An Introduction to Morse theory

Ke Liang Xiao
Supervised by Dr. Hans Boden

1 Introduction

Morse theory takes the perspective of analyzing a complicated space from the lens of its real-valued functions. These functions provide a more intuitive description of the global behaviour of a manifold through analogs to more natural concepts. For example, consider the height function that measures the water-level of a landscape. As water-level rises, we would see the topology of the landscape remain relatively unchanged until the water reaches either a peak or trough, at which point portions of the landscape start disappearing. Conveniently, these two terrains are natural representations of the object of study, critical points. It turns out we can extract key topological information about a manifold through analyzing the local behaviour of a differentiable function near its critical points. For instance, one can determine the homotopy type of a manifold and how it changes accordingly to the critical points of a "Morse" function. It may come off as surprising, but all the interesting behaviour comes from neighbourhoods of these critical points!

To give some motivation, let us consider the height function once again but applied to the torus positioned vertically. It is obvious there exist four critical points; a minimum, two saddles and a maximum. Suppose we label the height of each critical point as (a),(b),(c) and (d), arranged in increasing order. It can be easily deduced that the subset of the torus with height between (a) and (b) is homotopy equivalent to a single point; the reason is similar to how an open disk deformation retracts to its origin. What is less obvious is the portion of the torus with height between (b) and (c) is homotopy equivalent to the subset with height between (a) and (b) with a 1-cell attached. Similarly, the entire torus is homotopic to its subset below the maximal height (d) but with a 2-cell attached [Milnor, 1963]. The sudden appearance of the n-cells and its dimension may seem random, but it is intimately linked to the critical points of our chosen function. This result follows from one of the fundamental theorems of Morse theory. With these ideas in mind, we will now study the relationship between a manifold and its real-valued functions. We will follow the ideas of Marston Morse as presented in John Milnor's book [Milnor, 1963]. As with any introduction to Morse theory we will begin at the gateway; Morse's Lemma.

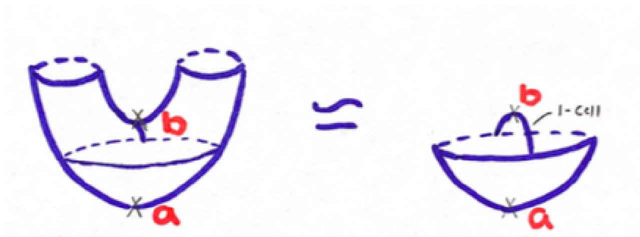

Figure 1: homotopy equivalence between subsets of the torus with an attached 1-cell.

2 Morse's Lemma

Before we get to the actual theory we should first review some general definitions and concepts that we will need to discuss topological properties of manifolds. For simplicity we will assume that our manifolds are embedded in euclidean space, this way we do not have to fuss over charts and atlases.

Lemma 2.1. *Let M and N be manifolds. Suppose $f : M \to N$ is a smooth map between manifolds, then the induced map df maps tangent vectors at x to tangent vectors at $f(x)$, that is $df(x) : TM_x \to TN_{f(x)}$.*

Proof. The proof follows from an application of the chain rule and the commutativity of derivatives. □

We are specifically interested in real-valued functions. In this case, if $f : M \to \mathbb{R}^n$ is a differentiable function and x is a point in M, then $df(x)$ is represented by

$$\left(\frac{\partial (f \circ \phi^{-1})}{\partial x_1}(0), \ldots, \frac{\partial (f \circ \phi^{-1})}{\partial x_n}(0) \right),$$

where $\phi : U \subset M \to \mathbb{R}^n$ is a local parametrization with $\phi(x) = 0$. We will often write $\frac{\partial f}{\partial x_i}$ in place of $\frac{\partial (f \circ \phi^{-1})}{\partial x_i}$ for simplicity of notation.

Definition 2.2. *Let f be a real-valued smooth function on a manifold M. We say x_0 is a critical point of f, if $df_{x_0} : TM_{x_0} \to \mathbb{R}$ is zero. Under local coordinates this translates to all first-order partial derivatives $\frac{\partial f}{\partial x_i}$ equating to zero at x_0.*

Definition 2.3. *If $x_0 \in M$ is a critical point of $f : M \to \mathbb{R}$ and $H(x_0)$ denotes the 2nd derivative matrix (referred to as the hessian), we say x_0 is a nondegenerate critical point of f, if $H(x_0)$ is non-singular.*

Similarly to df, $H(x_0)$ is represented by the $n \times n$ matrix

$$\left(\frac{\partial^2 (f \circ \phi^{-1})}{\partial x_i \partial x_j} \right)_{i,j=1}^{n}.$$

Once again, we will often write $\frac{\partial^2 f}{\partial x_i \partial x_j}$ in place of $\frac{\partial^2 (f \circ \phi^{-1})}{\partial x_i \partial x_j}$.

Remark. Non-degeneracy of critical points is invariant under local parametrization.

Definition 2.4. *Let x_0 be a critical point of a smooth real-valued function f. The index λ is the number of negative eigenvalues of $H(x_0)$.*

The entire foundation of Morse theory is built upon the information critical points and their indices may provide us. An example of their significance is displayed in Morse's Lemma. We will see how focusing on a neighbourhood of a critical point will allow us to express any smooth function explicitly as a polynomial under local coordinates. Before we can get to the proof, the following lemma will be very useful.

Lemma 2.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Suppose $f(0) = 0$ then there exist smooth functions $g_1, \dots, g_n : \mathbb{R}^n \to \mathbb{R}$ with $g_i(0) = \frac{\partial f(0)}{\partial x_i}$ such that*

$$f(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i g_i(x_1, \dots, x_n).$$

Proof. Fix $(x_1, \dots, x_n)$ and consider $f(tx_1, \dots, tx_n)$. The chain rule tells us

$$\frac{df(tx_1, \dots, tx_n)}{dt} = \sum_{i=1}^{n} x_i \frac{\partial f(tx_1, \dots, tx_n)}{\partial x_i}.$$

Combining the above with the fundamental theorem of calculus we can re-express f as

$$\begin{aligned} f(x_1,\ldots,x_n) &= \int_0^1 \frac{df(tx_1,\ldots,tx_n)}{dt}dt \\ &= \sum_{i=1}^{n}\int_0^1 x_i\frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}dt \\ &= \sum_{i=1}^{n} x_i\int_0^1 \frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}dt. \end{aligned}$$

Define $g_i(x_1,\ldots,x_n) = \int_0^1 \frac{\partial f(tx_1,\ldots,tx_n)}{\partial x_i}dt$. It is easy to check g_i satisfies our desired property at $t = 0$. □

Theorem 2.6 (Morse's Lemma)**.** *Let $f : M \to \mathbb{R}$ be a smooth function on an n-dimensional manifold M. Let p be a nondegenerate critical point of f, then there exist local coordinates $(x_1,\ldots,x_n)$ in a neighbourhood of p such that*

$$f(x_1,\ldots,x_n) = f(p) - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2,$$

with λ being the index of p.

Proof. The proof of Morse's lemma uses an inductive argument. We want to show that for all k in the set {1,. . . ,n+1} there exist coordinates $(y_1,\ldots,y_n)$ on a neighbourhood U_k of 0 such that

$$f(y_1,\ldots,y_n) = \pm y_1^2 \pm \cdots \pm y_{k-1}^2 + \sum_{i,j\geq k}^{n} y_iy_jh_{i,j}(y_1,\ldots,y_n),$$

where h_{ij} are smooth functions. We start off with the base case, $k = 1$. Since f is a smooth function on M, for all charts (U,ϕ), $f\circ\phi^{-1}$ is smooth. Let $(x_1,\ldots,x_n)$ be the local coordinates given by ϕ. We may assume that $\phi(p_0) = 0$. Translating f by $-f(p)$, we may also assume $f(p) = 0$. For simplicity of notation we will write $f\circ\phi^{-1}(x_1,\ldots,x_n) = f(x_1,\ldots,x_n)$. By Lemma 2.5 there exist n smooth functions g_i such that

$$f(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_ig_i(x_1,\ldots,x_n).$$

Since p is a critical point and $g_i(0) = \frac{\partial f(0)}{\partial x_i}$, we can apply the lemma again to each g_i to get smooth functions h_{ji} such that

$$g_i(x_1,\ldots,x_n) = \sum_{j=1}^{n} x_jh_{ji}(x_1,\ldots,x_n).$$

From the proof of the lemma, we have by construction

$$h_{ji} = \int_0^1 \frac{\partial g_i(tx_1, \ldots, tx_n)}{\partial x_j} dt,$$

but

$$\begin{aligned} g_i(x_1, \ldots, x_n) &= \int_0^1 \frac{\partial f(tx_1, \ldots, tx_n)}{\partial x_i} dt \\ \left.\frac{\partial g_i}{\partial x_j}\right|_{x=0} &= \left.\frac{\partial}{\partial x_j}\left(\int_0^1 \frac{\partial f(tx_1, \ldots, tx_n)}{\partial x_i} dt\right)\right|_{x=0} \\ &= \left.\int_0^1 \frac{\partial^2 f(tx_1, \ldots, tx_n)}{\partial x_j \partial x_i} dt\right|_{x=0} \\ &= \frac{\partial^2 f(0, \ldots, 0)}{\partial x_j \partial x_i}, \end{aligned}$$

so

$$h_{ji}(0, \ldots, 0) = \frac{\partial^2 f(0, \ldots, 0)}{\partial x_j \partial x_i}.$$

Plugging this into f we get

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} x_i x_j h_{ij}(x_1, \ldots, x_n).$$

We may assume $h_{ij} = h_{ji}$ by replacing h_{ij} with $(h_{ij}+h_{ji})/2$, then $(h_{ij}(0))_{i,j=1}^{n} = H(p)$. But since p is nondegenerate, $H(p)$ is non-singular hence has only non-zero eigenvalues. If we diagonalize $H(p)$ on a neighbourhood of 0, f can be written with no cross terms and all the square terms will have non-zero coefficients. This proves the base case.

