lies on the boundary, then it is determined up to the identifications $-\frac{1}{2} + it \sim \frac{1}{2} + it$ on the vertical boundary and $z \sim -1/z$ on the circular part.

Notice that since $z \mapsto z+1$ and $z \mapsto -1/z$ are holomorphic functions, the domain $\mathcal{M}$ with the given identifications possesses a natural complex structure, with the exception of the two 'conic' points i and $\frac{\pm 1+\sqrt{3}i}{2}$, where the total angle after making identifications collapses to π and $2\pi/3$, respectively. This can be relieved by introducing the coordinates $w = (z-i)^2$ and $w = (z - \frac{1+\sqrt{3}i}{2})^3$ near those points. However, it turns out to be more useful to keep the conic points and consider $\mathcal{M}$ as a complex surface with two conical singularities somewhat similar to the standard cone (1.3). It is called the *modular surface*, and plays an extraordinarily important role in number theory and the theory of group representations.

We have thus encountered a very interesting phenomenon: the collection of classes of Riemann surfaces on the torus (up to holomorphic equivalence), is itself naturally endowed with the structure of a Riemann surface! The presence of a complex structure on this collection of equivalence classes, called *Teichmüller space*, is a simple, albeit highly non-trivial, manifestation of a general phenomenon seen throughout different areas of mathematics, wherein the set of invariants of a structure of a certain kind itself possesses a similar structure.

Lecture 20.

a. Differentiable functions on real surfaces. In various aspects of the study of surfaces, an important role is played by the class of 'nice' functions on a given surface. For complex (Riemann) surfaces, the natural class to consider is the set of compex-valued holomorphic functions, while for real smooth surfaces, one considers differentiable real-valued functions. There is an important difference here; in the complex case, we deal with functions of one (complex) variable, and so the dimensions of the domain and the range are same, while in the real case, we consider functions of two (real) variables. In the complex case, then, the level set of a given value z, that is, the set of points on the surface at which f takes the value z, is generally a discrete set of points, while in the real case, the level set is usually a smooth curve. In particular, this allows the possibility of 'building up' a real smooth surface by considering the level sets of a sufficiently nice function; this procedure, which we will do later on in this lecture, is one of the basic constructions of Morse theory.

Definition 3.13. Given a function $f\colon S \to \mathbb{R}$ on a smooth surface, we say that f is *differentiable* if its coordinate representation $f \circ \phi^{-1}\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable for every chart $\phi\colon U \to \mathbb{R}^2$.

We first note that if f is differentiable in one coordinate chart on a neighbourhood, then it is differentiable in any other chart on that same neighbourhood. Indeed, if we have two charts $\phi\colon U \to D^2$ and $\psi\colon V \to D^2$, the coordinate representation of f using ϕ is given by

$$f_U = f \circ \phi^{-1}\colon D^2 \to \mathbb{R}$$

and the representation using ψ is

$$\begin{aligned} f_V &= f \circ \psi^{-1} \\ &= (f \circ \phi^{-1}) \circ (\phi \circ \psi^{-1}) \\ &= f_U \circ (\phi \circ \psi^{-1}) \end{aligned}$$

The transition map $\phi \circ \psi^{-1}$ is smooth and has smooth inverse, so f_V is differentiable on $\psi(U \cap V)$ iff f_U is differentiable on $\phi(U \cap V)$.

Definition 3.14. Given a chart $\phi\colon U \to D^2$ and a function $f\colon S \to \mathbb{R}$, the point $p \in U$ is a *critical point* for f if the gradient $\nabla(f \circ \phi^{-1})$ vanishes at p. If the gradient is nonzero at p, we say that p is a *regular point.*

Differentiating the above formula relating f_V and f_U, we have

$$\nabla f_V = D(\phi \circ \psi^{-1})\nabla f_U$$

where $D(\phi \circ \psi^{-1})$ is the Jacobian of the transition map. By the axioms of a smooth manifold, this has nonzero determinant and hence is invertible, so $\nabla f_V = 0$ if and only if $\nabla f_U = 0$. We have proved the following:

Lemma 3.15. *The critical points of a differentiable function are independent of the particular choice of coordinate chart.*

We now show that away from its critical points, any function can be made to assume a standard form by choosing an appropriate coordinate chart.

