Introduction to Morse theory

Mendel Keller

December 12, 2018

Abstract

Morse theory allows us to construct a manifold M as a cell complex using a special kind of map $f : M \to \mathbb{R}$ called a Morse function. For each critical point $p \in M$ of f, we add a λ-cell to our construction, where λ is the index of f at p. This cell structure is helpful in computing the homology of M.

1 Introduction

A function $f : M \to \mathbb{R}$ is called Morse (definition 14) if all of its critical points are nondegenerate (definition 13). Morse functions provide useful tools in studying the topology of manifolds. In particular, a Morse function f provides a cell-decomposition (definition 3) of the manifold in which cells one-to-one correspond to critical points (definition 11) of f. That is, given a manifold M and morse function f, we can define a collection of sub-manifolds M^a (definition 16) of M. We will see (theorem 18) that for $a < b$ we will have that M^a is homotopy equivalent to M^b whenever there is no critical value (definition 11) of f in $[a, b]$. For a critical value $c \in \mathbb{R}$ we will have that M^c is homotopy equivalent to $M^{c-\epsilon} \cup e^\lambda$. That is, $M^{c-\epsilon}$ with a λ-cell (definition 2) attached to it, where λ is the index (definition 15) of p, the critical point with critical value c. We can then apply this to computing the homology groups of M.

Remark 1. For some vague intuition to help understand what's going on in all this, you can imagine continuously varying the input to f. In regular regions of M, the values f maps to will flow nicely, all going in one direction. In order to have some critical point of f, that is some point $p \in M$ where f can switch direction, we need some structural properties of M. The smoothness of f doesn't allow for inputs lying in a normal region to suddenly change direction, but rather if f changes direction, it's only because M is changing direction. This is why we can use a function on M to learn things about M.

The layout of the paper is as follows. In section 2 we introduce some topological preliminaries, in particular the notion of a cell complex. In section 3 we define some calculus notions for manifolds and the index of a point, which are vital in describing Morse theory. Section 4 contains a theorem which tells us that in regions without critical points the structure of the manifold remains unchanged. Section 5 contains a useful theorem about how we can rewrite a Morse function near a critical point. We use this theorem in section 6, where we present the main result of the paper, that critical points of a Morse function indicate the addition of a cell. In section 7 we see some worked examples of constructing a manifold using Morse theory. In section 8 we explore topological consequences of Morse theory.

My treatment in this paper follows [1] chapter 1, with some adjustments to the notation. Further treatment can be found in [2].

2 Cell complex

A cell complex is a particularly nice class of topological space which can be constructed inductively from simple parts. And this is the way which we would like to view manifolds using Morse theory.

2.1 Definitions

Before we go any further, let us introduce the notion of a cell complex. Essentially a cell complex is a topological space which is built out of solid spheres, which can be thought of as some n-dimensional blob. The good thing about these blobs is that everything about their topology is trivial, as they share their homology with a point. The first step in defining a cell complex is defining a cell.

Definition 2. An n-cell e^n is the set of points $\{x \in \mathbb{R}^n : |x| \leq 1\}$ this then has boundary $e^n \supset \partial e^n = \{x \in \mathbb{R}^n : |x| = 1\}$.

Note that in fact this is simply the n-dimensional solid unit sphere and its boundary is the corresponding hollow unit sphere S^{n-1}. A cell complex is then a topological space built out of cells.

Definition 3. A topological space X is a *cell complex* if it can written as the union of cells $X = \bigcup_{n,i} e_i^n$ in such a way that the boundary of a cell is contained in the lower dimensional cells $\partial e_i^n \subset \bigcup_{k<n} e_i^k$. We call $X^n := \bigcup_{k \leq n} e_i^k$ the n-skeleton of X.

A cell complex X can then be constructed with a set of n-cells for each n, and an attaching map $\partial e^n \to X^{n-1}$ for each cell. This is an inductive construction, passing from the n-skeleton to the $n+1$-skeleton.

