t_0+\epsilon}$ is obtained from $S_{t_0-\epsilon}$ by adding a disc as a new connected component.

Index 2 This is the reverse of the index 0 case. The surface $S_{t_0+\epsilon}$ is formed by attaching a disc to a boundary component of $S_{t_0-\epsilon}$.

index 1 This is a little more subtle. The local picture is to consider

$$\{u^2 - v^2 \leq \pm\epsilon\},$$

as shown in [DIAGRAM 10]. This is equivalent to adding a strip to the boundary [DIAGRAM 11]. Thus $S_{t_0+\epsilon}$ is formed by adding a strip to the boundary of $S_{t_0-\epsilon}$.

We can now see a stratgy to prove the Classification Theorem. What we should do is to prove a more general theorem, classifying surfaces with boundary (not necessarily connected, and including the case of empty boundary). Suppose we have any class of model surfaces with boundary which is closed under the three operations associated to index 0,1,2 critical points explained above. Then it follows that our original closed surface S must lie in this class, since it is obtained by a sequence of these operations.

For $r \geq 0$ let $\Sigma_{g,r}$ be the surface with boundary (possibly empty) obtained by removing r disjoint discs from Σ_g. Similarly for $\Xi_{h,r}$. Now we aim to prove that the class of disjoint unions of copies of the $\Sigma_{g,r}$ and $\Xi_{h,r}$ (for any collection of g's h's and r's) is closed under the three operations above. That is if X is in our class (a disjoint union of copies of $\Sigma_{g,r}$ and $\Xi_{h,r}$ and we obtain X' by performing of the operations then X' is also in our class. This does require a little thought, and consideration of various cases.

Index 0 This is obvious, since $\Sigma_{1,1}$ is the disc so we can include a new disc component in our class.

Index 2 Also obvious. We cap off some boundary component with a disc turning some $\Sigma_{g,r}$ into a $\Sigma_{g,r-1}$ or a $\Xi_{h,r}$ into a $\Xi_{h,r-1}$.

Index 1 This requires more work. There are various cases to consider.

Case 1. The ends of the attaching strip lie on different components of X.

Case 2. The ends of the attaching strip lie on the same component of X.

Now Case 2 subdivides into

Case 2(i) The ends of the strip lie on the same boundary component of a component of X.

Case 2(ii) The ends of the strip lie on the different boundary components of one common component of X.

Further, each of these cases 2(i), 2(ii) subdivide because there are two ways we can make the attaching, diferring by a twist, in just the same fashion as when we form a Mobius band from a disc above. (The reader may like to think through why we do not need to make this distinction in Case 1.)

Now let us get to work.

Case 1. Let the relevant components of X be A and B say. Then we can write $A = A' \setminus \text{disc}, B = B' \setminus \text{disc}$ for some A', B' and $\setminus$ disc means that the operation of removing a disc, so of course the boundaries of the indicated discs contain the attaching regions. Then we can see that the manifold we get when we attach a strip is $(A' \sharp B') \setminus \text{disc}$. [DIAGRAM 12].

Case 2(i). Let the component of X to which we attach the strip be $A = A' \setminus \text{disc}$. Then in the untwisted case we get, after the strip attachment, $A' \setminus \text{disc} \setminus \text{disc}$. [DIAGRAM 13]

In the twisted case we get $(A' \sharp \mathbf{R}P^2) \setminus \text{disc}$. [DIAGRAM 14]

Case 2(ii). Here the distinction between the “twisted” and “untwisted” attachments are more subtle.Suppose again that the component of X where we attach the disc has the form $A = A' \setminus \text{disc} \setminus \text{disc}$ with the two indicated discs corresponding to the attaching regions. Choose a path Γ in A between points in the two attaching regions. [DIAGRAM 15]. Then we will say the twisted case is when the union of the attached strip and a strip about Γ in A form a Mobius band, and that the untwisted case is when the union forms an ordinary band. Then a little thought shows that the operation takes $A' \setminus \text{disc} \setminus \text{disc}$ to $A' \sharp T^2 \setminus \text{disc}$ in the untwisted case and to $A' \sharp K$ in the twisted case. However we have seen that K is equivalent to Ξ_2 so we can write the new surface as $A' \sharp \mathbf{R}P^2 \sharp \mathbf{R}P^2 \setminus \text{disc}$.

At this point we have the proof of a result, although not quite the one we want. Our agrument shows that any connected surface S is equivalent to a connected sum $\Sigma_g \sharp \Xi_h$.

This holds because the class of disjoint unions of surfaces with boundary of the form $\Sigma_g \sharp \Xi_h \setminus r$ discs is closed under the operations above.

To complete the proof of the classification Theorem stated we need

Lemma The surfaces $T^2 \sharp \mathbf{R}P^2$ and $\mathbf{R}P^2 \sharp \mathbf{R}P^2 \sharp \mathbf{R}P^2$ are equivalent.

Given this we see that if $h > 1$ the surface $\Sigma_g \sharp \Xi_h$ is equivalent to Ξ_{h+2g} so the the result we obtained above implies the stronger form.

To prove the Lemma it suffices to show that $K \sharp \mathbf{R}P^2$ is equivalent to $T^2 \sharp \mathbf{R}P^2$ since we know that K is equivalent to $\mathbf{R}P^2 \sharp \mathbf{R}P^2$. Now $\mathbf{R}P^2 \setminus \text{disc}$ and $T^2 \setminus \text{disc}$ are pictures together in [DIAGRAM 16] From this one can see easily that $(T^2 \sharp \mathbf{R}P^2) \setminus \text{disc}$ is pictured in [DIAGRAM 17].

Similarly a little thought shows that $K \setminus \text{disc}$ is as in [DIAGRAM 18]. (That is, it is similar to the torus case but with one strip twisted.) So $(K \sharp \mathbf{R}P^2) \setminus \text{disc}$ is pictured in [DIAGRAM 19].

We can deform this picture into the other by sliding handles around the boundary as shown in [DIAGRAMS 20, 21]. When we attach the disc this gives an equivalence between $K \sharp \mathbf{R}P^2$ into $T^2 \sharp \mathbf{R}P^2$ as desired

2.2 Discussion: the mapping class group

This topological classification of surfaces has been known for many years and, while our discussion above is completely informal, a fully rigorous proof is not really difficult by modern standards. From this one might be tempted to think that the subject of surface topology is a closed, fully understood, area. One might be further tempted to think that the analogous classification problem in higher dimensions—the topological classification of manifolds—should not be too much harder. However the second of these notions is certainly false and the first is false if one broadens the conception of surface topology slightly. Moreover these two issues are tightly connected as we will now explain.

Suppose one tries to implement the same strategy to classify 3-dimensional manifolds. Then it is not hard to show that any close 3-manifold can be built up from standard pieces in a similar fashion to what we have discussed above. More precisely, any closed 3-manifold has a *Heegard decomposition*. This is defined as follows. Take the standard picture of the surface Σ_g in $\mathbf{R}^3$ and let N_g be the 3-dimensional region enclosed by the surface. So N_g is a 3-manifold with boundary Σ_g. Now let N_g' be another copy of N_g with boundary Σ_g'; another copy of Σ_g. Let $\phi : \Sigma_g \to \Sigma_g'$ be a homeomorphism. Then we can obtain a 3-manifold M_ϕ by gluing N_g to N_g' along their boundaries by ϕ. More precisely we define M_ϕ to be the quotient of the disjoint union $N_g \cup N_g'$ by the equivalence relation which identifies $x \in \Sigma_g \subset N_g$ with $\phi(x) \in \Sigma_g' \subset N_g'$. Then a Heegard decomposition of a 3-manifold M is a homeomorphism $M \cong M_\phi$ for some ϕ, and , as we have said, any M arises in this way, determined up to equivalence by a ϕ. Of course if we fix some standard identification between Σ_g and Σ_g' as a reference then we can regard ϕ as a self-homeomorphism from Σ_g to itself.

Now the point is that the apparent simplicity of this description of 3-manifolds is illusory, because the set of self-homeomorphisms of a surface Σ_g is *enormously complicated* (at least once $g \geq 2$). These self-homeomorphisms obviously form a group and there is a natural notion of equivalence (isotopy) such that the set of equivalence classes of self-homeomorphisms modulo isotopy forms a countable discrete group Γ_g called the "mapping class group" of genus g. The complication which we refer to really resides in the complexity of this group. Looking back at the clasification of surfaces from this perspective we can see that what made the argument run so smoothly in that case is that the analogous group associated to the 1-dimensional manifold—the circle– is very simple. The group of self-homeomorphisms of the circle modulo isotopy has just two elements, realised by the identity and a reflection map. This means that when we talked about "attaching a disc to a circle boundary" say, the meaning was essentially unambiguous. (However there is an issue lurking here because we do have two distinct ways of attaching surfaces along circle boundaries, and we should really have kept track of this throughout our discussion above. In the end it turns out that this does not matter, because it happens that for any surface with boundary there is a self-homeomorphism of the surface inducing the non-trivial map on any given boundary component. This was much the same issue as that which the reader was invited to consider when attaching twisted or untwisted bands in Case 1 above.)

Expressing our main point in another way: the complexity of surface topology arises not from the relatively easy fact that any orientable surface S is equivalent to some Σ_g but from the fact that there is a vast set of essentially different choices of equivalence between Σ_g and S, any two differing by an element of the mapping class group.

To illustrate these remarks we introduce "Dehn twists". Let S be an orientabl surface and $C \subset S$ be an embedded circle. Since S is orientable there is a neighbourhood N of C which we can identify with a standard band or cylinder $S^1 \times [-1,1]$. We define a homeomorphism $\phi_0 : S^1 \times [-1,1] \to S^1 \times [-1,1]$ as follows. Regard S^1 as the unit circle in $\mathbf{C}$ and fix a function $f(t)$ on $[-1,1]$ which is equal to 0 for $t \leq -1/2$ say and to 2π for $t \geq 1/2$ say. Then set

$$\phi_0(e^{i\theta}, t) = (e^{i(\theta + f(t))}, t).$$

The choice of f means that ϕ_0 is the identity near the boundary of the cylinder. Now if we identify the cylinder with the neighboourhhod N in S we can regard ϕ_0 as a homeomorphism from N to N, equal to the identity near the boundary. Define $\phi : S \to S$ by

$$\phi(x) = x \text{ if } x \notin N, \phi(x) = \phi_0(x) \text{ if } x \in N.$$

The fact that ϕ_0 is the identity near the boundary means that ϕ is a homeomorphism from S to itself, the "Dehn twist" around C [DIAGRAM 22]. Of course the construction depends on various particular choices: the function f, the neighboorhood N and the identification of N with the cylinder, but up to isotopy the map ϕ is independent of these choices and we get a well-defined element of the mapping class group. We will see later that these Dehn twists, and the mapping class group generally, arise naturally in questions of complex analysis and geometry.

Part II

Basic Theory

Chapter 3

Basic definitions

3.1 Riemann surfaces and holomorphic maps

Definition 3 *A Riemann surface is given by the following data.*

- *A Hausdorff topological space X.*
- *A collection of open sets $U_\alpha \subset X$, where α ranges over some index set, which cover X (i.e. $X = \bigcup_\alpha U_\alpha$).*
- *For each α a homeomorphism*

$$\psi_\alpha : U_\alpha \to \tilde{U}_\alpha,$$

where $\tilde{U}_\alpha$ is an open set in $\mathbf{C}$ with the property that for all α, β the composite map $\psi_\alpha \circ \psi_\beta^{-1}$ is HOLOMORPHIC *on its domain of definition.*

The maps ψ_α are referred to as "charts", or "co-ordinate charts" or just "local co-ordinates", and the entire collection of data $(U_\alpha, \tilde{U}_\alpha, \psi_\alpha)$ is called an "atlas" of charts.

The reader who has never encountered this kind of notion before may find the definition hard to digest, so a few remarks are in order.

First, we define ψ_β^{-1} to be the obvious homeomorphism from $\tilde{U}_\beta$ to U_β so $\psi_\alpha \circ \psi_\beta^{-1}$ is well defined as a map

$$\psi_\alpha \circ \psi_\beta^{-1} : V_{\alpha,\beta} \to V_{\beta,\alpha}$$

where $V_{\alpha,\beta} = \psi_\beta(U_\alpha \cap U_\beta)$ and $V_{\beta,\alpha} = \psi_\alpha(U_\alpha \cap U_\beta)$. Since $V_{\alpha,\beta}$ and $V_{\beta,\alpha}$ are open sets in $\mathbf{C}$ the notion of a holomorphic map, as specified in the definition, makes sense. Notice that, interchanging α and β, it is a consequence of the definition that $\psi_\alpha \circ \psi_\beta^{-1}$ is a homeomorphism from $V_{\alpha,\beta}$ to $V_{\beta,\alpha}$ with a holomorphic inverse.

The second remark is that in practice, when working with Riemann surfaces, one rarely sees this bulky collection of data explicitly. Suppose we have a point p in X. This lies in at least one U_α, so we choose one. The map ψ_α is just a complex-valued function defined on a neighbourhhod of p, and we will normally denote this by a symbol such as z. Then in making calculations near to p we label points by the corresponding value of the variable z in $\mathbf{C}$, so we are effectively working in the traditional notation of complex analysis. On the other hand we might have chosen a different co-ordinate chart, ψ_β, which we might call w. Thus the map $\psi_\alpha \circ \psi_\beta^{-1}$ expresses, in more classical notation, z as a holomorphic function of w. The key feature of Riemann surface theory is that we have to study the behaviour of our calculations and constructions under such a holomorphic change of variable to obtain results which are independent of the choice of co-ordinate chart.

The third remark has more mathematical content. The main ideas embodied in the definition are not specific to the particular case at hand. If we take the same definition but replace the word *HOLOMORPHIC* by another appropriate condition (****) on maps between open sets in $\mathbf{C}$ then we get a definition of another kind of mathematical object. The main instances we need are

- Taking (****) to be *SMOOTH* we get the definition of a *smooth surface.* Here a smooth map between open sets in $\mathbf{C}$ is one which is differentiable infinitely often (C^∞) in the sense of two real variables, identifying $\mathbf{C}$ with $\mathbf{R}^2$.

- Taking (****) to be *SMOOTH WITH POSITIVE JACOBIAN* we get the definition of an *oriented smooth surface.* Here the Jacobian is, as usual, the determinant of 2×2 matrix of first derivatives of the map.

But there are many other interesting possibilities. For example, we could take (****) to be *SMOOTH WITH JACOBIAN* 1, to get the notion of a "surface with an area form". More generally, there is no need to restrict to two real dimensions. If we modify our definition to allow $\tilde{U}_\alpha$ to be open sets in $\mathbf{R}^n$ (for fixed n) and if we fix a suitable condition on maps between open sets in

$\mathbf{R}^n$ we get the definition of a corresponding class of *n-dimensional manifolds*. Taking smooth maps, we get the notion of a n-dimensional smooth manifold. So a smooth surface is the same as a 2-dimensional smooth manifold. Slightly more sophisticated, if $n = 2m$ and we identify $\mathbf{R}^{2m}$ with $\mathbf{C}^m$ we may consider the condition that a map between open sets in $\mathbf{C}^m$ be *holomorphic* in the sense of several complex variables. Then we get the notion of an m-dimensional *complex manifold*. So a Riemann surface is the same thing as a 1-dimensional complex manifold.

Exercise 1. Show that a Riemann surface is naturally an oriented smooth surface.

Exercise 2. Suppose $F = F(x, y, z)$ is a smooth function on $\mathbf{R}^3$ with $F(0, 0, 0) = 0$ and that the partial derivative $F_z = \frac{\partial F}{\partial z}$ does not vanish at $(0, 0, 0)$. Then one can show that there is a smooth function $f(x, y)$, defined on a disc $D = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < r^2\}$ and taking values in an interval $I = (-\epsilon, \epsilon) \subset \mathbf{R}$, such that the intersection of $F^{-1}(0)$ with the cylinder $D \times I$ is the graph of f. (This result is analogous to Theorem 1, and is another instance of the implicit function theorem.) Using this, prove the following result. Let F be a smooth function on $\mathbf{R}^3$ and $S = F^{-1}(0)$. Suppose that at each point of X at least one of the partial derivatives F_x, F_y, F_z does not vanish. Then X is naturally an oriented smooth surface.

Carrying on with the theory, we consider maps between Riemann surfaces.

Definition 4 *Let X be a Riemann surface with an atlas $(U_\alpha, \tilde{U}_\alpha, \psi_\alpha)$ and let Y be another Riemann surface with atlas $(V_i, \tilde{V}_i, \phi_i)$. A map $f : X \to Y$ is called holomorphic if for each α and i the composite $\phi_i \circ f \circ \psi_\alpha^{-1}$ is holomorphic on its domain of definition.*

Here $\phi_i \circ f \circ \psi_\alpha^{-1}$ is a map from $\psi_\alpha(U_\alpha \cap f^{-1}(V_i))$ to $\tilde{V}_i$; these are open sets in $\mathbf{C}$ so the condition of being holomorphic is the usual one of complex analysis.

Again, this definition has obvious variants which we will not spell out in detail. For example we get the definition of a smooth map between smooth surfaces, a smooth function on a smooth surface, a smooth map from $\mathbf{R}$ to a surface etc.

Now of course we say that two Riemann surfaces X and Y are *equivalent* if there is a holomorphic bijection $f : X \to Y$ with holomorphic inverse. We will often treat equivalent Riemann surfaces as identical. For example this allows us to remove the dependence of the definition of a Riemann surface on the particular choice of atlas. If we have one atlas $(U_\alpha, \tilde{U}_\alpha, \psi_\alpha)$ we can always concoct another one, for example by adding in some extra charts. Strictly, according to our definition, this gives a different Riemann surface, with same underlying set X. However the surfaces that arise are all equivalent, since the identity map gives a holomorphic equivalence between them.

3.2 Examples

3.2.1

First, any open set in $\mathbf{C}$ is naturally a Riemann surface. Familiar examples are the unit disc $D = \{z : |z| < 1\}$ and the upper half-plane $H = \{w \in \mathbf{C} : Im(w) > 0\}$. These Riemann surfaces are equivalent, via the well known map

$$z = \frac{w-i}{w+i}.$$

Next we consider the *Riemann sphere* S^2. As a set this is $\mathbf{C}$ with one extra point ∞. The topology is that of the "one point compactification": i.e. open sets in S^2 are either open sets in $\mathbf{C}$ or unions

$$\{\infty\} \cup (\mathbf{C} \setminus K)$$

where K is a compact subset of $\mathbf{C}$. Alternatively (as our notation has already suggested), we can define the topology by an obvious identification of $\mathbf{C} \cup \{\infty\}$ with the unit sphere in $\mathbf{R}^3$. We make S^2 into a Riemann surface with an atlas of two charts:

$$U_0 = \{z \in \mathbf{C} : |z| < 2\}\ ,\ U_1 = \{z \in \mathbf{C} : |z| > 1/2\} \cup \{\infty\}.$$

We take $\tilde{U}_0 = \tilde{U}_1 = U_0$ and we let $\psi_0 : U_0 \to \tilde{U}_0$ be the identity map. We define ψ_1 by $\psi_1(\infty) = 0$ and $\psi_1(z) = 1/z$ for $z \in \mathbf{C}, |z| > 1/2$. Then the maps $\psi_0 \circ \psi_1^{-1}$ and $\psi_1 \circ \psi_0^{-1}$ are each the map $z \mapsto 1/z$ from the annulus $\{z 1/2 < |z| < 2\}$ to itself, so these are both holomorphic and the condition of the definition is realised.

This example is perhaps confusing in its simplicity but we have spelled it out in detail to illustrate how the definition works. Notice that the Riemann sphere is an example of a *compact* Riemann surface.

3.2.2 Algebraic curves

. This is a much more extended example, in which we cover some important theory. We begin with "affine curves".

Let $P(z, w)$ be a polynomial in two complex variables. Define X, as a topological space, to be the set of zeros of P in $\mathbf{C}^2$,

$$X = \{(z, w) \in \mathbf{C}^2 : P(z, w) = 0\}.$$

Let us suppose that P has the following property. For each point (z_0, w_0) of X, at least one of the partial derivatives P_z, P_w does not vanish. Then we can make X into a Riemann surface in the following way. Suppose (z_0, w_0) is a point of X where P_w does not vanish. Then according to Theorem 1 we can find small discs D_1, D_2 and a holomorphic map $f : D_1 \to D_2$ such that $X \cap (D_1 \times D_2)$ is the graph of f—points of the form $(z, f(z)$. We make a co-ordinate chart with $U_\alpha = X \cap (D_1 \times D_2)$, with $\tilde{U}_\alpha = D_1$ and with ψ_α the restriction of the projection from $D_1 \times D_2$ to D_1. Symmetrically, if (z_1, w_1) is a point of X where P_z does not vanish we can find discs B_1, B_2 say and a holomorphic map $g : B_2 \to B_1$ describing $X \cap (B_1 \times B_2)$ as the set of points of the form $(g(w), w)$. Clearly then we can find a collection of charts, either of the first kind or the second kind, which cover X. We have to check that the "overlap maps" between the charts are holomorphic. Now between charts of the first kind the overlap map will be the identity map on a suitable intersection of discs in $\mathbf{C}$. Likewise for the two charts of the second kind. Between a chart of the first kind, say $U_\alpha, \tilde{U}_\alpha, \psi_\alpha$ as above, and a chart of the second kind, the overlap map will be given by the composite

$$z \mapsto (z, f(z)) \mapsto f(z),$$

i.e. by the holomorphic map f.

The preceding discussion is crucial in understanding the historical roots and the significance of the notion of a Riemann surface. Our Theorem 1 made precise the idea of an algebraic function, defined locally, and leads to the question of understanding how the different local pictures fit together. Now we can say, roughly speaking that this is encoded in the topology of the Riemann surface $X \subset \mathbf{C}$ which is described locally by the branches of the algebraic function.

The addition of the point at infinity turns the non-compact Riemann surface $\mathbf{C}$ into the compact Riemann sphere. We extend this idea to the algebraic curves considered above, defining *projective curves.*

Recall that complex projective space $\mathbf{CP}^n$ is the quotient of $\mathbf{C}^{n+1} \setminus \{0\}$ by the equivalence relation which identifies vectors $v, \lambda v$ in $\mathbf{C}^{n+1} \setminus \{0\}$ for any λ in $\mathbf{C} \setminus \{0\}$. A point in $\mathbf{CP}^n$ can be represented by homogeneous coordinates $[Z_0, \ldots, Z_n]$, with the understanding that $[\lambda Z_0, \ldots, \lambda Z_n]$ represents the same point.

Exercise Prove that $\mathbf{CP}^n$ is compact, in its natural topology.

Let U_0 be the subset of $\mathbf{CP}^n$ consisting of points with the co-ordinate $Z_0 \neq 0$. Since, in this case,

$$[Z_0, Z_1, \ldots, Z_n] = [1, Z_1/Z_0, \ldots, Z_n/Z_0],$$

we can identify U_0 with $\mathbf{C}^n$. That is, a point $(z_1, \ldots, z_n)$ in $\mathbf{C}^n$ is identified with the point $[1, z_1, \ldots, z_n]$ in $\mathbf{CP}^n$. The complement of U_0 in $\mathbf{CP}^n$ is a copy of $\mathbf{CP}^{n-1}$, the "hyperplane at infinity". For example when $n = 1$, $\mathbf{CP}^{n-1}$ is a single point so $\mathbf{CP}^1 = \mathbf{C} \cup \{\infty\}$ and $\mathbf{CP}^1$ can be canonically identified with Riemann sphere. For general n and any $i \leq n$ we can define a subset $U_i \subset \mathbf{CP}^n$ as the set of points where the co-ordinate Z_i does not vanish. Then

$$\mathbf{CP}^n = U_0 \cup U_1 \ldots \cup U_n$$

and each U_i is a copy of $\mathbf{C}^n$. When making calculations around a point in $\mathbf{CP}^n$ we can always choose a U_i containing that point, and then perform our calculations in $\mathbf{C}^n$. (In fact, $\mathbf{CP}^n$ is an n-dimensional complex manifold, with charts furnished by the U_i, although we will not make explicit use of this notion.)

Now let p be a *homogeneous polynomial* in the variables $Z_0, \ldots, Z_n$. This means that p can be written as

$$p(Z_0, \ldots, Z_n) = \sum_{i_0, \ldots, i_n} a_{i_0 \ldots i_n} Z_0^{i_0} \ldots Z_n^{i_n},$$

where for each term in the sum $i_0 + \ldots i_n = d$ for some fixed integer d, the *degree* of p. Equivalently

$$p(\lambda Z_0, \ldots, \lambda Z_n) = \lambda^d p(Z_0, \ldots Z_n).$$

Thus the equation $p(Z_0, \ldots, Z_n) = 0$ defines a subset of $\mathbf{CP}^n$ in the obvious way. For example, if p is the polynomial $p(Z_0, Z_n) = Z_0$ then this zero-set would just be the hyperplane at infinity considered above. (More generally, if p has degree 1 then the zero set is a copy of $\mathbf{CP}^{n-1}$ in $\mathbf{CP}^n$.) The upshot of this is that if we have a collection of homogeneous polynomials p_1, p_r (of any degrees $d_1, \ldots, d_r$) then we define a subset V of $\mathbf{CP}^n$ as the intersection of the zero sets of the p_i. Such a set in $\mathbf{CP}^n$ is called a *projective algebraic variety*, and their study is the field of projective algebraic geometry. Notice that V is compact, as a closed subset of the compact space $\mathbf{CP}^n$.