Assume for k in $\{1, 2, \ldots, n\}$ there exist local coordinates $(y_1, \ldots, y_n)$ and smooth functions h_{ij} on a neighbourhood of 0 such that

$$f(y_1, \ldots, y_n) = \pm y_1^2 \pm \cdots \pm y_{k-1}^2 + \sum_{i,j \geq k}^{n} y_i y_j h_{ij}(y_1, \ldots, y_n).$$

We may assume $h_{kk}(0) \neq 0$ as we can apply a linear change of coordinates on the last $n-k+1$ variables to remedy this. A sketch of the procedure follows. Since $H(p)$ is non-singular there must exist at least one non-zero

$h_{ij}(0)$ for $i, j \geq k$ else $H(p)$ would be singular. If $i = j$ we can swap rows and columns so that $h_{ii}(0)$ is on the kth diagonal. If $h_{ij}(0)$ is a non-diagonal entry, define new variables $\widetilde{y_i} = y_i + y_j$ and $\widetilde{y_j} = y_i - y_j$. Under these new coordinates we have

$$\begin{aligned} 2y_i y_j h_{ij}(y) &= \frac{\widetilde{y_i}^2 + \widetilde{y_j}^2}{2} h_{ij}(y_1, \ldots, \frac{\widetilde{y_i} + \widetilde{y_i}}{2}, .., \frac{\widetilde{y_i} - \widetilde{y_i}2}{2}, \ldots, y_n) \\ &= \widetilde{y_i}^2 \widetilde{h_{ii}}(y_1, \ldots, \widetilde{y_i}, \ldots, \widetilde{y_j}, \ldots, y_n) + \widetilde{y_j}^2 \widetilde{h_{jj}}(y_1, \ldots, \widetilde{y_i}, \ldots, \widetilde{y_j}, \ldots, y_n). \end{aligned}$$

We can simply swap rows and columns after to get the desired result. By continuity of h_{kk} there exist a smaller neighbourhood such that $h_{kk} \neq 0$. Define new coordinates $(v_1, \ldots, v_n)$ on this smaller neighbourhood as follows

$$v_i(y) = \begin{cases} y_i & i = j \\ \sqrt{|h_{kk}|} \left(y_k + \sum_{i>k} \frac{y_i h_{ik}(y)}{h_{kk}(y)} \right) & i \neq k. \end{cases}$$

Notice $\frac{\partial v_i}{\partial y_j} = \delta_{ij}$. Therefore, the jacobian matrix is 0 for all entries off the diagonal and 1 for all entries on the diagonal with the exception of $J_{kk} = \frac{\partial v_k}{\partial y_k}(0)$. It can be shown that $\frac{\partial v_k}{\partial y_k}(0) \neq 0$, thus the determinant of the jacobian is non-zero. By inverse function theorem $(v_1, \ldots, v_n)$ are local coordinates on M. We will assume without loss of generality that $h_{kk} > 0$ then

$$\begin{aligned} v_k^2 &= h_{kk} \left(y_k + \sum_{i>k} \frac{y_i h_{ik}(y)}{h_{kk}(y)} \right)^2 \\ &= h_{kk} y_k^2 + 2 \sum_{i>k} y_k y_i h_{ik}(y) + \frac{\left(\sum_{i>k} y_i h_{ik}(y) \right)^2}{h_{kk}(y)}. \end{aligned}$$

We will now verify that f may be expressed in the desire form under these new coordinates. Observe that

$$
\begin{aligned}
f &= \pm y_1^2 \pm \cdots \pm y_{k-1}^2 + \sum_{i,j\geq k}^{n} y_i y_j h_{i,j}(y_1,\ldots,y_n) \\
&= \pm y_1^2 \pm \cdots \pm y_{k-1}^2 + \left(y_k^2 h_{kk} + 2\sum_{i>k} y_k y_i h_{ik} + \sum_{i,j>k} y_i y_j h_{ij} \right) \\
&= \pm y_1^2 \pm \cdots \pm y_{k-1}^2 + \left(y_k^2 h_{kk} + 2\sum_{i>k} y_k y_i h_{ik} + \sum_{i,j>k} y_i y_j h_{ij} \right) + \\
&\frac{1}{h_{kk}} \left(\sum_{i>k} y_i h_{ik} \right)^2 - \frac{1}{h_{kk}} \left(\sum_{i>k} y_i h_{ik} \right)^2 \\
&= \pm v_1^2 \pm \cdots \pm v_{k-1}^2 + v_k^2 + \left(\sum_{i,j\geq k+1} v_i v_j h_{ij} - \frac{1}{h_{kk}} \left(\sum_{i\geq k+1} v_i h_{ik} \right)^2 \right) \\
&= \pm v_1^2 \pm \cdots \pm v_{k-1}^2 + v_k^2 + \sum_{i,k\geq k+1} v_i v_j \widetilde{h_{ij}}.
\end{aligned}
$$

The case $h_{kk} < 0$ gives similar results except the sign of v_k^2 is flipped. Therefore, f takes the desired form by induction. To wrap up the proof we just need to show that λ is the index of f at p. Assume

$$f(x_1,\ldots,x_n) = -x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2.$$

Notice $\frac{\partial f}{\partial x_i} = -2x_i$, so $\frac{\partial^2 f}{\partial x_i x_j} = 2\delta_{ij}$, implying the hessian matrix of f is diagonal with λ many -2's and $(n-\lambda)$ many 2's.

$$H(p) = \begin{bmatrix} -2 & & & & & \\ & \ddots & & & & \\ & & -2 & & & \\ & & & 2 & & \\ & & & & \ddots & \\ & & & & & 2 \end{bmatrix},$$

therefore the dimension in which $H(p)$ is negative definite is λ and the dimension in which $H(p)$ is positive definite is $n-\lambda$. If there was a subspace V in which $H(p)$ is negative definite but with dimension bigger than λ, this subspace would have to intersect the positive definite subspace, but that implies $H(p)$ is not negative definite on V. We can then conclude λ is indeed the index of p.

$\square$

A general property of nondegenerate critical points is that they are isolated. Now that we have Morse's Lemma, this becomes a very simple proof.

Corollary 2.7. *The nondegenerate critical points of a smooth real-valued function on a manifold are isolated.*

Proof. Suppose $p \in M$ is a nondegenerate critical point of f. By Morse's Lemma there exist local coordinates on a neighbourhood V of p such that $p = (0, \ldots, 0)$ and

$$f(x_1, \ldots, x_n) = f(p) - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2.$$

Clearly $\frac{\partial f}{\partial x_i}(x) = 0$ if and only if $x_i = 0$. Therefore, p is the only critical point in V. $\square$

Corollary 2.8. *If $f : M \to \mathbb{R}$ is a smooth function such that all critical points of f are nondegenerate and M^a is compact for all $a \in \mathbb{R}$, then the set of critical values of f has no limit point.*

Proof. We first observe the following result of Corollary 2.7; if M^a is compact then it may only contain finitely many critical points of f in its interior. Suppose for contradiction that M^a contains infinitely many nondegenerate critical points, then there exists $\{p_1, p_2, \ldots\}$ a sequence critical points contained in M^a. Since each p_i is isolated there exists an open set U_i containing p_i and no other critical point. Now within each U_i we may construct a smaller open set V_i such that $\overline{V_i} \subset U_i$ (we may do this as we can choose U_i to be homeomorphic to an ϵ-ball in $\mathbb{R}^n$ and choose V_i to be homeomorphic to an $\epsilon/2$-ball). Let $V = \bigcup_{i=1}^{\infty} \overline{V_i}$, then V^c is open with countably infinite holes. To fill in these holes we join V^c with $\bigcup_{i=1}^{\infty} U_i$, which is an open cover of M^a, but it contains no finite sub-cover giving us our contradiction.

Returning to critical values, if we assume $a \in \mathbb{R}$ is a limit point of the set of critical values, then for all $\epsilon > 0$, there must exist at least countably infinite distinct critical values in $(a - \epsilon, a + \epsilon)$, which in turn imply the existence of countably infinite critical points in $M^{a+\epsilon}$, but this contradicts our result above. $\square$

3 Critical Points and Homotopy Type

We can now begin exploring how critical points are related to the homotopy type of the manifold. It turns out that regions that do not contain any critical points are essentially identical, and our attention should only be on neighbourhoods that contain critical points. To explore this we are going to need the following concepts.