Remark 4. A cell complex can be constructed inductively. A cell complex X starts with an underlying set of points X^0. Then, once built up to the X^{n-1} skeleton, the X^n skeleton is obtained by taking some collection of n-cells $e_1^n, \ldots, e_k^n$ which are attached to X^{n-1}. This is done via attaching maps $f_i : \partial e_i^n \to X^{n-1}$. After including the boundary of the n-cell into the $n-1$ skeleton, the n cell is then attached to its boundary in a continuous fashion. We proceed inductively in this way. Attaching some collection of k many n-cells to the X^{n-1} skeleton, until the space X is completely built up.

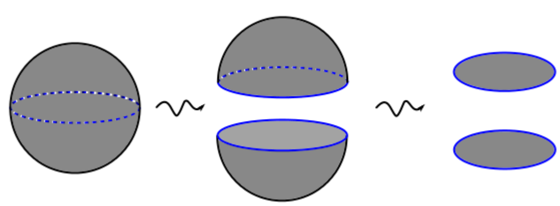
This inductive construction can be a useful tool in proofs, and also helps build intuition about a space.

2.2 Examples

Some example cell complexes will help demonstrate the point.

Example 5. The n-sphere S^n can be constructed as a single point and an n-cell, with the map $\partial e^n \to \{*\}$ being the constant map, the only available map.

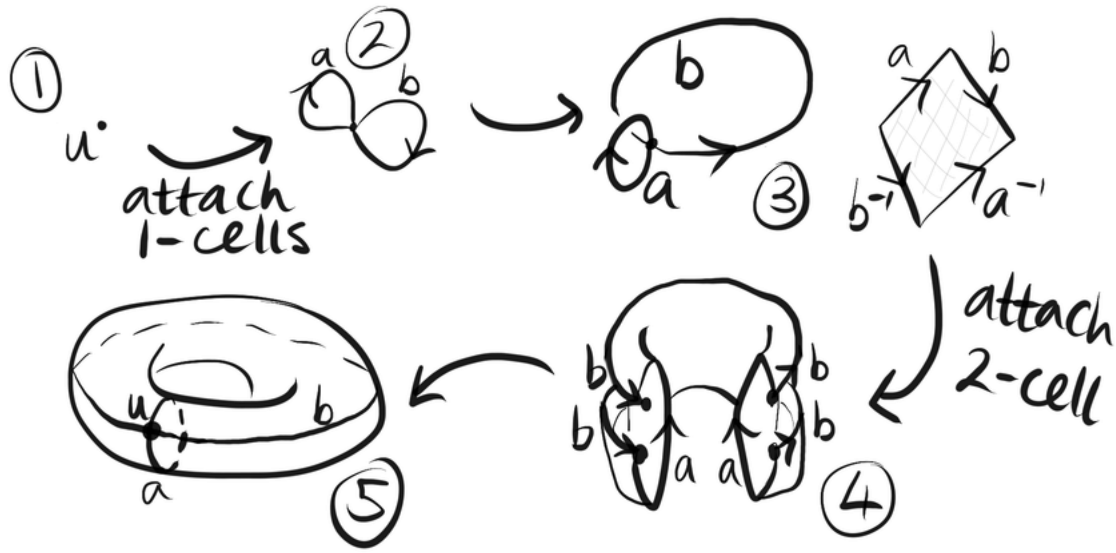
Here, by collapsing the boundary to a point, we shape the n-cell in such a way that moving across it now loops around, which is a sphere-like property. This construction is perhaps easiest to visualize in the case of S^1, where we take a 1-cell $[0,1]$ and identify the points $\{0,1\}$ as being the same. There is also an alternative cellular construction of the n-sphere.



Example 6. We can also build S^n inductively from S^{n-1} by attaching two n-cells. Here we simply identify ∂e^n with S^{n-1}.

The above example is perhaps best understood in the case of S^2, pictured above. Here we are attaching two disks to the circle S^1, each disk being one hemisphere and S^1 being the equator. As can be seen from these two example constructions of the n-sphere, the cellular structure of a cell complex needn't be unique.

