An overview of Morse Theory

Shrivathsa Pandelu

November 21, 2021

Contents

1 Morse Lemma

Let M be a smooth manifold without boundary, and let $f\colon M \to \mathbb{R}$ be a smooth function. The critical points of f are those points $p \in M$ where $df_p = 0$. If $(U, x^1, \dots, x^n)$ is a chart about p, then p is a critical point if $\partial f/\partial x^i = 0 \forall i$. This notion is of course chart independent. Now, suppose p is a critical point, then we may define the Hessian of f at p locally as the matrix $(\partial^2 f/\partial x^1 \partial x^j)_{i,j}$. We need to check that this definition is chart independent.

Let $v, w \in T_pM$ and let $\tilde{v}, \tilde{w}$ be local extensions of v, w respectively (such extensions exist for example by considering "constant" vector fields in a local chart). Then their Lie Bracket satisfies

$$0 = [\tilde{v}, \tilde{w}]_p f = \tilde{v}_p(\tilde{w} f) - \tilde{w}_p(\tilde{v} f).$$

It follows that the bilinear function $(v, w) \mapsto \tilde{v}_p(\tilde{w} f)$ is symmetric and independent of the local extensions $\tilde{v}, \tilde{w}$: it is independent of $\tilde{v}$ because $\tilde{w}_p(\tilde{v} f) = \tilde{v}_p(\tilde{w} f) = v(\tilde{w} f)$ and the right side is independent of $\tilde{v}$.

By a local calculation it is easy to see that the matrix corresponding to this bilinear form in the chart above is precisely the Hessian of f at p. We say that p is a non-degenerate critical point if p is a critical point and the Hessian H is non-degenerate, i.e., if $H(v, w) = 0 \forall w$, then $v = 0$.

In general, given a bilinear form H on a (real) vector space V, the index of H is the maximal dimension of subspaces on which H is negative definite and the nullity of H is the dimension of the nullspace of H, i.e., all those v for which $H(v, w) = 0 \forall w$. It is a theorem of linear algebra that there is a basis of V such that H has the matrix

$$\begin{bmatrix} -I_r & & \\ & I_s & \\ & & 0 \end{bmatrix}$$

where r is the index, $r + s$ is the rank of H.

Lemma 1. (*Morse lemma*) *Let p be a non-degenerate critical point of f. Then there is a local coordinate system $(U, y^1, \dots, y^n)$ around p with $y^i(p) = 0 \forall i$ and*

$$f = f(p) - (y^1)^2 - \cdots - (y^k)^2 + (y^{k+1})^2 + \cdots + (y^n)^2$$

on U where k is the index of f at p.

Proof sketch. In the vector space case we use an argument similar to the Gram-Schmidt procedure. Here we do the same, but shrink the neighbourhood so all the steps (particularly those of divisions and square roots) are well defined. To start, one uses Hadamard lemma (this is the version in [1]) which says the following : if f is a smooth function on a convex neighbourhood V of 0 in $\mathbb{R}^n$, with $f(0) = 0$, then

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

for some smooth functions g_i with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

We apply the Hadamard lemma twice(because p is critical, the derivatives vanish) to $f - f(p)$ to obtain a chart where

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

Now, we can assume $h_{ij} = h_{ji}$ by replacing h_{ij} with $\frac{1}{2}(h_{ij} + h_{ji})$. Proceed with diagonalization. k is well defined as it is the index of the Hessian at p. $\square$

As a corollary, non-degenerate critical points of f are isolated. A function f is said to be Morse if all its critical points are non-degenerate. If M is compact, it follows that a Morse function has finitely many critical points (this is not true for non-Morse functions: consider the height function on a torus lying on the plane).

2 Pseudo-gradients

Let M be a manifold, $f\colon M \to \mathbb{R}$ a Morse function. A pseudo-gradient for f is a vector field X on M such that

- $df(X) \leq 0$ everywhere with equality only at critical points of f.
- In a Morse chart around critical points, X agrees with the negative of the usual Euclidean gradient of f.