To make things more concrete, we will now suppose that $n = 2$ and consider a single homogeneous polynomial $p(Z_0, Z_1, Z_2)$ of degree d. We denote its zero set in $\mathbf{CP}^2$ by $\overline{X}$. Let P be the corresponding inhomogeneous polynomial in two variables z, w:

$$P(z, w) = p(1, z, w).$$

By definition, the intersection of $\overline{X}$ with $U_0 = \mathbf{C}^2$ is the zero set X of P that we considered before: an affine algebraic curve. Thus

$$\overline{X} = X \cup (\overline{X} \cap L_\infty),$$

where $L_\infty = \mathbf{CP}^2 \setminus U_0$ is the line at infinity. There is an exceptional case when P is the polynomial Z_0^d, in which case X is empty (since $P(z, w) = 1$) and $\overline{X} = L_\infty$, but otherwise $\overline{X} \cap L_\infty$ will be a finite set of points and $\overline{X}$ is a *compactification* of X obtained by adjoining this finite set.

Now suppose that the polynomial $P(z, w)$ obtained from p satisfies the condition of the previous subsection: that P_z, P_w do not both vanish at any point of X. Then we have made X into a Riemann surface. We can repeat the discussion, replacing Z_0 by Z_1 and Z_2. If the partial derivatives of the corresponding inhomogeneous polynomials satisfy the relevant non-vanishing conditions then we make $\overline{X} \cap U_1$ and $\overline{X} \cap U_2$ into Riemann surfaces. It is easy to check that the three Riemann surface structures are equivalent on their common regions of definition $\overline{X} \cap U_i \cap U_j$, and thus we make $\overline{X}$ into a *compact* Riemann surface.

Exercise. Show that if $p(Z_0, Z_1, Z_2)$ is a homogeneous polynomial of degree d then

$$\frac{\partial p}{\partial Z_0} + \frac{\partial p}{\partial Z_1} + \frac{\partial p}{\partial Z_2} = dp.$$

Use this to prove the following. If for each non-zero (Z_0, Z_1, Z_2) where $p(Z_0, Z_1, Z_2) = 0$ at least one of $\frac{\partial p}{\partial Z_i}$ does not vanish then $\overline{X}$ is a Riemann surface.

It is not hard to extend this discussion to Riemann surfaces obtained as algebraic varieties in $\mathbf{CP}^n$ for larger n, but we will not go through this here.

To see how this construction works, consider the polynomial

$$P(z, w) = z^3 - zw^2 + 10z^2 + 3w + 16.$$

This has real co-efficients so, to aid our geometric intuition, we can consider the corresponding *real agebraic curve* $X_{\mathbf{R}}$ in $\mathbf{R}^2$, which we can sketch. It takes some labour to work out an accurate picture of this, but one thing we can read off easily is the asymptotic behaviour. Informally, when z, w are large, the leading, cubic, terms in P should dominate the other terms, so we expect that the curve has asymptotes given by the zeros of $z^3 - zw^2$. Factorising this as

$$z^3 - zw^2 = z(z - w)(z + w),$$

we expect that the curve has asymptotic lines $z = 0, z = \pm w$, and this is indeed the case. Now consider the homogeneous polynomial of degree 3 corresponding to P:

$$p(Z_0, Z_1, Z_2) = Z_1^3 - Z_1 Z_2^2 + 10 Z_1^2 Z_0 + 3 Z_2 Z_0^2 + 16 Z_0^3.$$

This defines a subset $\overline{X}$ of $\mathbf{CP}^2$ as above and one can check that the condition of the Exercise above is satisfied, so $\overline{X}$ is a Riemann surface. Now $\overline{X}$ meets the line at infinity L_∞ at the points $[0, Z_1, Z_2]$ which satisfy $P(0, Z_1, Z_2) = 0$. But

$$p(0, Z_1, Z_2) = Z_1^3 - Z_1 Z_2^2,$$

So $\overline{X} \cap L_\infty$ consists of three points $Z_1 = 0, Z_1 = \pm Z_2$ which of course correspond exactly to the asymptotic lines we saw in the affine picture. (One can carry over the entire projective space construction to the case of real co-efficients, getting a real projective curve $\overline{X}_{\mathbf{R}} \subset \overline{X}$, and in this case all the points of $\overline{X} \cap L_\infty$ lie in $\overline{X}_{\mathbf{R}}$ and so were apparent as asymptotes of our sketch of $X_{\mathbf{R}}$.)

What we see from this example–and which of course holds more generally–is that the projective space construction gives us a systematic way to discuss the asymptotic phenomena of affine curves.

3.2.3 Quotients.

We begin with a very simple case. Consider $2\pi\mathbf{Z}$ as an subgroup of $\mathbf{C}$ under addition and form the quotient set $\mathbf{C}/2\pi\mathbf{Z}$. First, this has a standard quotient topology in which it is clearly homeomorphic to a cylinder $S^1 \times \mathbf{R}$. Second, we can make $\mathbf{C}/2\pi\mathbf{Z}$ into a Riemann surface in a very simple way. For each point z in $\mathbf{C}$ we consider the disc D_z centred on z and with radius $1/2$. Clearly if z_1, z_2 are two points in D_z and if

$$z_1 = z_2 + 2\pi n,$$

for $n \in \mathbf{Z}$, then we must have $n = 0$ and $z_1 = z_2$. (Since $1/2 < \pi$.) What this means is that the projection map $\pi : \mathbf{C} \to \mathbf{C}/2\pi Z$ maps D_z bijectively to the quotient space. We use this to construct a chart about $\pi(z) \in \mathbf{C}/2\pi Z$, taking $U = \pi(D_z)$, $\tilde{U} = D_z$ and ψ_z the local inverse of π. Then we cover $\mathbf{C}/2\pi\mathbf{Z}$ by some collection of charts of this form. The overlap maps between the charts will have the just have the shape

$$z \mapsto z + 2\pi n,$$

for appropriate $n \in \mathbf{Z}$, and these are certainly holomorphic. In fact the Riemann surface $\mathbf{C}/2\pi\mathbf{Z}$ that we construct this way is equivalent to $\mathbf{C} \setminus \{0\}$, with the equivalence induced by the map $z \mapsto e^{iz}$.

Now let Λ be a *lattice* in $\mathbf{C}$; that is a discrete additive subgroup. To be concrete we can consider the lattice

$$\Lambda = \mathbf{Z} \oplus \mathbf{Z}\lambda,$$

where λ is some fixed complex number with positive imaginary part. We can repeat the discussion above without essential change, all we need to do is to choose the radius r of D_z so that

$$2r < min_{n,m}|n + \lambda m|$$

where n, m run over the integers, not both zero. Since $|n + \lambda m| \geq Im(\lambda)$ if $m \neq 0$ and $|n + \lambda m| \geq 1$ if $m = 0$ and $n \neq 0$, it suffices to take

$$2r < min(1, Im(\lambda)).$$

In this way we see that $\mathbf{C}/\Lambda$ is a Riemann surface, clearly homeomorphic to the torus $S^1 \times S^1$ and in particular *compact*.

The construction of the Riemann surface structures in the examples above is rather trivial, but we have gone through it at some length because precisely the same ideas apply more generally. Suppose a group Γ acts on a Riemann surface X by holomorphic automorphisms. Suppose that the following condition (*) holds. Around each point p of X we can find an open neighbourhood N such that if $q_1, q_2 \in N$ and $g \in \Gamma$ with $\gamma(q_1) = q_2$ then we must have $\gamma = 1$ and $q_1 = q_2$. Then we can go through exactly the same construction to endow the the quotient set X/Γ with a Riemann surface structure. Notice that the condition considered above implies that Γ acts *freely* on X. Conversely in many situations the freeness of the action will imply that this property holds.

With this theory in place we can easily write down some examples which lead to very interesting Riemann surfaces, very important in Number Theory. Consider the upper half plane H. The holomorphic automorphisms of H are given by Mobius maps

$$z \mapsto \frac{az+b}{cz+d}.$$

where a, b, c, d are real and $ad - bc = 1$. In other words

$$Aut(H) = PSL(2, \mathbf{R}),$$

the quotient of the group $SL(2, R)$ of 2×2 real matrices of determinant 1 by the subgroup $\{\pm 1\}$. So if we have a suitable subgroup $\Gamma \subset PSL(2, \mathbf{R})$ we can construct a Riemann surface H/Γ. Fix an integer $a > 1$ and let $\tilde{\Gamma}_a$ be the set of matrices M with integer entries, with determinant 1 and with $M = \pm 1$ modulo a. Dividing by ± 1 we get a subgroup $\Gamma_a \subset PSL(2, \mathbf{R})$.

Exercise. Show that if $a > 4$ then Γ_a acts freely on H. Study the stabilisers of points in H for $a \leq 4$.

Exercise. Suppose that Γ is any subgroup of $PSL(2, \mathbf{Z}) \subset PSL(2, \mathbf{R})$ (obvious definition). Show that if Γ acts freely on H then the condition (*) above holds. Deduce that, for $a > 4$, H/Γ_a is a Riemann surface.

Chapter 4

Maps between Riemann surfaces

4.1 General properties

The foundation for this Chapter is provided by two simple Lemmas from complex analysis.

Lemma 1 *Let f be a holomorphic function on an open neighbourhood U of 0 in $\mathbf{C}$ with $f(0) = 0$. Suppose that the derivative $f'(0)$ does not vanish. Then there is another open neighbourhood $U' \subset U$ of 0 such that f is a homeomorphism from U' to its image $f(U') \subset \mathbf{C}$ and the inverse map is also holomorphic.*

The proof of this is very similar to that of Theorem 1. (The Lemma is an instance of the general "inverse function theorem" while Theorem 1 is an instance of the general "implicit function theorem".) Thus we choose a small disc D_ϵ about 0 and use the fact that the number of roots of $f(z) = w$ in D_ϵ is given by the contour integral

$$\int_{\partial D_\epsilon} \frac{f'(z)}{f(z) - w} dz,$$

provided there are no roots on the boundary ∂D_ϵ. The argument then runs parallel to that for Theorem 1 and we leave the details to the reader. (The Lemma can be also be deduced from a slightly more general form of Theorem 1, where we take the function of two variables $P(z, w) = f(z) - w$.)

Lemma 2 *Let f be a holomorphic function on an open neighbourhood U of 0 in $\mathbf{C}$ with $f(0) = 0$, but with f not identically zero. There is a unique integer $k \geq 1$ such that on some smaller neighborhood U' of 0 we can find a holomorphic function g with $g'(0) \neq 0$ and $f(z) = g(z)^k$ on U'.*

To see this we consider the power series expansion of f about 0 and let k be the order of the first non-zero term:

$$f(z) = a_k z^k + a_{k+1} z^{k+1} + \ldots, \quad a_k \neq 0.$$

Thus

$$f(z) = a_k z^k (1 + b_1 z + b_2 z^2 + \ldots),$$

where $b_i = a_{k+i}/a_k$. If z is sufficiently small there is a well-defined holomorphic function

$$h(z) = (1 + b_1 z + b_2 z^2 + \ldots)^{1/k},$$

(more precisely, we need $|\sum b_i z^i| < 1$). Then $f(z) = g(z)^k$ where

$$g(z) = a_k^{1/k} z h(z),$$

taking any choice of the root $a_k^{1/k}$. The derivative of g at 0 is $a_k^{1/k}$, hence nonzero, so we have established the existence asserted in the Lemma. Uniqueness of k is also clear. Note that $k = 1$ if and only if $f'(0) \neq 0$ and otherwise $k - 1$ is the multiplicity of the zero of f' at $z = 0$.

These simple Lemmas yield a complete local description of holomorphic maps between Riemann surfaces.

Proposition 3 *Let X and Y be connected Riemann surfaces and $F : X \to Y$ a non constant holomorphic map. For each point x in X there is a unique integer $k = k_x \geq 1$ such that we can find charts around x in X and $F(x)$in Y in which F is represented by the map $z \mapsto z^k$.*

To spell out in more detail the statement, we mean that there are is a chart $(U, \tilde{U}, \psi)$ about $x \in X$, with $\psi(x) = 0 \in \tilde{U} \subset \mathbf{C}$ and a chart $(V, \tilde{V}, \phi)$ about $F(x) \in Y$, with $\phi(F(x)) = 0 \in \tilde{V} \subset \mathbf{C}$ such that the composite $\phi \circ F \circ \psi^{-1}$ is equal to the map $z \mapsto z^k$ on its domain of definition.

To prove the proposition, we begin by choosing arbitrary charts about x and $F(x)$. In these charts, F is represented by a holomorphic function, which we denote by f. We apply Lemmalem:locmodel to write f as g^k. Then the derivative of g at 0 does not vanish so we can can apply Lemma 1 to see that, after restricting the domain of definition, g gives a holomorphic homeomorphism with holomorphic inverse. Thus we can change the chart about x by composing with g to get a new chart having the desired property. Again the uniqueness of $k = k_x$ is clear.

To get a straightforward global theory it is natural to impose some conditions on the holomorphic maps we wish to study. A good class to work with is that of *proper* holomorphic maps. Recall that a map $F : S \to T$ between topological spaces S, T is called proper if for any compact set $K \subset T$ the preimage $f^{-1}(K)$ is also compact. Note that if S itself is compact then any map F is proper, since $F^{-1}(K)$ is a closed subset of S, hence compact.

Recall also that subset Δ of topological space S is *discrete* if for any point $\delta \in \Delta$ there is an neighbouhood U of δ in S such that $\Delta \cap U = \{\delta\}$.

Proposition 4 *Let $F : X \to Y$ be a non-constant holomorphic map between connected Riemann surfaces.*

1. *Let $R \subset X$ be the set of points x where $k_x > 1$, then R is a discrete subset of X.*

2. *If F is proper then the image $\Delta = F(R)$ is discrete in Y.*

3. *If F is proper then for any y in Y the pre-image $f^{-1}(y)$ is a finite subset of X.*

The first item follows from the fact that, in local charts, R is given by the set of zeros of the derivative–using the standard fact that the zeros of a nonconstant holomorphic function are discrete. The other two items are straightforward exercises using the definition of properness.

Suppose, again, that $F : X \to Y$ is a proper, non constant holomorphic map between Riemann surfaces, with Y connected. For each $y \in Y$ we define an integer $d(y)$ by

$$d(y) = \sum_{x \in F^{-1}(y)} k_x.$$

The sum runs over a finite set, by item (3) of the previous Proposition. Notice that if $y \notin \Delta$ then $d(y)$ is just the number of points in $F^{-1}(y)$, and in general we will refer to $d(y)$ as the number of points in $f^{-1}(y)$ counted with multiplicity.

Proposition 5 *The integer $d(y)$ does not depend on $y \in Y$.*

First, observe that this is true in the special case when $X = Y = \mathbf{C}$ and F is the map $F(z) = z^n$ for some $n \geq 1$. The general case can be reduced to this using the local description, in Proposition prop:chartmod, of holomorphic maps. Fix $y \in Y$. We can find charts $U_x \subset X$ about each point $x \in f^{-1}(y)$ and a corresponding $V_x \subset Y$ about y, with respect to which F is expressed locally as $z \mapsto z^{k_x}$. Let V be the intersection of the V_x; this is an open neighbouhood of y in Y since there are only finitely many $x's$. Using the properness of F we can arrange that $F^{-1}(V)$ is contained in the union of the U_x's. Thus for another point $y' \in V$ we can study the set $F^{-1}(y')$ using the local models. It follows from the special case we began with that $d(y') = d(y)$. Thus $d(y)$ is locally constant on Y and hence consntant, since Y is connected.

The upshot of this is that we have defined an integer invariant, the *degree*, of a proper holomorphic map between connected Riemann surfaces. This is just the integer $d(y)$ for any y in the target space. (In the special case of a constant map we define the degree to be 0.) While we have defined this in a holomorphic setting it is in fact essentially a topological invariant: see the next Chapter.

While the proofs above are not difficult the results give us a striking corollary. First some terminology. A *meromorphic function* F on a Riemann surface is a holomorphic map to the Riemann sphere which is not identically equal to the ∞. In local charts this agrees with ordinary notion of a meromorphic function, having a Laurent series

$$\sum_{i=-n}^{\infty} a_i z^i.$$

The *poles* of F are just the points in $F^{-1}(\infty)$ and if x is a pole its *order* is the integer k_x.

Corollary 1 *Let X be a compact connected Riemann surface. If there is a meromorphic function on X having exactly one pole, and that pole has order 1, then X is equivalent to the Riemann sphere.*

Let $F : X \to S^2$ be the given meromorphic function. It is proper, since X is compact. The hypotheses imply that the degree of F is 1 (computing using $y = \infty$). This means that for any $y \in S^2$ there is exactly one point x in $f^{-1}(y)$ (and $k_x = 1$). Thus F is a bijection. The inverse map is continous (since the image under F of a closed set in X is compact in S^2 and hence closed), so F is a homeomorphism. It also follows from Lemmalem:invmap that the inverse map is holomorphic.

4.2 Monodromy and the Riemann Existence Theorem

4.2.1 Digression in algebraic topology

To take our study further we recall some algebraic topology. Let $F : P \to Q$ be a map between topological spaces.

Definition 5 *F is a local homeomorphism if around each point x in P there is an open neighbourhood U such that $F|_U$ is a homeomorphism to its image in Q*

Definition 6 *F is a covering map if around each point $y \in Q$ there is a open neighbourhood V such that $F^{-1}(V)$ is a disjoint union of open sets U_α in P and $F|_{U_\alpha}$ is a homeomoprhism from U_α to V.*

Clearly a covering map is a local homeomorphism but the converse is not true in general. (Exercise: give an example). However we have

Proposition 6 *A proper local homeomorphism is a covering map.*

In fact, a proper local homeomorphism is the same as a *finite* covering map, where the number of points in $f^{-1}(y)$ is finite for each $y \in Q$.

We need to recall the relation between these notions and the *fundamental group*. Let Q be a topological space and $q_0 \in Q$ a "base point". The fundamental group $\pi_1(Q, q_0)$ consists of homotopy classes of loops based at q_0. One often drops the base point from the notation. We give some examples which will be important for us.

- If Q is $\mathbf{C}$ or the disc D then $\pi(Q)$ is trivial.

- If Q is the punctured plane $\mathbf{C} \setminus \{0\}$ then $\pi_1(Q) = \mathbf{Z}$.
- If Q is a multiply punctured plane $\mathbf{C} \setminus \{z_1, \dots, z_n\}$ then $\pi_1(Q)$ is the free group on n generators.
- If Q is the torus T^2 then $\pi(Q) = \mathbf{Z} \times \mathbf{Z}$.
- If Q is the standard compact surface of genus g then $\pi_1(Q)$ is a group with $2g$ generators $a_1, b_1, \dots a_g, b_g$ and a single relation
$$[a_1, b_1][a_2, b_2] \dots [a_g, b_g] = 1$$
where $[a, b]$ denotes the commutator $aba^{-1}b^{-1}$.

The connection between the fundamental group and coverings is the following.

Proposition 7 *Let Q be a connected and locally simply connected space and q_0 a base point in Q. There is a one-to-one correspondence between:*

- *Equivalence classes of coverings $F : P \to Q$ where P is connected.*
- *Conjugacy classes of subgroups of $\pi_1(Q, q_0)$*

Some amplification. First, a space Q is "locally path connected" if any point has a simply connected neighbourhood–but we can ignore this since the property certainly holds for the Riemann surfaces we shall be concerned with. Second, coverings $F : P \to Q$ and $F' : P' \to Q$ are equivalent if there is a homeomorphism $g : P \to P'$ such that $F = F' \circ g$. Third, and most important, the correspondence is realised in the following way. Any map $F : P \to Q$ induces a homomorphism of fundamental groups $F_* : \pi_1(P, p_0) \to \pi_1(Q, F(p_0))$. The subgroup corresponding to a covering $F : P \to Q$ is the image of $F_*(\pi(P), p_0) \to \pi_1(Q, q_0)$ for any choice of $p_0 \in F^{-1}(q_0)$. Different choices of p_0 change the subgroup by conjugation.

To construct the covering corresponding to a subgroup of $\pi_1(Q)$ we can begin with the case of the trivial subgroup. The covering $G : \tilde{Q} \to Q$ which corresponds to this is characterised by the property that $\pi_1(\tilde{Q})$ is trivial. As a set, we define $\tilde{Q}$ to be the pairs (q, A) where q is a point in Q and A is a homotopy class of paths in Q from q_0 to q. Then we define $G(q, A) = q$. We refer to standard text books for the details of how to put a topology on

$\tilde{Q}$ such that G is a covering map. We also have an action of $\pi_1(Q, q_0)$ on $\tilde{T}$ given by concatenating a path with a loop based at q_0, and

$$Q = \tilde{Q}/\pi_1(Q, q_0).$$

Granted this, for any subgroup $H \subset \pi_1(Q, q_0)$ we define the associated covering space S_H to be $\tilde{Q}/H$. We then have a factorisation of the universal covering as

$$\tilde{T} \to S_H = \tilde{T}/H \to T = \tilde{T}/\pi_1(T, t_0).$$

The basic idea required in the proofs of the assertions above is that of *lifting* of paths, and homotopies of paths. Let $F : P \to Q$ be any map and $\gamma : [0, 1] \to Q$ a path. Given a pont $p_0 \in P$ with $F(p_0) = \gamma(0)$, a lift of γ starting at p_0 is just a path $\tilde{\gamma} : [0, 1] \to P$ with $F \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = p_0$.

Proposition 8 1. *If $F : P \to Q$ is a local homeomorphism then a path lift (with a given initial point) is unique, if it exists. If F is a covering map then path lifts (with a given initial point) always exist.*

2. *If $F : P \to Q$ is a local homeomorphism and γ_0, γ_1 are paths in Q with the same endpoints which are homotopic (rel. end points) through liftable paths, with lifts $\tilde{\gamma}_s$ having the same initial point in P. Then $\tilde{\gamma}_s(1) = \tilde{\gamma}_0(1)$ for all $s \in [0, 1]$.*

4.2.2 Monodromy of covering maps

Now let us return to a proper holomorphic map $F : X \to Y$ of connected Riemann surfaces, with degree $d \geq 1$. It follows immediately from Lemma * that the restriction of F to $X \setminus R$ is a local homeomorphism. This restriction need not be a proper map, but if we set $R^+ = F^{-1}(\Delta) = F^{-1}(F(R))$ then the restriction of F to $X \setminus R^+$ is proper, as one can easily check from the definition. So we have a covering map

$$F : X \setminus R^+ \to Y \setminus \Delta,$$

This is classified by a subgroup $H \subset \pi_1(Y \setminus \Delta)$ (or more precisely a conjugacy class of subgroups). There is another way to think about this algebro-topological data. It follows from the definitions that H has finite index in $\pi_1(Y \setminus \Delta)$, indeed the index is just the number of sheets of the cover which is the degree d. In general, the subgroups of index d in a group π correspond to transitive permutation representations

$$\rho : \pi \to \mathcal{S}_d,$$

where $\mathcal{S}_d$ denotes the symmetric group of permutations of $\{1, \dots, d\}$. (Here *transitive* means taht the image of ρ acts transitively on the set.) Thus in our situation our proper holomorphic map F yields a transitive representation $\rho : \pi_1(Y \setminus \Delta) \to \mathcal{S}_d$, determined up to conjugacy. This is the *monodromy* of the covering and we can give a more intuitive description of it as follows. Suppose we have a loop γ in $Y \setminus \Delta$ beginning and ending at y_0. We label the points in the $F^{-1}(y_0)$ by $1, \dots, d$. Now we move around the loop γ and "transport" the points, with their labelling, in $F^{-1}(\gamma(t))$ continously around in X to match. When we return to y_0 we recover the same set $F^{-1}(y_0)$ but the labelling may have changed. This change is given by a permutation in $\mathcal{S}_d$ which is $\rho([\gamma])$, where $[\gamma]$ denotes the homotopy class of the loop γ.

This point of view is close to that traditionally adopted when introducing Riemann surfaces: regarding them as formed from sheets over domains in $\mathbf{C}$ joined along "cuts". For a very simple example, consider the Riemann surface X defined by the equation $w^2 = f(z)$ where

$$f(z) = (z - z_1)(z - z_2) \dots (z - z_{2n}),$$

and with $F : X \to \mathbf{C}$ the projection to the z factor. Then $\Delta = \{z_1, \dots, z_{2n}\}$ and $\pi_1(\mathbf{C} \setminus \Delta)$ is generated by $2n$ loops $\gamma_1, \dots, \gamma_{2n}$ where γ_i is a standard loop going once around z_i. The degree d is 2 and the representation ρ maps each generator γ_i to the non-trivial element of $\mathcal{S}_2$(a transposition of the two objects). In traditional language we make cuts along n disjoint paths joining z_{2i-1} to z_{2i} for $i = 1, \dots, n$. Then we take two copies of the cut plane and form $X \setminus R$ by gluing these along the cuts. More generally we can express the procedure as saying that we make cuts so that ρ becomes trivial on π_1 of the cut plane, then ρ is just the combinatorial data required to specify the gluing along the cuts.