Example 7. A slightly more involved example can be given by the torus $S^1 \times S^1$. Here we attach two circles to a point and than wrap a disk around the skeleton they provide as illustrated below. Notice here that the attaching map is complicated and actually wraps around each circle twice, in order to attach the disk to both sides.



2.3 utility

These types of construction of a space are particularly useful in Algebraic topology which focuses heavily and computationally relies on the way spaces are built out of simpler parts, and there are hardly any types of part simpler than n-cells. In particular the computation of homology closely parallels the n-cell construction and can be used to define it, which makes homology of cell complexes particularly simple. We explore this further in section 8.

A robust introduction to the notion of a cell complex can be found in [3] chapter 0.

3 Multivariable derivatives

We now introduce some calculus machinery in terms of which we will later state the Morse theorem.

3.1 Gradient

When we have a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$, we generalize the notion of derivative to that of a gradient. The gradient is a vector which points in the direction of greatest increase of f.

Definition 8. Given a real valued function $f : \mathbb{R}^n \to \mathbb{R}$ the *gradient* of f is a function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\nabla f(x_1, \ldots, x_n) = (\frac{\partial f}{\partial x_1}(x_1, \ldots, x_n), \ldots, \frac{\partial f}{\partial x_n}(x_1, \ldots, x_n))$.

The gradient also defines a notion closer to the typical notion of derivative, one which is fundamental in studying tangent spaces on manifolds, that of directional derivative.

Definition 9. Given a tangent vector v at $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ the *directional derivative* of f at x in the direction of v is given by $\nabla f \cdot v$.

The directional derivative gives us the derivative of f considered as a function $\mathbb{R} \to \mathbb{R}$ where we restrict the inputs of f to the subspace of $\mathbb{R}^n$ spanned by v.

When we work with manifolds things are a bit more complicated because we aren't actually insides $\mathbb{R}^n$, so we need to be a bit more careful in how we construct things. But here we can simply choose a chart and define things in terms of it. That is, using a chart diffeomorphism $\psi : U \to \mathbb{R}^n$ we get a function $f \circ \psi^{-1}$ going from a subset of $\mathbb{R}^n$ to $\mathbb{R}$ in terms of which we can define notions like the gradient and partial derivatives, and then translate that back over to M using ψ. In this way we define the gradient for a manifold.

Definition 10. Given a riemannian metric $g = \langle , \rangle$ on a manifold M, we can simply take ∇f to be defined as the vector field on M satisfying $\langle \nabla f, v \rangle_p = v(f)_p$. Where $v(f)_p$ is the derivative of f at p in the direction of v.

It is easy to see how this definition of ∇f is the natural generalization of the gradient to manifolds, as this is motivated by definition 9.

3.2 Critical points

A critical point of a multivariable function, or a function on a manifold, is just the higher dimensional analogue of the one dimensional case, and tells us similar things about function behavior.

Definition 11. A point $x \in M$ is a *critical point* of a function $f : M \to \mathbb{R}$ if $\nabla f(x) = 0$. We then call $c = f(x) \in \mathbb{R}$ a *critical value* of f.

As in the one dimensional case, a critical point often signals a local minimum or maximum. Again as in the 1d case, a critical point can also be something else. The higher dimensional analogue of an inflection point is called a *saddle point*. In order to study those however, an multivariable analogue of higher order derivative is needed.

3.3 Hessian

The Hessian is often used to study critical points of a function, as it is the multivariable version of a second derivative.

Definition 12. The Hessian H of a function f at a point $x \in M$ is the matrix defined by having i, j-th entry $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

In our case, we use the Hessian to classify critical points as degenerate and nondegenerate.

Definition 13. A critical point $p \in M$ of a function $f : M \to \mathbb{R}$ is called *non-degenerate* if its Hessian H_p has nonzero determinant.

This is the higher dimensional analogue of the second derivative test. Broadly, the Hessian provides a picture of the quadratic behavior of a function, analogous to the way the second derivative does. Therefore, it is the Hessian we look to when a point does not have local linear behavior, i.e. is a critical point.