Lemma 3.16. *Given a differentiable function $f\colon S \to \mathbb{R}$ and a regular point $p \in S$, there exists a chart $\phi\colon U \to D^2$ around p in which $f_U(x,y) = f(\phi^{-1}(x,y)) = f(p) + x$.*

Proof. Take any coordinates (u,v) around p; because p is not a critical point, we may assume without loss of generality that $\frac{\partial f}{\partial u} \neq 0$. (Here we are abusing notation by using f to stand for both the function $S \to \mathbb{R}$ and its coordinate representation $D^2 \to \mathbb{R}$).

Then by the Implicit Function Theorem, we may write v as a function of f and u, and hence we can use these as our coordinates. □

The next exercise establishes a similar result in the complex case.

Exercise 3.10. Given a holomorphic function f on a Riemann surface and a point p such that $f'(p) \neq 0$ for some choice of local coordinate, show that one can find a holomorphic chart ϕ around p such that $f(w) = f(p) + \phi(w)$.

So much for the regular points. But what happens at the critical points? We cannot hope for a single standard sort of chart around critical points in the same manner as we just obtained for regular points, because critical points of f have various properties which must remain invariant under changes of coordinates. For example, some critical points are isolated, while others are not. For the time being, we consider only isolated critical points; that is, points $p \in S$ such that for some neighbourhood U, p is the only critical point contained in U.

Even so, there are various possibilities. We typically use critical points as a tool to optimise the value of f; we may find that a particular critical point is a local maximum, a local minimum, or neither, and this classification is independent of our choice of coordinates. In the one-dimensional case, we classified critical points by looking at the second derivative; in two dimensions, the object of interest is the *Hessian matrix*

$$D^2 f(p) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(p) & \frac{\partial^2 f}{\partial x \partial y}(p) \\ \frac{\partial^2 f}{\partial y \partial x}(p) & \frac{\partial^2 f}{\partial y^2}(p) \end{pmatrix}$$

Note that the form of this matrix will only be meaningful if p is a critical point, since otherwise the Hessian vanishes in the coordinate system specified by the above lemma.

Recall from linear algebra that given a symmetric 2×2 matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

such as the one above, we can either use A to define a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ bx + cy \end{pmatrix} \tag{3.5}$$

or to define a quadratic form $\mathbb{R}^2 \to \mathbb{R}$ by

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}^T A \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= ax^2 + 2bxy + cy^2 \end{aligned}$$

For the Hessian, it is the latter meaning which is relevant here, rather than the more familiar use as a linear transformation. For a linear transformation, the matrix A transforms under a change of coordinates to the matrix $C^{-1}AC$, where C is the matrix specifying the new coordinates; for a quadratic form, A becomes instead $C^T AC$.

It is a basic property of the determinant that $\det C^T = \det C$, and so $\det(C^T AC) = \det(C)^2 \det A$. Thus the sign of the determinant is preserved by changes of coordinates. Assuming the matrix $D^2 f(p)$ is nondegenerate, we have three possibilities:

(1) $\det D^2 f(p) > 0$ and $D^2 f(p)$ is positive definite. Then p is a local minimum for f.

(2) $\det D^2 f(p) > 0$ and $D^2 f(p)$ is negative definite. Then p is a local maximum for f.