We summarise our work so far. Starting with a proper, nonconstant, holomorphic map between connected Riemann surfaces X, Y we get a degree d, a discrete set $\Delta \subset Y$ and a transitive permutation representation] (up to conjugacy) $\rho : \pi_1(Y \setminus \Delta) \to \mathcal{S}_d$. The next result, *Riemann's existence theorem*, shows that we can go in the other direction.

Theorem 2 *Let Y be a connected Riemann surface and Δ a discrete subset of Y. Given $d \geq 1$ and a transitive permutation representation $\rho : \pi_1(Y \setminus \Delta) \to \mathcal{S}_d$ there is a connected Riemann surface X and a proper holomorphic map $F : X \to Y$ which realises ρ as its monodromy homomorphism. Morover X and F are unique up to equivalence.*

First, the theory of covering spaces recalled above gives us a covering map $F_0 : X_0 \to Y \setminus \Delta$. It is easy to see that there is a unique way to make X_0 into a Riemann surface such that the map is holomorphic. At the end of the proof, the Riemann surface X_0 will correspond of course to $X \setminus R^+$, so what we need to see is how to "fill in " the points of R^+ lying over Δ. Let y_1 be a point of Δ and choose a small disc D_1 in Y about y_1, not containing any other points of Δ. The boundary of the disc D_1 defines an element of $\pi_1(Y \setminus \Delta)$ (or, more precisely a conjugacy class). The homomorphism ρ maps this to a permutation σ of $(1, \ldots, d)$. Now σ may not act transitively on $(1, \ldots, d)$. This coresponds to the fact that $F_0^{-1}(\Delta \setminus \{y_1\}$ may not be connected. If we write σ as a product of disjoint cycles then it is easy to see from the definitions that the cycles naturally correspond to the components of $F_0^{-1}(\Delta \setminus \{y_1\})$. Thus if Z is one such connected component, corresponding to a cycle of length d' the restriction of F_0 to Z gives a connected covering of $\Delta \setminus \{y_1\}$, determined by the homomorphism which maps the generator of $\pi_1(\Delta \setminus \{y_1\})$ to the d' cycle in $\mathcal{S}_{d'}$. But we know a covering which realises this data: if we identify $D_1 \setminus \{y_1\}$ with the standard punctured unit disc $D^* \subset \mathbf{C}$ it is given by the map $z \mapsto z^{d'}$ from D^* to D^*. So we conclude that Z is equivalent as a Riemann surface to D^* by an isomorphism $\phi : D^* \to Z$ say.

We now define a set X^+ by

$$X^+ = X_0 \cup_\phi D,$$

where D is the unit disc in $\mathbf{C}$ and the notation means that we identify $z \in D^* \subset D$ with $\phi(z) \in Z \subset X_0$. We make X^+ into a Riemann surface as follows. We take an atlas of charts in X^+ to be an atlas for X_0 with one further chart, the inverse of the obvious map from D to X^+ arising from the definition. There is then a unique way to introduce a topology on X^+ making all these charts homomorphisms, but we have to check that this topology is Hausdorff, i.e. that for any two points a, b in X^+ there are disjoint open sets U, V containing a, b respectively. If a and b lie in the copy of X_0 in X^+ this is clear: we just take corresponding open sets in X_0. So suppose a lies in X_0 and b is the point corresponding to 0 in D. Then $F_0(a)$ is not equal to y_1 in

Y so we can find a small neighbourhod N of $F_0(a)$ in $Y \setminus \Delta$ which is disjoint from a smaller disc $D_2 \subset D_1$ containing y_1. The open sets $F_0^{-1}(N)$ and

$$\{0\} \cup \phi^{-1}(F_0^{-1}(D_2)),$$

are disjoint in X^+ and contain b, a respectively. This shows that X^+ is indeed a Riemann surface. Moreover the map F_0 obviously extends to a holomorphic map from X^+ to Y. (The point of dealing carefully with the Hausdorff condition here is this. Suppose W is any Riemann surface and $\phi : D^* \to W$ is a holomorphic map. Then we can form the set $W \cup_\phi D^*$ as above and equip it with charts, but in general the resulting space may not be Hausdorff—it will be fail to be Hausdorff precisely when ϕ extends to a holomorphic map from D to W—that is when the "new " point we are trying to add in was already there!)

We repeat the procedure above for each point of Δ and for each cycle of the corresponding monodromy. This gives us a Riemann surface X with a holomorphic map F to Y and one checks that this map is indeed proper.

4.2.3 Compactifying algebraic curves

This construction has an important application to algebraic curves. Suppose $P(z, w)$ is a polynomial in two complex variables and consider again the set X of solutions to the equation $P(z, w) = 0$ in $\mathbf{C}^2$ with the projection map $\pi : X \to \mathbf{C}$ onto the z-factor. Suppose that P is an *irreducible* polynomial, i.e. it cannot be written as $P = QR$ for nonconstant polynomials Q, R. Let S be the set of points in X where both partial derivatives P_z, P_w vanish. We will show in Chapter 15 that the irreducibility of P implies that there are only finitely many points in S. The proof of * shows that the complement $X \setminus S$ is a Riemann surface. Now we let F be the finite subset of $\mathbf{C}$ defined by those values of z for which the term in P of highest degree in w (a polynomial in z) vanishes. Put

$$S^+ = \pi^{-1}(\pi(S) \cup F) \subset X.$$

The set S^+ is again finite, for there are only finitely many points in $\pi(S) \cup F$ and if z_0 is such a point the set $\pi^{-1}(z_0)$ consists of the roots of the polynomial equation in $P(z_0, w) = 0$ in the single variable w. This has only many finitely roots unless the polynomial vanishes identically, which would imply that

$(z - z_0)$ divides P. Now let E be the discrete subset $\pi(S) \cup F \cup \{\infty\}$ of the Riemann sphere S^2. We get a proper holomorphic map

$$\pi : X \setminus S^+ \to S^2 \setminus E.$$

Applying the theory above, we have a ramification set $\Delta \subset S^2 \setminus E$ and we can recover $X \setminus S^+$ from the monodromy homomorphism $\rho : \pi_1(S^2 \setminus (\Delta \cup E)) \to \mathcal{S}_d$. On the other hand this data also defines a *compact* Riemann surface X^*, containing $X \setminus S^+$ as a dense open set mapping holomorphically to S^2.

Now recall that on the other hand we have a compact set $\overline{X} \subset \mathbf{CP}^2$ defined by the homogeneous polynomial corresponding to P. This contains X and hence $X \setminus S^+$, again as dense open sets.

Proposition 9 *The inclusion of $X \setminus S^+$ in $\overline{X}$ extends to a holomorphic map from X^* to $\mathbf{CP}^2$, mapping onto $\overline{X}$.*

Here a holomorphic map from a Riemann surface to $\mathbf{CP}^2$ is defined in the obvious way as a continous map which is holomorphic with respect to the three charts $W_i \equiv \mathbf{C}^2$ covering $\mathbf{CP}^2$.

To prove the Lemma it suffices to work in the affine plane $\mathbf{C}^2$. What we need to show is that when we attach discs to $X \setminus S^+$ the inclusion of the punctured disc extends meromorphically over 0. Thus the proposition boils down to the following.

Lemma 3 *Suppose P is an irreducible polynomial in two variables and n is a positive integer. Suppose f is a holomorphic function on the punctured disc $D \setminus \{0\}$ with $P(z^n, f(z)) = 0$ for all $z \in D \setminus \{0\}$. Then f is a meromorphic function.*

The irreduciblity of P implies, as above, that there are only finitely many roots $w_1, \dots w_N$ say of the equation $P(0, w) = 0$. Thus when $|z|$ is small, $f(z)$ must be close to one of the w_i. We recall a result from complex analysis: if f has an essential singularity at 0 then for all w in $\mathbf{C}$ and $\epsilon, \delta > 0$ there is z with $0 < |z| < \delta$ and $|f(z) - w| < \epsilon$. This clearly does not hold in our case, when w is not one of the w_i and ϵ, δ are sufficiently small. Thus f is meromorphic as asserted.

Given the Lemma, we know that for suitable m, $z^m f(z)$ is holomorphic and non vanishing for small z. Then teh map $z \mapsto [z^n, f(z), 1]$ from the punctured disc to $\mathbf{CP}^2$ is equal to the map $z \mapsto [z^{n+m}, z^m f(z), z^m]$ and extends to a holomorphic map of the disc to $\mathbf{CP}^2$.

The conclusion is that we have associated a compact Riemann surface X^* to any irreducible polynomial. This is called the "normalisation" of the projective curve $\overline{X}$.

Examples

1. Suppose P is the polynomial $w^2 - z^2(1-z)$. Both partial derivatives vanish at the point $(0,0)$. To help our intuition we can sketch the corresponding real curve as below.

The origin is a singular point, where two branches of the curve cross. Following through the constructions one finds that X^* is equivalent to the Riemann sphere and the map from X^* to $\mathbf{CP}^2$ is given by $\tau \mapsto [1, 1-\tau^2, \tau - \tau^3]$. (The point at infinity in S^2 maps to $[0,0,1]$.) Thus we obtain X^* by separating the two braches of the curve passing through the origin.

2. Let P be the polynomial $w^2 - z^3$. The real curve looks like this:

It has a "cusp" singularity at the origin. Again the normalisation is the Riemann sphere with the map $\tau \mapsto [1, \tau^2, \tau^3]$.

4.2.4 The Riemann surface of a holomorphic function

Throughout this Chapter we have emphasised the case of "proper" maps, which most naturally arise when one considers algebraic functions. We will say a little more now about the general case. Suppose $F : X \to Y$ is a holomorphic map between connected Riemann surfaces, without branch points, and let $\Psi : X \to \mathbf{C}$ be any holomorphic function. Then F is a local homeomorphism so given a point $x_0 \in X$ we can define a holomorphic function $\psi_0 = \Psi \circ F_{x_0}^{-1}$ on a neighbourhood of $y_0 = F(x_0)$ in Y, where $F_{x_0}^{-1}$ denotes a local inverse mapping y_0 to x_0. More generally, given a path

$\gamma : [0,1] \to Y$ a lift $\tilde{\gamma}$ of γ starting at x_0 defines an analytic continuation of ψ_0 along γ (as defined in Chapter 1).

The converse to this construction is expressed in the following

Proposition 10 *Suppose given a point y_0 in a connected Riemann surface Y and a holomorphic function ψ_0 defined on a neighborhood of y_0. The there is a Riemann surface X, a holomorphic map $F : X \to Y$ without branch points mapping, a point x_0 in $F^{-1}(y_0)$, and a holomorphic function Ψ on X such that*

- *ψ_0 can be analytically continued along a path γ in Y if and only if γ has a lift to X starting at x_0.*
- *The analytic continuation of ψ_0 along γ has ψ_1 equal to $\Psi \circ F_{\tilde{\gamma}}^{-1}$ in a neighbourhood of $\gamma(1)$.*

The Riemann surface X is the "Riemann surface of the function germ ψ_0" and the data about all possible analytic continuations of ψ_0 is encoded in the holomorphic map $F : X \to Y$.

The construction of the Riemann surface X associated to ψ_0 is a variant of the construction of covering spaces. We define an equivalence realtion on analytic continuations of ψ_0 as follows. Let γ, γ' be paths starting at y_0. Recall that analytic continuations are given by one parameter families of holomorphic functions ψ_t, ψ'_t say. We say these are equivalent if $\gamma(1) = \gamma'(1)$ and $\psi_1 = \psi'(1)$ on some neighbourhood of this point in Y. Now we define X to be the set of equivalence classes of analytic continuations of ψ_0, with a map $F : X \to Y$ induced by $(\gamma_t, \psi_t) \mapsto \gamma(1)$. We define $\Psi : X \to \mathbf{C}$ to be the map induced by $(\gamma_t, \psi_t) \mapsto \psi_1(\gamma(1))$. Then one has to check that there is a natural (and in fact unique) way to introduce a Riemann surface structure on X such that F and Ψ are holomorphic and F is a local homeomorphism. We leave the interested reader to work out the details, or consult suitable texts.

This construction could be generalised. In one direction we can replace holomorphic functions by maps to a third Riemann surface Z. This makes essentially no difference. In the other direction we can sometimes extend X by including suitable ramification points, as we did in the case of proper maps.

Chapter 5

Calculus on surfaces

In this Chapter we develop the theory of *differential forms* on smooth surfaces and Riemann surfaces. This will be our main technical tool in the proofs of the major structural results in the following sections. Most of the material will be familiar to readers who have taken a standard course on manifold theory but for those who have not we give a fairly self-contained treatment.

5.1 Smooth surfaces

5.1.1 Cotangent spaces and 1-forms

Lemma 4 *Let f be a smooth, real-valued, function define on an open neighbourhood U of 0 in $\mathbf{R}^2$ and let $\gamma_1 : (-\epsilon_1, \epsilon_1) \to U$ and $\gamma_2 : (-\epsilon_2, \epsilon_2) \to U$ be smooth maps (for some $\epsilon_1, \epsilon_2 > 0$), with $\gamma_1(0) = \gamma_2(0) = 0$. Let $\chi : U \to V$ be a diffeomorphism to another open set in $\mathbf{R}^2$ with $\chi(0) = 0$. Set $\tilde{\gamma}_i = \chi \circ \gamma_i$ and $\tilde{f} = f \circ \chi^{-1}$.*

- *If both partial derivatives $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}$ vanish at 0 then the same is true of $\tilde{f}$.*

- *If the derivatives $\frac{d\gamma_i}{dt}$ at $0 \in \mathbf{R}$ are equal then the the same is true of $\tilde{\gamma}_i$.*

The assertions here follow immediately from the chain rule for partial derivatives. Now let S be a smooth surface, p be a point in S, f be a smooth function on S and $\gamma_i : (-\epsilon_i, \epsilon_i) \to S$ $(i = 1, 2)$ be a pair of smooth paths with $\gamma_i(0) = p$. We say that f is *constant to first order* at p if the derivative of the function representing f in a local co-ordinate chart about p vanishes at the point corresponding to p. By the first item of the Lemma this notion is independent of the choice of co-ordinate chart. Similarly, we say that γ_1, γ_2 are equal to the first order at p if the derivaties of the paths representing them in a local chart are equal.

Definition 7
- *The tangent space TS_p of S at p is the set of equivalence classes of maps $\gamma : (-\epsilon, \epsilon) \to S$ with $\gamma(0) = p$ under the equivalence relation $\gamma_1 \sim \gamma_2$ if γ_1 and γ_2 are equal to first order at p.*
- *The (real) cotangent space T^*S_p is set of equivalence classes of smooth functions on an open neighbourhood of p in S under the equivalence relation $f_1 \sim f_2$ if $f_1 - f_2$ is constant to first oder at p.*

The cotangent space T^*S_p has a natural vector space structure induced from that on the smooth functions on open neighboorhoods of p. From the definition, if U is an open neighbourhood of p there is a map from $C^\infty(U)$ to T^*S_p which we denote by

$$f \mapsto [df]_p.$$

Let x_1, x_2 be local co-ordinates about p. They are smooth functions so we have elements $[dx_1]_p, [dx_2]_p \in T^*S_p$. If f is any smooth function on a neighbourhood of p we write $f = f(x_1, x_2)$, making the usual notation supressing explicit dependence on the co-ordinate charts. Then the reader we readily verify, from the definition, that

$$[df]_p = \frac{\partial f}{\partial x_1}[dx_1]_p + \frac{\partial f}{\partial x_2}[dx_2]_p.$$

One sees from this that $[dx_1]_p, [dx_2]_p$ form a basis for the vector space T^*S_p. If $\gamma : (-\epsilon, \epsilon) \to S$ is a smooth path with $\gamma(0) = 0$ and f is a function on a neighboourhood of p the composite $f \circ \gamma$ is defined, as a real valued function, on some possibly smaller interval, and one checks that the derivative is independent of teh choice of f or γ in the equivalence classes defining the tangent space and cotangent space. Thus this derivative induces a map

$$TS_p \times T^*S_p \to \mathbf{R}.$$

Again we leave it to the reader to check that there is a unique vector space structure on TS_p with respect to which this is a bilinear map. In fact the map induces a duality between TS, T^*S— there is a canonical isomorphism

$$T^*S_p = Hom(TS_p, \mathbf{R}).$$

Now define the *cotangent bundle* T^*S to be set

$$T^*S = \bigcup_{p \in S} T^*S_p.$$

A smooth 1-form α on S is a map $\alpha : S \to T^*S$ with $\alpha(p) \in T^*S_p$ for all $p \in S$ and which varies smoothly with p in the following sense. In local co-ordinates (x_1, x_2) about a point p_0 we can write

$$\alpha = \alpha_1 dx_1 + \alpha_2 dx_2,$$

where α_1, α_2 are functions of x_1, x_2, and we have dropped the $[\ ]_p$ from our notation. Then we require that α_1, $alpha_2$ are smooth functions of the local co-ordinates. Again one needs to check that this is notion is independent of the co-ordinate system. We do this in detail because it gives an opportunity to illustrate how to compute with these forms. Suppose y_1, y_2 is another system of local co-ordinates, so x_1, x_2 are smooth functions of y_1, y_2 with partial derivatives $\frac{\partial x_i}{\partial y_j}$. Then by the chain rule, applying the definition, we have

$$dx_i = \frac{\partial x_i}{\partial y_1} dy_1 + \frac{\partial x_i}{\partial y_2} dy_2.$$

Thus if α is represented locally by $\alpha_1(x_1, x_2)dx_1 + \alpha_2(x_1, x_2)dx_2$ in the x_1, x_2 co-ordinates, it is represented by

$$\alpha_1(x_1(y_1, y_2), x_2(y_1, y_2)) \left(\frac{\partial x_1}{\partial y_1} dy_1 + \frac{\partial x_1}{\partial y_2} dy_2 \right) + \alpha_2(x_1(y_1, y_2), x_2(y_1, y_2) \left(\frac{\partial x_2}{\partial y_1} dy_1 + \frac{\partial x_2}{\partial y_2} dy_2 \right)$$

in the y_1, y_2 co-ordinates. The co-efficients of dy_1, dy_2 are obviously smooth functions of y_1, y_2, as required. An example of a smooth 1-form is furnished by the derivative of a function. If f is a function on S then we define a 1-form df by

$$df(p) = [df]_p.$$

In local co-ordinates this is just

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2.$$

There is another important notion, which leads to the same formula *. Suppose $F : S \to Q$ is a smooth map between surfaces. Then a moments thought about the definitions will show the redaer how to define, for any $p \in S$, natural linear maps

$$dF : TS_p \to TQ_{F(p)},$$

$$dF^*_p : T^*Q_{F(p)} \to T^*S_p,$$

compatible with the dual pairings between tangent and cotangent spaces. Suppose α is a smooth 1-form on Q. Then we define the pull-back form $F^*(\alpha)$ by

$$F^*(\alpha)(p) = dF^*(\alpha(F(p)).$$

Then $F^*(\alpha)$ is a smooth 1-form on S. If now x_1, x_2 are local co-ordinates about $F(p)$ in Q and y_1, y_2 are local co-ordinates about p in S the formula * gives the local representation of $F^*(\alpha)$, where F is locally represented by the functions $x_i(y_j)$.

We sum up the work so far. Write Ω^0_S for the smooth functions on S and Ω^1_S for the smooth 1-forms. Then we have defined

$$d : \Omega^0_S \to \Omega^1_S,$$

with the following properties

- $d(fg) = fdg + gdf$, where fg denotes the pointwise product;
- if $F : S \to Q$ is smooth then $d(F^* f) = F^*(df)$, where $f \in \Omega^0_Q$ and $F^*(f) = f \circ F$.

A difficulty with 1-forms is that they do not perhaps have a very obvious geometric meaning and it takes time, on first encountering the notion, to become comfortable working with them. (The dual notion, of a *vector field* is probably more intuitively appealing.) One important property is that 1-forms are the objects which can naturally be integrated over 1-dimensional sets. There are two slightly different notions here. One is to work with smooth paths

$$\gamma : [0, 1] \to S.$$

Suppose first that the image of γ lies inside some local co-ordinate system x_1, x_2. If α is a 1-form on S we define

$$\int_\gamma \alpha = \int_0^1 \alpha_1 \frac{d\gamma_1}{dt} + \alpha_2 \frac{d\gamma_2}{dt}.$$

Here $\gamma_1(t), \gamma_2(t)$ are the x_1 and x_2 co-ordinates of the local representation of γ and α_1, α_2 are the co-efficients of dx_1, dx_2 in the local represention of α. Thus in a less compressed notation in * we would write $\alpha_i(\gamma_1(t), \gamma_2(t))$. One checks that this is independent of the local co-ordinate system. Thus if the image of γ does not lie in a single chart one can define the integral by breaking up $[0,1]$ into subintervals and proceeding in the obvious way. The integral has another invariance property. Suppose $\psi : [0,1] \to [0,1]$ mapping 0 to 0 and 1 to 1. Then we get another smooth path $\gamma \circ \psi$ and

$$\int_{\gamma\circ\psi} \alpha = \int_\gamma \alpha.$$

Essentially this expresses the fact that the integral depends only on the image of γ. (A more sophisticated way to fomrulate these definitions is to introduce the notion of a 1-form on an interval in $\mathbf{R}$—or indeed on any smooth manifold—and the integral of a 1-form over an interval. Then the map γ gives a pull-back form $\gamma^*(\alpha)$ which we integrate over $[0,1]$.)

Now suppose that C is an *oriented embedded curve* in a surface S. Then we can define the integral of α over C by decomposing C into pieces which can be parametrised by smooth paths as above. All of this is essentially the same as the definition of contour integrals in elementary complex analysis, so we will not dwell on the details.

5.1.2 2-forms and integration

Next we want to define smooth 2-forms on a surface. To motivate the definitions here we consider the following question. Given a 1-form α on a surface S when can it be written as $\alpha = df$ for some function f? Consider first the case when S is $\mathbf{R}^2$, so

$$\alpha = \alpha_1 dx_1 + \alpha_2 dx_2,$$

where α_1, α_2 are arbitrary smooth functions of x_1, x_2. Writing $\alpha = df$ means finding a function f with

$$\frac{\partial f}{\partial x_1} = \alpha_1 \ , \ \frac{\partial f}{\partial x_2} = \alpha_2.$$

The symmetry of second partial derivatives means that an obvious necessary condition is that the function

$$R = \frac{\partial \alpha_1}{\partial x_2} - \frac{\partial \alpha_2}{\partial x_1}$$

vanishes everywhere. The converse is a classical result ("criterion for an exact differential") which the reader has very likely encountered: if $R = 0$ then we can find an f. Let us recall the proof. Given α_1, α_2 we define a functions f_1, f_2 by

$$f_2(x_1, x_2) = \int_0^{x_1} \alpha_1(t, 0)dt + \int_0^{x_2} \alpha_2(x_1, t)dt;$$
$$f_1(x_1, x_2) = \int_0^{x_2} \alpha_2(0, t)dt + \int_0^{x_1} \alpha_1(t, x_2)dt.$$

By construction $\frac{\partial f_i}{\partial x_i} = \alpha_i$. But *Stokes' Theorem*, applied to a rectangle V with vertices at $(0, 0), (x_1, 0), (0, x_2), (x_1, x_2)$ shows that

$$f_1(x_1, x_2) - f_2(x_1, x_2) = \int_V R,$$

so our hypothesis that $R = 0$ shows that $f_1 = f_2$ and the proof is complete. What we see from this argument in the plane is that the three notions

- The criterion for an exact differential,
- Integration over 2-dimensional regions,
- Stokes Theorem

are tightly bound together, and the definition of a 2-form is framed to allow us to extend these notions to surfaces.

Let E be a real vector space. We define $\Lambda^2 E^*$ to be the set of bilinear maps

$$B : E \times E \to \mathbf{R}.$$

which are skew-symmetric, $B(e, f) = -B(f, e)$. We define a "wedge product"

$$\wedge : E^* \times E^* \to \Lambda^2 E^*,$$

by

$$(\alpha \wedge \beta)(e, f) = \alpha(e)\beta(f) - \beta(e)\alpha(f).$$

So the wedge product is linear in each variable and $\alpha \wedge \beta = -\beta \wedge \alpha$. Now suppose that E has dimension 2 then (exercise for reader) $\Lambda^2 E^*$ is a 1-dimensional real vector space and if α_1, α_2 is a basis for E^* the wedge product $\alpha_1 \wedge \alpha_2$ furnishes a basis element in $\Lambda^2 E^*$.

We apply this algebra to the case when $E = TS_p$, so $E^* = T^*S_p$. Thus for each point p in a surface S we have a 1-dimensional space $\Lambda^2 T^* S_p$. If

x_1, x_2 are local coordinates around p we get a basis element $dx_1 \wedge dx_2$ for $\Lambda^2 T^* S_p$. One often omits the wedge product symbol to write this as $dx_1 dx_2$. We now proceed as before. We define a smooth 2-form ρ on S to be a map from S to the union

$$\bigcup_{p \in S} \Lambda^2 T^* S_p$$

such that $\rho(p)$ lies in $\Lambda^2 T^* S_p$ and varies smootjly with p in the following sense. In local co-ordinates we can write

$$\rho = R(x_1, x_2) dx_1 dx_2,$$

and we require that R be a smooth function. Applying the definitions one finds that in a different system of co-ordinates y_1, y_2 this same 2-form is represented by

$$R(x_1(y_1, y_2), x_2(y_1, y_2)) J(y_1, y_2) dy_1 dy_2),$$

where

$$J(y_1, y_2) = \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1}.$$

The reader will recognise this as the usual *Jacobian*: the determinant of the matrix of partial derivatives $\frac{\partial y_i}{\partial x_j}$. Again, this formula can be read in a different way. If $F : S \to Q$ is a smooth map and ρ is a 2-form on Q there is a natural way to define a pulled back form $F^*(\rho)$ on S and the formula * expresses this in local co-ordinates.