Having defined degeneracy, we are now ready to actually define our objects of study, morse functions.

Definition 14. A function $f : M \to \mathbb{R}$ is called a *Morse function* if every critical point of f is nondegenerate.

3.4 Index of a critical point

In general, our picture of a critical point should be that there is no local linear behavior. Therefore the next object of interest is the quadratic behavior, and the hessian provides it. We can then imagine a bunch of parabolas crossing through the critical point, this being something like a frame on which the rest of the function hangs. The next step in analyzing the critical point then becomes seeing how many of these parabolas are facing upwards and how many are facing downwards. This is just like the second derivative test, except there we expect a single definite answer, and here it might depend on which direction we are considering. To talk about this we introduce the index.

Definition 15. The *index* λ of a critical point $p \in M$ is the maximal dimension of a subspace $V \subset T_pM$ such that $v^T H_p v < 0$ for all $v \in V$. We also call λ the index of H_p.

The index λ depends only on H_p because if H_p is negative definite on two subspaces $v, W \subset T_pM$ then it is negative definite on $V + W = \{v + w \in T_pM : v \in V, w \in W\} \subset T_pM$ so that there is a single maximal subspace of T_pM on which H_p is negative definite. The notion of index gives us detailed information about the behavior of f in a neighborhood of p. That is, it describes the quadratic shape of f near p.

4 Homotopy equivalence

Before explaining how critical points provide a cellular construction of M, we must see that noncritical points leave our cellular construction untouched. In order to do so we will use our function f to construct a diffeomorphism at points with nonzero gradient. First, we define a series of submanifolds using f, then we use the flow of the gradient in noncritical regions to define an isotopy which will give us a diffeomorphism.

4.1 Submanifolds

We consider a series of submanifolds which are cut out by a function f. We will use these to study what f can teach us about M.

Definition 16. Given a manifold M and smooth $f : M \to \mathbb{R}$, we define M^a as $\{x \in M : f(x) \leq a\} \subset M$.

Sometimes, M^a is a submanifold of M, indeed this holds whenever a is not a critical value of f.

We now consider how M^a changes as we vary a. For most values of a, a variation will not essentially change M^a. That is, the manifold obtained will be fundamentally the same if we increase a by a little, except it will have some more points, but no qualitative change in manifold shape will occur in most cases. That is, except near critical points.

4.2 Gradient

Recall that a smooth vector field v on a manifold M defines an isotopy $\rho : M \times \mathbb{R} \to M$. We state this notion formally.

Definition 17. Given a vector field v on a manifold M, the *flow* of v is a locally defined family of diffeomorphisms on $U \subset M \times \mathbb{R}$, that is, a map $\rho : U \to M$, satisfying $\rho(0) = \text{id}$ and $\frac{d}{dt}\rho_t = v \circ \rho_t$.

In our case, we choose the vector field ∇f as defined in definition 10. That is, the vector field such that the inner product with it by some $v \in T_pM$ is the same as the evaluation of the directional derivative, i.e. $\langle \nabla f, v \rangle = v(f) : M \to \mathbb{R}$. We then use its flow to show that M^a and M^b are diffeomorphic.

4.3 Diffeomorphism

So now we construct our diffeomorphism $M^a \cong M^b$, relying on the fact that no critical value of f in $[a, b]$.

Theorem 18. *Let f be a smooth function $f : M \to \mathbb{R}$. Then if there is no critical value of f in $[a, b] \subset \mathbb{R}$, then M^a is diffeomorphic to M^b.*

Proof. The nonexistence of a critical value in $[a, b]$ gives us that ∇f is nonevanishing on $f^{-1}[a, b]$. We can then take a vector field v that is given by $v = \nabla f / \langle \nabla f, \nabla f \rangle$ on the compact set $f^{-1}[a, b]$, and vanishing outside of a compact neighborhood of $f^{-1}[a, b]$. We then have a isotopy ρ obtained from v which satisfies $\frac{d\rho}{dt} = v \circ \rho$. Notice however that $\frac{d\rho}{dt}(f) = \langle \nabla f, v \rangle = \frac{\langle \nabla f, \nabla f \rangle}{\langle \nabla f, \nabla f \rangle} = 1$. This then tells us that $\rho_{b-a} M^a = M^b$, so that M^a is diffeomorphic to M^b. ∎

4.4 Deformation retract

We further get that M^a and M^b agree on their algebraic topological invariants since they are homotopy equivalent. It is therefore worth defining the notion of deformation retraction.