Definition 3.1. *Given a manifold M, a Riemannian metric on M is a collection of inner products $\langle \cdot, \cdot \rangle_x$ such that for every x in M $\langle \cdot, \cdot \rangle_x$ is an inner product on the tangent space TM_x of M.*

Definition 3.2. *Let M be a manifold. We say*

$$\rho : \mathbb{R} \times M \to M$$

is a 1-parameter group of diffeomorphisms on M if

1. *For all $t \in \mathbb{R}$, the map $\rho_t : M \to M$ is a diffeomorphism from M to itself and $\rho_t(x) = \rho(t, x)$,*
2. *$\rho_0(x) = x$ for all $x \in M$,*
3. *For all $t, s \in \mathbb{R}$, $\rho_{t+s} = \rho_t \circ \rho_s$.*

Definition 3.3. *Given ρ a 1-parameter group of diffeomorphisms, the induced vector field X on M is given by: for all $f : M \to \mathbb{R}$ smooth,*

$$X_q(f) = \lim_{t \to 0} \frac{f(\rho_t(q)) - f(q)}{t}.$$

The vector field X is said to generate the group ρ.

Lemma 3.4. *Let X be a smooth vector field on M that vanishes outside a compact set, then X generates a unique 1-parameter group of diffeomorphisms on M that satisfies the following differential equation:*

$$\frac{d\rho_t(q)}{dt} = X_{\rho_t(q)}.$$

The proof of Lemma 3.4 will be omitted as it requires some ODE theory.

Theorem 3.5. *Let f be a smooth real-valued function on M. If $a < b$ such that $f^{-1}[a, b]$ is compact and contains no critical points, then M^a is diffeomorphic and a deformation retraction of M^b.*

Proof. The idea of the proof is to make use of the fact that $f^{-1}[a, b]$ contains no critical points, hence the vector field of ∇f would go uninterrupted. We can then slide M^b along the flow lines of the gradient down to M^a.

Let $\langle \cdot, \cdot \rangle_x$ be a Riemannian metric on the manifold M. We consider the gradient function ∇f that maps points in M to a tangent vector in its tangent space. More formally, ∇f is characterized by the identity:

$$\langle X_q, \nabla f(q) \rangle_q = X_q(f),$$

for any smooth vector field X. Recall $X_q(f)$ is the directional derivative of f along the vector X_q and is also a smooth real-valued function on M.

Define a new smooth function $P : M \to \mathbb{R}$ from ∇f by

$$P(q) = \begin{cases} \frac{1}{\langle \nabla f(q), \nabla f(q) \rangle_q} & \text{if } q \in f^{-1}[a, b] \\ 0 & \text{if } q \notin C \text{ for some compact set C containing } f^{-1}[a, b]. \end{cases}$$

Since ∇f only vanishes at critical points of f and $\langle \nabla f(q), \nabla f(q) \rangle_q$ is non-zero in $f^{-1}[a, b]$ we may define a smooth vector field X by

$$X_q = P(q) \nabla f(q).$$

That is X normalizes the vector $\nabla f(q)$ for all q in $f^{-1}[a, b]$ and X vanishes outside a compact set. By Lemma 3.4 X generates a unique 1-parameter group of diffeomorphisms; $\rho : \mathbb{R} \times M \to M$ such that

$$X_{\rho_t(q)} = \frac{d\rho_t}{dt}.$$

Notice if we fix a point q in M, $\rho_t(q)$ gives a curve that crosses q at time $t = 0$. Recall the velocity vector of a path ρ_t is given by the equality

$$\left\langle \nabla f, \frac{d\rho_t}{dt} \right\rangle_q = \frac{d}{dt} \left(f \circ \rho_t \right),$$

then for $\rho_t(q)$ in $f^{-1}[a, b]$ we have

$$\frac{d}{dt} \left(f \circ \rho_t \right) = \left\langle \nabla f, \frac{d\rho_t}{dt} \right\rangle_q = \langle \nabla f, X_q \rangle_q = \left\langle \nabla f, \frac{\nabla f}{\langle \nabla f . \nabla f \rangle} \right\rangle = 1.$$

This implies f grows linearly in the direction $\frac{d\rho_t}{dt}$ and for all q in $f^{-1}[a, b]$ there exist an interval $I_q \subset \mathbb{R}$ such that

$$f(\rho_t(q)) = f(q) + t,$$

for all t in I_q. We now claim ρ_{b-a} is our desired diffeomorphism. Fortunately, diffeomorphism is given to us for free by property of 1-parameter groups. We just need to show $\rho_{b-a}(M^a) = M^b$.

We first show points in M^a gets mapped to points in M^b. Let $q \in M^a$, then by by definition $f(q) \leq a$. Applying linear growth of f we have

$$f(\rho_{b-a}(q)) = f(q) + (b - a) \leq b.$$

Therefore, $\rho_{b-a}(q)$ is indeed contained in M^b. A similar argument works for the reverse direction as well. Take $y \in M^b$, then there exists $x \in M$ such that $\rho_{b-a}(x) = y$, reapplying our equality $f(y) = f(x) + (b - a)$ allow us to conclude $f(x) \leq a$ and $\rho_{b-a}^{-1}(M^b) = M^a$. This concludes the proof of M^a being diffeomorphic to M^b. Conveniently enough our construction to show diffeomorphism has provided us with some insight of continuous trajectories on M. For deformation retraction, define $r_t : M^b \to M^b$ by

$$r_t(q) = \begin{cases} q & q \in M^a \\ \rho_{t(a-f(q))}(q) & q \in f^{-1}[a, b]. \end{cases}$$

Clearly $r_0 = \mathrm{Id}_{M^b}$ and as t approaches 1 r_t slides points in $f^{-1}[a, b]$ down to the boundary of M^a. Thus, M^b deformation retracts down to M^a. □

Remark. It may seem like the only necessity of $f^{-1}[a, b]$ being compact is so that we may generate our desired vector field. But this condition cannot be omitted. Consider the case where we remove a point p from M, where p originally lies in $f^{-1}[a, b]$. The locally euclidean property of M would allow us to construct a sequence converging to p, but after removal of p this sequence would not have a limit point in $f^{-1}[a, b]$. Hence, $f^{-1}[a, b]$ would no longer be compact. The hole would then prevent M^b from deformation retracting down to M^a.

Theorem 3.6. *Let $f : M \to \mathbb{R}$ be a smooth function and let p be a nondegenerate critical point with index λ and critical value c. If there exists $\epsilon > 0$ such that $f^{-1}[c - \epsilon, c + \epsilon]$ is compact and contains no other critical point other than p, then there exists $\delta > 0$ such that $M^{c+\delta}$ has the same homotopy type as $M^{c-\delta}$ union a λ-cell.*

Proof. Under the assumption of the theorem, let $p \in M$ be our critical point and $f(p) = c$. By Morse's Lemma there exists a neighbourhood of p with local coordinates $(u_1, \dots, u_n)$ such that $p = (0, \dots, 0)$ and

$$f(u_1, \dots, u_n) = f(p) - u_1^2 - \cdots - u_\lambda^2 + u_{\lambda+1}^2 + \cdots + u_n^2.$$

For convenience, let us denote

$$\xi = u_1^2 + \cdots + u_\lambda^2$$
$$\eta = u_{\lambda+1}^2 + \cdots + u_n^2,$$

so for $u \in U$

$$f(u) = c - \xi(u) + \eta(u).$$

Since $f^{-1}[c-\epsilon, c+\epsilon]$ is compact we may choose $\delta > 0$ such that

1. $f^{-1}[c-\delta, c+\delta] \subset f^{-1}[c-\epsilon, c+\epsilon]$,
2. $\psi(\overline{B_{\sqrt{2\delta}}(0)}) \subset U$, where ψ is the diffeomorphic map from $V \subset \mathbb{R}^n$ to $U \subset M$.

Note $f^{-1}[c-\delta, c+\delta]$ is also compact and contains no other critical points other than p. Define the λ-cell e^λ as

$$e^\lambda = \left\{ (u_1, \ldots, u_n) \in U \mid \sum_{i=1}^{\lambda} u_i^2 \leq \delta,\ u_{\lambda+1} = \cdots = u_n = 0 \right\}.$$

Clearly e^λ is contained in U. In the context of this theorem we may think of e^λ as the handle that attaches to $M^{c-\delta}$. This is due to the fact that e^λ intersect $M^{c-\delta}$ along its boundary $\partial(e^\lambda)$. To see this, let $x \in \partial(e^\lambda)$, then under the local coordinates on U, $x = (x_1, \ldots, x_n)$ and $\xi(x) = \sum_{i=1}^{\lambda} x_i^2 = \delta$. Plugging this into our formula $f(x) = c - \xi(x) = c - \delta$, which shows $x \in M^{c-\delta}$. For the reverse direction, let $x \in M^{c-\delta} \cap e^\lambda$ then $\eta(x) = 0$ and

$$f(x) = c - x_1^2 \cdots - x_\lambda^2 \leq c - \delta,$$

implying $x_1^2 + \cdots + x_\lambda \geq \delta$, but $x \in e^\lambda$ so $x_1^2 + \cdots + x_\lambda^2 = \delta$. Therefore, $M^{c-\delta} \cap e^\lambda = \partial(e^\lambda)$.