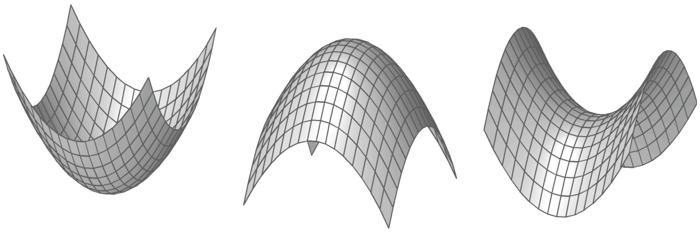

Figure 3.11. Three nondegenerate critical points.

(3) $\det D^2 f(p) < 0$. Then p is a saddle; neither a minimum nor a maximum.

We can now make a linear change of coordinates which brings the quadratic part of the function to a particularly simple form, so that the graph is as shown in Figure 3.11. In all cases the remainder term will be $o(x^2 + y^2)$.

(1) In the first case, there exists a local coordinate system in which $f(x, y) = f(0, 0) + x^2 + y^2 + o(x^2 + y^2)$.

(2) In the second case, there exists a local coordinate system in which $f(x, y) = f(0, 0) - (x^2 + y^2) + o(x^2 + y^2)$.

(3) In the third case, there exists a local coordinate system in which $f(x, y) = f(0, 0) + x^2 - y^2 + o(x^2 + y^2)$.

Exercise 3.11. Prove that any critical point p with $\det D^2 f(p) \neq 0$ is isolated from other critical points.

In fact, the consideration of the behavior of a function near a non-degenerate critical point is made more convenient by a useful technical result called the *Morse lemma*, which states that under an appropriate choice of local coordinates, the error term in the above representation can be eliminated. We present the proof in the most interesting case, that of a saddle, as a series of exercises.

Exercise 3.12. Let p be a non-degenerate saddle point of the function f. Show that locally, the level set $\{\, (x, y) \mid f(x, y) = f(p) \,\}$ is a union of two smooth curves which are tangent at the origin to the diagonals $y = x$ and $y = -x$.

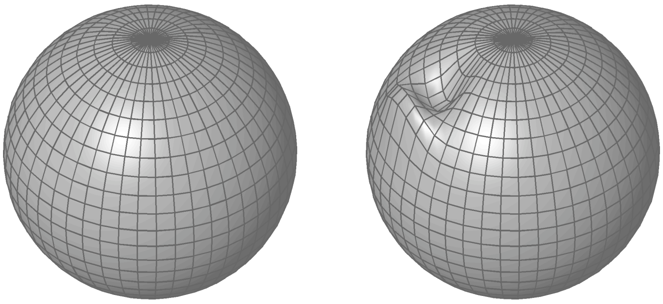

Figure 3.12. Two spheres with different height functions.

Exercise 3.13. Under the same assumption, show that there exist local coordinates (x', y') such that locally, the set $\{ (x, y) \mid f(x, y) = f(p) \}$ is a union of the diagonals $y' = x'$ and $y' = -x'$ themselves.

Exercise 3.14. Show that there exists a smooth map in a neighbourhood of p which is the identity on the diagonals $y' = x'$ and $y' = -x'$, and which maps the curves $f = c$ to hyperbolas $x'y' = c$ for every constant c.

For the other two cases, we will need only a weaker statement which parallels that of Exercise 3.12.

Exercise 3.15. Let p be a non-degenerate minimum of the function f. Show that there exists $\varepsilon > 0$ such that for any c with $f(p) < c < f(p) + \varepsilon$, the level set $f(x, y) = c$ is locally a smooth curve which intersects every ray in the (x, y) coordinates at a single point, and which is transversal (not tangential) to those rays.

b. Morse functions. Given a compact surface S and a smooth function $f \colon S \to \mathbb{R}$, basic topological arguments imply that f achieves its maximum and minimum on S; since the gradient of f in any coordinate representation vanishes at each of these, f must have at least two critical points.