Definition 19. We say that a topological subspace $X \subset Y$ is a *retraction* of Y if there is a continuous map $f : Y \to Y$ such that $f|_X = id_X$ and $f(Y) \subset X$.

More important is the notion of a deformation retract.

Definition 20. We say that a topological subspace $X \subset Y$ is a *deformation retract* of Y if there is a continuous map $f : Y \times [0, 1] \to Y$ such that: (i) $f(t, x) = x$ for all $t \in [0, 1]$ and $x \in X$ (ii) $f(-, 0) = \mathrm{id}_Y$ and (iii) $f(Y, 1) \subset X$.

The notion of a deformation retract is that there is a way to smoothly deform Y into X. If X is a deformation retract of Y, then they are *homotopy equivalent*, which means that they share homotopy and homology groups. This is often useful in algebraic topology as it may be easier to compute the homology of a subspace X than the full space Y or vice versa.

Theorem 21. *Let everything be as in theorem 18, then M^a is a deformation retract of M^b*

Proof. Define $\phi_t : M^b \to M^b$ by $\phi_t(q) = q$ if $q \in M^a$ and $\phi_t(q) = \rho_{t(a-f(q))}(q)$ if $a \leq f(q) \leq b$ (here ρ is the same as in the proof of theorem 18). Then $\phi_0 = \mathrm{id}$ and ϕ_1 is the identity on M^a and maps $M^b - M^a$ to M^a. So we then have that M^a is a deformation retract of M^b. This then tells us that M^a is homotopy equivalent to M^b. ∎

So now we have that all the interesting structural properties of M are things that happen as we pass over a critical point of f. That is, the only place where we might not get homotopy equivalence is if we have M^a and M^b with a critical value c of f in $[a, b]$. If that's the case, then M^b will be some modification of M^a, and by looking at critical points of f on M we can determine all of these modifications and thus all of the interesting topological structure of M. But before we get to that we will need some machinery which allows us to pick nice coordinates.

5 Choosing co-ordinates

As we've done a bunch of times this semester, we make a useful choice of coordinates. Here, we want to choose a local chart $(\mathcal{U}, u_1, \ldots, u_n)$ in the neighborhood of a critical point p. As usual, the idea is to express our function in a simple way in terms of these coordinates. In particular we have that the function does not have linear behavior in a small neighborhood of p, but it does have quadratic behavior. Some of the quadratic behavior is positive, and some of it negative. So that we want a local chart $(\mathcal{U}, u_1, \ldots, u_n)$ so that $f = f(p) + \sum_i^k u_i^2 - \sum_{i=k+1}^n u_i^2$ on $\mathcal{U}$. That is, the local chart splits nicely into the positive and negative local quadratic behavior of f near p.

5.1 Breaking a function into components

Before we get on to showing that we can express a function nicely in the neighborhood of a critical point, we will first need a helpful little lemma.

Lemma 22. *Let f be a C^∞ function in a convex neighborhood V of 0 in $\mathbb{R}^n$, with $f(0) = 0$. Then $f(x_1, \ldots, x_n) = \sum_i^n x_i g(x_1, \ldots, x_n)$ for some collection of functions $g_i \in C^\infty(V)$, with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.*

Proof. Since V is convex and $f(0) = 0$ we can write $f(x_1, \ldots, x_n)$ as the integral along a straight line from 0 to $(x_1, \ldots, x_n)$ of its derivative. So that we can simply take g_i to be the integral $\int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n)dt$. Clearly we have that $g_i(0) = \frac{\partial f}{\partial x_i}(0)$, since this is the integral of a constant from 0 to 1. It is also clear that g_i is C^∞ given that f is. ■

Now that we have a way o rewrite functions, we will use this trick twice to write a function f as a quadratic.