The remainder of this proof may be broken down into two deformation retractions; $M^{c+\delta}$ deformation retracts to $M^{c-\delta} \cup H$ for some subset H, then $M^{c-\delta} \cup H$ deformation retract to $M^{c-\delta} \cup e^\lambda$. The idea is to introduce a new function F that agrees with f everywhere but in a small neighbourhood of p, where points are perturbed so that $F < f$. The reasoning for this is we want to push the critical point p outside of our compact set so that Theorem 3.5 may be applied.

First define a smooth function $\mu : \mathbb{R} \to \mathbb{R}$ with the following properties

1. $\mu(0) > \delta$,

2. $\mu(x) = 0$ for $x \geq 2\delta$,

3. $-1 < \mu'(x) \leq 0$ for all $x \in \mathbb{R}$.

Let $F : M \to M$ be defined as

$$F(x) = \begin{cases} f(x) & x \notin U \\ f(x) - \mu(\xi(x) + 2\eta(x)) & x \in U. \end{cases}$$

Notice F coincides with f everywhere but in the ellipsoid $\xi + 2\eta \leq 2\delta$. We claim two important properties of F; it preserves both $M^{c+\delta}$ and the critical points of f.

Claim 1. $F^{-1}(-\infty, c+\delta] = M^{c+\delta}$.

Proof of claim 1. Since $F = f$ everywhere but within the ellipsoid we just need to check $\{x \in M \mid \xi(x) + 2\eta(x) \leq 2\delta\} \subseteq M^{c+\delta} \cap F^{-1}(-\infty, c+\delta]$. This follows from the fact $\xi \geq 0$, then for all x in the ellipsoid

$$F \leq f = c - \xi + \eta \leq c + \frac{\xi}{2} + \eta \leq c - \delta.$$

Therefore, $M^{c+\delta} = F^{-1}(-\infty, c+\delta]$.

Claim 2. F preserves the critical points of f.

Proof of claim 2. Again we only need to check for points in the ellipsoid. Recall $-1 < \mu' \leq 0$, then

$$\frac{\partial F}{\partial \xi} = -1 - \mu'(\xi + 2\eta) < 0$$
$$\frac{\partial F}{\partial \eta} = 1 - 2\mu'(\xi + 2\eta) \geq 1.$$

Chain rule then tells us

$$\frac{\partial F}{\partial x_i} = \frac{\partial F}{\partial \xi}\frac{\partial \xi}{\partial x_i} + \frac{\partial F}{\partial \eta}\frac{\partial \eta}{\partial x_i}.$$

For i between 1 and λ,

$$\frac{\partial F}{\partial x_i} = \frac{\partial F}{\partial \xi} 2x_i.$$

For i between $\lambda + 1$ and n,

$$\frac{\partial F}{\partial x_i} = \frac{\partial F}{\partial \eta} 2x_i.$$

Both of which are only equal to 0 on the origin. Hence, the only critical points of F in the ellipsoid is p.

Combining claim 1 and $F \leq f$ we see that $F^{-1}[c-\delta, c+\delta] \subseteq f^{-1}[c-\delta, c+\delta]$, so $F^{-1}[c-\delta, c+\delta]$ is our desired compact set. Furthermore, Claim 2 tell us this set does not contain any critical points as

$$F(p) = f(p) - \mu(0) = c - \mu(0) < c - \delta,$$

which follows from $\mu(0) > \delta$. We may now apply Theorem 3.5 to conclude $F^{-1}(-\infty, c+\delta] = M^{c+\delta}$ deformation retracts to $F^{-1}(-\infty, c-\delta]$. Now it remains to show $F^{-1}(-\infty, c-\delta]$ deformation retracts to $M^{c-\delta} \cup e^{\lambda}$.

Note since $F \leq f$ it holds in general $M^a \subseteq F^{-1}(-\infty, a]$ for all $a \in \mathbb{R}$. For notational convenience we will denote $F^{-1}(-\infty, c-\delta] = M^{c-\delta} \cup H$ where

$$H = \overline{F^{-1}(-\infty, c-\delta] \setminus M^{c-\delta}}.$$

We have shown before that e^{λ} is not contained in $M^{c-\delta}$ as it only intersects along its boundary, so we require e^{λ} to be contained in H. Let $x \in e^{\lambda}$, then $\eta(x) = 0$. Using the fact $\frac{\partial F}{\partial \eta} < 0$ we have

$$F(x) < F(p) = c - \mu(0) < c - \delta.$$

This implies $e^{\lambda} \subseteq F^{-1}(-\infty, c-\delta]$, but $e^{\lambda} \not\subset M^{c-\delta}$ so e^{λ} is contained in H (points on $\partial(e^{\lambda})$ lie in $M^{c-\delta}$ but defining H as the closure allows the boundary to be contained in H as well).

Claim 3. $M^{c-\delta} \cup H$ deformation retracts down to $M^{c-\delta} \cup e^{\lambda}$.

Proof of claim 3. The region $M^{c-\delta} \cup H$ may be decomposed into 3 cases. We will define $r_t : M^{c-\delta} \cup H \to M^{c-\delta} \cup H$ separately for each case. Keep in mind we well define our deformation retraction so that $r_1 = \mathrm{Id}$ and $r_0(M^{c-\delta} \cup H) = M^{c-\delta} \cup e^{\lambda}$. Figure 2 below displays the 3 cases and their deformation retractions.

Case 1 ($\xi \leq \delta$)

In this region we want to slide points down to the λ-cell. We do this by defining

$$r_t(x_1, \ldots, x_\lambda, x_{\lambda+1}, \ldots, x_n) = (x_1, \ldots, x_\lambda, tx_{\lambda+1}, \ldots, tx_n).$$

Clearly $r_1 = \mathrm{Id}$ and $r_0(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, 0, \ldots, 0) \in e^{\lambda}$. Observe for $x \in F^{-1}(\infty, c-\delta]$, as t decreases $\eta(x)$ decreases and $\xi(x)$ remain constants under the path given by r_t. Due to the fact $\frac{\partial F}{\partial \eta} > 0$, we see that as η decreases so does F. Therefore,

$$F(r_{t_1}(x)) < F(r_{t_2}(x)) < F(r_1(x)) = F(x) \leq c - \delta,$$

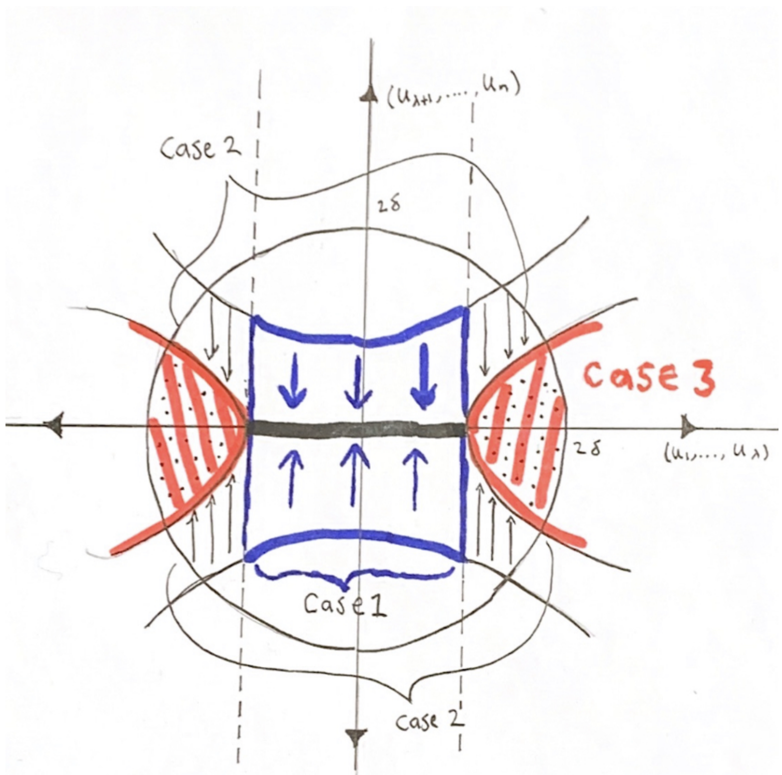

Figure 2: 2δ ball centered at p. Black bold line represents e^λ.

for all $t_1 < t_2$. This shows r_t maps subsets of $F^{-1}(-\infty, c-\delta]$ to itself.