We can easily construct an example where f has no other critical points aside from these two; consider the sphere $S^2 = \{ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1 \}$ and the height function $f \colon (x, y, z) \mapsto z$. Then

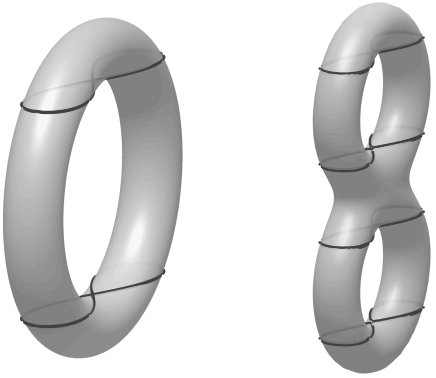

Figure 3.13. Defining a Morse function on a sphere with one or two handles.

f has a maximum at the north pole $(0, 0, 1)$, a minimum at the south pole $(0, 0, -1)$, and no other critical points.

If we perturb the sphere slightly, as shown in Figure 3.12, we will introduce a new pair of local extrema; one local maximum and one local minimum. Along with these we will create two saddle points, so that all in all the perturbed sphere has six critical points; two maxima, two saddles, and two minima.

Another interesting example is given by the standard torus of revolution standing sideways as shown in Figure 3.13, again with the height function $f\colon (x, y, z) \mapsto z$. Now f has one maximum and one minimum, along with two saddles at $(0, 0, \pm 1)$. A similar procedure yields a smooth function on the sphere with m handles having one maximum, one minimum, and $2m$ saddles; the case $m = 2$ is shown, with critical levels drawn for the four saddle points.

Definition 3.17. Let S be a smooth surface and $f\colon S \to \mathbb{R}$ a smooth function. f is called a *Morse function* if every critical point p of f is *nondegenerate*; i.e. the Hessian matrix $D^2 f(p)$ is invertible.

It follows from the definition that every critical point of a Morse function is either a maximum, a minimum, or a saddle.

Exercise 3.16. Represent the second surface shown in Figure 3.13 (or one homeomorphic to it) as a regular level set of a smooth function, and prove that the height function is indeed a Morse function with one minimum, one maximum, and four saddles.

We will find that looking at the level sets of a Morse function $f\colon S \to \mathbb{R}$ and how they change from one level to another reveals a great deal of information about the surface S. In fact, we can describe a procedure to reconstruct S (up to diffeomorphism) from knowledge of just the critical points of f.

First suppose that for a particular $c \in \mathbb{R}$ the level set $f^{-1}(c) \subset S$ has no critical points (that is, c is a regular value). Then by the same argument used to establish that the level set $F^{-1}(c)$ is a surface (2-dimensional manifold) whenever c is a regular value of $F\colon \mathbb{R}^3 \to \mathbb{R}$, we can deduce from the Implicit Function Theorem and the Inverse Function Theorem that $f^{-1}(c)$ is a 1-dimensional submanifold of S. Since every compact 1-dimensional manifold is a disjoint union of circles, it follows that $f^{-1}(c)$ has this form.

Now what happens if c is a critical value? Let $p \in f^{-1}(c)$ be a critical point; then by the Morse lemma we may choose local coordinates around p such that f takes a standard form.[4] There are three possibilities:

(1) p a local minimum, $f = c + x^2 + y^2$. Then for c' slightly smaller than c, the level set $f^{-1}(c')$ does not contain any points near p. For $c' = c$, it contains just one point, p, and for c' slightly greater than c, $x^2 + y^2 = c' - c$ defines a circle. Thus as we increase the value of c' through c, a circle is born around the critical point p.

(2) p a local maximum, $f = c - (x^2 + y^2)$. The reverse of the above process occurs; the circle which exists for $c' < c$ shrinks to a point at $c' = c$ and then vanishes for $c' > c$. As we increase the value of c' through c, a circle dies around p.

[4] The use of the Morse lemma in our considerations is convenient, but not essential. In the minimum and maximum cases, we only need Exercise 3.15, while Exercise 3.12 suffices for the case of a saddle.