Given a critical point p, by Morse lemma, there is a chart $(U, x^1, \ldots, x^n)$ such that on $Ux^i(p) = 0$ and

$$f = f(p) - (x^1)^2 - \cdots - (x^k)^2 + (x^{k+1})^2 + \cdots + (x^n)^2$$

where k is the index of f at p. Denote by V_-, V_+ to be the span of $\{x^1, \ldots, x^k\}, \{x^{k+1}, \ldots, x^n\}$ respectively intersected with U. Note that we only require X to agree with $-\text{grad} f$ on some Morse chart around critical points.

If M has a Riemannian metric, then it is known that any function f has gradient, which is given by "dualising" the differential of f, see [3]. However, on a general manifold, we do not have a canonical choice of a Riemannian metric, so it is not a priori guaranteed that a pseudo gradient exists. However, just as every manifold can be given a Riemannian metric, we can construct a pseudo-gradient for any given f. Both the results involve a local construction patched up using a partition of unity.

Intuitively, the gradient is the direction of maximum increase of f, so a trajectory along a pseudo-gradient corresponds to travelling along a line where f decreases. The trajectories at critical points are constants, and under the assumption that M is compact, since every flow is complete, we expect that the trajectories start and end at critical points, because f cannot keep decreasing. We shall prove this soon.

Theorem 1. (*Existence of pseudo-gradient*) *Given a compact manifold M and a morse function f, a pseudo-gradient of f exists.*

Proof. Observe that the morse lemma implies that the critical points of f are isolated. Since the set of critical points is a closed set, by compactness, it is finite. Let $c_1, \ldots, c_r$ be the critical points. Around each, take a Morse chart $(U_i, \phi_1)_{1 \le i \le r}$. Next, include some more charts $(U_i, \phi_i)_{r < i \le N}$ covering M. We may shrink these other charts so that they contain no critical points (there are finitely many critical points). On each U_i, we have the function $f \circ \phi_i^{-1}$, and it's corresponding (pullback of) negative gradient X_i on $\phi_i^{-1}(U_i)$. Let ψ_i be a partition of unity subordinate to the given cover, and extend X_i to $\tilde{X}_i$ by using ψ_i and set $X = \sum \tilde{X}_i$. We claim that X is a pseudo-gradient for f.

By construction, each critical point is in a single U_i, so on a smaller Morse chart X agrees with the negative gradient. Next, at each point $x \in M$,

$$(df)_m(X) = \sum \phi_i(m) df_m(X_i(m)) \le 0$$

and if it is equal to zero, then each term must be zero, which means that either m is a critical point or $\phi_i(m) = 0 \forall i$ which is impossible. Therefore, equality holds only at the critical points. Thus, X is a pseudo-gradient for f. □

Given a smooth vector field, we have the corresponding flows $\phi_t(x)$ for each $x \in M$. Since M is compact, $\phi_t(x)$ is defined for all $t \in \mathbb{R}$. Given $a \in M$, define

$$W^s(a) = \{x \in M | \lim_{t \to \infty} \phi_t(x) = a\},$$

$$W^u(a) = \{x \in M | \lim_{t \to -\infty} \phi_t(x) = a\}.$$

$W^s(a)$ is called the stable manifold of a and $W^u(a)$ the unstable manifold. The flow for X is a path where f is decreasing, so if a point is in the stable manifold of a, then $\phi^t(x)$ is a path of decreasing f from x to a.

2.1 The standard ball

Suppose p is a critical point of a morse function f on a compact manifold M. Let $(U, x^1, \ldots, x^n)$ be a Morse chart around p where $f(x) = f(p) - (x^1)^2 - \cdots - (x^k)^2 + (x^{k+1})^2 + \cdots + (x^n)^2$ with k, the index of f at p. Denote by Q the quadratic form $-||x_-||^2 + ||x_+||^2$ where x_- is the vector $(x^1, \ldots, x^k)$ and x_+ is $(x^{k+1}, \ldots, x^n)$. The standard balls around p is defined as

$$U(\epsilon, \eta) = \{x : |Q(x)| \le \epsilon, ||x_-||^2 ||x_+||^2 \le \eta(\epsilon + \eta)\}.$$

The idea is as follows. Given a pseudo-gradient for f, we want to analyse its flow. Since M is compact, the flows are complete, i.e., they exist for all time. However, by definition/construction, the value of f decreases along the flow lines, but because M is compact we cannot expect f to decrease forever. So, intuitively, we expect the flow lines to start from and end at critical points.