We write Ω^2_S for the set of 2-forms on a surface S.

Now these 2-forms provide a natural "home" for the expression appearing in the critrion for an exact differential above. We have

Lemma 5 *There is a unique way to define an* $\mathbf{R}$*-linear map*

$$d : \Omega^1_S \to \Omega^2_S$$

such that

- *If* $\alpha_1 = \alpha_2$ *on an open set* $U \subset S$ *then* $d\alpha_1 = d\alpha_2$ *over* U.
- *If* f *is a function on* S *then* $ddf = 0$.
- *If* f *is a function on* S *and* α *is a* 1*-form on* S *then*

$$d(f\alpha) = df \wedge \alpha + f d\alpha.$$

To prove this we first check the uniqueness. Suppose that we have an operator satisfying the conditions of the Lemma. By the first condition we can calculate $d\alpha$ in local co-ordinates. Then we have

$$d(\alpha_1 dx_1 + \alpha_2 dx_2) = d\alpha_1 \wedge dx_1 + d\alpha_2 \wedge dx_2,$$

using the second and third conditions (the second condition gives $ddx_i = 0$). Explicitly

$$d(\alpha_1 dx_1 + \alpha_2 dx_2) = \left(\frac{\partial \alpha_2}{\partial x_1} - \frac{\partial \alpha_1}{\partial x_2} \right) dx_1 dx_2.$$

On the other hand, if we take this formula as the definition of $d\alpha$ one can check that it is independent of the choice of co-ordinate system.

What we see in the course of the proof is that in local co-ordinates $d\alpha$ is just $R dx_1 dx_2$ where R is the function, discussed above, which enters into the criterion for an exact differential. So we can reformulate that result as saying that for a 1-form α on the surface $S = \mathbf{R}^2$ we can find a function with $\alpha = df$ if and only if $d\alpha = 0$.

We now turn to integration. Suppose S is an *oriented* surface and ρ is a 2-form with compact support and supported in the domain of a co-ordinate chart on S. Write $\rho = R(x_1, x_2) dx_1 dx_2$ in these local co-ordinates. Then we define the integral of ρ on S by the following, apparently tautological, formula

$$\int_S \rho = \int_{\mathbf{R}^2} R(x_1, x_2) dx_1 dx_2,$$

where on the right hand side we mean the ordinary Lebesgue integral on the compactly-supported functions on $\mathbf{R}^2$. If y_1, y_2 is another oriented chart then the Jacobian J relating the two is poistive by definition and the fact that we get the same value of the integral just expresses the usual transformation law for multiple integrals. To define the integral more generally we use the following lemma, which we will also need later.

Lemma 6 *Let K be a compact subset of a surface S and let $U_1, \dots, U_n$ be open sets in S with $K \subset U_1 \cup \dots U_n$. Then there are smooth, non-negative, functions $f_1, \dots, f_n$ on S, each of compact support and with the support of f_i contained in U_i, such that $f_1 + \dots + f_n = 1$ on K.*

To prove this we begin with the case when $n = 1$. First consider the very special case when $n = 1$, $S = U_1$ is the unit disc in $\mathbf{R}^2$ and K is the closed disc of radius 1/2. Then we take a non-negative function $f(x_1, x_2) = F(\sqrt{x_1^2 + x_2^2})$ where F is a function of one variable with $F(r) = 1$ for $r \leq 1/2$ and $F(r) = 0$ for $r \geq 3/4$, say.

Now consider the general case when $n = 1$. For each point $p \in K$ we take a local co-ordinate chart mapping a disc D_p about p to the open unit disc in $\mathbf{R}^2$ and the closure $\overline{D_p}$ to the closed unit disc in $\mathbf{R}^2$. Let $\frac{1}{2}D_p$ be the preimage of the half-sized open disc. We can suppose (scaling the chart) that the closure of D_p is contained in U_1. The set of open discs $\frac{1}{2}D_p$ as p ranges over K forms an open cover of K, so we can find a finite subcover, corresponding to points $p_1, \ldots, p_N$ say. Then for each $j \leq N$ we have a function, g_j say, of compact support on D_{p_j} and equal to 1 on the closure of $\frac{1}{2}D_{p_j}$ using the very special case above. We extend g_j by zero to regard it as a function on S. Now $g = \sum g_j$ has the following properties

- $g \geq 1$ on K, since each point of K lies in at least one disc $\frac{1}{2}D_{p_j}$ on which $g_j = 1$;

- g has compact support contained in U_1, since the support of g is the (finite) union of the supports of the g_i which are contained in the compact discs $\overline{\frac{1}{2}D_{p_j}} \subset U_1$.

Now take a smooth non-negative function H of one variable with $H(t) = 1$ if $t \geq 1$ and $H(t) = 0$ if $t \leq 1/2$. Then $f_1 = H \circ g$ has the desired property. (i.e. $f_1 = 1$ on K and the support of f_1 is a compact subset of U_1.

Finally we consider the general case, when $K \subset U_1 \cup \ldots \cup U_n$. Proceeding just as before we get discs $\frac{1}{2}U_{p_i} \subset U_{p_i}$ for $i = 1, \ldots N$ where

- $K \subset \frac{1}{2}U_{p_1} \cup \ldots \frac{1}{2}U_{p_N}$
- for each j there is an $i(j)$ such that the closed (compact) disc $\overline{D_{p_j}}$ is contained in $U_{i(j)}$.

Now for $i = 1, \ldots n$, let

$$K_i = \bigcup_{i(j)=i} \overline{\frac{1}{2}D_{p_j}}.$$

$$N_i = \bigcup_{i(j)=i} D_{p_j}$$

and

$$J_i = \bigcup_{i(j)=i} \overline{D_{p_j}}.$$

Thus the K_i and J_i are compact, N_i is open, we have

$$K_i \subset N_i \subset J_i \subset U_i$$

and $K \subset \bigcup_i K_i$. Applying the discussion above to each $J_i \subset U_i$ we find smooth functions h_i on S with $h_i = 1$ on J_i and with h_i compactly supported in U_i. Thus if $h = \sum_{i=1}^n h_i$ we have $h \geq 1$ on $J_1 \cup \ldots \cup J_n$. Let N be the open subset of S

$$N = N_1 \cup \ldots \cup N_n.$$

Applying the previous case again, we can find a function A of compact support in N and with $A = 1$ on K. Thus $h \geq 1$ on the support of A so the ratio A/h extends to a smooth function on S. Finally put

$$f_i = \frac{Ah_i}{h}.$$

Then f_i has compact support in U_i and $\sum f_i = 1$ on K since $A = 1$ there.

Given this Lemma and any 2-form ρ of compact support on an oriented surface S we proceed as follows. We cover $K = supp(\rho)$ by open sets U_i, each the domain of a local coordinate chart. By the compactness of K we may can do this with a finite collection of sets U_i. Then let f_i be a system of functions as in the Lemma. For each i the support of $f_i\rho$ is contained in a co-ordinate chart and we can define the integral of $f_i\rho$ as above. Now we define

$$\int_S \rho = \sum_i \int_S f_i\rho.$$

Of course this formula *must* hold true if we are to have an integral with the obvious linearity properties since $\sum f_i = 1$ on $supp(\rho)$ implies that

$$\rho = \sum f_i\rho.$$

Conversely one readily shows that the linearity of the Lebesgue integral implies that that this value of the integral of ρ is independent of the choice of the functions f_i.

While one must be careful to distinguish between 2-forms and functions, on an oriented surface there is a well-defined notion of a *positive* 2-form, just one which is given in local co-ordinates by $R dx_1 dx_2$ with $R \geq 0$. By the definition of an oriented surface and the transformation law for 2-forms, this is independent of the choice of co-ordinate system. If ρ is a positive 2-form of compact support then the integral of ρ is positive. With the usual conventions we can define the integral of any positive 2-form ρ, not necessarily of compact support, taking values in the extended real numbers $\mathbf{R} \cup \{+\infty\}$, by

$$\int_X \rho = \sup \int_X \chi\rho,$$

where χ runs over the smooth, compactly supported, functions on X with $0 \leq \chi \leq 1$ everywhere.

Notice that it is not really necessary that the 2-forms we integrate are smooth. We can define the notion of a *continuous* 2-form in an obvious way and the discussion above applies equally well. If ρ is any 2-form we can define a continuous, positive 2-form $|\rho|$ by the requirement that at each point $|\rho| = \pm\rho$. Thus for any 2-form ρ whatsoever on a surface S we can define the integral

$$\int_S |\rho|$$

taking values in $[0, \infty]$. This notion will be a convenience later. By an *area form* on an oriented surface S we mean a strictly positive 2-form. If we have a fixed area form then we can identify the 2-forms on S with functions, and the notion above becomes the usual notion of integration of functions with respect to a measure. However there atre certain important reasons why we do *not* want to assume that our surface support such area forms (see the discussion in Chapter 10 below).

The final item is the general form of Stokes' Theorem.

Proposition 11 *If ρ is a compactly supported 1-form on an oriented surface with boundary S then*

$$\int_{\partial S} \alpha = \int_S d\alpha.$$

To sum up we now have on any surface:

- Spaces of $0, 1, 2$ forms and the *exterior derivative*

$$\Omega^0_S \to \Omega^1_S \to \Omega^2_S.$$

- The integral $\int_C \alpha$ of a 1-form S over a curve a curve C in a surface S the

- The wedge product $\Omega^1_S \times \Omega^1_S \to \Omega^2_S$.
- If S is oriented, the integral $\int_S \rho$ of a compactly supported 2-form ρ.
- Stokes' Theorem, as above.

5.2 de Rham cohomology

5.2.1 Definition and examples

Let S be a smooth surface. We define the de Rham cohomology groups $H^i(S)$, for $i = 0, 1, 2$, to be the cohomology of the sequence

$$\Omega^0 \to \Omega^1 \to \Omega^2.$$

That is

$$H^0(S) = ker(d : \Omega^0_S \to \Omega^1_S,$$
$$H^1(S) = ker(d : \Omega^1_S \to \Omega^2_S)/Im(d : \Omega^0_S \to \Omega^1_S),$$
$$H^2(S) = \Omega^2_S/Im(d : \Omega^1_S \to \Omega^2_S).$$

Clearly $H^0(S) = \mathbf{R}$, the constant functions, if S is connected. The classical criterion for an exact differential discussed above amounts to the statement that if S is diffeomorphic to $\mathbf{R}^2$ then $H^1(S) = 0$. It is also clear that, for such S, the cohomology group $H^2(S)$ also vanishes, since any function $R(x_1, x_2)$ can be written as $\frac{\partial \alpha_1}{\partial x_2} - \frac{\partial \alpha_2}{\partial x_1}$ for some α_1, α_2. (In fact we can take $\alpha_2 = 0$ and

$$\alpha_1(x_1, x_2) = \int_0^{x_1} R(t, x_2)dt.)$$

Examples

1. Consider the 2-sphere S^2. We write this as the union of two open sets $S^2 = U \cup V$ where U and V are slightly enlarged upper and lower hemispheres, intersecting in a annulus around the equator. So U and v are each diffeomorphic to $\mathbf{R}^2$. Let α be a 1-form on S^2 with $d\alpha = 0$. Then by the previous discussion we can find functions f_U, F_V on U, V respectively such that df_U, df_V are the restrictions of α to U, V. Thus $d(f_U - f_V) = 0$ on $U \cap V$ and, since $U \cap V$ is connected, this means that $f_U - f_V$ is constant on the intersection, say $f_U - f_V = c$. There is no loss in supposing that this

constant is zero, since we can change f_U to $f_U - c$ without changing df_U. But if $f_U = f_V$ on $U \cap V$ they arise as the restrictions of a function f on S^2 to U and V, and $\alpha = df$. Hence $H^1(S^2) = 0$.

2. Consider the torus T and take standard angular co-ordinates $\theta, \phi \in [0, 2\pi)$. Let $\gamma_1, \gamma_2 \subset T$ be the standard embedded circles corresponding to $\theta = 0, \phi = 0$ respectively. Then the map

$$\alpha \mapsto (\int_{\gamma_1} \alpha, \int_{\gamma_2} \alpha)$$

induces a linear map from $H^1(T)$ to $\mathbf{R}^2$ since the integral of df around the γ_i vanishes for any function f on T. The forms $d\theta$ and $d\phi$ show that this map is surjective. We claim that the map is also injective, so $H^1(T) = \mathbf{R}^2$. For if $\alpha = Pd\theta + Qd\phi$ is a closed 1-form with integral 0 around γ_2 then for any fixed ϕ we have, by Stokes' Theorem,

$$\int_0^{2\pi} P(\theta, \phi) d\theta = 0.$$

This means that the indefinite integral

$$f(\theta, \phi) = \int_0^\theta P(u, \phi) du,$$

defines a smooth function on T with $\frac{\partial f}{\partial \theta} = P$. Thus $\tilde{\alpha} = \alpha - df$ is a closed 1-form of the form $\tilde{Q} d\phi$. But the closed condition impllies that Q is constant and if the integral around γ_1 is zero this constant be zero and $\alpha = df$.

3. Consider the cylinder $C = (-1, 1) \times S^1$ and let δ be the circle $\{0\} \times S^1$. As before, δ defines a linear map from $H^1(C)$ to $\mathbf{R}$ which is clearly surjective. To see that this map is also injective we proceed as in the first example, writing $c = U \cup V$ with open sets U, V diffeomorphic to $\mathbf{R}^2$. This time,$U \cap V$ has two components. If α is a closed 1-form on C then we obtain functions f_U, f_V just as in the first example, but we cannot deduce that $f_U - f_V$ is constant, although it is constant on each component of $U \cap V$. Pick points p, q on δ in the two components of $U \cap V$. Then a moments thought shows that

$$(f_U - f_V)(p) - (f_U - f_V)(q) = \pm \int_\delta \alpha.$$

So if the integral of α around δ is zero then $f_U - f_V$ is constant and we can proceed as in Example 1.

3. Consider the standard surface Σ of genus 2 formed from the connected sum of two copies T, T' of the torus. Thus in T we have a pair of standard loops γ_1, γ_2 as above, and likewise γ_1', γ_2' in T'. We form the conected sum by removing two open discs D, D' from T, T' and adding a copy of the cylinder C. We can suppose that the discs do not meet the circles γ_i, γ_i'. Then in an obvious way we get 4 circles, which we also denote by γ_i, γ_i' in Σ. Integration around these circles defines a linear map from $H^1(\Sigma)$ to $\mathbf{R}^4$ and we claim that this is an isomorphism.

To see that the map is injective we argue as follows. Suppose α is a closed 1-form on Σ and the integral of α around the four loops vanishes. Let δ be the loop in the cylinder $C \subset \Sigma$ as in Example 3. Then by Stokes' Theorem, the integral of α around δ vanishes, since δ is obviously the boundary of a portion of Σ. By Example 3 we can write the restriction of α to C as dg for some function g on C. Let P be a smooth function supported in C and equal to 1 on δ (see the Lemma below). Then $\tilde{\alpha} = \alpha - d(Pg)$ is in the same class in $H^1(\Sigma)$ as α and $\tilde{\alpha}$ vanishes on a neighbourhood of C. This means that $\tilde{\alpha}$ defines 1-forms β, β' on T, T' respectively, in an obvious way, vanishing near the centres of the discs D, D'. The integrals of β, β' around the circles in T, T' are still zero so we know that $\beta = df, \beta' = df'$ for some functions f, f' on T, T' with f, f' constant near the centres of the discs. Then the same argument as before shows that we cannot suppose f, f' match up to give a function f on Σ with $df = \tilde{\alpha}$.

To see that the map is surjective we argue as follows. Given any four real numbers we can find closed 1-forms β, β' on T, T' respectively realising these as their integrals around teh circles. By applying our result for $\mathbf{R}^2$ we can write the restrictions of β, $beta'$ to neighbourhoods of $\overline{D}, \overline{D'}$ as dg, dg' for functiosn g, g' defined on these neighboorhoods. Arguing as in the previous paragraph, we find $\tilde{\beta}, \tilde{\beta}'$ in the same cohomology classes as β, β' and vanishing over these discs. Then β, β' define a 1-form α on Σ in an obviuous way, having the given integrals around the γ_i, γ_i'.

5. It should now be clear how to show that for the standard closed oriented surface Σ_g of genus g—the connected sum of g copies of the torus– we have

$$H^1(\Sigma_g) = \mathbf{R}^{2g},$$

the isomorphism being realised by integration over $2g$ standard circles in Σ_g.

6. We now take a more abstract point of view. Let S be a connected

smooth surface and s_0 a base point in S. One can show that elements of $\pi_1(S, s_0)$ can be represented by smooth loops $\gamma : [0, 1] \to S$. For such loops the integral around γ is defined and induces a map

$$\int_\gamma : H^1(S) \to \mathbf{R}.$$

This integral depends only on the homotopy class of the loop and is additive with respect to the product in π_1 so we get a linear map

$$H^1(S) \to Hom(\pi_1(S, s_0), bR).$$

This map is an isomorphism, at least if S satsisfies the technical condition of being "paracompact".

5.2.2 Cohomology with compact support and Poincaré duality

There is a variant of the definition of cohomology in which we consider the forms $\Omega_S^{i,c}$ of *compact support*. Then we define cohomology groups $H_S^{i,c}$ in just the same fashion. Thus, for example, if S is connected but not compact then $H_S^{0,c} = 0$, since the non-zero constants do not have compact support. If S is oriented, the map from $\Omega_S^{2,c}$ to $\mathbf{R}$ defined by integration over S induces a linear map

$$\int_S : H_S^{2,c} \to \mathbf{R},$$

since the integral of $d\alpha$, for $\alpha \in \Omega_S^{1,c}$, vanishes by Stokes' Theorem.

Proposition 12 *If S is a connected, oriented, smooth surface then the map $\int_S$ is an isomorphism from $H_S^{2,c}$ to $\mathbf{R}$.*

The proof uses Lemmalem:partun, and the following Lemma.

Lemma 7 *Proposition * is true for the case when $S = \mathbf{R}^2$.*

To prove this, suppose $\rho = R(x_1, x_2)dx_1dx_2$ is a 2-form of compact support on bR^2 with integral 0. Choose a function ψ on $\mathbf{R}$ of compact support and with

$$\int_{-\infty}^{\infty} \psi(t)dt = 1.$$

Let

$$r(x_1) = \int_{-\infty}^{\infty} R(x_1, t)dt.$$

Now write

$$\tilde{R}(x_1, x_2) = R(x_1, x_2) - r(x_1)\psi(x_2).$$

So $\tilde{R}$ also has compact support in $\mathbf{R}^2$. For each x_1 we have

$$\int_{-\infty}^{\infty} \tilde{R}(x_1, t)dt = 0.$$

Define

$$P(x_1, x_2) = \int_{-\infty}^{x_2} \tilde{R}(x_1, t)dt.$$

Then P has compact support and

$$\frac{\partial P}{\partial x_2} = \tilde{R}(x_1, x_2).$$

Put

$$Q(x_1, x_2) = \psi(x_2) \int_{-\infty}^{x_1} r(t)dt.$$

Then Q has compact support and

$$\frac{\partial Q}{\partial x_1} = \psi(x_2)r(x_1).$$

Thus

$$R = \frac{\partial P}{\partial x_2} + \frac{\partial Q}{\partial x_1},$$

or in other words $\rho = d\alpha$ where α is the compactly supported form

$$\alpha = -Pdx_1 + Qdx_2.$$

We can now dispose of the proof of Proposition *. It is obvious that the map $\int$ is surjective so what we need to show is that if ρ is a 2-form of a compact support on a connected, oriented, surface and the integral of ρ is zero then $\rho = d\alpha$ for some α of compact support. Clearly we can choose a compact connected set K containing the support of ρ and we can cover K by a finite number of open sets $U_1, \ldots U_n$, each the image of a disc under a local chart. We use induction on n. If $n = 1$ then we are reduced to the case when the surface is $\mathbf{R}^2$ considered above. So suppose $n > 1$. Write

$V = U_2 \cup \ldots \cup U_n$, so K is contained in $U \cup V$. If either $K \cap U$ or $K \cap V$ is empty then we are done by the inductive hypothesis, so suppose these two sets are not empty. Then since K is connected there is a point p in $K \cap U \cap V$ and in particular in $U \cap V$. Clearly we can choose a 2-form τ with compact support contained in the open set $U \cap V$. Now apply the Lemma to choose functions f_1, f_2 supported in U and V respectively and with $f_1 + f_2 = 1$ on K. Then

$$\rho = f_1\rho + f_2\rho,$$

and $f_1\rho, f_2\rho$ have compact support in U, V respectively. Let

$$I = \int_S f_1\rho = -\int_S f_2\rho.$$

Then $f_1\rho - I\tau$ and $f_2\rho + I\tau$ are 2-forms with compact support in U and V respectively and with integral 0. By the inductive hypthesis we can find a 1-form α of compact support on U with $d\alpha = f_1\rho - I\tau$ and likewise a 1-form β of compact support in V with $d\beta = f_2\rho + I\tau$. Then $\rho = d(\alpha + \ beta)$ and the proof is complete.

Now let γ be a loop in an oriented surface S. Integration around γ yields a linear map

$$I_\gamma : H^1(S) \to \mathbf{R}.$$

On the other hand given any closed 1-form θ of compact support we get a linear map

$$J_\theta : H^1(S) \to \mathbf{R}$$

defined by

$$J_\theta([\phi]) = \int_S \theta \wedge \phi.$$

(By Stokes' theorem the integral on the right hand side is unchanged if we take a different representative ϕ for the same cohomology class.)

Proposition 13 *For any loop γ there is a compactly supported form θ such that $J_\theta = I_\gamma$.*

To prove this we choose a subdivision $0 = t_0 < t_1 < t_2 < \ldots < t_N = 1$ of the unit interval such that for each i with $0 \leq i \leq N-1$ there is a co-ordinate chart U_i (diffeomorphic to a disc) containing $\gamma([t_i, t_{i+1}]).Choose$. We can also easily arrange that $\gamma(t_i)$ are distinct except for $i = 0$ and N. (But see the remark at the end of the proof). Now choose small discs D_i containing $\gamma(t_i)$ such that

$$\overline{D_i} \subset U_{i-1} \cap U_i.$$

and with $D_N = D_0$, but with $D_i \cap D_j$ empty unless $\{i,j\} = \{0,N\}$. Let ρ_i be a 2-form supported in the interior of D_i with integral 1, and with $\rho_N = \rho_0$. Then $\rho_1 - \rho_0$ is a compactly supported form of integal zero on U_0 and we can find a compactly supported form θ_0 on U_0 such that

$$\rho_1 - \rho_0 = d\theta_0.$$

Similarly for $i \leq N-1$ we find θ_i such that

$$\rho_{i+1} - \rho_i = d\theta_i.$$

Now put

$$\theta = \theta_0 + \theta_1 + \ldots + \theta_{N-1}.$$

Then

$$d\theta = (\rho_1 - \rho_0) + (\rho_2 - \rho_1) + \ldots + (\rho_N - \rho_{N-1})$$

and this vanishes since $\rho_N = \rho_0$. We claim that $J_\theta = I_\gamma$. For let $[\phi]$ be a class in $H^1(S)$ we can choose a representative that vanishes on the discs D_j. Now on one of the open sets U_i we can write $\phi = df_j$ for some function f, since U_i is diffeomorphic to a disc and $H^1(U_i) = 0$. Thus

$$\int_S \theta_i \phi = \int_S \theta_i \wedge df_i = \int_S f_i(\rho_{i+1} - \rho_i).$$

Now since ϕ vanishes on all the discs D_j the function f_i is constant on D_{i+1} and D_i, which contain the supports of ρ_{i+1}, ρ_i respectively. Thus

$$\int_S \theta_i \wedge \phi = f_i(\gamma(t_{i+1}) - f_i(\gamma(t_i).$$

On thge other hand the integral of ϕ over the portion of the path parametrised by the sub-interval $[t_{i+1}, t_i]$ is the same, since $\phi = df_i$ over the image of the path. Summing over i finished the proof. (Remark: it is not really necessary that the points $\gamma(t_i)$ are distinct–all we need do is choose $\rho_i = \rho_j$ for any pair pair with $\gamma(t_i) = \gamma(t_j)$.)

Now the linear map J_θ depends only on the class of θ in $H^1_c(S)$. In other words we have a bilinear pairing

$$H^1_c(S) \times H^1(S) \to \mathbf{R}).$$

In particular if S is compact, so H^1 and H^1_c are the same thing, we have a bilinear form on $H^1(S)$ which is obviously skew-symmetric. Suppose Φ is any class in $H^1(S)$. If Φ is not zero we know that we can find a loop γ such that $I_\gamma(\Phi)$ is non-zero. By the Proposition above we can find a 1-form θ such that $J_\theta(\Phi)$ is non-zero. In other words, this bilinear form is nondegenerate. Since a vector space which supports a nendegenerate skew symmetric form must be even dimensional we get

Corollary 2 *For any compact oriented surface, the de Rham cohomology $H^1(S)$ is even dimensional.*

5.3 Calculus on Riemann surfaces

5.3.1 Decomposition of the 1-forms

Now let X be a Riemann surface, so *a fortiori* a smooth oriented surface. Thus for each point p in X we have a tangent space TX_p—a 2-dimensional real vector space. We also have a cotangent space

$$T^*X_p = Hom_{\mathbf{R}}(TX, \mathbf{R}),$$

such that the derivative of any real valued function on X yields an element of T^*X_p. We may just as well consider the complex cotangent space

$$T^*X_p^{\mathbf{C}} = Hom_{\mathbf{R}}(TX, \mathbf{C}),$$

such that the derivative of any complex valued function on X yields an element of $T^*X_p^{\mathbf{C}}$.