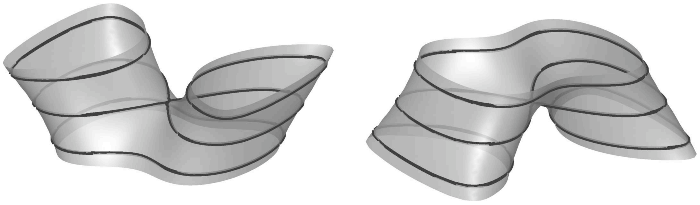

Figure 3.14. Level sets $f^{-1}(c')$ passing through a saddle point.

(3) p a saddle, $f = c + x^2 - y^2$. For $c' < c$, the (local) level set is a hyperbola opening left and right; for $c' = c$ it is two lines intersecting at p, and for $c' > c$ it is a hyperbola opening up and down. At the global level, we know that between critical points, the level sets are unions of circles, so there are two possibilities, as illustrated in Figure 3.14; as we pass through c, two circles may join and become one, or one circle may split and become two.

With these in mind, we may reconstruct S by increasing c through the range of f; this is the central idea of *Morse theory*, which has very powerful applications in a more general setting than we will consider here. Although the process is much more complicated in higher dimensions, the techniques developed from this theory are involved in the proof of the generalization of the famous Poincaré conjecture for manifolds of dimension ≥ 5, one of the landmark achievements of mathematics in the third quarter of the twentieth century.[5] The very rough outline of the method is to start from a Morse function on a given manifold which satisfies the assumptions of the Poincaré conjecture—i.e. has certain invariants identical to those of a sphere—and modify it to decrease the number of critical points until only one maximum and one minimum remain.

c. The third incarnation of Euler characteristic. At a more down-to-earth level, we will now show how to use Morse functions to describe a third incarnation of the Euler characteristic χ for surfaces.

[5] This brought a Fields Medal to Stephen Smale in 1966; the solution of the conjecture in the two remaining dimensions—first in dimension four, and then in the original three—resulted in two more Fields Medals later.

If we count the various sorts of critical points on the surfaces we have examined so far (using the height function as our Morse function each time), we have the following:

Surface	maxima	saddles	minima	χ
sphere	1	0	1	2
(perturbed) sphere	2	2	2	2
torus	1	2	1	0
sphere with m handles	1	$2m$	1	$2-2m$

Note that in each case, the Euler characteristic χ is equal to the alternating sum of the three columns; in fact, this is true in general.

Theorem 3.18. *For any Morse function $f\colon S \to \mathbb{R}$, the Euler characteristic is related to the number of critical points by the formula*

$$\chi = \text{(\# of maxima)} - \text{(\# of saddles)} + \text{(\# of minima)} \tag{3.6}$$

Before proving the theorem, we describe the general method and examine what happens in the case of the torus. We proceed by examining the *sublevel sets*

$$S_c = f^{-1}((-\infty, c]) = \{\, x \in S \mid f(x) \leq c \,\}$$

Let m and M be the minimum and maximum values, respectively, assumed by f on S. Then for $c < m$, we have $S_c = \emptyset$, and for $c \geq M$, $S_c = S$. The real story is what happens in between m and M...

The next observation to make is that nothing interesting happens at non-critical levels. This is the content of the following lemma, which intuitively looks quite plausible, although a rigorous proof requires certain tools which we will not develop until later (see Lecture 36(b)).