To prove our intuition, we would like to know how the flow behaves near critical points. To this end, on a neighbourhood as above, the pseudo-gradient X matches with the actual negative gradient (shrink U if necessary), so in coordinates we have

$$X = \sum_{i=1}^{k} 2x^i \partial_i + \sum_{i=k+1}^{n} -2x^i \partial_i$$

Then by the uniqueness of solutions to ODEs, the flow starting from $(c^1, \ldots, c^n) \in U$ is given by

$$\gamma : t \mapsto (c^1 e^{2t}, \ldots, c^k e^{2t}, c^{k+1} e^{-2t}, \ldots, c^n e^{-2t})$$

as long as it lies in U. This is a curve of the form $||x_-||^2 ||x_+||^2 = \text{const}$.

On U, consider the continuous map $\theta\colon x \mapsto (||x_-||, ||x_+||) \in \mathbb{R}^2$. In $\mathbb{R}^2$ we consider the region $S = |u^2 - v^2| \le \epsilon, u^2 v^2 \le \eta(\epsilon + \eta)$ so thatl $U(\epsilon, \eta)$ is $\theta^{-1}(S)$. Because θ is continuous, the boundary of $U(\epsilon, \eta)$ is the inverse of the boundary of S, and this boundary divides U into 2 regions. It is easy to verify that the boundary is given by

$$\partial_{\pm} U = \{x : Q(x) = \pm\epsilon, ||x_{\mp}||^2 = \eta\}$$
$$\partial_0 U = \{x : ||x_-||^2 ||x_+||^2 = \eta(\epsilon + \eta)\}$$

Our vector field is $(2x_-, -2x_+)$ and the flow through $c = (c_-, c_+)$ is given by $(e^{2t}c_-, e^{-2t}c_+)$. Under θ, the flow is mapped to $(e^{2t}||c_-||, e^{-2t}||c_+||)$. In dimension 2, the flows of $(u, -v)$ are precisely the level sets of $(u, v) \mapsto u^2v^2$ (this is not true in higher dimensions), so the quadratic form $-u^2 + v^2$ strictly decreases along the level sets of u^2v^2.

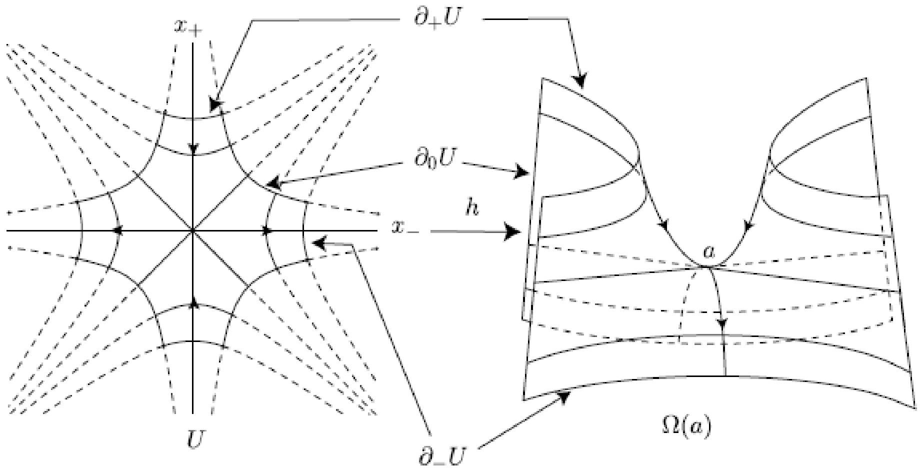

Figure 1: The standard ball, h is the diffomorphism from the Morse chart to the neighbourhood $\Omega(a)$ around critical point a. Figure taken from [2].