Case 2 ($\delta \le \xi \le \eta + \delta$)
If we were to rearrange the inequality we would get $-\xi \le -\delta \le \eta - \xi$. Suppose x is an element of such set, then $f(x) = c - \xi + \eta \ge c - \delta$. Therefore, $\delta \le \xi \le \eta + \delta$ is contained in H and not $M^{c-\delta}$. Within this region we want to slide points down to the boundary of $M^{c-\delta}$, to do so define r_t as

$$r_t(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, s_t x_{\lambda+1}, \ldots, s_t x_n),$$

where

$$s_t = t + (1-t)\sqrt{\frac{\xi - \delta}{\eta}}.$$

Again, it is obvious that $r_1 = \mathrm{Id}$. To check r_0 maps to the desired set let

$x \in \{x \in M \mid \delta \leq \xi \leq \eta + \delta\}$, then

$$\begin{aligned} f(r_0(x)) &= c - \xi(r_0(x)) + \eta(r_0(x)) \\ &= c - \xi(x) + \frac{\xi(x) - \delta}{\eta(x)}\eta(x) \\ &= c - \delta. \end{aligned}$$

Thus, r_0 maps $\{x \in M \mid \delta \leq \xi \leq \eta + \delta\}$ to $M^{c-\delta}$. We have one more thing to check in this case; continuity of s_t as ξ approaches δ and η approaches 0. Rearranging the inequality we have $0 \leq \xi - \delta \leq \eta$ implying as η approaches 0, $\xi - \delta$ must approach 0 at a faster rate than η, hence

$$0 \leq \lim_{\xi \to \delta, \eta \to 0} \frac{\xi - \delta}{\eta} \leq 1.$$

This completes Case 2.

Case 3 $(\eta + \delta \leq \xi)$
Let $x \in \{z \in M \mid \eta + \delta \leq \xi\}$, then $\eta(x) - \xi(x) \leq -\delta$, plugging into f gives

$$f(x) = c - \xi + \eta \leq c - \delta.$$

Therefore, $x \in M^{c-\delta}$, we should then simply take $r_t = \text{Id}$ for all points in this subset.

This wraps up the proof of the second deformation retractions. To summarize $M^{c+\delta}$ deformation retracts to $F^{-1}(-\infty, c-\delta]$, which in turn deformation retracts to $M^{c-\delta} \cup e^{\lambda}$ as desired.

□

By applying similar ideas to those seen in the proof of Theorem 3.6, we may generalize the result to obtain the following:

Theorem 3.7. *If $p_1, \ldots, p_k$ are all nondegenerate critical points with critical value c and indices $\lambda_1, \ldots, \lambda_k$ respectively, then for δ sufficiently small $M^{c+\delta}$ has the same homotopy type as $M^{c-\delta} \cup_{\rho_1} e^{\lambda_1} \cup \cdots \cup_{\rho_k} e^{\lambda_k}$, where each ρ_i is an attachment map.*

So far we have only discussed local topological properties of manifolds but we have yet to be able to classify the entire space. In the next result Morse theory is applied to determine the homeomorphism type of any compact manifold that admits a smooth function with exactly two nondegenerate critical points.

Theorem 3.8 (Reeb's Theorem)**.** *If M is a compact n-dimensional manifold and f is a smooth function on M with only two critical points, both of which are nondegenerate, then M is homeomorphic to the n-sphere.*

Proof. Since M is compact, by Extreme Value Theorem M attains both its max and min, call these points q and p respectively. Since every single point has a neighbourhood homeomorphic to an open subset in $\mathbb{R}^n$, p and q must be the two nondegenerate critical points. We may normalize f so that $f(p) = 0$ and $f(q) = 1$. By Morse's Lemma there exist an open set U_p of p such that for p has local coordinates 0 and for all $x \in U_p$

$$f(x) = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Similarly, there exist an open neighbourhood U_q of q such that

$$f(x) = 1 - x_1^2 - x_2^2 - \cdots - x_n^2.$$

There then exist $\epsilon > 0$ such that

$$M^\epsilon = \overline{\psi_p(B_\epsilon^p(0))} \quad \text{and} \quad f^{-1}[1-\epsilon, 1] = \overline{\psi_q(B_\epsilon^q(0))}.$$

Where $\psi_{p/q}$ is the diffeomorphism that maps an open in $\mathbb{R}^n$ set to $U_{p/q}$. Since there are no other critical points in $f^{-1}[\epsilon, 1-\epsilon]$, Theorem 3.5 tells us M^ϵ is diffeomorphic to $M^{1-\epsilon}$. This implies $M = M^{1-\epsilon} \cup f^{-1}[1-\epsilon, 1]$ or rather M is the union of two sets homeomorphic to the closed n-disks such that their intersection lines up along the boundary. Recall the n-sphere S^n may be constructed by taking two n-disks B_n and glueing their boundaries together via the map Id $|_{\partial B_n}$. We may then glue the boundary of $M^{1-\epsilon}$ to $f^{-1}[1-\epsilon, 1]$ in a similar fashion call this attachment map ρ. Define the homeomorphism $g : \overline{\psi_p(B_\epsilon^p(0))} \bigcup_{\text{Id}} \overline{\psi_q(B_\epsilon^q(0))} \to M^{1-\epsilon} \bigcup_\rho f^{-1}[1-\epsilon, 1]$ by

$$g(x) = \begin{cases} \psi_p(x) & x \in B_\epsilon^p(0) \\ \psi_q(x) & x \in B_\epsilon^q(0) \\ \rho(\psi_q(x)) & x \sim_{\text{Id}} y. \end{cases}$$

Therefore, M is homeomorphic to the n-sphere. □

As culmination of the theorems and ideas we've established thus far, we can now relate the homotopy type of a manifold to a corresponding CW-complex.

Theorem 3.9. *If f is a smooth real-valued function on a compact manifold M and f has only nondegenerate critical points, then M has the homotopy type of a CW-complex with a λ-cell for each critical point of index λ.*

The proof makes use of the following two lemmas and theorem, only one of which we will present with a proof.

Lemma 3.10. *Let ρ_0 and ρ_1 be homotopy maps from ∂e^λ to a space X, then the identity map on X extends to a homotopy equivalence between $X \cup_{\rho_0} e^\lambda$ and $X \cup_{\rho_1} e^\lambda$.*

Unlike most proofs we have presented thus far that came from Milnor's text, the following may be found in [Whitehead, 1949].

Proof. Let ρ_t be the homotopy between ρ_0 and ρ_1. Define $k : X \cup_{\rho_0} e^\lambda \to X \cup_{\rho_1} e^\lambda$ by

$$\begin{cases} k(x) = x & x \in X \\ k(tu) = 2tu & 0 \leq t \leq \frac{1}{2} \text{ and } u \in \partial e^\lambda \\ k(tu) = \rho_{2-2t}(u) & \frac{1}{2} < t \leq 1 \text{ and } u \in \partial e^\lambda. \end{cases}$$

Notice that for all u along the boundary of e^λ, u is related to $\rho_0(u)$. From our construction $k(u) = \rho_0(u) = k(\rho_0(u))$, so k is well-defined. Similarly, define $q : X \cup_{\rho_1} e^\lambda \to X \cup_{\rho_0} e^\lambda$ by

$$\begin{cases} q(x) = x & x \in X \\ q(tu) = 2tu & 0 \leq t \leq \frac{1}{2} \text{ and } u \in \partial e^\lambda \\ q(tu) = \rho_{2t-1}(u) & \frac{1}{2} < t \leq 1 \text{ and } u \in \partial e^\lambda. \end{cases}$$

Then $q \circ k$ is homotopic to the identity on $X \cup_{\rho_0} e^\lambda$ and the homotopy is given by

$$\begin{cases} h_s(x) = x & x \in X \text{ for all } s \in [0,1] \\ h_s(tu) = (4-3s)tu & 0 \leq t \leq \frac{1}{4-3s} \text{ and } u \in \partial e^\lambda \\ h_s(tu) = \rho_{\frac{1}{2}(4-3s)(1-t)}(u) & \frac{2-s}{4-3s} \leq t \leq 1 \text{ and } u \in \partial e^\lambda, \end{cases}$$

concluding the proof of the lemma. □

Lemma 3.11. *Let $\rho : \partial e^\lambda \to X$ be an attaching map and $f : X \to Y$ be any homotopy equivalence, then f extends to a homotopy equivalence between $X \cup_\rho e^\lambda$ and $Y \cup_{f \circ \rho} e^\lambda$.*

Theorem 3.12 (Cellular Approximation Theorem)**.** *Let X and Y be CW-complexes, if $f : X \to Y$ is continuous then f is homotopic to an attaching map.*

We will now proceed with the proof of Theorem 3.9.

Proof. Since M is compact by Corollary 2.8, M has only finitely many critical values, which we will denote as $c_1 < \cdots < c_k$. Note c_1 is the minimum of f and $M^{c_k} = M$. The proof follows from induction. We will show for each $a \in \mathbb{R} \setminus \{c_1, \dots, c_k\}$, M^a has the homotopy type of a CW-complex.