5.2 Local quadratic expression

We now come to a central theorem of morse theory: that we can rewrite f in terms of squares of coordinates.

Theorem 23. *If f is a C^∞ function on a manifold M and p is a nondegenerate critical point of f, then there exists a local coordinate system $(\mathcal{U}, u_1, \ldots, u_n)$ centered around p such that $f(x) = f(p) - u_1(x)^2 - \cdots - u_k(x)^2 + u_{k+1}(x)^2 + \cdots + u_n(x)^2$ on $\mathcal{U}$, where k is the index of p.*

Proof. We work with the function $\hat{f}(x) = f(x) - f(p)$ instead, and once we've written $h(x) = u_1(x)^2 + \cdots + u_k(x)^2 - u_{k+1}(x)^2 - \cdots - u_n(x)^2$ we add back the $f(p)$ to get f back. I will just refer to $\hat{f}$ by f, since this turns out to be inconsequential. We have that $\forall i : x_i(p) = 0$ (if not we can adjust things by a translation to make this the case) so we then apply lemma 22 to get $f(x) = \sum g_i(x)x_i$. But since p is a critical point, we have that $g_i(0) = \frac{\partial f}{\partial x_i}(0) = 0$, so that g_i also satisfies the assumptions of lemma 22 and we have $f(x) = \sum_i \sum_j x_i x_j h_{ij}(x)$. We can then rewrite this sum to get a symmetric matrix which we can then diagonalize by the spectral theorem. In order to make this symmetric, simply have $\hat{h}_{ij} = \frac{1}{2}(h_{ij} + hji)$ and so we then get $f(x) = \sum_i \sum_j x_i x_j h_{ij}(x) = \sum_i \sum_j x_i x_j \hat{h}_{ij}(x)$. Once we diagonalize, we get that $f(x) = \sum_i x_i^2 g_i(x)$ for some smooth g_i. But we can just swallow the g_i into the coordinate function, and get the desired form for f. Note in particular that $h_{ij}(0)$ is equal to the Hessian at p. ■

So that, as stated earlier, we now look at f as only determined by quadratic behavior near p, since p is a critical point so that there is no local linear behavior. We have also untangled the coordinates so that there is no interaction between the directions to move in to get downwards motion with respect to f and the directions which yield upwards motion with respect to f.

6 Critical points

We now come to our desired result. That critical points of f correspond to the additions of λ-cells to M^a. If we imagine f as giving the height of a point of M, then we can imagine the directions along which f

decreases as facing downwards. Thus they collectively define some subset of M which covers that which lies below it. This is the sort of picture to keep in mind in the coming construction. Our reasoning here relies fundamentally on the expression for f we obtained in theorem 23.

Theorem 24. *Let $f : M \to \mathbb{R}$ be Morse, and let p be a nondegenerate critical point of f with index λ. If $f(p) = c$ and $f^{-1}[c - \epsilon, c + \epsilon]$ is compact and contains no other critical point of f for some $\epsilon \in \mathbb{R}$. Then $M^{c+\epsilon}$ has homotopy type of $M^{c-\epsilon}$ with a λ-cell attached.*

The proof is fairly complicated and so we only provide a general outline.

Theorem 25. *Given an n-dimensional manifold M and morse function $f : M \to \mathbb{R}$ with critical point $p \in M$ we can choose local coordinates $(\mathcal{U}, u_1, \dots, u_n)$ centered at p so that $f(x) = f(p) - u_1(x)^2 - \cdots - u_\lambda(x)^2 + u_{\lambda+1}(x)^2 + \cdots + u_n(x)^2$.*

Proof. Choose coordinates as in theorem 23. We will have then that $u_i(p) = 0$. ∎

We then use these coordinates to express the existence of a λ-cell. In particular we take the downwards facing coordinates to comprise the cell.