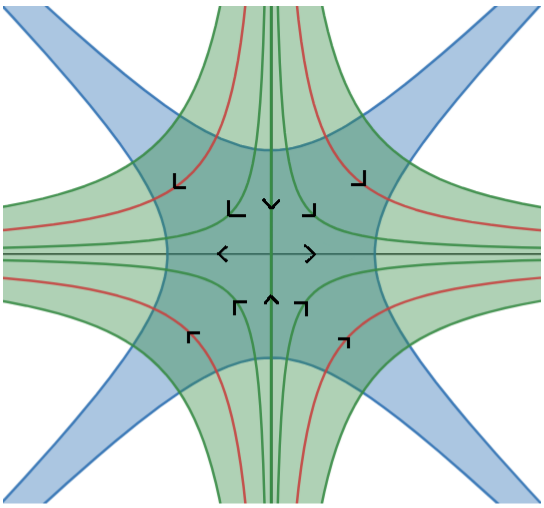

Figure 2: Here we see the regions in the standard ball and the flow lines are marked with arrows. Figure made using Desmos: [4].

With all this background, suppose γ is any arbitrary flow to X that enters $U(\eta, \epsilon)$ (for small enough η, ϵ this standard ball is contained inside a usual Euclidean ball). Then we look at the image of this curve under θ. By looking at the flows of $(u, -v)$ in $\mathbb{R}^2$, we conclude that either γ must reach the origin in infinite time, or it reaches the $Q = -\epsilon$ region. And any point on the $Q = -\epsilon$ region cannot enter $U(\eta, \epsilon)$. From this analysis we conclude that if γ enters and leaves any $U(\eta, \epsilon)$ then it cannot re-enter the same region. From here we get the following theorem.

Theorem 2. *Let M be compact and γ a trajectory for X, then there exist critical points c, d of f such that $\lim_{t\to\infty} \gamma(t) = c, \lim_{t\to-\infty} \gamma(t) = d$.*

Proof. We consider the limit towards $t = \infty$, the other end is similar. There are finitely many critical points, say $c_1, \dots c_r$, and suppose none of these are a limit point. Then around each c_i we get a $U_i(\eta_i, \epsilon_i)$ which γ doesn't enter or enters and leaves (by the analysis above, it cannot oscillate within the bounds of the standard ball). Let Ω be the union of the interiors of these standard balls. Observe that $df(X)$ is a continuous negative function which vanishes precisely at the c_i, therefore outside Ω, there is a positive number ϵ_0 such that $df(X) \leq -\epsilon_0$.

Since f decreases along γ, and, by assumption, there is a time t_0 such that for $t \geq t_0, \gamma \cap \Omega = \emptyset$, we conclude that

$$f(\gamma(t)) = f(\gamma(t_0)) + \int_{t_0}^{t} df_{\gamma(s)}(X(\gamma(s))ds \leq f(\gamma(t_0)) - \epsilon_0(t - t_0) \to -\infty$$

which is impossible. Therefore, γ must tend towards critical points as $t \to \pm\infty$. □

2.2 Stable and unstable manifolds

Since M is compact, X generates a one-parameter family of diffeomorphisms $\phi_t \colon M \to M$ corresponding to the flows along X. We defined the sets

$$W^s(a) = \{x \in M | \lim_{t\to\infty} \phi_t(x) = a\},$$
$$W^u(a) = \{x \in M | \lim_{t\to-\infty} \phi_t(x) = a\}.$$

In this subsection we prove that these are actually submanifolds of M. We will prove that the stable manifold is a manifold, the unstable part is similar. On a chart such as U as above, the points that terminate at p are the points corresponding $x_- = 0$ and these lie on a sphere S^{n-k-1} of the form $||x_+||^2 = \epsilon, x_- = 0$ in U (for a small ϵ). Consider

$$\begin{aligned} \Phi \colon S^{n-k-1} \times \mathbb{R} &\to M \\ (x, t) &\mapsto \phi_t(0, x) \end{aligned}$$

The image of Φ is precisely $W^s(p) \setminus \{p\}$ because any (non constant) flow line terminating at p has to pass through the boundary of $U(\epsilon, \eta)$ for some η and since it terminates at p, it must actually pass through S^{n-k-1}, which means, by the properties of a one-parameter family of diffeomorphisms, it is in the image of Φ.

We claim that Φ is an embedding. If $\Phi(z_1, t_1) = \Phi(z_2, t_2)$, then $\phi_{t_1-t_2}(0, z_1) = (0, z_2)$ but this is impossible unless $(z_1, t_1) = (z_2, t_2)$ as z_1, z_2 are on the sphere and the flow is radially inward. Therefore Φ is injective.