By a *complex structure* on a real vector space V we mean a $\mathbf{R}$ linear map $J : V \to V$ with $J^2 = -1$.

Lemma 8 *There is a unique way to define a complex structure on TX_p such that the derivative of any holomorphic function, defined on a neighbourhood of p in X, is complex linear.*

This follows immediately from the definition of a holomorphic function.

Now let V be a real vector space and J be a complex structure on V. We say that an $\mathbf{R}$-linear map A from V to $\mathbf{C}$ is complex linear if $A(Jv) = iAv$ for all v, and complex antilinear if $A(Jv) = -iA(v)$ for all v.

Lemma 9 *Any* $\mathbf{R}$*-linear map from* V *to* $\mathbf{C}$ *can be written in a unique way as a sum of a complex linear and antilinear maps*

For the existence we write $A = A' + A''$ where

$$A'(v) = \frac{1}{2}(A(v) - iA(Jv)), A''(v) = \frac{1}{2}(A(v) + iA(Jv)),$$

and check that A', A'' are complex linear and antilinear respectively. Uniqueness is similarly easy.

Putting this together, we see that we can write the complex cotangent space as a direct sum

$$T^*X_p^{\mathbf{C}} = T^*X_p' \oplus T^*X_p'',$$

in such a way that if f is a local holomorphic function then the derivative of f lies in T^*X_p' and the derivative of $\overline{f}$ lies in T^*X_p''. Now we can decompose the complex 1-forms on X into corresponding pieces

$$\Omega^1_{X,\mathbf{C}} = \Omega^{1,0}_X \oplus \Omega^{0,1}_X,$$

where elements of $\Omega^{1,0}$ lie in T^*X_p' for each p and $\Omega^{0,1}$ is the complex conjugate.

We now decompose the exterior derivative operators according to this decomposition of the forms, so we get a diagram

$$\begin{tikzcd}
\Omega^{0,1} & \Omega^2 \\
\Omega^0 \arrow[r] & \Omega^{1,0}
\end{tikzcd}$$

Let us see this more explicitly, in a complex local co-ordinate $z = x + iy$. Thus x, y are real co-ordinates as considered in the previous section. We have

$$dz = dx + idy \ , \ d\overline{z} = dx - idy,$$

and these form basis elements for T^*X', T^*X'' respectively. So a $(1,0)$ form is expressed locally as αdz and a $(0,1)$ form as $\beta\overline{z}$ for functions α, β. If f is a complex valued function then

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy.$$

We write
$$dx = \frac{1}{2}(dz + d\overline{z}), dy = \frac{1}{2i}(dz - d\overline{z}),$$
so
$$df = \frac{1}{2}(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y})dz + \frac{1}{2}(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y})d\overline{z}.$$
This means that
$$\partial f = \frac{\partial f}{\partial z}dz \ , \ \overline{\partial} f = \frac{\partial f}{\partial \overline{z}}d\overline{z},$$
where we define
$$\frac{\partial f}{\partial z} = \frac{1}{2}(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}) \ , \ \frac{\partial f}{\partial \overline{z}} = \frac{1}{2}(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}).$$

The equation $\overline{\partial} f = 0$ is, in this local co-ordinate, just the Cauchy-Riemann equation, as of course it should be since by definition a function is holomorphic if and only if $\overline{\partial} f = 0$. If f is a holomorphic function then we have, in local coordinates,
$$df = \partial f = f'(z)dz,$$
where f' denotes the usual derivative of complex analysis.

Now consider the operators $\partial, \overline{\partial}$ on $\Omega^{0,1}$ and $\Omega^{1,0}$. By following through the definitions we find that, in our local co-ordinate
$$\partial(A d\overline{z}) = \frac{\partial A}{\partial z} dz d\overline{z} = 2i\frac{\partial A}{\partial z} dx dy,$$
$$\overline{\partial}(B dz) = \frac{\partial B}{\partial \overline{z}} d\overline{z} dz = -2i\frac{\partial B}{\partial \overline{z}} dx dy.$$

Definition 8 *A* $(1,0)$*-form* β *is a holomorphic* 1*-form if* $\overline{\partial}\beta = 0$.

Thus in local co-ordinates a holomorphic 1-form has the shape Bdz where B is a holomorphic function.

Suppose $S \subset X$ is a compact surface with boundary and α is a holomorphic 1-form on a neighbourhhod of S. Then α is in particular a closed 1-form so Stokes' Theorem gives
$$\int_{\partial S} \alpha = 0.$$
This is one version of Cauchy's Theorem on a Riemann surface.

We define a *meromorphic* 1-form α on X in the obvious way: a holomorphic 1-form on $X \setminus D$, where D is a discrete subset of X, which can be written locally as $f(z)dz$ where f is a meromorphic function. (Of course ones needs to check that this is independent of the choice of local chart.) The points of the minimal such set D are the *poles* of the meromorphic 1-form. Let p be such a pole and let C be a small loop in X encircling p. We define the *residue* of α at p to be

$$Res_p(\alpha) = \frac{1}{2\pi i}\int_C \alpha.$$

Clearly this is the same as writing, in a local co-ordinate z centred at p,

$$\alpha = f(z)dz,$$

where $f(z) = \sum_{-k}^{\infty} a_j z^j$ is a meromorphic function and taking the usual residue a_{-1} of f and one can just as well take this as the definition.

Proposition 14 *Suppose α is a meromorphic 1-form on a compact Riemann surface X. Then the sum of the residues of α, running over all the poles, is zero.*

To see this, we let S be the complement in X of a union of small discs about the poles and apply Cauchy/Stokes'.

5.3.2 The Laplace operator and harmonic functions

On a Riemann surface we have a natural second order differential operator. We define

$$\Delta = 2i\overline{\partial}\partial : \Omega^0 \to \Omega^2.$$

Then in local co-ordinates

$$\Delta f = 2i\frac{1}{4}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) f(dzd\overline{z}) = -\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) dxdy.$$

Thus if, in a given local co-ordinate system, we identify the 2-forms with functions using the area form $dxdy$ the operator Δ becomes the standard Laplace operator (up to a sign convention). A function satisfying the differential equation is called a *harmonic function.* If f is a holomorphic function then the real and imaginary parts of f are harmonic since

$$\overline{\partial}\partial(f \pm \overline{f}) = -\partial\overline{\partial} f \pm \overline{\overline{\partial}(\overline{\partial} f)} = 0 \pm 0 = 0.$$

Conversely we have

Lemma 10 *If ϕ is a real-valued harmonic function on a neighbourhood N of a point p in a Riemann surface X then there is an open neighbourhood $U \subset N$ of p and a holomorphic function f on U with $\phi = \Re(f)$.*

Being local, this is not really different from the corresponding result for functions on open sets in $\mathbf{C}$ which the reader has very likely encountered in a standard complex analysis course. However it may be helpful to see how the proof works in our notation.

Let A be the real 1-form $i\overline{\partial}\phi + \overline{(i\overline{\partial}\phi)}$. Then the hypothesis that $\overline{\partial}\partial\phi = 0$ shows that $dA = 0$. Thus if U is a small disc about p (or any open set with $H^1(U) = 0$) we can find a real-valued function ψ with $A = d\psi$. This means that $\partial\psi = -i\partial\phi$ and $\overline{\partial}\psi = i\overline{\partial}\phi$. Then

$$\overline{\partial}(\phi + i\psi) = \overline{\partial}\phi + i\overline{\partial}\psi = 0.$$

So we can take $f = \phi + i\psi$.

We will also want to use the "maximum principle" occasionally.

Lemma 11 *Suppose ϕ is a non-constant, real valued, holomorphic function on a connected open set U in a Riemann surface X. Then for a point x in U there is a point x' in U with $\phi(x') > \phi(x)$.*

This can be seen by writing ϕ, near to x as the real part of a holomorphic function and then using the fact that holomorphic functions are open maps.

5.3.3 The Dirichlet norm

Let X be a Riemann surface and α be a $(1,0)$ form on X. We consider the 2-form $i\alpha \wedge \overline{\alpha}$. In a local complex co-ordinate $z = x + iy$, if $\alpha = pdz$ then

$$i\alpha \wedge \overline{\alpha} = i|p|^2 dz d\overline{z} = |p|^2 dx dy.$$

So $i\alpha \wedge \overline{\alpha}$ is a positive 2-form. We define

$$\|\alpha\|^2 = \int_X i\alpha \wedge \overline{\alpha},$$

taking values in $[0, \infty]$. Of course, if α has compact support the integral is finite and defines a norm on the space of compactly supported $(1,0)$ forms.

This norm (on the compactly supported forms) is derived from a Hermitian inner product

$$\langle \alpha, \beta \rangle = \int_X i\alpha \wedge \overline{\beta}.$$

If we have an area form ω on X we can define a pointwise norm, the function characterised by

$$i\alpha \wedge \overline{\alpha} = |\alpha|^2 \omega.$$

Then we can write, tautologically,

$$\|\alpha\|^2 = \int_X |\alpha|^2 \omega,$$

This is perhaps a more familar point of view. However the key point is that the "L^2-norm"on $(1,0)$ forms is actually independent of the choice of an area form.

We can identify the real 1-forms with the $(1,0)$ forms by mapping a real 1-form A to its $(1,0)$ component $A^{1,0}$. Thus we define

$$\|A\|^2 = \|A^{1,0}\|^2.$$

Again this norm is associated to a (real valued) inner product $< , >$ on the compactly supported real 1-forms

$$\langle A, B \rangle = i \int_X A^{0,1} \wedge B^{1,0}.$$

In Chapter 10 we will need the following simple result.

Lemma 12 *Let A, B be real 1-forms on a Riemann surface X. Then*

$$\int_X |A \wedge B| \leq \|A\| \|B\|.$$

(This needs to be interpreted in the obvious way: if either $\|A\|$ or $\|B\|$ is infinite then the statement is vacuous, if both are finite then the left hand side is also finite and the stated inequality holds.) This is, at bottom, very elementary. Suppose first that A and B are supported inside some local coordinate chart. Thus we can write $A = Pdz + \overline{P}d\overline{z}, B = Qdz + \overline{Q}d\overline{z}$, for complex-valued functions P, Q. Then

$$A \wedge B = (P\overline{Q} - Q\overline{P}) dz d\overline{z} = \Im(P\overline{Q}) dx dy.$$

Thus
$$\int_X |A \wedge B| = \int_{\mathbf{C}} |\Im(P\overline{Q})| dxdy.$$
By the Cauchy-Schwartz inequality
$$\int_X |A \wedge B| \leq \left(\int_{\mathbf{C}} |P|^2\right)^{1/2} \left(\int_{\mathbf{C}} |Q|^2\right)^{1/2} = \|A\|\|B\|.$$

Now suppose that A and B are any forms with compact support, both supported in some compact set $K \subset X$. It is convenient to explain the proof by choosing an arbitrary area form ω over a neighbourhood of K. The proof is then essentially the same. We have, pointwise over K,
$$|A \wedge B| \leq |A||B|\omega$$
and the proof is just the Cuachy-Schwartz inequality (for the functions $|A|, |B|$ with respect to the measure defined by ω). In the general case, by the definituion of the integral it suffices to prove that for any function χ of compact support with $0 \leq \chi \leq 1$ we have
$$\int_X \chi |A \wedge B| \leq \|A\|\|B\|.$$
Since $\chi|A \wedge B| = |(\chi A) \wedge B|$ and (obviously)
$$\|\chi A\| \leq \|A\|$$
, we can reduce to teh case when A has compact support (replacing A by χA). Suppose A is supported in a compact set J. Then we can choose a function η, with $0 \leq \eta \leq 1$, equal to 1 on J and with compact support. Replacing B be ηB, which has compact support, does not change the wedge product with A, and (obviously) $\|\eta B\| \leq \|B\|$. Thus we can reduce to the case of forms of compact support, considered above.

Suppose f and g are real-valued functions, at least one of compact support, on X. We define the *Dirichlet inner product* to be
$$\langle f, g \rangle_D = \langle df, dg \rangle.$$
Likewise we define the *Dirichlet norm* by
$$\|f\|_D = \|df\|,$$
with our usual conventio that this could be $+\infty$. The following will be crucial in Chapters 9 and 10.

Lemma 13 *If at least one of f, g have compact support then*

$$\langle f, g\rangle_D = \int_X g\Delta f = \int_X f\Delta g.$$

This really amounts to little more than the elementary identity, for functions f, g (at elast one of compact support) on $\mathbf{C}$,

$$\int_{\mathbf{C}} f\left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) = -\int_{\mathbf{C}} \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y}\right) dxdy,$$

which is derived immediately by integration-by-parts. In our notation, on a general Riemann surface, the proof becomes

$$\langle f, g\rangle_D = \langle df, dg\rangle = i\int_X \partial f \wedge \overline{\partial} g = i\int_X \partial(f\overline{\partial} g) - f\partial\overline{\partial} g) = \int_X f\Delta g,$$

where of course we have used Stokes' Theorem for the vanishing of the integral of $\partial(f\overline{\partial} g)$.

Chapter 6

Elliptic functions and integrals

In this Chapter we study Riemann surfaces of genus 1. On the one hand, the constructions will give an important model for the more general theory we develop in Part II. On the other hand, the constructions involve classical topics in mathematics, which relate the abstractions of Riemann surface theory to their origin in concrete calculus problems.

6.0.4 Elliptic integrals

The problem which gives this subject is the name was that of finding the arc length between two points on an ellipse. However we will take as our model problem that of finding the motion of a pendulum under gravity. In suitable units and in an obvious notation, the motion is defined by the energy conservation condition:

$$\dot{\theta}^2 - \cos\theta = E,$$

where E is constant. Thus

$$\dot{\theta} = \sqrt{E + \cos\theta},$$

and

$$t = \int \frac{d\theta}{\sqrt{E + \cos\theta}}.$$

So the problem reduces to doing this indefinite integral. Writing $u = \cos\theta$ this is transformed into

$$t = \int \frac{du}{\sqrt{(E+u)(1-u^2)}}.$$

More generally, we may consider

$$\int \frac{du}{\sqrt{f(u)}}$$

where f is any polynomial. In the case when f is quadratic we know how to perform these integrals in terms of elementary functions, the question is what kind of functiosn arise for polynomials of higher degree, and particularly for cubic polynomials such as that arising in the pendulum problem.

We now have the language required to interpret this in a better way. Suppose $f(z)$ is a polynomial of degree n, with distict roots $z_1, \dots, z_n$. Let $X \subset \mathbf{C}^2$ be the set of solutions of the equation $w^2 = f(z)$. The condition that f has distinct roots means that the partial derivatives of $w^2 - f(z)$ do not both vanish anywhere on X, so X is a Riemann surface. By construction we have a pair of holomorphic functions z, w on X with derivatives dz, dw. Since $w^2 - f(z)$ vanishes on X we have an identity

$$2wdw = f'(z)dz.$$

Now dz/w is a holomorphic 1-form away from the points where $w = 0$. In punctured neighbourhoods of such points we can write

$$\frac{dz}{w} = 2\frac{dw}{f'(z)},$$

and $f'(z)$ does not vanish since f has simple roots. So we conclude that dz/w extends to a *holomorphic* 1-form α on X. Moreover we see that α does not vanish anywhere on X.

To summarise so far: when we write an expression such as

$$\int_{z_0}^{z_1} \frac{dz}{\sqrt{f(z}},$$

what we shoudl really mean is that we choose a path γ on the Riemann surface X, running from a point with $z = z_0$ to a point with $z = z_1$ and form

$$\int_\gamma \alpha.$$

Now recall that we defined a compactification X^* of X. We want to consider the extension of α to X^*. This is a good exercise in the theory developed in Chapter 2. We consider two cases, depending on whether n is odd or even. To construct X^* we need to examine the monodromy of the covering around a large circle in $\mathbf{C}$ (corresponding to a small circle around $\infty \in S^2$). The monodromy lies in the group $\mathcal{S}_2$ of permutations of the two sheets. Clearly if n is even the monodromy is trivial and if n is odd it is the nontrivial element of $\mathcal{S}_2$. Consider the odd case. We form X^* by attaching a single disc to X, adjoining one extra point P say. If

$$f(z) = z^n + a_1 z^{n-1} + \ldots + a_n,$$

then in terms of a standard co-ordinate τ on the disc

$$z = \tau^{-2}, w = \sqrt{f(z)} = \tau^{-n}\sqrt{1 + a_1\tau^2 \ldots + a_n\tau^{2n}},$$

where the square root is well-defined for small τ. Then we have

$$\frac{dz}{w} = \left(-2\frac{d\tau}{\tau^3}\right)\left(\frac{\tau^n}{\sqrt{1 + a_1\tau^2 + \ldots}}\right) = \frac{-2}{\sqrt{1 + a_1\tau^2 + \ldots}}\tau^{n-3}d\tau.$$

So we conclude that, if n is odd, α extends to a meromorphic 1-form on X^* and that:

- If $n = 1$, α has a pole of order 2 at the point P;
- if $n = 3$ then α is holomorphic near P and does not vanish at P;
- If $n > 3$ then α is holomorphic near P and has a zero of order $n - 3$ at P.

The even case is similar. Now we adjoin two extra points P^+, P^- say to form X^* and the reader can check, as above, that α is meromorphic on X^* with

- If $n = 2$ then α has simple poles at $P_\pm$;
- If $n = 4$ then α is holomorphic near $P_\pm$ and does not vanish at $P_\pm$;
- if $n > 4$ then α is holomorphic near $P_\pm$ and has zeros of order $(n-4)/2$ at $P_\pm$.

We want to focus attention on the cases when $n = 3$ or 4 so α is holomorphic on X^* and does not vanish anywhere. In fact there is no real disticntion between $n = 3, 4$ since we can transform one case to the other by a Mobius transformation of the Riemann sphere: in either case X^* is a double cover of the sphere with four brach points and the distinction is just whether we choose ∞ to be a branch point ($n = 3$) or not ($n = 4$). We change point of view slightly and prove a general classification theorem.

Theorem 3 *Let X be a compact Riemann surface and let α be a holomorphic 1-form on X with no zeros. Then there is a lattice $\Lambda \subset \mathbf{C}$ and an isomorphism $\iota : \mathbf{C}/\Lambda \to X$ such that $\pi^* \iota^*(\alpha) = du$, where u is the identity function on $\mathbf{C}$ and $\pi : \mathbf{C} \to \mathbf{C}/\Lambda$ is the projection map.*

First we sketch the idea of the proof, which is quite simple. We try to define an indefinite integral of the holomorphic 1-form α. We can perform the integral along a path in X but the value depends on the end points since we have a choice of homotopy class of paths. This indeterminacy means that the indefinite integral is not defined as a $\mathbf{C}$-valued function but it is defined as a map to $\mathbf{C}/\Lambda$ for a suitable Λ and this map is the inverse of the desired isomorphism ι.

Now for the detailed proof. Consider the univeral cover $p : \tilde{X} \to X$. The lift $p^*(\alpha)$ is a holomorphic 1-form on $\tilde{X}$ and, since $\tilde{X}$ is simply connected the integral of $p^*(\alpha)$ along paths depends only on the endpoints. So we get a holomorphic map $F : \tilde{X} \to \mathbf{C}$ with $dF = p^*(\alpha)$. Since α has no zeros the map F is a local homeomorphism. We claim that F is in fact a covering map. For each point $x \in X$ we can find a radius $r > 0$ and an injective holomorphic map $j_x : D_r \to X$ where D_r is the r-disc $\{u : |u| < r\}$ in $\mathbf{C}$ such that $j_x(0) = x$ and $j_x^*(du) = \alpha$. That is, j_x is the inverse of a locally-defined indefinite integral of α. Since x is compact we can, by a simple argument, find a single r which works for all $x \in X$. Now suppose $\tilde{x}$ is a point in $\tilde{X}$. Since the disc is simply connected we can lift $j_{p(\tilde{x})}$ to get an injective map

$$\tilde{j}_{\tilde{x}} : D_r \to \tilde{X},$$

with $\tilde{j}_{\tilde{x}}(0) = \tilde{x}$ and $\tilde{j}_{\tilde{x}}^*(du) = p^*(\alpha)$. Let $\Delta_{\tilde{x}}$ be the image under $\tilde{j}_{\tilde{x}}$ of the disc $D_{r/2}$ of radius $r/2$. Then by construction $F(\Delta_{\tilde{x}})$ is the $r/2$-disc $D_{F(\tilde{x}),r/2}$ in $\mathbf{C}$ centred on $F(\tilde{x})$. Now we observe that, for $\tilde{x}, \tilde{y} \in \tilde{X}$,

$$\tilde{y} \in \Delta_{\tilde{x}} \Leftrightarrow \tilde{x} \in \Delta_{\tilde{y}}.$$

For if $\tilde{y}$ is in $\Delta_{\tilde{x}}$, so $\tilde{y} = \tilde{j}_{\tilde{x}}(v)$ say for some $|v| < r/2$ then the whole set $\Delta_{\tilde{y}}$ can be described as

$$\tilde{j}_{\tilde{x}}(\{w : |v - w| < r/2\}),$$

and this obviously contains $\tilde{j}_{\tilde{x}}(0) = \tilde{x}$. Now let z be any point in $\mathbf{C}$ and consider the disc $D_{r/2,z}$ of radius $r/2$ centred on z. Suppose $\tilde{y}$ is a point of $F^{-1}(D_{r/2,z})$. Then z lies in the $r/2$ disc centred on $F(\tilde{y})$ so there is a point $\tilde{x}$ in $\Delta_{\tilde{y}}$ with $F(\tilde{x}) = z$. Equally, by the remark above $\tilde{y}$ lies in $\Delta_{\tilde{x}}$ where $F(\tilde{x}) = z$. So we have

$$F^{-1}(D_{r/2,z}) = \bigcup_{\tilde{x} \in F^{-1}(z)} \Delta_{\tilde{x}}.$$

Suppose that $\tilde{x}_1$ and $\tilde{x}_2$ are two points in $F^{-1}(z)$ and that $\Delta_{\tilde{x}_1} \cap \Delta_{\tilde{x}_2}$ is not empty. Then there is a point $\tilde{y}$ in the intersection. Then by the remark above $\tilde{x}_1, \tilde{x}_2$ both lie in $\Delta_{\tilde{y}}$ but this is a contradiction to the fact that F is injective on $\Delta_{\tilde{y}}$. So we conclude that the union above is a disjoint union, hence F is indeed a covering map.

Now since $\mathbf{C}$ is simply connected it has no non-trivial connected coverings and we conclude that F is an isomorphism from $\tilde{X}$ to $\mathbf{C}$. But we know that X is the quotient of $\tilde{X}$ by an action of $\pi_1(X)$ on $\tilde{X}$. So we conclude that X is isomorphic to the quotient of $\mathbf{C}$ by a group of holomorphic automorphisms. By the classification of these quotients we see that the only posibility is that $X = \mathbf{C}/\Lambda$ for some lattice Λ. The identification of the form α follows from the construction.

To sum up: if f is a cubic polynomial with distinct roots and X^* is the compact Riemann surface associated to the equation $w^2 = f(z)$ then there is a lattice $\Lambda \subset \mathbf{C}$ and an isomorphism

$$\iota : \mathbf{C}/\Lambda \to X^*.$$

This can also be regarded as a Λ-periodic map from $\mathbf{C}$ to X^* which can be written as a pair of meromorphic functions $z(u), w(u)$ on $\mathbf{C}$ with

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

The map has the property that it pulls the holomorphic form $dz/w = dz/\sqrt{f(z)}$ back to the constant form du on $\mathbf{C}$, or equivalently

$$\frac{dz}{du} = w = \sqrt{f(z)}.$$

6.0.5 The Weierstrasse function

We now make a fresh start with a lattice Λ in $\mathbf{C}$. We ask the question: can we find a meromorphic function on $\mathbf{C}/\Lambda$? Since $\mathbf{C}/\Lambda$ is compact there are no non-trivial holomorphic functions, so we need to allow poles. Moreover, since $\mathbf{C}/\Lambda$ is not homeomorphic to the Riemann sphere we must have more than one simple pole (or a multiple pole) by Corollary *. We can see this more directly as follows. Let P be a parallelogram forming the standard fundamental domain and let Γ be the boundary of P. A meromorphic function F on $\mathbf{C}/\Lambda$ yields a *doubly periodic* meromorphic function $\tilde{F}$ on $\mathbf{C}$. There is no loss in supposing that the no pole of $\tilde{F}$ lies on Γ. Then Cauchy's Theorem implies that

$$\int_\Gamma \tilde{F} du$$

is the sum of the residues of the poles in P. But the double periodicity means that the integrals around opposite of Γ cancel so we see that the sum of these residues is zero. In particular we cannot have a single, simple pole.