Lemma 3.19. *Given a Morse function $f\colon S \to \mathbb{R}$ and $a, b \in \mathbb{R}$ such that every $c \in (a, b)$ is a regular value ($f^{-1}(c)$ contains no critical points), then S_c and $S_{c'}$ are diffeomorphic for every $c, c' \in (a, b)$.*

Thus for the torus shown in Figures 3.13 and 3.15, with inner radius 1 and outer radius 2, all the action happens at $f(x) = \pm 1, \pm 3$. In between those points, the boundary of S_c is the level set $f^{-1}(c)$, which we know to be a disjoint union of circles. The four critical

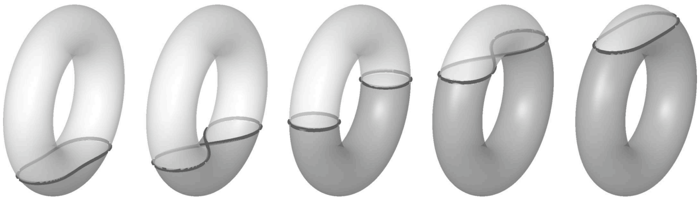

Figure 3.15. Sublevel sets on the vertical torus.

points run through the four possibilities enumerated in our earlier discussion:

1. At $c = -3$, a circle is born, so the empty set is replaced by a disc.
2. At $c = -1$, one circle splits into two, so the disc is replaced by a cylinder.
3. At $c = 1$, the two circles rejoin and become one, so the cylinder is replaced by a torus with a hole.
4. At $c = 3$, the circle dies, so the hole is filled with a cap, and we obtain the entire torus.

Proof of Theorem 3.18. It follows from Lemma 3.19 that between critical levels, the changes in S_c are only quantitative, not qualitative, and have no effect on the Euler characteristic; in order to prove the theorem, therefore, it suffices to examine the change in χ as we pass through each of the various sorts of critical points. To accomplish this, we first extend the definition of χ to allow non-connected manifolds; this will allow examples with $\chi > 2$, which is impossible in the connected case.

Now there are three cases to examine. If $f^{-1}(c)$ contains a local minimum of f, then passing through c corresponds to adding a new disc, as we saw, and hence increases χ by one. Similarly, passing through a local maximum corresponds to filling in a hole with a disc, which involves adding a face and leaving the number of edges and faces unchanged, and so also increases χ by one.

It remains only to show that passing through a saddle point decreases χ by one. Figure 3.16 shows the sublevel sets $S_{c'}$ (viewed from

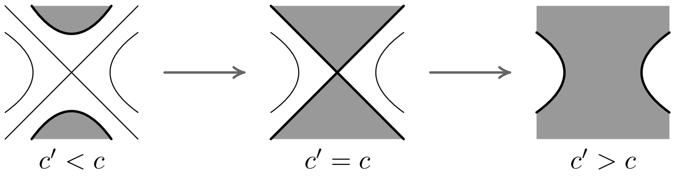

Figure 3.16. Sublevel sets near a saddle.

above) for values of c' near the critical value c. Upon passing through the saddle, the number of edges and vertices remains the same, but two faces which previously were separate are joined into one. Hence the alternating sum $\chi = V - E + F$ is decreased by one. □

If we carry out this construction a bit more carefully, we can actually obtain a complete classification of smooth surfaces using Morse functions as our tool; this was in fact the inspiration for the proof we gave of the classification theorem for compact orientable surfaces (Theorem 2.15), and is a 'baby version' of the arguments used in higher dimension, like those on which the afore-mentioned proof of the Poincaré conjecture in dimensions five and above is based.

Exercise 3.17. Consider the function $f(x, y) = \sin(4\pi x)\cos(6\pi y)$ on the standard flat torus $\mathbb{R}^2/\mathbb{Z}^2$.

(1) Prove that it is a Morse function, and calculate the number of minima, saddles, and maxima.

(2) Describe the evolution of the sub-level sets $f^{-1}((-\infty, c))$ as c varies from the lowest minimum value to the highest maximum value.

Lecture 21.

a. Functions with degenerate critical points. Having successfully used the ideas of Morse theory to reconstruct the surface S and run across our old friend, the Euler characteristic, we would now like to extend the same ideas and techniques to the case where our function $f\colon S \to \mathbb{R}$ may fail to be Morse by having degenerate critical points.