Next, the pseudo-gradient is not tangential to the sphere since the sphere is a level set of f. Thus, the push forward of the tangent space from a point $(x, t) \in S^{n-k-1} \times \mathbb{R}$ is injective because ϕ_t is a diffeomorphism (so the map is injective on tangent space of S^{n-k-1}) and X is not tangential at $t = 0$, so it continues to be not in the span of the push forward of the tangent space at any time (because ϕ_t is a diffeomorphism).

More precisely, let $v_1, \dots, v_{n-k-1}$ be a basis for the tangent space of S^{n-k-1} at x. Under Φ, the push forward of these vectors is $(\phi_t)_* v_1, \dots, (\phi_t)_* v_{n-k-1}$ because it's just the restriction of ϕ_t to S^{n-k-1}. Furthermore,

$$\Phi_*(\partial/\partial t) = X(\Phi(x, t)) = (\phi_t)_*(X|_{(0,x)})$$

where $(0, x)$ is the point in U. This equality follows from the fact that ϕ_t is a flow of X. Since $X|_{(0,x)}$ is not tangential to the sphere (because the sphere is a level set of f), and ϕ_t is a diffeomorphism, it follows that Φ_* is injective.

So, we have proved that Φ is a $1-1$ immersion. To show that it is an embedding, we need to show that Φ is a homeomorphism onto its image. First, since it is an immersion, it is a local embedding, i.e., around every $(x_0, t_0) \in S^{n-k-1} \times \mathbb{R}$, there is a neighbourhood V such that $\Phi|_V$ is an embedding. The problem lies in showing that this image is open in $\Phi(S^{n-k-1} \times \mathbb{R})$.

We have $W^s(p) \cap U = \{x \in U : x_- = 0\}$ which is a submanifold of U, hence a submanifold of M. Given a $q \in S^{n-k-1}$, it is easy to see that Φ is a local embedding (look at it in terms of coordinates), so there is a neighbourhood V of $(q, 0) \in S^{n-k-1} \times \mathbb{R}$ and a neighbourhood W of $(0, q) \in U$ such that $\Phi\colon V \to W \cap W^s(p)$ is a diffeomorphism.

Given a $T \neq 0$, let l_T denote the translation of $\mathbb{R}$ by T. We have the commutative diagram

$$\begin{tikzcd} S^{n-k-1} \times \mathbb{R} \arrow[r, "\Phi"] \arrow[d, "1\times l_T"'] & M \arrow[d, "\phi_T"] \\ S^{n-k-1} \times \mathbb{R} \arrow[r, "\Phi"] & M \end{tikzcd}$$

Restricting this to V gives the following commutative diagram

$$\begin{tikzcd} V \arrow[r, "\Phi"] \arrow[d, "1\times l_T"'] & W \cap W^s(p) \arrow[d, "\phi_T"] \\ V' \arrow[r, "\Phi"] & \phi_T(W) \cap W^s(p) \end{tikzcd}$$

where V' is the image of V under $1 \times l_T$. Because ϕ_T is a diffeomorphism and $W^s(p)$ is invariant under ϕ_T,

$$\phi_T(W \cap W^s(p)) = \phi_T(W) \cap \phi_T(W^s(p)) = \phi_T(W) \cap W^s(p).$$

In this diagram, all maps except the bottom one are diffeomorphisms, therefore the bottom map is a diffeomorphism. Since V' is a neighbourhood of $(q, T) \in S^{n-k-1} \times \mathbb{R}$ and $\phi_T(W)$ is a neighbourhood of $\phi_T(q)$, it follows that Φ is a local embedding onto its image, i.e., around each q there are neighbourhoods around $q, \Phi(q)$ such that the restriction is an embedding. Because Φ is injective, it is a homeomorphism onto its image, i.e., it is an embedding.