Lemma 26. *Let the setup be as above then we define e^λ to be the points x of $\mathcal{U}$ such that $\sum_{i=1}^{\lambda} u_i(x)^2 \leq \epsilon$ and $u_{\lambda+1}(x) = \cdots = u_n(x) = 0$, and then e^λ is diffeomorphic to a λ-cell.*

Proof. Clearly we have that e^λ is the same as a λ-cell, since it's defined in the same way except with ϵ in place of 1, and this will be diffeomorphic by a simple linear shrinking or expanding. e^λ is also defined as a subset of M instead of $\mathbb{R}^n$, but by assumption it's diffeomorphic to the relevant subset of $\mathbb{R}^n$. ∎

So now we have found the cell we want, and are well on our way to proving the Morse theorem. One thing to show is that the boundary of a newly added cell sits inside what we've built so far.

Theorem 27. *The intersection $M^{c-\epsilon} \cap e^\lambda$ is the boundary $\partial e^\lambda = \{x \in \mathcal{U} : \sum_{i=1}^{\lambda} u_i(x)^2 = \epsilon, u_{\lambda+1} = \cdots = u_n = 0\}$.*

Proof. In order for x to be in $M^{c-\epsilon}$, we need to have $f(x) = f(p) - u_1(x)^2 - \cdots - u_\lambda(x)^2 + u_{\lambda+1}(x)^2 + \cdots + u_n(x)^2 \leq c - \epsilon = f(p) - \epsilon$, which means that $u_1(x)^2 - \cdots - u_\lambda(x)^2 + u_{\lambda+1}(x)^2 + \cdots + u_n(x)^2 \leq -\epsilon \Rightarrow u_1(x)^2 - \cdots - u_\lambda(x)^2 \leq -\epsilon \Rightarrow \sum_i^\lambda u_i(x)^2 \geq \epsilon$. But in order for x to be in e^λ we need $\sum_{i=1}^{\lambda} u_i(x)^2 \leq \epsilon$ so that in fact we have that $M^{c-\epsilon} \cap e^\lambda = \{x \in \mathcal{U} : \sum_{i=1}^{\lambda} u_i(x)^2 = \epsilon\}$, which is just to say the boundary of e^λ. ∎

So thus far we have shown:

Theorem 28. *There is a submanifold of $M^{c+\epsilon}$ which is topologically the cell complex $M^{c-\epsilon} \cup e^\lambda$.*

This is because as above we have a λ-cell e^λ attached to $M^{c-\epsilon}$ by attaching map mapping its boundary into $M^{c-\epsilon}$ as given by the inclusion map which is given by theorem 27. Now the only think to state is that adding of a λ-cell is the only thing that happens in $[c - \epsilon, c + \epsilon]$, this is shown using an appropriate deformation retract, which we don't bother sketching here, as it is tedious and unenlightening. So we state without proof.

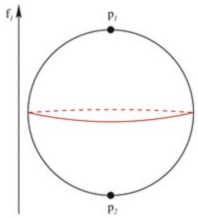
Theorem 29. *There is a deformation retract of $M^{c+\epsilon}$ onto $M^{c-\epsilon} \cup e^\lambda$.*

This now completes the demonstration of our theorem.

7 Examples

To see what's really going on here, we construct some examples.