Following the considerations above we seek a meromorphic function with one *double* pole, and we obtain this through the famous Weierstrasse construction. We define $\wp = \wp_\Lambda$ on $\mathbf{C}$ by

$$\wp(u) = \frac{1}{u^2} + \sum_{\lambda \in \Lambda \setminus \{0\}} \left(\frac{1}{(u-\lambda)^2} - \frac{1}{\lambda^2} \right).$$

For any u in $\mathbf{C} \setminus \Lambda$ the sum on the right hand side of this expression converges. For when $|\lambda|$ is large (and u is fixed)

$$\frac{1}{(u-\lambda)^2} - \frac{1}{u^2} = O(|\lambda|^{-3}),$$

We can compare the sum, over large λ in the lattice, with the double integral

$$\int_{|\lambda|>1} |\lambda|^{-3} dp dq,$$

where $\lambda = p + iq$, to see that the sum converges absolutely. It follows easily that the formula above defines a Λ-periodic meromorphic function on $\mathbf{C}$ with double poles at the points of Λ and no other poles. This then descends to yield a meromorphic function, which we still call $\wp$, on $\mathbf{C}/\Lambda$ with one double pole. Note from the form of the construction that $\wp$ is an even function: $\wp(-u) = \wp(u)$.

Now $\wp$ has a Laurent expansion about 0:

$$\wp(u) = \frac{1}{u^2} + 0 + au^2 + bu^4 + \dots,$$

where the vanishing of the coefficient of u^0 follows from the shape of the construction. This gives

$$\wp''(u) = \frac{6}{u^4} + 2a + \dots,$$

so

$$\wp'' - 6\wp^2 = -10a + \dots,$$

is a *holomorphic* Λ-periodic function, hence constant. Thus $\wp'' - 6\wp^2 = 10a$. We can rewrite this identity as

$$\frac{d}{du}\left(\wp'^2\right) = \frac{d}{du}\left(4\wp^3 - 20a\wp\right)$$

so $\wp'^2 = 4\wp^3 - 20a\wp + a'$, say, with a' another constant. Adopting conventional notation, $\wp$ satisfies an equation

$$\frac{d\wp}{du}^2 = 4\wp^3 - g_2\wp - g_3,$$

for certain constants g_2, g_3 depending on the lattice.

Exercise Show that

$$g_2 = 60 \sum_{\lambda \in \Lambda'} \lambda^{-4}$$

$$g_3 = 160 \sum_{\lambda \in \Lambda'} \lambda^{-6}$$

where Λ' denotes $\Lambda \setminus \{0\}$.

Now this just expresses the fact that $\mathbf{C}/\Lambda$ arises as the Riemann surface associated to an equation $w^2 = f(z)$ for cubic f so our conclusion is that the classes of Riemann surfaces obtained in these two ways are identical. Returning finally to our starting point we see that the solution of the pendulum equation can be written as

$$\theta = \cos^{-1}(\wp_\Lambda(t + t_0)),$$

for a suitable lattice Λ.

Exercise. Show that the lattice Λ associated to the pendulum equation is *rectangular*, generated by 1 and iq for some $q > 0$, and that for the physical solutions the imaginary part of the constant of integration t_0 must be $q/2$.

Chapter 7

Applications of the Euler characteristic

We have seen that the genus of a compact oriented smooth surface S can be defined as one half the dimension of the de Rham cohomology group $H^1(S)$. There are many other possibilities. In particular another way of defining the genus, in some respects more elementary, uses the *Euler characteristic*. This can be done via triangulations of the surface. We do not want to take the time to develop the theory of triangulations in detail but we will describe this approach slightly informally here and then develop some applications. The reader with a suitable background in topology will know how to make the discussion more rigorous and in any case we will be able to derive the corollaries as simple consequences of the more advanced theory (the Riemann-Roch formula) later. Howver we give this more elementary discusion now, since we do not want the reader to gain the impression that these essentially topological results depend essentially on the rather deeper analysis in Part III.

7.1 The Euler characteristic and meromorphic forms

7.1.1 Topology

Suppose we have a surface, posibly with boundary, which is *triangulated* in the manner indicated.

Then the Euler characteristic of the triangulation is defined to be

$$\chi = V - E + F$$

where V, E, F are the number of vertices, edges and faces respectively. The first basic fact is that number is independent of the choice of triangulation, hence defines an integer $\chi(S)$ the Euler characteristic of S. It is not hard to check, for example by defining explicit triangulations, that for the model surfaces

$$\chi(\Sigma_{g,r}) = 2 - 2g - r, \chi(\Xi_{h,r}) = 2 - h - r.$$

We can define the genus of a closed oriented surface by $g = 1 - \frac{\chi(S)}{2}$ and of course the next thing we need to know is that this coincides with our previous definition. If we are willing to accept the classification of surfaces of course there is no need to have any theory here, since we just need to check for the model surfaces. But is preferable from many points of view (for example, extensions to higher dimensions) to understand the result independent of the clasification of surfaces. In any case, let us assume it from now on.

Now suppose S is a compact oriented surface and that α is a real 1-form on S. Suppose that the set $\Delta \subset S$ where α vanishes is discrete. Given any point p of Δ we choose local co-ordinates centred on p and represent α locally as

$$\alpha = \alpha_1 dx_1 + \ alpha_2 dx_2.$$

Our hypothesis asserts that for small r the only zero of the vector-valued function (α_1, α_2) on the closed r-disc about the origin is at the origin itself. Thus the restriction of this function to a circle of radius r gives a map from the circle to $\mathbf{R}^2 \setminus \{0\}$ which has an integer winding number. It not hard to check that this is indpendent of teh choice of r and the local co-ordinate system. We define the multiplicity $m_p(\alpha)$ of the zero p of α to be this winding number.

Proposition 15 *In the situation above*

$$\sum_{p \in \Delta} m_p(\alpha) = -\chi(S).$$

We sometimes call the sum on the left hand side of this formula the "number of zeros of α(counted with multiplicity)". Of course there are many different ways of building up this theory. For example if one shows that the sum of zeros (counted with multiplicity) is independent of teh choice of α then one can use it to *define* the Euler characteristic, and hence the genus. To relate this to the count in the definition by triangulations one can consider a standard 1-form on a triangle given by df where f is the function indicated by the following picture.

If we have a triangulation of S then we can define a 1-form on S which restricts to this model on each triangle. There is then one zero for each vertex, one for each edge and one for each face: the multiplicities are all $+1$ for the first and third case and -1 for the second. So the "count" of zeros gives precisely the count of vertices, edges and faces, with the right signs. In any case what we will assume known is that Proposition * holds true for any smooth 1-form α with discrete zero-set and where $\chi(S) = 2 - 2g$ with g the genus defined in Chapter *.

7.1.2 Meromorphic forms

Now suppose that X is a compact Riemann surface and that α is a holomorphic 1-form on X, not identically zero. We associate to this the real 1-form $A = \alpha + \overline{\alpha}$ In a local co-ordinate z we write $\alpha = f(z)dz$; the zeros of A are the zeros of f hence discrete. Moreover the multiplicity of a zero is a positive integer, equal to the multiplicity of the zero of f in the usual sense. Hence in this case Proposition * says that the total number of zeros counted with these positive multiplicities, is $2g - 2$. In particular if $g = 0$ there can be no such α and if $g = 1$, the situation considered in the previous Chapter, a non-trivial holomorphic form is nowhere vanishing.

We can extend this discussion to *meromorphic* 1-forms. To do this we fix an area form ω on X. This means that we can define a Hermitian metric on T^*X':

$$\xi \wedge \overline{\xi} = |\xi|^2\omega.$$

Suppose α is a meromorphic 1-form on X. Choose a real-valued function p on $\mathbf{R}$ with $p(t) = 1$ for small t and $p(t) = t^{-1}$ for large t. Now define

$$\tilde{\alpha} = p(|\alpha|^2\alpha,$$

away from the poles of α and $\tilde{\alpha} = 0$ at the poles of α. Locally, around a pole of α, we have

$$\tilde{\alpha} = \frac{1}{|f(z)|^2} f(z) R dz = \overline{f(z)} R dz$$

where R is a smooth strictly positive function. Thus tilde α is smooth and its zero set is the union of the zeros and poles of α. It is clear that the zeros of $\tilde{\alpha}$ corresponding to the poles of α have multiplicity equal to minus the order of the pole. Thus we have

Proposition 16 *If α is a nontrivial meromorphic 1-form on a compact Riemann surface X then the number of zeros of α minus the number of poles of α, counted with multiplicity is equal to $2g - 2$.*

7.2 Applications

7.2.1 The Riemann-Hurwitz formula

Suppose that $f : X \to Y$ is a nonconstant holomorphic map between connected compact Riemann surfaces. In Chapter 4 we have associated a multiplicity k_x to each point of X, equal to 1 except for a finite set of ramification points. We define the total ramification index to be

$$R_f = \sum_{x \in X} k_x - 1.$$

So this is really a finite sum. We have also defined the degree $d \geq 1$ of the map.The following result, the *Riemann Hurwitz* formula, is very useful for calculations

Proposition 17 *The genus g_X of X and the genus g_Y of Y are related*

$$2 - 2g_Y = d(2 - 2g_X) - R_f.$$

One way of proving, which we sketch, is to show (or assume depending on taste) that there is a triangulation of Y such that each branch point of f is a vertex. Then the triangulation can be "lifted" to a triangulation of Y. Each face or edge of the triangulation of Y gives rise to d faces or edges of the triangulation of X. Likewise each vertex of the triangulation of X which is *not* branch point gives rise to d vertices in the triangulation of X. On the other hand a branch point y in Y gives rise to only

$$d - \sum_{x \in f^{-1}(y)} (k_x - 1)$$

vertices in the triangulation of X and so the formula follows from the counting formulae for the Euler characteristics.

We can give another proof if we suppose that there is a meromorphic 1 form β on Y. (In Part II we shall see that these always exist.) Then $f^*(\beta)$ is a meromorphic 1 form on X. A pole or zero of β which is not a branch point gives rise to d poles or zeros of $f^*(\beta)$ on X with same multiplicity. On the other hand suppose that x is a ramification point so in local co-ordinates the map f can be represented as $z \mapsto w = z^k$, where $k = k_x > 1$. If β is the given in the w co-ordinate by $g(w)dw$ for some meromorphic g the pull-back $f^*(\beta)$ is

$$kz^{k-1} g(z^k) dz.$$

If g has a zero of order $l \in \mathbf{Z}$ (where a negative value of l indiactes a pole in the obvious way) then $z^{k-1} g(z^k)$ has a zero of order $kl + k - 1$. So the contribution to the count of zeros/poles of $f^*(\beta)$ from the points x in $f^{-1}(y)$ is

$$\sum_{x \in f^{-1}(y)} (k_x l + k_x - 1) = dl + \sum_{x \in f^{-1}(y)} (k_x - 1),$$

since we know that

$$d = \sum_{x \in f^{-1}(y)} (k_x - 1).$$

Then applying * to β and $f^*(\beta)$ we obtain the Riemann-Hurwitz formula.

7.2.2 The degree-genus formula

Now suppose that X is a smooth complex curve of degree d in $\mathbf{CP}^2$. Recall that this means that X is defined by a homogeneous polynomial $p(Z_0, Z_1, Z_2)$ and that not all of the partial derivatives $\frac{\partial p}{\partial Z_i}$ vanish at any point of X.

Proposition 18 *The genus of X is given by*

$$g_X = \frac{1}{2}(d-1)(d-2).$$

For example we have already seen that when $d = 1$ or 2 the Riemann surface X is equivalent to the Riemann sphere (genus 0), and when $d = 3$ to a complex torus $\mathbf{C}/\Lambda$ (genus 1). Notice that this shows that some compact Riemann surfaces *cannot* be realised as smooth curves in $\mathbf{CP}^2$, since not all integers can be expressed as $(d-1)(d-2)/2$.

There are various ways of obtaining the formula in the Proposition. We will establish the result by contructing a meromorphic form on X and counting the poles and zeros. Suppose for simplicity that the curve X meets the line at infinity in d distinct points. (It is easy to show that this can be arranged by a suitable linear transformation of the Z_i.) Let $P(z, w)$ be the polynomial in 2 variables defining the corresponding affine curve X_0. We follow the same construction that we used, in a special case, in Chapter 5. Thus dz, dw represent holomorphic 1-forms on X_0 and the identity $P(z, w) = 0$ on X_0 yields

$$P_z dz + P_w dw = 0.$$

At points where P_z, P_w are both non-zero we have

$$\frac{dz}{P_w} = -\frac{dw}{P_z}.$$

Since, by hypothesis, there are no points on X_0 where P_z, P_w both vanish we obtain a non-vanishing holomorphic 1-form θ on X_0 equal to dz/P_w or $-dw/P_z$ at the points where these are defined. We have to check that θ is a meromorphic 1-form on X, i.e. that it has at worst poles at the d points of intersection with the line at infinity, and then count the zeros or poles at these points.

7.2.3 Real structures and Harnack's bound

In Chapter 2 we discussed nonorientable surfaces, indeed this was the most interesting case from the point of view of the classification theorem, but since then they have dropped out of the picture, mainly because any Riemann surface is oriented. However non-orientable surfaces do arise naturally in certain questions, as we will now illustrate.

In Chapter 3 we have defined the notion of *holomorphic maps* between Riemann surfaces. One can just as well define *antiholomorphic maps*: given in local complex coordinates by antiholomorphic functions (a function f on an open set in $\mathbf{C}$ is antiholomorphic if $\overline{f}$ is holomorphic. The composite of antiholomorphic maps is holomorphic and the composite of a holomorphic map and an antiholomorphic map is antiholomorphic. We say that a *real structure* on a Riemann surface X is an antiholomrphic map $\sigma : X \to X$ with $\sigma \circ \sigma$ equal to the identity. The *real points* $X_{\mathbf{R}}$ of such a pair (X, σ) are defined to be the fixed points of σ. For example the maps $\sigma_0, \sigma_1 : \mathbf{C} \setminus \{0\} \to \mathbf{C} \setminus \{0\}$ given by

$$\sigma_0(z) = \overline{z}\ ,\ \ \sigma_1(z) = -1/\overline{z},$$

are two different real structures on $\mathbf{C} \setminus \{0\}$. The real points being the real axis in one case and the empty set in the other. Each of these extends to a real structure on the Riemann sphere and $S^2_{\mathbf{R}}$ is a copy of a circle in one case and empty in the other.

Proposition 19 *Let (X, σ) be a Riemann surface with a real structure and x be a point of $X_{\mathbf{R}} \subset X$. There is a local holomorphic co-ordinate z around x in which σ is given by the map $z \mapsto \overline{z}$.*

The proof is an exercise for the reader.

If (X, σ) is a surface with a real structure we can form the quotient space X/σ. Using the proposition above it is not hard to show

Proposition 20 *The space X/σ is a surface with boundary, where the boundary of X/σ can be identified with $X_{\mathbf{R}}$.*

(More precisely we should say that X/σ can be endowed with structure of a surface with boundary.) The quotient X/σ may be orientable or non-orientable and the boundary may or may not be empty. For example with the two real structures σ_0, σ_1 on S^2 above S^2/σ_0 is a disc and S^2/σ_1 is the real projective plane $\mathbf{RP}^2$.

Now suppose that X is a *compact* connected Riemann surface with a real structure σ. Then X/σ is a compact connected surface with boundary. We have

Proposition 21 *The Euler characteristics of X and X/σ satisfy*

$$\chi(X) = 2\chi(X/\sigma).$$

Just as with the Riemann-Hurwitz formula various proofs are possible. In terms of triangulations, if we choose a triangulation of X/σ we can lift to a triangulation of X. Then to each face of X/σ correspond two faces of X and likewise to each vertex and edge which do *not* lie in the boundary. On the other hand each edge and vertex in the boundary of X/σ correspond to just one edge or vertex in X. But each component of the boundary of X/σ is a circle, so clearly the numbers of edges and vertices in the boundary are equal. Then the result follows from simple counting.

Now for any surface with boundary Y with r boundary components we have

$$\chi(Y) \leq 2 - r,$$

with equality in the case when Y is a sphere with r discs removed. If we accept the classification of surfaces we can read this assertion off from that: an independent proof is also possible of course. In any case we obtain the conclusion

Proposition 22 *Let X be a compact connected Riemann surface of genus g. If σ is a real structure on X then the number of components of $X_{\mathbf{R}}$ is at most $g+1$.*

This follows immediately from the discussion above, since if r is the number of components of $X_{\mathbf{R}}$ we have

$$1 - g = \frac{1}{2}\chi(X) = \chi(X/\sigma) \leq 2 - r$$

since r is also the number of boundary components of X/σ.

Equality holds in the case when σ is the reflection map acting on the standard picture of a surface of genus g, and the quotient is a sphere with $g+1$ discs removed.

.

To give a concrete application of this, suppose that P is a polynomial in two variables with *real co-efficients*. Then, regarded as a complex polynomial,

P defines an affine curve in $\mathbf{C}^2$ and a projective curve X in $\mathbf{CP}^2$. Suppose for simplicity that these are smooth. Complex conjugation of the co-ordinates induces a real structure on X and the set $X_{\mathbf{R}}$ just corresponds to the points of the corresponding real projective curve in $\mathbf{RP}^2$. So we have *Harnack's bound*

Proposition 23 *Let p be a homogeneous polynomial with real co-efficients and degree d in three variables with the property that the corresponding complex projective curve is smooth. Let $\Gamma \subset \mathbf{RP}^2$ be the real projective curve defined by p. Then the number of components of Γ is at most $\frac{1}{2}(d-1)(d-2)+1$.*

(It is not hard to remove the hypothesis on the smoothness of the complex projective curve.)

7.2.4 Modular curves

In Chapter 3 we introduced the modular curves X_a, the quotients of the upper half plane by the group Γ_a. From the definition it was not clear how to say much about these. We will now see how to understand the topology of these Riemann surfaces, for simplicity we restrict to the case when a is prime.

Proposition 24 *If p is a prime number, not equal to 2, then there is a compact Riemann surface $\overline{X}_p$ of genus*

$$g = 1 + \left(\frac{(p-6)(p^2-1)}{24} \right)$$

with a subset $\Delta \subset \overline{X}_a$ containing $(p^2-1)/2$ points, such that X_p is equivalent to the complement $\overline{X}_p \setminus \Delta$.

Part III

Deeper Theory

Chapter 8

Meromorphic functions and the Main Theorem for compact Riemann surfaces

This Section does not contain much technical content but is crucial in explaining our strategy for proving the fundamental structural results about Riemann surfaces. For motivation, we can review the discussion of surfaces of genus 1 in the previous section. Suppose we have a general compact Riemann surface X of genus 1. We can show that X is equivalent to one of the family of surfaces studied in the previous Chapter if we can show *either*:

- That there is a meromorphic function with a double pole (or two single poles) on X. This then represents X as a two sheeted cover of the Riemann sphere with four branch points.

or

- That there is a nowhere-vanishing holomorphic 1-form on X. Then we can apply Theorem * to see that X is a complex torus C/Λ.

This motivates our task, which is to get a good understanding of the existence of meromorphic functions, holomorphic 1-forms and the relationship between these for general compact Riemann surfaces.

Recall that on any Riemann surface X we have the "square" of differential operators *. We define complex vector spaces

$$H^{0,0}_X = \mathrm{Ker}\overline{\partial} : \Omega^0 \to \Omega^{0,1}.$$

$$H_X^{1,0} = \mathrm{Ker}\overline{\partial} : \Omega^{1,0} \to \Omega^2.$$
$$H_X^{0,1} = \mathrm{Coker}\overline{\partial} : \Omega^0 \to \Omega^{0,1}.$$
$$H_X^{1,1} = \mathrm{Coker}\overline{\partial} : \Omega^{1,0} \to \Omega^2.$$

Thus $H_X^{0,0}, H_X^{1,0}$ are the spaces of holomorphic functions and holomorphic 1-forms respectively. The significance of the other groups is not so clear. What we want to do now is to explain how $H_X^{0,1}$ arises naturally when one attempts to construct meromorphic functions.

Let p be a point in X. We ask the question: is there a meromorphic function on X with a simple pole at p and no other poles? Let z be a local co-ordinate centred on p. Thus $\frac{1}{z}$ can be regarded as a meromorphic function on some open neighbourhood U of p. We introduce a cut-off function β; a smooth function supported in U and equal to 1 near p. Then $\beta\frac{1}{z}$ can be thought of as a function on $X \setminus p$, extending by zero outside U. Finding a meromorphic function with a pole at p is equivalent to finding a smooth function g on X such that $g + \beta\frac{1}{z}$ is holomorphic on $X \setminus p$. Now

$$A = \overline{\partial}(\beta\frac{1}{z}) = (\overline{\partial}\beta)\frac{1}{z}$$

has compact support in $X \setminus \{p\}$ since β equals 1 near p. So we can regard A as a $(0,1)$ form on X, extending by zero over p. Thus our problem is equivalent to solving the equation

$$\overline{\partial}g = -A$$

for the given element A of $\Omega_X^{0,1}$ and the unknown $g \in \Omega_X^0$. By definition, a solution exists if and only if the class $[A]$ in the quotient $H_X^{0,1} = \mathrm{Coker}\overline{\partial} = \Omega^{0,1}/\mathrm{Im}\overline{\partial}$ is zero. In particular, a solution will exist if $H_X^{0,1} = 0$. Even if a solution does not exist, the class $[A]$ is, up to multiplication by a non-zero scalar, a well-defined element of $H_X^{0,1}$ associated to the point p in X. For if ϕ is any smooth function on $X \setminus \{p\}$ which restricts to a meromorphic function with a pole at p on some neighbourhood of p then for a suitable choice of $\lambda \in \mathbf{C}$ the difference $\phi - \lambda\beta\frac{1}{z}$ extends to a smooth function on X (holomorphic near p) so

$$[\overline{\partial}\phi] = \lambda[A] \in H_X^{0,1}.$$

Now suppose that we have d distinct points $p_1, \ldots p_d$ in X. We ask if we can find a meromorphic, but not holomorphic, function on X with poles at some, or all, of the p_i and no other poles. We follow through the same procedure as before, choosing local co-ordinates around the p_i to get $(0,1)$ forms A_i, which we can take to be supported in small disjoint annuli around the p_i if we like. The same argument as before shows that we can find the desired meromorphic function if there are scalars λ_i, not all zero, such that

$$\lambda_1[A_1] + \ldots + \lambda_d[A_d] = 0 \in H^{0,1}_X.$$

Given such a linear relation we get a meromorphic function with poles at the points p_j for which $\lambda_j \neq 0$. In particular we have

Proposition 25 *Suppose $H^{0,1}_X$ has finite dimension h. Then given any $h+1$ points $p_1, \ldots, p_h$ on X there is a non-holomorphic meromorphic function on X with simple poles at some subset of the $p_1, \ldots, p_h$.*

This is just because there must be a non-trivial linear relation between any $h+1$ elements of $H^{0,1}_X$.

The discussion aboves shows how we can cast our problem in terms of the spaces $H^{0,1}_X$ but it does not by itself get as very far. To go further we need some much deeper input and we will formulate this in terms of the following "Main Theorem for compact Riemann surfaces" (not standard terminology):

Theorem 4 *Let X be a compact connected Riemann surface and let ρ be a 2-form on X. there is a solution f to the equation $\Delta f = \rho$ if and only the integral of ρ over X is zero.*

We will give a proof of this Theorem in the next section, but let us see some consequences first.

8.1 Consequences of the main theorem

The relation between the "Dolbeault cohomology" $H^{i,j}$ and the de Rham cohomology H^i can be summarised as follows We have the following natural maps

- A map $\sigma : H^{1,0} \to \overline{H^{0,1}}$ induced by $\alpha \mapsto \overline{\alpha}$.
- A bilinear map
$$B : H^{1,0} \times H^{0,1} \to \mathbf{C},$$

defined by

$$B(\alpha, [\theta]) = \int_X \alpha \wedge \theta.$$

(This is well defined since changing the representative θ to $\theta + \overline{\partial} f$ changes the integral by

$$\int_X \alpha \wedge \overline{\partial} f = -\int_X \overline{\partial}(f\alpha),$$

which vanishes by Stokes' Theorem.)

- A map $i : H^{1,0} \to H^1$ defined by mapping a holomorphic (hence closed) 1-form to its cohomology class.
- A map $\nu : H^{1,1} \to H^2$ defined to be the natural map indced from the inclusion $\mathrm{Im} : \overline{\partial} : \Omega^{0,1} \to \Omega^2 \subset \mathrm{Im} : d : \Omega^1 \to \Omega^2$,

Theorem 5 *Let X be a compact connected Riemann surface.*

1. *The map σ induces an isomorphism from $H^{1,0}$ to $\overline{H^{0,1}}$.*
2. *The pairing B induces an isomorphism $H^{0,1} \cong (H^{1,0})^*$.*
3. *The map $H^{1,0} \oplus H^{0,1} \to H^1$ defined by*

$$(\alpha, \theta) \mapsto i(\alpha) + \overline{i(\sigma^{-1}(\theta))}$$

is an isomorphism.