So, $W^s(p) \setminus \{p\}$ is an embedded submanifold. Using the diffeomorphism $(0, 1) \to \mathbb{R} : s \mapsto \ln(s/1 - s)$, we get the embedding $\Psi\colon S^{n-k-1} \times (0, 1) \to M$. In local coordinates

$$\Psi\colon (x_{k+1}, \dots, x_n, s) \mapsto (0, \dots, 0, x_{k+1}((1 - s)/s)^2, \dots, x_n((1 - s)/s)^2)$$

This map extends to $S^{n-k-1} \times (0, 1]$ and from there we can quotient out $S^{n-k-1} \times \{1\}$ to get a the diffeomorphism $D^{n-k} \cong W^s(p)$. A similar analysis (or using $-f$) holds for $W^u(p)$, thus $W^s(p), W^u(p)$ are embedded submanifolds diffeomorphic to discs with

$$\dim W^u(p) = \operatorname{codim} W^s(p) = \operatorname{Ind}(p).$$

Soon we shall see the stable and unstable manifolds on a torus corresponding to a height function, see Figure 4.

3 Morse Theorems

Given a real valued function f on M, let $M^a = f^{-1}(-\infty, a]$.

Theorem 3. (*First Morse theorem*) *Let f be a smooth real valued function on a manifold M (not necessarily compact; without boundary). Let $a < b$ and suppose $f^{-1}[a, b]$ is compact and without critical points. Then M^a is diffeomorphic to and a deformation retract of M^b and furthermore, the inclusion $M^a \hookrightarrow M^b$ is a homotopy equivalence.*

Proof sketch. The idea is to push down along a pseudo-gradient of f. Let ρ be a smooth function satisfying

$$\rho(x) = \begin{cases} -\frac{1}{df_x(X)} & x \in f^{-1}[a,b] \\ 0 & \text{outside a relatively compact neighbourhood of } f^{-1}[a,b] \end{cases}$$

Such a ρ exists by using bump functions. Note that [1] uses a Riemannian metric on M to get hold of a gradient instead of using a pseudo-gradient as in [2]. Let $Y = \rho X$. Since Y has compact support, it's flow is defined for all time (see [1] for an explanation), so let ϕ_t denote the t-time flow of Y and these form a one-parameter family of diffeomorphisms of M.

For a fixed $q \in M$, it can be verified that the t-derivative of $f(\phi_t(q))$ is 1 as long as $\phi_t(q) \in f^{-1}[a,b]$. From here, it follows that $\phi_{b-a}\colon M \to M$ carries M^a diffeomorphically onto M^b because the derivative is 1, so f varies linearly (one also sees that ϕ_{a-b} is the inverse along which f decreases with unit speed). A deformation retract is given by the one parameter family $r_t\colon M^b \to M^a$

$$r_t(q) = \begin{cases} q & x \in f(q) \leq a \\ \phi_{t(a-f(q))}(q) & a \leq f(q) \leq b \end{cases}$$

Note that by the choice of ρ, points outside a neighbourhood of $f^{-1}[a,b]$ are stationary under ϕ_t, so these flows migrate points in M^a to M^b by "stretching" a neighbourhood of the level set of a. □

Remark. The compact-ness hypothesis on $f^{-1}[a,b]$ cannot be removed as seen in this example from [1] where M^a, M^b are not diffeomorphic. Essentially, the puncture obstructs the flow.

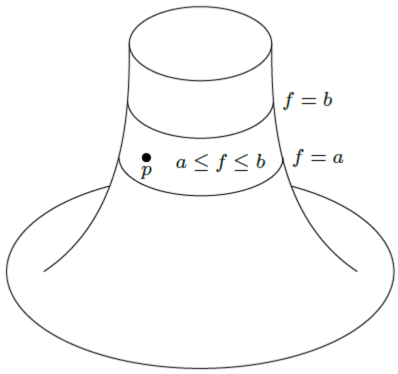

Figure 3: $f^{-1}[a,b]$ is not compact and M^a, M^b are not diffeomorphic. Figure taken from [1].

Corollary. (*Reeb's theorem*) *Let M be a compact manifold (without boundary) admitting a Morse function f with exactly 2 critical points, then M is homeomorphic to the sphere.*

Proof. Being compact M admits a maximum and minimum and these are critical points of index $n, 0$ respectively. Since connected components are closed, hence compact, we conclude that M is connected. By composing with a diffeomorphism, we may assume $f(M) = [0,1]$. By Morse lemma, for sufficiently small $\epsilon > 0$ the sets $f^{-1}[0,\epsilon], f^{-1}[1-\epsilon, 1]$ are discs of dimension n.