For the base case, let $c_1 < a < c_2$. By Theorem 3.7, for a sufficiently small δ, $M^{c_1+\delta}$ is homotopy equivalent to $M^{c_1-\delta}$ union a collection of 0-cells. But $M^{c_1-\delta}$ is just the empty set. Applying Theorem 3.5 we get M^a has the homotopy type of a collection of 0-cells. For the inductive step assume $c_{i-1} < a < c_i$ and M^a has the homotopy type of a CW-complex K. Let $h' : M^a \to K$ be the homotopy equivalence. By Theorem 3.7 there exists $\delta > 0$ such that $M^{c_i+\delta}$ is homotopy equivalent to

$$M^{c_i-\delta} \cup_{\rho_1} e^{\lambda_1} \cup \cdots \cup_{\rho_n} e^{\lambda_n}.$$

Applying Theorem 3.5 once again there exists a homotopy equivalence $h : M^{c_i-\delta} \to M^a$. By Cellular Approximation theorem for each $j \in \{1, \dots, n\}$, $h' \circ h \circ \rho_j : \partial e^{\lambda_j} \to K$ is homotopic to an attachment map

$$\psi_j : \partial e^{\lambda_j} \to (\lambda_j - 1) \text{ skeleton of K}.$$

Therefore,

$$K \cup_{\psi_1} e^{\lambda_1} \cup \cdots \cup_{\psi_n} e^{\lambda_n}$$

is a CW-complex. Notice that $h' \circ h$ is a homotopy equivalence from $M^{c_i-\delta}$ to K, thus for each attachment map ρ_j Lemma 3.11 gives

$$M^{c_i-\delta} \cup_{\rho_j} e^{\lambda_j} \simeq K \cup_{h' \circ h \circ \rho_j} e^{\lambda_j}.$$

But $h' \circ h \circ \rho_j$ is homotopic to ψ_j so an application of Lemma 3.10 yields

$$K \cup_{h' \circ h \circ \rho_j} e^{\lambda_j} \simeq K \cup_{\psi_j} e^{\lambda_j}.$$

Building upon this construction for each $j \in \{1, \dots, n\}$ we see that

$$M^{c_i+\delta} \simeq M^{c_i-\delta} \cup_{\rho_1} e^{\lambda_1} \cup \cdots \cup_{\rho_n} e^{\lambda_n} \simeq K \cup_{\psi_1} e^{\lambda_1} \cup \cdots \cup_{\psi_n} e^{\lambda_n},$$

so $M^{c_i+\delta}$ has the homotopy type of a CW-complex. Given $a' \in \mathbb{R}$ such that $c_i < a' < c_{i+1}$, we may conclude $M^{a'}$ has the homotopy type of a CW-complex. □

The result above is not actually limited to compact manifolds, the only requirement is that M^a is compact for each $a \in \mathbb{R}$. Continuing from the end of the proof for the compact case, the theorem of Combinatorial Homotopy would allow one to extend the theorem to the case of infinitely many critical points in M [Milnor, 1963].

4 Existence of Morse Functions

The theory we have developed so far has been dependent on functions with nondegenerate critical points, therefore such functions deserve a proper name; If $f : M \to \mathbb{R}$ is a smooth function and all critical points of f are nondegenerate, then f is a Morse function. It will certainly be helpful to know such functions do in fact exist on our manifold, else Morse theory would seem quite vacuous. In this section we establish that most smooth functions are actually Morse, and functions that are not, may be used to construct new functions that are Morse [Guillemin and Pollack, 1974]. Fortunately most of the heavy lifting has already been done for us by Sard's Theorem.

Theorem 4.1 (Sard's Theorem)**.** *Let M and N be manifolds and $f : M \to N$ be a smooth map between manifolds. Then the set of critical values of f has measure 0.*

So far we have only talked about critical values, we say $f(x) \in Y$ is a regular value if $f(x)$ is not a critical value. More precisely we define regular values as:

Definition 4.2. *Let $f : M \to N$ be a smooth map of manifolds. A point y in the image of f is a regular value of N if for all $x \in M$ such that $f(x) = y$, the map $df_x : TM_x \to TN_y$ is surjective.*

So a critical value is a point y where the derivative map is not surjective. In the case of real-valued functions, the derivative map is not surjective when it is identically 0. We may then equivalently restate Sard's Theorem as; almost every point in Y is a regular value.

Remark. The term "almost" is used just as it is in Measure Theory, that is the complement set has measure 0.

As before, we will assume our manifold M is embedded in euclidean space $\mathbb{R}^N$. Under the usual coordinates of $\mathbb{R}^N$, let $a = (a_1, \ldots, a_N)$ be an N-tuple in $\mathbb{R}^N$, then given any smooth real-valued function f on M we will define a new function

$$f_a = f + a_1x_1 + \cdots + a_Nx_N.$$

The objective is then to show that for almost all a in $\mathbb{R}^N$, f_a is Morse. Let us start with the simpler case of a smooth map between an open set U in $\mathbb{R}^N$ to $\mathbb{R}$.

Lemma 4.3. *Let $f : U \subseteq \mathbb{R}^N \to \mathbb{R}$ be a smooth function on an open set U, then for almost all $a \in \mathbb{R}^N$, f_a is Morse on U.*

Proof. Let $a = (a_1, \ldots, a_N)$ be an N-tuple in $\mathbb{R}^N$ and let $g : U \subseteq \mathbb{R}^N \to \mathbb{R}^N$ be defined as

$$g = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}\right),$$

then

$$dg(x) = H(x).$$

Since $f_a = f + a_1x_1 + \cdots + a_Nx_N$, we have that the gradient of f_a is

$$\begin{aligned} df_a(x) &= \left(\frac{\partial f}{\partial x_1}(x) + a_1, \ldots, \frac{\partial f}{\partial x_N}(x) + a_N\right) \\ &= g(x) + a. \end{aligned}$$

Therefore, the set of critical points of f_a is the set of $p \in U$ such that $g(p) = -a$. If we assume that $-a$ is a regular value, then $dg(p)$ is surjective or equivalently $H(p)$ is non-singular. But the hessian matrix of f coincides with the hessian matrix of f_a, so non-degeneracy of critical points of f_a relies on the condition that $-a$ is a regular value of g. Sard's Theorem tells us that almost every a is a regular value of g. Hence, for almost all $a \in \mathbb{R}^N$, f_a is Morse. □

Translating this result to a manifold will be a bit more tricky, we will need to establish some new concepts and lemmas before we may tackle it. Consider $\mathbb{R}^N$ and suppose $N = k + l$, where k and l are integers then $\mathbb{R}^N = \mathbb{R}^k \times \mathbb{R}^l$. Let $c \in \mathbb{R}^k$, we denote the vertical slice as $V_c = \{c\} \times \mathbb{R}^l$. Let A be a subset of V_c, then $A = \{c\} \times U$ for some subset U of $\mathbb{R}^l$, we say A has measure 0 in V_c if U has measure 0 in $\mathbb{R}^l$.

The following two lemmas will be important in the proof Fubini's Theorem on measure 0 and their proofs may be found in the appendix.

Lemma 4.4. *Let $S_1, \ldots, S_n$ be an open cover of the closed interval $[a, b]$, then there exists another cover $Q_1, \ldots, Q_m$ such that each Q_j is contained in some S_i and*

$$\sum_{i=1}^{m} \mathit{length}(Q_i) < 2(b - a).$$

Lemma 4.5. *Let A be a compact subset of $\mathbb{R}^N$ and let $c \in \mathbb{R}$. Suppose $A \cap V_c \subseteq \{c\} \times U$ for U open in $\mathbb{R}^{N-1}$ then for a sufficiently small interval I containing c, $A \cap V_I \subseteq I \times U$.*

Theorem 4.6 (Fubini's Theorem on Measure 0)**.** *Let A be a closed subset of $\mathbb{R}^N$ such that $A \cap V_c$ has measure 0 in V_c for all $c \in \mathbb{R}^k$ then A has measure 0 in $\mathbb{R}^N$.*

Proof. We will prove the theorem for the case $k = 1$ and $l = N - 1$. For $k > 1$, the argument follows from induction. Since A is closed we may assume A is made up of countably many compact sets (just take $A_n = [-n, n] \cap A$). Therefore, it suffices to show a compact set A with the above property has measure 0.

Let $\epsilon > 0$. Assuming A to be compact, we have that A is also bounded so there exists $I = [a, b]$ such that $A \subseteq I \times \mathbb{R}^{N-1}$. Let $c \in I$, by the hypothesis $A \cap V_c$ being compact and having measure 0 we may cover $A \cap V_c$ with finitely many open $(n-1)$ dimensional rectangles, denoted by $S'_1(c), \ldots, S'_{N_c}(c)$ with total volume less than ϵ. Notice

$$A \cap V_c \subseteq \{c\} \times \bigcup_{i=1}^{N_c} S'_i(c),$$

by Lemma 4.5, there exists an open interval J_c such that

$$A \cap V_{J_c} \subseteq J_c \times \left(\bigcup_{i=1}^{N_c} S'_i(c) \right).$$

This holds for all $c \in I$ and clearly J_c is an open over of I. By compactness of I and Lemma 4.4, we may assume the existence of finitely many open intervals $J'(c_1), \ldots, J'(c_m)$ that cover I with total length less than $2(b - a)$. Note each $J'(c_i) \subseteq J_c$ for some c, therefore

$$\bigcup_{i=1}^{m} \left(J'(c_i) \times \left(\bigcup_{j=1}^{N_{c_i}} S_j(c_i) \right) \right)$$

covers A and possesses total volume less than $2\epsilon(b - a)$. We may conclude A has measure 0 in $\mathbb{R}^N$. □

We now have all the tools to show existence of Morse functions on a manifold.