4. *The map $\nu : H^{1,1} \to H^2$ is an isomorphism.*

The proofs are entirely straightforward applications of the main theorem. To show that σ is surjective we start with any class $[\theta]$ in $H^{0,1}$. We want to find a representative $\theta' = \theta + \overline{\partial} f$ such that $\partial\theta = 0$. For this means that $\alpha = overline\theta'$ is a holomorphic 1-form and $[\theta] = -\sigma(\alpha)$. Thus we want to solve the equation

$$\partial\overline{\partial} f = -\partial\theta.$$

Since $\partial\overline{\partial} = \frac{1}{2} i\Delta$ the main theorem tells us that we can solve this equation provided the integral of $\partial\theta$ vanishes, but this is so by Stokes' theorem.

Now the composite of the map i with the bilinear pairing B is up to factor the Hermitian form

$$\langle\alpha, \beta\rangle = \int_X \alpha \wedge \overline{alpha}$$

which we know is positive definite. It follows that the map i must be injective and in turn that B is a dual pairing. We leave the other parts as an exercise for the reader.

We see in particular from the Corollary that both $H^{1,0}, H^{0,1}$ are complex vector spaces of dimension g. Thus the genus, which was initially a topological invariant appears also as teh crucial numerical invariant of the complex geometry of a Riemann surface.

We can give some simple consequences of the above.

Corollary 3 *Any compact Riemann surface X with $H^1(X) = 0$ is equivalent to the Riemann sphere.*

Corollary 4 *Any compact Riemann surface with $g = 1$ is equivalent to a torus $\mathbf{C}/\Lambda$.*

Corollary 5 *Let X be a compact riemann surface of genus g and let $p_1, \ldots, p_g$ be distinct points on X. Then there is a non-constant meromorphic function on X with poles at some subset of the p_i.*

Chapter 9

Proof of the Main Theorem

9.1 Discussion and motivation

We will now embark on the proof of our main analytical result, thm:mainthm for compact Riemann surface. Before getting to work we give some preliminary discussion. The Theorem really consists of the three statements, on a compact connected Riemann surface X,

- If there is a solution to the equation $\Delta\phi = \rho$ then the integral of ρ is zero.
- If there is a solution it is unique up to the addition of a constant.
- Conversely if ρ is a form of integral 0 we can find a solution ϕ.

The first and second of these three statements are very easy to prove, so the real content is the third statement. The first statement follows immediately from Stokes' Theorem, since for any ϕ,

$$\int_X \Delta\phi = 2i \int_X \overline{\partial}\partial\phi = 2i \int_X d(\partial\phi) = 0.$$

The second statement is equivalent to the assertion that the only harmonic functions—solutions of the equation $\Delta f = 0$—are the constants. One can see this in two ways: either by the *maximum principle* for harmonic functions or by considering the *Dirichlet integral*. For the first one considers a point in X where f is maximal, which exists by the compactness assumption, and applies the maximum principle at that point. For the second one writes

$$\int_X |df|^2 = \int_X f\Delta f = 0,$$

when $\Delta f = 0$. Thus df vanishes everywhere on X and f is a constant. These two proofs of the uniqueness are both very simple but the two approaches are manifestations of two different approaches—via the maximum principle or Dirichlet integral— which can be taken to the whole theory. To illustrate this consider a problem which is closely related to that considered our Theorem, the solution of the Dirichlet boundary value problem. Here we consider a bounded domain $\Omega \subset \mathbf{C}$ with smooth boundary $\partial\Omega$ and a given function g on the boundary. The problem is to solve the equation $\Delta\phi = 0$ in Ω with the boundary condition that ϕ has a continuous extension to $\overline{\Omega}$, equal to g on the boundary. Supposing that a solution ϕ exists one can show that it is characterised by two different extremal properties.

1. For each x in Ω,
$$\phi(x) = \min\{\psi(x) : \psi|_{\partial\Omega} = g, \Delta\psi \geq 0\}.$$
2. The function ϕ minimises $\int_\Omega |\nabla\psi|^2$ over all functions ψ on $\overline{\Omega}$ with $\psi|_{\partial\Omega} = g$.

Conversely, one can prove the existence of a solution ϕ by showing that such extremal functions exist. Both of these approaches have their own merits and generalise in different ways. The line we will take in the proof of the Main Theorem, and the further results in the next Chapter, will follow the Dirichlet integral approach, closest to the heuristic arguments originally employed by Riemann.

The equation $\Delta\phi = \rho$ is, in local co-ordinates, the *Poisson equation* which may be familiar from potential theory in $\mathbf{R}^n$. In this vein, one can obtain some physical intuition into why the main theorem should be true, as follows. In this discussion we will anticipate a result proved in Chapter 13, that an oriented surface in $\mathbf{R}^3$ is naturally a Riemann surface. Suppose the Riemann surface X arises this way. We get a standard area form on X, so we can identify the 2-forms and the functions. Think of this surface in $\mathbf{R}^3$ as being made of a thin metal sheet and the function ϕ as being the temperature distribution over the sheet. The function ρ represents some externally imposed source or removal of heat, varying over the surface. Then the Poisson equation $\Delta\phi = \rho$ is the equation for a steady-state temperature distribution, and the content of our Theorem is that if the integral of ρ is zero—so there is no overall gain or loss of heat—then such a temperature distribution exists. One can also think about this in terms of the time-dependent heat equation

$$\frac{\partial \phi}{\partial t} = \rho - \Delta\phi.$$

Here ϕ is now a function on $X \times [0, \infty)$ while we still suppose that ρ is constant in time. One can prove analytically, in line with ones physical intuition, that if the integral of ρ is zero then for any initial temperature distribution this heat equation has a solution which converges as $t \to \infty$ to a solution of . the Poisson equation.

One case in which the Theorem can be proved easily is that of a torus. Suppose for example that X is the quotient of $\mathbf{C}$ by the "square" lattice $2\pi\mathbf{Z} \oplus 2\pi i\mathbf{Z}$. We take standard real angular co-ordinates θ_1, θ_2 on X and identify functions with 2-forms in the obvious way. Any smooth function on X can be written as a c=double Fourier series

$$f = \sum_{n,m} f_{nm} e^{in\theta_1 + im\theta_2}.$$

The Laplacian of such a function is

$$\Delta f = \sum (n^2 + m^2) f_{n,m} e^{in\theta_1 + im\theta_2},$$

and the integral of f, with respect to the standard area form, is $4\pi^2 f_{00}$. Thus if ρ has integral zero we can write down the solution to the Poisso equation in the form

$$\phi = \sum_{(m,n)\neq(0,0)} \frac{1}{m^2 + n^2} \rho_{mn} e^{im\theta_1 + n\theta_2},$$

where ρ_{mn} are the Fourier coefficients ofρ.

As a final remark, it is worth pointing out that our Theorem fits into a wider setting of elliptic differential operators on compact manifolds. We do not pause to explain what is meant by an elliptic operator: suffice it to say that this is a class of linear differential operators which includes the Laplace operator Δ as a particular case. If $\mathcal{L}$ is any linear differential operator over a compact manifold, and if we choose appropriate volume forms etc., there is a formal adjoint operator $\mathcal{L}^*$, characterised by the fact that for any f, g

$$\langle \mathcal{L} f, g \rangle = \langle f, \mathcal{L}^* g \rangle,$$

where $\langle\ ,\ \rangle$ denotes the L^2 inner product. Then the main result is that if $\mathcal{L}$ is elliptic one can solve the equation

$$\mathcal{L}\phi = \rho$$

if and only if ρ is orthogonal, in the L^2 sense, to the kernel of $\mathcal{L}^*$. In the Riemann surface situation of Theorem *, if we choose a an area form on the surface to identify functions and 2-forms the formal adjoint of the Laplacian is the same operator, $\Delta^* = \Delta$, so the kernel of Δ^* consists of the constant functions and the condition that ρ be orthogonal to this kernel is just that the integral of ρ is zero.

9.2 The Riesz representation theorem

We will now begin the proof. As we have said above, our approach will hinge on the Dirichlet integral and an efficient way to build this in to the argument uses the language of Hilbert spaces. Recall that we have defined the Dirichlet norm and inner product on functions on X. The norm and inner product are unchanged if we modify our functions by the additions of constants. We let $C^\infty(X)/\mathbf{R}$ be the vector space obtained by dividing out by the constant functions, so the norm and inner product descend to this quotient. Then we have

Proposition 26 *The Dirichlet norm and inner product make $C^\infty(X)/\mathbf{R}$ into a pre-Hilbert space.*

(Our notation will sometimes blur the distinction between a function on X and the equivalence class in $C^\infty(X)/\mathbf{R}$ which it represents. Note that if we fix a metric, i.e. area form, on X we can identify $C^\infty(X)/\mathbf{R}$ with the space of functions on X of integral zero.)

Now suppose that ρ is a 2-form on X. For any functions ϕ, ψ on X we have

$$\int_X \psi(\rho - \Delta\phi) = \int_X \psi\rho - \int_X \psi\Delta\phi = \int_X \psi\rho - \int_X \nabla\phi.\nabla\psi = \int_X \psi\rho - \langle\phi, \psi\rangle_D.$$

By a simple, standard, argument the equation $\Delta\phi = \rho$ is equivalent to the condition that for all functions ψ we have

$$\int_X \psi(\rho - \Delta\phi) = 0.$$

Thus the content of

Theorem 6 *:mainthm* can be reformulated as the assertion that, when $\int_X \rho = 0$ there is a function ϕ such that

$$\int_X \psi\rho = \langle \psi, \phi \rangle_D,$$

for all $\psi \in C^\infty(X)$. To write this more compactly, define

$$\hat{\rho}(\psi) = \int_X \rho\psi.$$

Then, if the integral of ρ is zero, this induces a linear map

$$\hat{\rho} : C^\infty(X)/\mathbf{R} \to \mathbf{R},$$

and our problem is to find a ϕ such that

$$\hat{\rho}(\psi) = \langle \psi, \phi \rangle_D$$

for all ψ.

In this formulation, our problem falls into the class covered by the well-known Riesz representation theorem from Hilbert space theory.

Theorem 7 *Let H be a real Hilbert space and $\sigma : H \to \mathbf{R}$ be abounded linear map (so there is a constant C such that $|\sigma(x)| \leq C\|x\|$ for all $x \in H$). Then there is a $z \in H$ such that*

$$\sigma(x) = \langle z, x \rangle,$$

for all x in H.

For a proof see almost any elementary functional analysis text book

With all this background in place, we can see that the proof of * divides into two parts. First, we want to put ourself into the position where we can apply the Riesz represention theorem and for this we need a Hilbert space. Thus we let H be the abstract *completion* of our pre-Hilbert space $C^\infty(X)/\mathbf{R}$, under the Dirichlet norm $\| \; \|_D$. A point of H is an equivalence class of Cauchy sequences (ψ_i) in $C^\infty(X)/\mathbf{R}$ under the equivalence relation $(\psi_i) \sim (\psi_i')$ if $\|\psi_i - \psi_i'\|_D \to 0$. The crucial thing we need now is

Theorem 8 *The functional $\hat{\rho} : C^\infty(X)/\mathbf{R} \to \mathbf{R}$ is bounded: there is a constant C such that $|\hat{\rho}(\psi)| \leq C\|\psi\|_D$ for all ψ in $C^\infty(X)/\mathbf{R}$.*

Assuming this for the moment, it follows that $\hat{\rho}$ *extends* to a bounded linear map from H to $\mathbf{R}$ (which we still denote by $\hat{\rho}$. (The proof is just to observe that for any Cauchy sequence (ψ_i) in $C^\infty(X)/\mathbf{R}$ the sequence $\hat{\rho}(\phi_i)$ is Cauchy in $\mathbf{R}$ so we can define the extension of $\hat{\rho}$ by taking the limit.) So we can apply the Riesz representation theorem and we conclude that there is a ϕ in the completion H with $\hat{\rho}(\psi) = \langle \phi, \psi \rangle_D$ for all ψ. An object of this type is called a *weak solution* to our problem and the other part of the proof of Theorem * is to establish

Theorem 9 *If ρ is a smooth 2-form on X of integral zero then a weak solution ϕ in H of * is smooth, i.e. lies in the subset $C^\infty(X)/\mathbf{R}$ of H.*

9.3 The heart of the proof

The foundation of our proof of Theorem * will be a result from calculus of two real variables. Suppose Ω is a bounded, convex, open set in $\mathbf{R}^2$. (For our applications it suffices to consder the case of a circular disc). Let A be the area of Ω and d be its diameter.

Theorem 10 *Let ψ be a smooth function on an open set containing the closure $\overline{\Omega}$ and let $\overline{\psi}$ denote the average*

$$\overline{\psi} = \frac{1}{A} \int_\Omega \psi d\mu,$$

where $d\mu$ is the standard Lebesgue measure on $\mathbf{R}^2$. Then for $x \in \Omega$ we have

$$|\psi(x) - \overline{\psi}| \leq \frac{d^2}{2A} \int_\Omega \frac{1}{|x-y|} |\nabla \psi(y)| d\mu_y.$$

(Here the notation is supposed to indicate that the variable of integration on the right hand side is $y \in \Omega$.)

To prove this there is no loss in supposing that the point x is the origin in $\mathbf{R}^2$ (applying a translation in $\mathbf{R}^2$) and that $\psi(0)$ is zero (changing ψ by addition of a constant). We work in standard polar co-ordinates (r, θ) on the plane. Thus we can write

$$\overline{\psi} = \frac{1}{A} \int_0^{2\pi} \int_0^{R(\theta)} \psi(r, \theta) r dr d\theta,$$

where $R(\theta)$ is the length of the portion of the ray at angle θ lying in Ω. (Here we use the fact that Ω is convex.) Now if we introduce another radial variable ρ we can write, for each (r,θ)

$$\psi(r,\theta) = \int_0^r \frac{\partial \psi}{\partial \rho} d\rho,$$

using the fact that ψ vanishes at the origin. So now we have

$$\overline{\psi} = \frac{1}{A}\int_0^{2\pi}\int_0^{R(\theta)}\int_{\rho=0}^{r} \rho\frac{\partial \psi}{\partial \rho} r \frac{\partial \psi}{\partial \rho} d\rho dr d\theta.$$

We interchange the order of the r and ρ integrals, so

$$\overline{\psi} = \frac{1}{A}\int_0^{2\pi}\int_{\rho=0}^{R(\theta)}\left(\int_{r=\rho}^{R(\theta)} r dr\right)\frac{\partial \psi}{\partial \rho} dr d\theta.$$

The innermost integral is

$$\int_{r=\rho}^{R(\theta)} r dr = \frac{1}{2}(R(\theta)^2 - r^2)$$

which is positive and no larger than $\frac{R(\theta)^2}{2}$, while, by definition, $R(\theta) \leq d$. Thus

$$|\overline{\psi}| \leq \frac{d^2}{2A}\int_0^{2\pi}\int_0^{R(theta)} \frac{1}{\rho}|\frac{\partial \psi}{\partial \rho}|\rho d\rho d\theta.$$

The modulus of the radial derivative $\frac{\partial \psi}{\partial \rho}$ is at most that of the full derivative $\nabla\psi$, so switching back to a co- ordinate free notation we have

$$|\overline{\psi}| \leq \frac{d^2}{2A}\int_\Omega \frac{1}{|y|}|\nabla\psi_y| d\mu_y,$$

as required.

Corollary 6 *Under the hypotheses above*

$$\int_\Omega |\psi(x) - \overline{\psi}|^2 d\mu_x \leq \frac{d^2\pi}{2A}\int_\Omega |\nabla\psi|^2 d\mu.$$

To prove this, and for later use, we recall the notion of the *convolution* of functions on $\mathbf{R}^2$. The convolution of functions f, g is defined by

$$(f * g)(x) = \int_{\mathbf{R}^2} f(y)g(x-y)d\mu_y,$$

The operation * is commutative and associative and if $\| \|_T$ is any translation-invariant norm on functions on $\mathbf{R}^2$ we have

$$\|f * g|_T \leq \|f\|_{L^1}\|g\|_T,$$

where $\|f\|_{L^1}$ is the usual L^1 norm

$$\|f\|_{L^1} = \int_{R^2} |f| d\mu.$$

In particular this holds when $\| \|_T$ is the L^2 norm

$$\|g\|^2_{L^2} = \int_{\mathbf{R}^2} |g|^2 d\mu.$$

(Strictly we should specify what class of functions we are considering in the definition of the convolution, but this will be clear in the different contexts as they arise.)

To prove the corollary, we define

$$K(x) = \frac{d^2}{2A}\frac{1}{|x|} \text{ for } |x| < d,$$

and $K(x) = 0$ if $|x| \geq d$. This has a singularity at the origin but is nevertheless an integrable function and

$$\|K\|_{L^1} = 2\pi \frac{d^2}{2A}\int_0^d dr = \frac{d^3\pi}{A}.$$

Define a function g on $\mathbf{R}^2$ by

$$g(y) = |\nabla\psi|^2(y),$$

if $y \in \Omega$ and $g(y) = 0$ if $y \notin \Omega$. Then $K * g$ is a positive function on $\mathbf{R}^2$ and Theorem * asserts that for all $x \in \Omega$,

$$|\psi(x) - \overline{\psi}| \leq |(K * g)(x).$$

It follows that

$$\int_\Omega |\psi(x) - \overline{\psi}|^2 d\mu_x \leq \|K * g\|_{L^2}^2 \leq \|K\|_{L^1}^2 \|g\|_{L^2}^2 \leq \left(\frac{d^3\pi}{A}\right)^2 \|\nabla\psi\|_{L^2}^2,$$

as asserted.

We can now prove *. We begin with the case when ρ is supported in a single coordinate chart in our Riemann surface, which we identify with a bounded convex set Ω in $\mathbf{C} = \mathbf{R}^2$. Working in this local coordinate system we use the Lebesgue area form to identify functions and 2-forms, so ρ can be regarded as a function of integral zero supported on Ω. Likewise, a function ψ on X can be regarded as defining a function, which we also call ψ, on a neighbourhood of Ω in $\mathbf{C}$ and we can write

$$\hat{\rho}(\psi) = \int_\Omega \rho\psi d\mu.$$

Now since the integral of ρ is zero we also have

$$\hat{\rho}(\psi) = \int_\Omega \rho(\psi - \overline{\psi}) d\mu,$$

and by the Cauchy-Schwartz inequality

$$|\int_\Omega \rho(\psi - \overline{\psi}) d\mu| \leq \|\rho\|_{L^2(\Omega)} \|\psi - \overline{\psi}\|_{L^2(\Omega)}.$$

Using * we then deduce that

$$|\hat{\rho}(\psi)| \leq C \|\nabla\psi\|_{L^2(\Omega)},$$

where $C = d^3\pi A\|\rho\|_{L^2(\Omega)}$. Finally

$$\|\nabla\psi\|_{L^2(\Omega)} \leq \|\nabla\psi\|_{L^2(X)} = \|\psi\|_D$$

which completes the proof of * in this case.

To treat a general 2-form ρ on X, of integral zero, we recall from * that integration over X defines an isomorphism from $H^2(X)$ to $\mathbf{R}$, so we can write $\rho = d\theta$ for some 1-form θ on X. We fix a cover of X by a finite number of coordinate charts $U_\alpha \subset X$ of the kind considered above, and choose a partition of unity χ_α subordinate to this cover. Put $\rho_\alpha = d(\chi_\alpha\theta)$. Then each ρ_α is supported in the corresponding coordinate chart U_α and

$$\int_X \rho_\alpha = \int_X d(chi_\alpha\theta) = 0$$

On the other hand

$$\rho = d\theta = d((\sum \chi_\alpha)\theta) = \sum \rho_\alpha.$$

Our previous argument shows that each of the linear maps $\hat{\rho}_\alpha$ is bounded and so $\hat{\rho} = \sum \hat{\rho}_\alpha$ is also (as a finite sum of bounded linear maps).

9.4 Weyl's Lemma

Suppose that ϕ is an element of H which is a weak solution to our problem in the sense explained above. That is, we have a sequence of functions ϕ_i on X which is Cauchy with respect the Dirichlet norm and, for any ψ

$$\langle \phi_i, \psi \rangle \to \hat{\rho}(\psi),$$

as i tends to infinity. We want to see first that we can identify the abstract object ϕ with a function (up to a constant) on X, where initially this function will just be locally in L^2 (i.e. represented by an L^2 function, in the ordinary sense, in any local co-ordinate chart). To do this we consider first any fixed co-ordinate chart, identified with $\Omega \subset \mathbf{C}$, as above. We can suppose, after changing the ϕ_i by the addition of suitable constants, that the integrals of the ϕ_i over Ω vanish and then, by cor:poinc, we have

$$\|\phi_i - \phi_j\|_{L^2(\Omega)} \leq C\|\phi_i - \phi_j\|_D.$$

Hence ϕ_i gives a Cauchy sequence in $L^2(\Omega)$ which converges to an L^2 limit by the completeness of L^2. We claim now that this same sequence ϕ_i converges localy in L^2 over *all* of X. We will give the argument in a form which will work equally well in the generalisation considered in the next Chapter. Let A be the set of points x in X with the property that there is a co-ordinate chart around x in which ϕ_i converges to ϕ in L^2. Then A is non-empty by the preceding discussion and A is open in X from the nature of its definition. Since X is connected, the complement of A is not open, so either $A = X$ or there is a point y which is in the closure of A but not in A. But in the latter case we could find a coordinate neighbourhood Ω' about y and a sequence of real numbers c_i' such that $\phi_i - {}'_{i}$ converges in L^2 over Ω'. But now there is a point x in $A \cap \Omega'$ and on a small neighbourhood of x both ϕ_i and $\phi_i - c_i'$ converge in L^2. This means that c_i' tends to 0 as $i \to \infty$, so in fact y is in A after all, a contradiction.

To sum up we now have a function ϕ on X which is locally in L^2 and which is a *weak solution* to the equation $\Delta\phi = \rho$. We need to show that ϕ is smooth. Since smoothness is a local property we can fix attention on a single co-ordinate chart. What we need is a version of "Weyl's lemma"

Proposition 27 *Let Ω be a bounded open set in $\mathbf{C}$ and ρ be a smooth 2-form on Ω. Suppose is an L^2 function on Ω with the property that for any smooth function χ of compact support in Ω*

$$\int_\Omega \Delta\chi\phi = \int_\Omega \chi\rho.$$

Then ϕ is smooth and satisfies the equation $\Delta\phi = \rho$.

The proof will involve a number of steps. The first step is to reduce to the case when ρ is zero. Since smoothness is a local property it suffices to prove that ϕ is smooth over any given interior set Ω', where we suppose that the ϵ neighbourhood of Ω' is contained in Ω. Then we can choose a ρ' equal to ρ on a neighboorhoos of the closure of Ω' and of compact support in Ω. Suppose we can find some smooth solution ϕ' of the equation $\Delta\phi' = \rho'$ over Ω. Then $\psi = \phi - \phi'$ will be a weak solution of the equation $\Delta\psi = 0$ on Ω'. If we can prove that ψ is smooth then so will ϕ be.

To find the smooth solution ψ' we use the "Newton potential" in two dimensions:

$$K(x) = \frac{1}{2\pi}\log|x|.$$

Of course this is not defined at $x = 0$ but K is well-defined as a locally integrable function on $\mathbf{C}$. For any smooth function f of compact support in $\mathbf{C}$ the convolution

$$K * f(x) = \int K(y)f(x-y)d\mu_y,$$

is defined and $K * f$ is smooth.

Lemma 14 • *If σ has compact support in $\mathbf{C}$ then $K * (\Delta\sigma) = \sigma$.*

- *If f has compact support then $\Delta(K * f) = f$.*

This Lemma essentially expresses the stadard fact that convolution with K furnishes an inverse to the Laplace operator. To prove the first assertion we may, by translation invariance, calculate at the point $x = 0$. Then

$$(K * \Delta\sigma)(0) = \int \frac{1}{2\pi}\log(|y\ vert)(\Delta\sigma)_y d\mu_y.$$

Now $\Delta \log |y \ vert$ vanishes on $\mathbf{C} \setminus \{0\}$. We write the integral as the limit as δ tends to zero of the integral over the set where $|y| \geq \delta$.We then use Green's identity to write this as a boundary integral and then take the limit as δ tends to zero. The argument is rather stabdard so we do not give more details.

For the second part we write

$$(K * f)(x) = \int K(y) f(x-y) d\mu_y.$$

When we take the Laplacian with respect to x there is no problem in moving the differential operator inside the integral, since f is smooth and x does not appear inside the argument of K. Thus

$$\Delta(K * f) = \int K(y) \Delta_x f(x-y) d\mu_y,$$

where the notation means that we take the Laplcian with respect to x. But this is just the same as $K * \Delta f$, which is equal to f by the first part.