By Morse theorem, $M^{1-\epsilon}$ is diffeomorphic to M^ϵ, therefore, $M = M^{1-\epsilon} \cup f^{-1}[1-\epsilon, 1]$ is two discs glued along the boundary $f^{-1}(1-\epsilon)$. It is a standard result that this is homeomorphic to S^n. □

Theorem 4. (*Second Morse theorem*) *Let p be a non-degenerate critical point with index k. Let $f(p) = c$ and suppose $f^{-1}[c-\epsilon, c+\epsilon]$ is compact and without any other critical point for some $\epsilon > 0$. Then for all sufficiently small ϵ, the set $M^{c+\epsilon}$ has the same homotopy type as $M^{c-\epsilon}$ with a k-cell attached.*

See [1] or [2] for a proof. Intuitively, imagine that f is a height function (as in the next section), then if p has index k, then it is as if the manifold "grows" in $n-k$ directions (where double derivative is positive) and along the other directions f has peaked. So, upon crossing p we are "closing" of these k directions by attaching a handle.

4 Vertical Torus, a classic example

Let T denote a torus embedded in $\mathbb{R}^3$ with a parametrization given by

$$(u, v) \mapsto (r \cos u \cos v + R \cos v, r \cos u \sin v + R \sin v, r \sin u)$$

where $u, v \in [0, 2\pi]$ (closed interval to cover the torus), $R > r > 0$. We take a height function given by the projection onto the x-axis and an easy calculation shows that the critical points of this function are given when (u, v) is one of $(0, 0), (\pi, 0), (0, \pi), (\pi, \pi)$. One also finds that the indices are $2(\text{max}), 1, 1, 0(\text{min})$ respectively.

Label the critical points A, B, C, D as in the figure. The stable and unstable manifolds are depicted in the figure below. The flow lines move from one critical point to another.

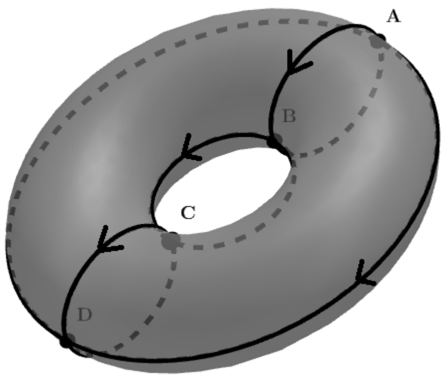

Figure 4: Flow lines on the torus viewed obliquely. Figure made using Geogebra [4].

At A, we have a 2-dimensional unstable manifold, consisting of flows going downwards. At B we have a one dimensional stable manifold consisting of flows originating from A and a one dimensional unstable manifold. Other points are similar. The exact relation and number of flows between critical points leads to Morse homology.

By Morse theorems, we get the following decomposition of the torus

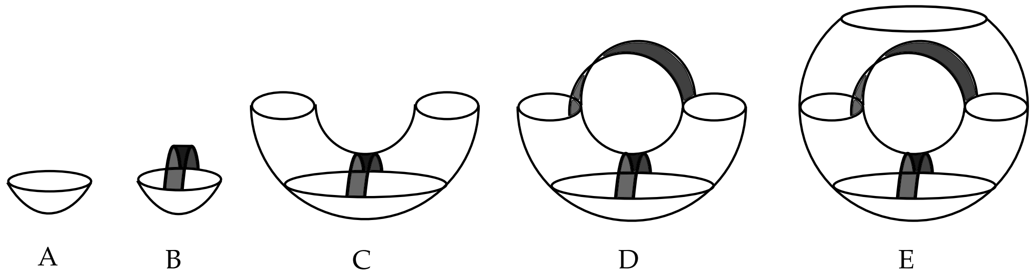

Figure 5: Decomposition of the torus

where A is a disc (homotopic to a point), and from A to B we add a 1 cell which results in a half-torus C. To C we add another 1-cell resulting in D which is homotopic to E. Finally to E we add a 2-cell and "close" the torus.

References

[1] Milnor, *Morse Theory*

[2] Audin and Damian, *Morse Theory and Floer Homology*

[3] Lee, *Riemannian Manifolds: an Introduction to curvature*

[4] Graphing calculator Desmos
Graphing calculator Geogebra