Theorem 4.7. *Assume M is a k-dimensional manifold embedded into $\mathbb{R}^N$. Let $f : M \to \mathbb{R}$ be any smooth function, then for almost all $a \in \mathbb{R}^N$, the function f_a is Morse on M.*

Proof. We are going to want to first parametrize M with a subset of the usual coordinate functions on $\mathbb{R}^N$ so that f_a is well-posed. Let $p \in M$ and $(x_1, \dots, x_N)$ be the usual coordinates on $\mathbb{R}^N$. Consider the dual space of $\mathbb{R}^N$ with its usual basis $\phi_1, \dots, \phi_N$, that is

$$\phi_i(x_1, \dots, x_N) = x_i.$$

Let $V = TM_p$ the tangent space of M at p. Then the set $\phi_1|_V, \dots, \phi_N|_V$ spans the dual space of V but is not linearly independent as $\mathrm{Dim}(TM_x) = k$. There will however exist a subset $\phi_{i_1}|_V, \dots, \phi_{i_k}|_V$ that forms a basis for the dual space of V. Now notice the derivative of the coordinate function $x_i : \mathbb{R}^N \to \mathbb{R}$ is $dx_i = \phi_i$. If we restrict x_i to the manifold M, this in turn restricts ϕ_i to the tangent space TM_p. For the prior indexed sub-collection $i_1, \dots, i_k$ the function $f : M \to \mathbb{R}^k$ by

$$f(x_1, \dots, x_N) = (x_{i_1}(x_1, \dots, x_N), \dots, x_{i_k}(x_1, \dots, x_N)),$$

has the following derivative

$$df = (dx_{i_1}, \dots, dx_{i_k}) = (\phi_{i_1}|_V, \dots, \phi_{i_k}|_V).$$

But $\phi_{i_1}|_V, \dots, \phi_{i_k}|_V$ forms a basis on TM_x so df is an isomorphism. By the Inverse Function Theorem $(x_{i_1}, \dots, x_{i_k})$ forms a local diffeomorphism on M at p.

We may then cover M with open sets U_x such that for each U_x there is some sub-collection $x_{i_1}, \dots, x_{i_k}$ mapping U_x to an open set in $\mathbb{R}^k$. Applying 2nd countability of $\mathbb{R}^N$ we may assume there are only finitely many U_i that covers M. Fix a U_i and for convenience assume $x_1, \dots, x_k$ forms a coordinate system on U_i. Let S_i denote the set of $a \in \mathbb{R}^N$ such that f_a is not Morse on U_i. Let $c \in \mathbb{R}^{N-k}$ and similarly to before define $V_c = \mathbb{R}^k \times \{c\}$. Let

$$f_{(0,c)} = f + c_1 x_{N-k+1} + \cdots + c_N x_N.$$

For any choice of $b \in \mathbb{R}^k$ define $f_{(b,c)} : U_i \to \mathbb{R}$ by

$$f_{(b,c)} = f_{(0,c)} + b_1 x_1 + \dots b_k x_k.$$

Since U_i is homeomorphic to an open set in $\mathbb{R}^k$ and non-degeneracy is preserved under diffeomorphic composition, Lemma 4.6 tells us for almost all

$b \in \mathbb{R}^k$, $f_{(b,c)}$ is Morse on U_i. Thus, we have just shown that $S_i \cap V_c$ has measure 0 for all $c \in \mathbb{R}^{N-k}$. By Fubini's Theorem S_i has measure 0. Finally observe that f_a has a degenerate critical point if and only if f_a has a degenerate critical point on some U_i then the set of $a \in \mathbb{R}^N$ in which f_a is not Morse is equal to the union of S_i hence must also be of measure 0. □

This establishes there are many Morse functions and our study of critical points is not restricted to any particular manifold.

5 Morse Inequalities

So far our focus has been on how critical points influences the topology of the manifold, but does the reverse hold true as well? If f is a smooth function, does properties of the manifold play a role in the critical points of f? Fortunately, such relation does exist. The topology of the manifold places a lower bound on the number of critical points of a given index for any smooth function. Morse's original treatment of the topic aimed to describe these relations through a collection of inequalities [Milnor, 1963]. In this section we will try to highlight some of these ideas.

Definition 5.1. *Let X be a topological space and let $Y \subseteq X$ such that (X, Y) is a good pair. The λth Betti-number is the rank of $H_\lambda(X, Y)$ denoted by $R_\lambda(X, Y)$.*

Definition 5.2. *The Euler characteristic of the pair (X, Y) is defined as*

$$\chi(X, Y) = \sum_{\lambda=1}^{\infty} (-1)^\lambda R_\lambda(X, Y).$$

Lemma 5.3. *The Betti-number R_λ is sub-additive whereas the Euler characteristic is additive. More precisely if $Z \subset Y \subset X$ then*

$$R_\lambda(X, Z) \leq R_\lambda(X, Y) + R_\lambda(Y, Z),$$

and

$$\chi(X, Z) = \chi(X, Y) + \chi(Y, Z).$$

Theorem 5.4 (Morse's Weak Inequalities)**.** *Let M be a compact manifold and f a smooth function. If C_λ denotes the number of critical points of f with index λ then*

$$R_\lambda(M) \leq C_\lambda,$$

and

$$\chi(M) = \sum_{\lambda} (-1)^{\lambda} C_{\lambda}.$$

Proof. By perturbing the function in a similar fashion as in the proof of Theorem 3.6, we may assume the critical values of f are all different. Furthermore, Corrollary 2.8 ensures there may only exist finitely many critical values. We may choose $a_1 < a_2 < \cdots < a_k$ such that M^{a_i} contains exactly i critical points and M^{a_i} deformation retracts down to $M^{a_{i-1}} \cup e^{\lambda}$ where λ is the index of the critical point in $f^{-1}(a_i, a_{i+2})$. Notice $M^{a_k} = M$ due to compactness. Then

$$H_j(M^{a_i}, M^{a_{i-1}}) \cong H_j(M^{a_{i-1}} \cup e^{\lambda}, M^{a_{i-1}}).$$

Choosing $X = M^{a_{i-1}} \cup e^{\lambda}$, $A = M^{a_{i-1}}$ and $B = e^{\lambda}$, we have A and B are open subsets of X and their union is X. Recall from Theorem 3.6, $e^{\lambda} \cap M^{a_{i-1}} = \partial e^{\lambda}$. By Excision Theorem,

$$\begin{aligned} H_j(M^{a_i}, M^{a_{i-1}}) &\cong H_j(M^{a_{i-1}} \cup e^{\lambda}, M^{a_{i-1}}) \\ &\cong H_j(e^{\lambda}, \partial e^{\lambda}) \\ &= \begin{cases} \mathbb{Z} & \text{if } j = \lambda \\ 0 & \text{if } j \neq \lambda. \end{cases} \end{aligned} \tag{$\star$}$$

Therefore, $\text{Rank}(H_j(M^{a_i}, M^{a_{i-1}})) = 1$ if and only if j is the index of the critical point and 0 else. Applying sub-additivity of R_{λ} on $\varnothing = M^{a_0} \subset M^{a_1} \subset M^{a_2} \subset \cdots \subset M^{a_k} = M$ gives us

$$R_{\lambda}(M) \leq \sum_{i=1}^{k} R_{\lambda}(M^{a_i}, M^{a_{i-1}}) = C_{\lambda},$$

which proves the first inequality. The second inequality follows from the additivity of the Euler characteristic and ($\star$).

$$\begin{aligned} \chi(M) &= \sum_{i=1}^{k} \chi(M^{a_i}, M^{a_{i-1}}) \\ &= \sum_{i=1}^{k} \sum_{\lambda} (-1)^{\lambda} R_{\lambda}(M^{a_i}, M^{a_{i-1}}) \\ &= \sum_{\lambda} C_{\lambda} \end{aligned}$$

□

The second inequality established by Morse is stronger than the one we just proved. It is referred to as simply Morse Inequalities.

Theorem 5.5 (Morse Inequalities)**.** *Let M be a compact manifold and f a smooth function on M then for each $\lambda \in \mathbb{N}$ we have*

$$R_\lambda(M) - R_{\lambda-1}(M) + R_{\lambda-2}(M) - \cdots \pm R_0(M) \leq C_\lambda - C_{\lambda-1} + C_{\lambda-2} - \cdots \pm C_0.$$

Proof. Let us first consider a topological space X with subspace Y and Z such that $Z \subset Y \subset X$. It will be convenient to define

$$S_\lambda(X,Y) = R_\lambda(X,Y) - R_{\lambda-1}(X,Y) + R_{\lambda-2}(X,Y) - \cdots \pm R_0(X,Y).$$

Claim. S_λ is sub-additive.