We have now reduced to the case when $\rho = 0$ so, changing notation, let us suppose that ϕ is a weak solution of $\Delta\phi = 0$ on Ω and seek to prove that ϕ is smooth on the interior domain Ω', with the ϵ-neighbourhood of Ω' contained in Ω. The argument now exploits the mean value property of smooth harmonic functions. This says that if ψ is a smooth harmonic function on a neighbourhood of a closed disc then the value of ψ at the centre of the disc is equal to the average value on the circle boundary. Fix a smooth function β on $\mathbf{R}$ with $\beta(r)$ constant for small r and vanishing for $r \geq \epsilon$ and such that

$$2\pi \int_0^\infty r\beta(r) dr = 1.$$

Now let B be the function $B(z) = \beta(|z|)$ on $\mathbf{C}$. Then B is smooth and has integral 1 over $\mathbf{C}$ (with respect to ordinary Lebesgue measure). Suppose first that ψ is a smooth harmonic function on a neighbourhood of the closed ϵ-disc centred at the origin. Then we have

$$\int_{\mathbf{C}} B(-z)\psi(z) d\mu_z = \int_0^\infty \int_0^{2\pi} r\beta(r)\psi(r,\theta) d\theta dr = 2\pi\psi(0) \int_0^\infty r\beta(r) dr = 1$$

where we have switched to polar coordinates and used the mean value property

$$\int_0^{2\pi} \psi(r,\theta) d\theta = \psi(0).$$

Now the integral above is just that defining the convolution $B * \psi$ at 0. By translation invariance we obtain the following

Proposition 28 *Let ψ be a smooth function on $\mathbf{C}$ and suppose that $\Delta\psi$ is supported in a compact set $J \subset \mathbf{C}$. Then $B * \psi - \psi$ vanishes outside the ϵ-neighbourhood of J.*

We see in particular from this that *if* our function ϕ on Ω is smooth we must have $B * \phi = \phi$ in Ω'. Conversely, for any L^2 function ϕ the convolution $B * \phi$ is smooth. So proving the smoothness of ϕ in Ω' is equivalent to establishing the identity $B * \phi = \phi$ in $Omega'$. To do this we proceed as follows. It suffices to show that for any smooth test function χ of compact support in Ω' we have

$$\langle \chi, \phi - B * \phi \rangle = 0,$$

where we are writing $\langle\ ,\ \rangle$ for the usual "inner product"

$$\langle f, g \rangle = \int f g d\mu.$$

We use the fact that for any functions f, g, h in a suitable class

$$\langle f, g * h \rangle = \langle g * f, h \rangle.$$

This follows by straightforward re-arrangements of the integrals. We will not bother to spell out conditions on the functions involved, since it will the validity of the identity will be fairly obvious in our applications below.

Put $h = K * (\chi - B * \chi) = K * \chi - B * K * \chi$. Now $K * \chi$ is a smooth function on $\mathbf{C}$ and $\Delta K * \chi = \chi$ by the Lemma above. Thus $\Delta K * \chi$ vanishes outside the support of χ, and hence by the Proposition above $B * K * \chi$ equals $K * \chi$ outside the ϵ-neighbourhood of the support of χ. Thus h has compact support contained in Ω. So we can use h as a test function in the hypothesis that $\Delta\phi = 0$ weakly, i.e. we have

$$\langle \Delta h, \phi \rangle = 0.$$

But $\Delta h = \Delta K * (\chi - B * \chi) = \chi - B * \chi$ by the Lemma above (since χ and $B * \chi$ have compact support). So we see that

$$\langle \chi - B * \chi, \phi \rangle = 0.$$

But applying the identity above again this gives

$$\langle \chi, \phi - B * \phi \rangle = 0,$$

as desired.

Chapter 10

The Uniformisation Theorem

10.1 Statement

In this chapter we prove

Theorem 11 *Let X be a connected, simply connected, non-compact Riemann surface. Then X is equivalent to either* $\mathbf{C}$ *or the upper half-plane* H.

Corollary 7 *Any connected Riemann surface is equivalent to one of*

- *The Riemann sphere S^2;*
- $\mathbf{C}$ *or* $\mathbf{C}/\mathbf{Z} = \mathbf{C} \setminus \{0\}$ *or* $\mathbf{C}/\Lambda$ *for some lattice* Λ*;*
- *A quotient H/Γ where $\Gamma \subset PSL(2, \mathbf{R})$ is a discrete subgroup acting freely on H,*

The Corollary follows because any Riemann surface is a quotient of its universal cover by an action of its fundamental group, and we have seen that the only compact simply connected Riemann surface is the Riemann sphere.

Our proof of thm:unif will follow the same general pattern as the one we have already given to classify compact simply connected Riemann surface, but the noncompactness will require some extra steps. Most of our work goes into the proof of an analogue of the "Main Theorem". To state this, recall that if ϕ is real-valued function on a non-compact space X and c is a real number we say that ϕ tends to c at infinity in X if for all $\epsilon > 0$ there is a compact subset K of X such that $|\phi(x) - c| < \epsilon$ if x is not in K. We say that ϕ tends to $+\infty$ at infinity in X if for all $A \in \mathbf{R}$ there is an compact set $K \subset X$ such that $\phi(x) > A$ if x is not in K (and we say that ϕ tends to $-\infty$ at infinity in X if $-\phi$ tends to $+\infty$).

Theorem 12 *Let X be a connected, simply-connected, non-compact Riemann surface. Then if ρ is a (real) 2-form of compact support on X with $\int_X \rho = 0$ there is a (real valued) function ϕ on X with $\Delta\phi = \rho$ and such that ϕ tends to 0 at infinity in X.*

We now show that thm:potdecay implies thm:unif. Choose a point $p \in X$ and a local complex co-ordinate z around p. Using the same notation as in Section 5, we put

$$A = \overline{\partial}(\frac{\beta}{z})$$

where β is a cut-off function, so A is a $(0, 1)$-form supported in an annulus around p. We put $\rho = \partial A$, so ρ is a complex-valued 2-form with integral zero, by Stokes' Theorem. By the result (applied to the real and imaginary parts of ρ) we can find a complex-valued function g with $\partial\overline{\partial}g = \rho$, and with the real and imaginary parts of g tending to 0 at infinity in X. Now let a be the real 1-form

$$a = (A - \overline{\partial}g) + (\overline{A} - \overline{\overline{\partial}g}).$$

By construction $\partial(A - \overline{\partial}f) = 0$ and this means that $da = 0$. So, since $H^1(X) = 0$, there is a real-valued function ψ with $a = d\psi$. This means that $A = \overline{\partial}g + \overline{\partial}\psi$. Hence

$$\overline{\partial}(\frac{\beta}{z} - (g + \psi)) = 0$$

on $X \setminus \{p\}$. Hence $f = g + \psi$ is a meromorphic function on X with a simple pole at p and the imaginary part of f tends to zero at infinity in X, since ψ is real. (Strictly the imaginary part of f is not a function on X since f has a pole, but the meaning should be clear—to be precise we could say that the imaginary part of f tends to 0 at infinity on $X \setminus D$ where D is an open disc about p.)

The meromorphic function f is a holomorphic map from X to the Riemann sphere $S^2 = \mathbf{C} \cup \{\infty\}$. Let H_+, H_- denote the (open) upper and lower half-planes in $\mathbf{C}$. Let $X_\pm$ be the preimages $f^{-1}(H_\pm$ in X. So X_+ and X_-

are open subsets of X and $X_+ \cup X_-$ is dense in X, since f is an open map. The restrictions of f gives holomorphic maps

$$f_\pm : X_\pm \to H_\pm.$$

We claim that f_+ and f_- are proper maps. For if B is a compact subset of H_+, say, there is an $\epsilon > 0$ such that $\Im(z) > \epsilon$ for all z in B. The fact that the imaginary part of f tends to zero at infinity implies that $f^{-1}(B)$ is a compact subset of X, but this is the same as $f_+^{-1}(B) \subset X_+$.

We know that f yields a local homeomorphism from a neighbourhod of p in X to a neighbourhood of ∞ in S^2. What we see first from this is that X_+, X_- are both non-empty. So we have degrees $d_+, d_- \geq 1$ of f_+, f_-. We claim that $d_+ = d_- = 1$. To see this we apply the condition that $\Im(f)$ tends to zero at infinity to find a compact set K in X such that $\Im(f)(x) < 1$ if x is not in K. Suppose the degree of f_+ is at least 2. Then for each integer $n \geq 1$ we can find a pair of points $x_n, \tilde{x}_n \in X_+$ such that $f(x_n) = f(\tilde{x}_n) = in$ and either $x_n, \tilde{x}_n$ are distinct or $x_n = \tilde{x}_n$ and the derivative of f vanishes at x_n. The choice of K means that $x_n, \tilde{x}_n$ lie in this compact set, so we can find a subsequence $\{n'\}$ such that $x_{n'}$ and $\tilde{x}_{n'}$ converge to limits $x, \tilde{x}$. Since the points in, regarded as points of the Riemann sphere, converge to ∞, we must have $f(x) = f(\tilde{x}) = \infty$ and since f has just one pole we must have $x = \tilde{x} = p$. But now we get a contradiction to the fact that f is a local homeomorphism, with non-vanishing derivative, on a neighbourhhod of p.

So now we know that f maps $X_\pm$ bijectively to $H_\pm$ in $\mathbf{C}$. We claim next that f is an injection from X to S^2. For if x_1, x_2 are distinct points of X with $f(x_1) = f(x_2) = Z \in S^2$ we can find disjoint open discs D_1, D_2 about x_1, x_2 and a neighbourhood N of $Z \in S^2$ such $f(D_1), f(D_2)$ each contain N. Now pick a point Z' in $N \cap H_+$. There are distinct points x_1', x_2' in D_1, D_2 which map to Z' and this contradicts the fact that f is injective on X_+.

Now we know that f maps X injectively to an open subset U of the Riemann sphere, containing $H_+ \cup H_- \cup \{infty\}$. That is

$$U = S^2 \setminus I$$

for some compact subset I of $\mathbf{R}$. Thus f yields an equivalence between X and this subset U. If I has more than one component then $\pi_1(U)$ is non trivial which would contradict the fact that X is simply connected. So we conclude that either

1. I is proper closed interval $[a, b]$ for $a < b$, or,

2. I is a single point in $\mathbf{R}$.

This completes the proof, since in the first case it is easy to write down a holomorphic equivalence between $S^2 \setminus [a,b]$ and the upper half plane and in the second case it is obvious that the complement of any point in the Riemann sphere is equivalent to $\mathbf{C}$.

10.2 Proof of the analogue of the Main Theorem

10.2.1 Set-up

We now turn to the proof of thm:potdec. In the proof we will make use of two facts which we state now.

Proposition 29 *Let X be a connected, simply connected, non-compact surface. Then for any compact set $K \subset X$ the complement $X \setminus K$ has exactly one connected component whose closure is not compact.*

(Remark: one says that a surface which satisfies the condition in the second sentence of this Proposition has "only one end". Thus the statement is that a simply connected surface has only one end.)

Our second fact involves calculus on surfaces. Recall from Chapter 5 the notion of the "modulus" $|\rho|$ of a 2-form.

Proposition 30 *Let S be a smooth oriented surface and let $F : S \to \mathbf{R}^2$ be a smooth map. Then for any compact set K in S*

$$\mu(F(K)) \leq \int_S |F^*(dx_1 dx_2)|,$$

where x_1, x_2 are standard co-ordinates on $\mathbf{R}^2$ and μ denotes Lebesgue measure on $\mathbf{R}^2$.

(Of course we need to know here that $F(K)$ is a Lebesgue measureable subset of $\mathbf{R}^2$ but this follows from general facts that will be discussed later.)

We hope that each of prop:oneend and prop:areabound are, in different ways, intuitively plausible, at least, and we postpone the proofs so that we can get on with the main argument.

To prove thm:potdecay we adopt the same strategy as for the proof of the "Main Theorem" in the compact case. There is one crucial additional step required (in subsection (10.2.3) below) but to set this stage for this we will need to prove a number of elementary, but slightly delicate, preliminary results (in subsection (10.2.2)).

First, just as before we can reduce to the case when the 2-form ρ is supported in a co-ordinate disc D about p (this occurs anyway for our application). We consider the vector space Ω^0_c of compactly supported real valued functions on X with the Dirichlet norm

$$\|f\|_D^2 = i\int \overline{\partial} f \wedge \partial f.$$

Notice that this is now a genuine norm, since the constant functions do not have compact support. The proof that the functional $\hat{\rho}$ is bounded goes through just as before. We let be the completion of Ω^0_c under this norm and the Riesz representation theorem gives an element ψ say of H such that $\hat{\rho}(f) = \langle f, \psi\rangle_D$. Just as before we can find a sequence f_i in Ω^0_c converging to ψ in H and a sequence of constants $c_i \in \mathbf{R}$ such that $\psi_i = f_i + c_i$ converges in L^2 over some co-ordinate disc. Again, the same argument as before shows that $f_i + c_i$ converges in L^2 over any co-ordinate disc. We should note however that ψ_i need not have compact support. The same argument as before shows that ψ is smooth and satisfies the desired equation $\Delta\phi = \rho$. What we achieve at this stage in the argument is summarised by the following Proposition.

Proposition 31 *Let ρ be a 2-form of integral 0 supported in a co-ordinate disc D. There is a smooth function ϕ on X which satisfies the equation $\Delta\phi = \rho$, and a sequence ϕ_i of smooth functions on X which have the following properties*

1. *There are real numbers c_i and compact sets $B_i \subset X$ such that $\phi_i = c_i$ outside B_i.*

2. *For any 1-form α of compact support on X the norms $\|(\phi - \phi_i)\alpha\|$ tend to zero as i tends to infinity.*

3. *The norms $\|d\phi - d\phi_i)\|$ tend to zero as i tends to infinity.*

What should be clear now is that the only essentially new thing required to prove thm:potdecay is to arrange that ϕ tends to zero at infinity in X. Of course it is equally good to arrange that ϕ tends to a finite constant c at infinity, since we can replace ϕ by $\phi - c$. Notice here that there is an exceptional case when the derivative of ϕ vanishes outside a compact set. In that case prop:oneend implies immediately that ϕ tends to a constant at infinity and we are done. So we can suppose that the derivative of ϕ does not vanish on any open set in $X \setminus supp(\rho)$.

We also note here

Lemma 15 *label:compeq Let J be any compact set in X. There is a sequence ϕ_i satisfying the conditions of prop:summup and with $\phi_i = \phi$ on J.*

To see this let χ be a smooth function of compact support, equal to 1 on J. Given ϕ_i and ϕ as in prop:summup, we define

$$\phi_i' = \chi\phi + (1-\chi)\phi_i = \phi_i + \chi(\phi - \phi_i).$$

Thus $f_i' = \phi_i' - c_i$ have compact support, since $\phi_i = \phi_i'$ outside the fixed compact set $supp(\chi)$ and $\phi_i' = \phi$ on Γ. We have

$$d\phi_i' - d\phi_i = \chi(d\phi - d\phi_i) + (d\chi)(\phi - \phi_i),$$

and it follows that

$$\|f_i' - f_i\|_D \to 0$$

as $i \to \infty$. This means that the sequence ϕ_i' has the same properties (as required in prop:summup) as ϕ_i.

10.2.2 Classification of behaviour at infinity

We begin with an elementary Lemma, which applies to functions on any non-compact space.

Lemma 16 *Suppose ϕ is a continous function on X then one of the following four statements holds*

1. *There is a constant $c \in \mathbf{R}$ such that ϕ tends to c at infinity in X;*
2. *ϕ tends to $+\infty$ at infinity in X;*
3. *ϕ tends to $-\infty$ in X;*

4. *There are real numbers α, β with $\alpha < \beta$ such that $\phi^{-1}((-\infty, \alpha])$ and $\phi^{-1}([\beta, \infty)$ are both noncompact subsets of X.*

To prove this, let

$$A^{-} = \{t \in \mathbf{R} : \phi^{-1}(-\infty, t] \text{ iscompact}\}.$$

Clearly if t_- is in A^- then any $t < t_-$ is also in A^-. Symmetrically, put

$$A^{+} = \{t \in \mathbf{R} : \phi^{-1}[t, \infty) \text{ iscompact}\}.$$

Then for any elements $t_\pm \in A_\pm$ we must have $t_- < t_+$ otherwise $\phi^{-1}(\mathbf{R}) = X$ would be compact, a contradiction.

Define $c_- = \sup A_-$ and $c_+ = \inf A_+$ with the convention that $c_- = -\infty$ if A_- is empty and $c_+ = +\infty$ if A_+ is empty. Thus, with obvious conventions involving $\pm\infty$, we have $c_- \leq c_+$.

Suppose $c_- = c_+$ is a finite value $c \in \mathbf{R}$. Then for any $\epsilon > 0$ we have $c \pm \epsilon \in A_\pm$ so $\phi^{-1}(-\infty, c - \epsilon]$ and $\phi^{-1}[c + \epsilon, \infty)$ are compact and $|\phi - c| \leq \epsilon$ outside the union of these two compact sets, which is compact. Thus ϕ tends to c at infinity in X. Likewise if $c_- = c_+ = \pm\infty$ we find that ϕ tends to $\pm\infty$ at infinity in X.

Suppose $c_- < c_+$. Then we can find real numbers α, β with $c_- < \alpha < \beta < c_+$ and by definition $\phi^{-1}(-\infty, \alpha]$ and $\phi^{-1}([\beta, \infty)$ are noncompact.

Now in our situation if ϕ tends to any real number c at infinity we can replace ϕ by $\phi - c$ to get a solution to our problem. Thus our task is to rule out the other three possibilities in Lemma *.

Lemma 17 *The function ϕ does not tend to $+\infty$ at infinity in X.*

Suppose it did and let C be the maximum value of ϕ on the compact set *supp*(ρ). Then the set $K = \phi^{-1}(-\infty, C + 1]$ is compact. The points in $K \setminus \mathit{supp}(\rho)$ where the derivative of ϕ vanishes form a discrete set (since a harmonic function is locally the real part of a holomorphic function). It follows that we can find some $t_0 \in (1/2, 1)$ such that the derivative of ϕ does not vanish on $\phi^{-1}(t_0)$ (since we only need to avoid a finite number of points). Then $K_0 = \phi^{-1}(-\infty, C + t_0]$ is a compact surface with boundary in X, containing the support of ρ in its interior. Now, by Stokes Theorem,

$$\int_{K_0} \Delta\phi = \int_{\partial K_0} d\phi.$$

On the one hand the integral on the left is zero (since it equals the integral of ρ) and on the other hand the boundary integral on the right is strictly positive since the derivative of ϕ does not vanish.

Of course the same argument shows that ϕ does not tend to $-\infty$, so our task is to rule out the fourth possibility in Lemma *.

For $\alpha \in \mathbf{R}$ we write $Y_\alpha = \phi^{-1}(\infty, \alpha]$.

Lemma 18 *Any compact connected component of Y_α intersects supp(ρ).*

This is a simple consequence of the maximum principle. Suppose Z is a compact connected component which does not meet *supp*(ρ). There is a point x in Z which minimises ϕ over Z and then the maximum principle implies that this can only happen if the derivative of ϕ vanishes near x, contrary to our assumption.

Remark One can also argue as follows. Suppose for simplicity that Z has a smooth boundary then

$$\int_Z \overline{\partial}\phi \wedge \partial\phi = \int_{\partial Z} (\phi - \alpha)\partial\phi = 0.$$

It is not hard to handle the case when the boundary of Z is not smooth, but the maximum principle argument is easier here.

Lemma 19 *For any point q in Y_α which is not in supp(ρ) and any neighbourhood N of q there is an open disc D_q centred on q and contained in N such that $Y_\alpha \cap D_q$ is path-connected.*

This is rather obvious. We write the function $\phi - \alpha$ as the real part of a holomorphic function in a neighbourhood of q. Then this function is given in a suitable holomorphic co-ordinate as z^k and the set

$$\{z \in \mathbf{C} : |z| < 1 : \Re(z^k) \leq 0\}$$

is path connected.

Proposition 32 *Suppose Y_α is not compact. Then there is non-compact path-connected subset of Y_α.*

Let $D' \subset D$ be an open interior disc whose closure lies in D and such that D' contains the support of ρ. In our local coordinate about p we can take D to correspond to $|z| < 1$ and D' to $|z| < r$, for some finxed $r < 1$. Let C be the circle corresponding to $|zvert = (1 + r)/2$. It follows from lem:obvious that we write the intersection $Y_\alpha \cap C$ as the union of finitely many pieces $C_1, \ldots C_N$ say such that any two points in the same C_j can be joined by a path in $Y_\alpha \setminus D'$. (We cover the compact set $Y_\alpha \cap C$ by finitely many small discs of the form D_q considered in the Lemma.) Now consider the path connected components of $Y_\alpha \setminus D'$. Notice that by lem:obvious these are the same as the connected components. If one of these is components is non-compact we are done. So suppose all these components are compact. Since D' has compact closure and Y_α is not compact, there must be infinitely many different compact components of $Y_\alpha \setminus D'$. The circle C divides X into two connected components one of which is a disc D'' containing p. Any compact component of $Y_\alpha \setminus D'$ must either intersect the circle C or lie entirely within the disc D'', for otherwise it lies in $X \setminus \overline{D''}$ and gives a compact component of Y_α which does not meet $supp(\rho)$. Contradicting Lemma *. The union of the components lying in D'' is contained in a compact set—the closed disc $\overline{D''}$. So there must be infinitely many different components which intersect C. But two of these must meet the same subset C_j and hence we get a contradiction since the points in C_j can be joined by paths in $Y_\alpha \setminus D'$.

Putting together our results from this subsection we have

Proposition 33 *Either ϕ tends to a finite limit at infinity in X, or there are real numbers $\alpha < \beta$ and non compact, closed, path connected subsets $Z_\alpha, Z_\beta \subset X$ such that $\phi(x) \leq \alpha$ for $x \in Z_\alpha$ and $\phi(x) \geq \beta$ for $x \in Z_\beta$.*

10.2.3 The main argument

Our task now is to show that the second alternative of prop:summupmore does not occur, so we suppose it does and argue for a contradiction. Fix points $x_\alpha \in Z_\alpha, x_\beta \in Z_\beta$ and a compact path-connected set Γ containing x_α and x_β (for example the image of some path between the two points. By lem:compeq there is no loss of generality in supposing that $\phi_i = \phi$ on Γ.

Our preparations are now complete. For each i neither of the noncompact sets Z_α, Z_β can be contained in the compact set $supp(f_i)$. Moreover they must intersect a noncompact component of $X \setminus supp(f_i)$ and by prop:oneend there is only one such component. Thus there is a path in $X \setminus supp(f_i)$ joining point y_α of Z_α to a point y_β of Z_β. Let $\gamma_i : [0,1] \to X$ be a loop in X starting and ending at x_α of the following form:

- On the interval $0 \leq t \leq 1/4$, γ_i traces out a path from x_α to x_β in the set Γ.
- On the interval $1/4 \leq t \leq 1/2$, γ_i traces out a path in Z_β from x_β to y_β.
- On the interval $1/2 \leq t \leq 3/4$, γ_i traces out a path in $X \setminus supp(\rho)$ from y_β to y_α.
- On the interval $3/4 \leq t \leq 1$, γ_i traces out a path in Z_α from y_α to x_α.

Since X is simply connected this loop is contractible, so we can find a compact set $K_i \subset X$, containing the image of γ_i, such that γ_i is contractible in K_i.

Now consider the smooth map $F : X \setminus \mathbf{R}^2$ defined by

$$F(x) = (\phi(x), \phi_i(x)).$$

Then the composite $F \circ \gamma$ is a loop in $\mathbf{R}^2$ with the properties that

- For $0 \leq t \leq 1/4$, $F \circ \gamma$ maps into the diagonal $\{(x,x)\}$ in $\mathbf{R}^2$.
- For $1/4 \leq t \leq 1/2$, $F \circ \gamma$ maps into the half plane $\{(x_1,x_2) : x_1 \geq \beta\}$.
- For $1/2 \leq t \leq 3/4$, $F \circ \gamma$ maps into the horzontal line $\{(x_i, c_i)\}$.
- For $3/4 \leq t \leq 1$, $F \circ \gamma$ maps into the half-plane $\{x_1, x_2) : x_1 \leq \alpha\}$.

Let Z be the subset of $\mathbf{R}^2$ given by

$$Z = \{(x_1,x_2) : \alpha < x_1 < \beta, min(x_1, c_i) < x_2 < max(x_1, c_i).$$

There are possible pictures for Z, depending whether c lies between α and β , above β or below α. Howver it is elementary to see that in any case the area satisfies

$$\mu(Z) \geq \frac{1}{4}(\beta - \alpha)^2.$$

The conditions on our loop $F \circ \gamma$ imply that it does not meet Z. Now we have

Lemma 20 *If P is any point of S the winding number of $F \circ \gamma$ around P is equal to ± 1.*

This is straightforward algebraic topology.

Corollary 8 *F maps the compact set $K \subset X$ onto Z.*

This follows because if P is not in the image $F(K)$ the composite of F with a contraction of γ would give a contraction of $F \circ \gamma$ in $\mathbf{R}^2 \setminus \{P\}$, contradicting the homotopy invariance of the winding number.

Now we can apply prop:areabound to deduce that the integral of the modulus of the form $F_i^*(dx_1 dx_2)$ must be at least $\frac{1}{4}(\beta^2 - \alpha^2) = \delta$ say. Writing

$$F_i^*(dx_1 dx_2) = d\phi \wedge d\phi_i$$

we can state a conclusion to our preceding arguments as follows.

Corollary 9 *If ϕ does not tend to a finite limit at infinity in X, then there is some $\delta > 0$ such that for each i*

$$\int_X |d\phi \wedge d\phi_i| \geq \delta.$$

Now we write

$$d\phi \wedge d\phi_i = (d\phi - d\phi_i) \wedge d\phi_i$$

and apply Lemma * from Chapter 5 to get

$$\int_X |d\phi \wedge d\phi_i| \leq \|d\phi - d\phi_i\| \| \| d\phi_i\|,$$

but this tends to zero by cor:summup and we have the desired contradiction.