7.1 Sphere



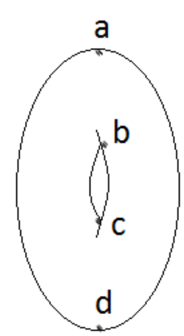
One example we can consider is the standard unit sphere S^n. In particular, consider S^2. The sphere is naturally seen as a subset of $\mathbb{R}^3$ centered at the origin. We can take a function $f : S^2 \to \mathbb{R}$ on the sphere which is the projection onto its z coordinate, $(x, y, z) \mapsto z$. We then have three types of points of $\mathbb{R}$, as listed in the table below. Here we imagine raising the xy-plane up from $-\infty$ to ∞, at some point passing through S^2 as we do so. As we do this, at each point we consider that part of S^2 which lies below our plane. At the start our horizontal cross section lies below the sphere so we just have the empty set. As we move upwards we eventually enter the sphere. When the plane is inside the sphere we then have a disk, since we do not have all of the sphere, so we do not have the part which closes the shape back up again. If we move up and down staying within the sphere, we only change the size of this disk, but keep the same fundamental shape. This disk is homotopy equivalent to a point, which is the 0 cell we obtain from p_2 in our diagram of the sphere here, that is because there are zero directions which we can move along to get a decreasing value for the height function near p_2. The last region is once we have passed through the sphere entirely, so that what lies below our plane is the whole of the sphere. We obtain this by adding a 2-cell when passing through the critical point p_1, from which there are two directions to move in which give decreasing values for the height function. This is the classical cell structure of the sphere, where we collapse the circle which is the boundary of a disk to a point. A similar construction holds for any S^n.

r	M^r
$r < -1$	$\emptyset$
$-1 < r < 1$	Disk
$1 < r$	Sphere

Table 1: Change in cell structure with height

7.2 Torus

Another classical example is the torus $S^1 \times S^1$. If we stand the torus up so that it is balancing on the xy plane and define f to spit out the z coordinate of a point. That is, we have f be the height function as in our previous example.



r	M^r
$r < d$	$\emptyset$
$d < r < c$	Disk
$c < r < b$	Tube
$b < r < a$	Torus with hole
$a < r$	Torus

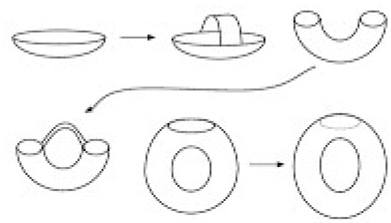
Table 2: Change in cell structure with height

We then find 4 critical points: (d) the bottom of the torus (c) the bottom of the inner ring (b) the top of the inner ring (a) the top of the torus.

(1) is a local (global) min, and (4) is a local (global) max, while (2) and (3) are both saddle points of index 1. That gives us the construction of one 0-cell, two 1-cells, and one 2-cell, as in the canonical cell complex construction of the torus we say in example 7.

We can see the step by step construction illustrated below. We start below the torus, so that our manifold is simply the empty set. When we pass through the critical point d we add a zero cell, since the height function is increasing in all directions near d. This zero cell is then homotopic to a disk. When we pass through the critical point c, we add a

1 cell, because one direction is decreasing while the other is decreasing in the vicinity of c. This then attaches the two opposite ends of our disk, giving us a tube, which is also homotopic to a circle in fact. When we come to b we have a second critical point of index 1. Here instead of attaching two ends of a disk we attach two ends of a tube, which gives us the shape seen below of a torus missing its top. When we finally pass through a then we obtain all of the Torus by plugging the hole with a 2 cell, since every direction is decreasing near a.



8 Applications

We then get from the cell construction the Morse inequality, which is used to compute the betti numbers of a manifold. Roughly speaking, the n-th betti number $b_n(M)$ of a manifold M is the number of n dimensional holes in M. For example, every S^n has exactly one nontrivial (that is to say besides the zeroth betti number which just counts the number of connected components of a manifold) nonzero betti number, which is $b_n(S^n) = 1$. This can be taken as a mental picture of what the betti numbers are meant to represent.

The morse inequalities states that $\forall k : \sum_{i=0}^{k}(-1)^{k-i}b_i(M) \leq \sum_{i=0}^{k}(-1)^{k-i}c_i(M)$ where $b_i(M)$ is the i-th betti number of M and $c_i(M)$ is the number of critical points of index i of M.

In this way, Morse functions can not just tell us how to decompose a manifold, but in fact give us certain restrictions on the number of holes that M has.

References

[1] John Milnor, *Morse theory.* (AM-51). Vol 51. Princeton university press, 2016.

[2] Dietmar Salamon, *Morse theory, the Conley index and Floer homology.* Bulletin of the London Mathematical Society 22.2 1990. pp 113-140.

[3] Allen Hatcher, *Algebraic Topology.* Cambridge university press. 2002.