Proof of claim. Consider the exact sequence for the triple (X,Y,Z)

$$\ldots \xrightarrow{\partial_{\lambda+1}} H_\lambda(Y,Z) \xrightarrow{i_\lambda} H_\lambda(X,Z) \xrightarrow{q_\lambda} H_\lambda(X,Y) \xrightarrow{\partial_\lambda} H_{\lambda-1}(Y,Z) \xrightarrow{i_{\lambda-1}} \cdots \to 0,$$

by Rank-Nullity Theorem we have $\mathrm{Rank}(i_\lambda) = R_\lambda(Y,Z) - \mathrm{Rank}(\partial_{\lambda+1})$. Therefore,

$$\begin{aligned}
\mathrm{Rank}(\partial_{\lambda+1}) &= R_\lambda(Y,Z) - \mathrm{Rank}(i_\lambda) \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + \mathrm{Rank}(q_\lambda) \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + R_\lambda(X,Y) - \mathrm{Rank}(\partial_\lambda) \\
&\vdots \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + R_\lambda(X,Y) - \cdots \pm R_0(X,Y) \geq 0.
\end{aligned}$$

Collecting the like terms and rearranging yields

$$S_\lambda(X,Z) \leq S_\lambda(X,Y) + S_\lambda(Y,Z),$$

completing the proof of the claim.

All there is left to do is apply sub-additivity of S_λ to $\varnothing \subset M^{a_1} \subset \cdots \subset M^{a_k} = M$ giving us

$$S_\lambda(M) \leq \sum_{i=1}^{k} S_\lambda(M^{a_i}, M^{a_{i-1}}).$$

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots \pm R_0(M) \leq \sum_{i=1}^{k} \sum_{j=0}^{\lambda} (-1)^j R_{\lambda-j}(M^{a_i}, M^{a_{i-1}}).$$

As before, $R_\lambda(M^{a_i}, M^{a_{i-1}}) = 1$ if and only if the critical point in the intersection has index λ and is equal to 0 otherwise. Thus,

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots \pm R_0(M) \leq C_\lambda - C_{\lambda-1} + C_{\lambda-1} - \cdots \pm C_0,$$

as desired. □

The following corollary gives a simple application of the Morse inqualities.

Corollary 5.6. *If* $C_{\lambda+1} = C_{\lambda-1} = 0$ *then* $R_\lambda(M) = C_\lambda$.

Proof. By Morse's weak inequalities, if $C_{\lambda+1} = C_{\lambda-1} = 0$ then $R_{\lambda+1}(M) = R_{\lambda-1} = 0$. By the strong inequality on $\lambda + 1$ we get

$$\begin{aligned} R_{\lambda+1}(M) - R_\lambda(M) + R_{\lambda-1}(M) - \cdots \pm R_0(M) &\leq C_{\lambda+1} - C_\lambda + C_{\lambda-1} - \cdots \pm C_0 \\ R_\lambda(M) + R_{\lambda-2}(M) - \cdots \pm R_0(M) &\geq C_\lambda + C_{\lambda-2} - \cdots \pm C_0. \end{aligned}$$

Applying Morse's inequalities on λ, we see that

$$R_\lambda(M) + R_{\lambda-2}(M) - \cdots \pm R_0(M) = C_\lambda + C_{\lambda-2} - \cdots \pm C_0.$$

Reiterating the argument above but on $\lambda - 1$ yields

$$R_{\lambda-2}(M) - R_{\lambda-3}(M) + \cdots \pm R_0(M) = C_{\lambda-2} - C_{\lambda-2} + \cdots \pm C_0.$$

Putting the two together, we conclude $R_\lambda(M) = C_\lambda$. □

6 Conclusion

To summarize, we have seen how smooth functions directly reflect the topology of the manifold, from classifying homotopy types to computing the homology of the manifold. The intimate relation between critical points and the manifold has indeed been shown to be quite resilient and advantageous. However, we have only scratched the surface; there are many more fascinating ideas and applications that arise from the techniques of Morse theory, including links to Lie theory and the calculus of variations on geodesics, just to name a few. Although we have only covered the fundamentals, these ideas provide a new perspective on the treatment of manifolds and topology as a whole, earning Morse theory a central role in the toolkit of all prospective geometers and topologists.

A potential next step would be to apply the ideas we have seen thus far to explore homology relations on algebraic varieties of complex projective space as described by the Lefschetz theorem [Milnor, 1963] or Milnor's own work on the discovery of the exotic spheres; manifolds that are homeomorphic to the 7-sphere but with a different smooth structure than the usual one [Milnor, 1956]. For a more applied route, one may instead study the connections between Morse theory and quantum mechanics. Interestingly, from a more physics theoretic approach, the concepts of supersymmetric quantum mechanics have been used to obtain an alternate proof of the Morse inequalities and to define "Morse homology" [Witten et al., 1982]. A similar approach was used to investigate the Morse homology of infinite dimensional spaces and gave birth to Floer homology, a more recent development that has led to numerous breakthrough results in symplectic geometry and low-dimensional topology.

The versatility of Morse theory is quite endless. In fact, since its inception in the 1940s, Morse theory has consistently re-emerged throughout the decades in various mathematical fields with new and unexpected results. Notably, Stephen Smale applied Morse theory to prove the famous generalized Poincare Conjecture for dimension $n \geq 5$ which states: every compact n-manifold is homeomorphic to the n-sphere if it is homotopy equivalent to the n-sphere [Guest, 2001]. It is even likely that Morse theory will show up once again at the fore-front of mathematics in the near future, so it is definitely worth keeping an eye out for!

Appendices

A Proof of Lemma 4.4

Proof. Since every open set may be expressed as the countable union of open intervals, we may express each S_i as

$$S_i = \bigcup_{k=1}^{\infty} (a_k^i, b_k^i).$$

Then, the union of all the open intervals of $S_1, \ldots, S_n$ forms an open cover of $[a, b]$. Applying compactness we may assume there exists a sub-cover of m open intervals, denoted by (a_j, b_j) for $j = 1, 2, \ldots, m$. Now let

$$\epsilon = \frac{b-a}{m}.$$

We may assume $a \in (a_1, b_1)$, $b \in (a_m, b_m)$ and the m intervals are indexed so that $b_j \in (a_{j+1}, b_{j+1})$. However, we have no way of guaranteeing this sub-cover satisfies our length condition so we will need an even smaller cover, which may be obtained from the following inductive construction. Let $Q_1 = (a - \epsilon_1, b_1)$ where $\epsilon_1 < \epsilon$ and $a - \epsilon_1 \in (a_1, b_1)$. We have from our indexing $b_1 \in (a_2, b_2)$ then we may choose ϵ_2 so that $b_1 - \epsilon_2 \in (a_2, b_2)$ and $\epsilon_2 < \epsilon$, therefore let $Q_2 = (b_1 - \epsilon_2, b_2)$. Repeat the construction for $j = 3, 4, \ldots, m-1$. For Q_m, choose ϵ_m so that $\epsilon_m < \epsilon$ and both $b + \epsilon_m$ and $b_{m-1} - \epsilon_m$ are contained in (a_m, b_m). Finally let $Q_m = (b_{m-1} - \frac{\epsilon_m}{2}, b + \frac{\epsilon_m}{2})$.

$$\begin{aligned}
\sum_{j=1}^{m} \text{length}(Q_j) &= (b_1 - a + \epsilon_1) + (b_2 - b_1 + \epsilon_2) + \ldots \\
&\quad + (b_{m-1} - b_{m-2} + \epsilon_{m-2}) + \left(b + \frac{\epsilon_m}{2} - b_{m-1} + \frac{\epsilon_m}{2}\right) \\
&= (b - a) + (\epsilon_1 + \cdots + \epsilon_m) \\
&< (b - a) + m\epsilon \\
&= 2(b - a).
\end{aligned}$$

Therefore, $Q_1, \ldots, Q_m$ is our desired cover. □

B Proof of Lemma 4.5

Proof. Suppose no such I exists, then for all I containing c, the set $A \cap V_I$ is not contained in $I \times U$. We may choose a sequence $\{c_i\}_{i=1}^{\infty}$ that converges to

c and for each c_i, there exists $x_i \in \mathbb{R}^{N-1}$ such that $(c_i, x_i) \in A$, but $x_i \notin U$. But A is compact, so there exists a limit point of $\{(c_i, x_i)\}_{i=1}^{\infty}$ in A. Since c_i converges to c, this limit point must be (c, x) for some x. However, this x cannot be contained in U as $\{x_i\}_{i=1}^{\infty} \not\subseteq U$ and U is open. Therefore, we found a point $(c, x) \in A \cap V_c$, but $(c, x) \notin \{c\} \times U$ giving us our contradiction. □

Remark. Notice the interval I may be open or closed.

References

[Guest, 2001] Guest, M. (2001). Morse theory in the 1990's. *arXiv preprint math/0104155*.

[Guillemin and Pollack, 1974] Guillemin, V. and Pollack, A. (1974). *Differential topology*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

[Milnor, 1956] Milnor, J. (1956). On manifolds homeomorphic to the 7-sphere. *Annals of Mathematics*, pages 399–405.

[Milnor, 1963] Milnor, J. (1963). *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J.

[Whitehead, 1949] Whitehead, J. (1949). On simply connected, 4-dimensional polyhedra. *Commentarii Mathematici Helvetici*, 22(1):48–92.

[Witten et al., 1982] Witten, E. et al. (1982). Supersymmetry and morse theory. *J. diff. geom*, 17(4):661–692.