A Brief Review on Morse Theory

Dongho Lee

February 13, 2021

Abstract

Morse theory is a way to see the topology of given manifold by analyzing a function on the manifold. In this note, we give brief overview on the Morse theory. In section 1, we introduce the basic ideas of Morse theory. In section 2, we introduce the construction of Morse homology. In section 3, we give an example of infinite dimensional analogy of Morse function, which is the energy functional on the path space of manifold. We hope that this note would help to read the note on Floer homology, which will be uploaded in the future.

A nice reference for Morse theory is a beautiful book of John Milnor, [Mil1]. I also found that [Mat] was helpful. There's also some stuffs concerning with Morse homology in the first few chapters of [AD].

Contents

1 Morse Functions

1.1 Motivating Example: Height Function on the Surfaces

In this note, every manifold and function is assumed to be smooth. If there's a subtlety, for example a manifold with corners, we will usually ignore the problem since such problem can be somehow solved.

Consider the sphere, torus and genus 2 surface embedded in $\mathbb{R}^3$ as in the Figure 1. Denote these by S_0, S_1, S_2. Define $h_i : S_i \to \mathbb{R}$ by $h(x, y, z) = z$, which could be seen as a *height functions* on the surfaces. One can easily check that h_0, h_1, h_2 has 2, 4, 6 critical points each. For example, the function on a sphere has global minimum and maximum, which correspond to two critical points.

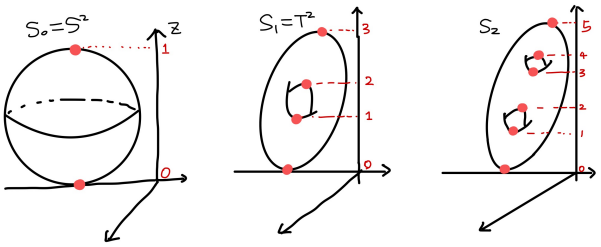

Figure 1: Height Functions on Surfaces

Consider the sublevel set $h^{-1}(-\infty, c] := S^c$. (We omit the subscript for now.) Then one can see that the topology of sublevel set does not change as c passes through regular values, and it changes at critical values. For example, let's see the torus case. The critical values are 0, 1, 2 and 3. We see that, as illustrated in Figure 2,

- If $c < 0$, S^c is an empty set.
- If $0 < c < 1$, S^c is a disk.
- If $1 < c < 2$, S^c is a cylinder.
- If $2 < c < 3$, S^c is a torus with a puncture.
- If $c > 3$, S^c is a torus.

This looks like attaching cells to CW complex, which describes the topology of given manifold.

Let's see one more case with 4 critical points, so called a *Morse heart*. This could be seen as a sphere, which is deformed a bit from its standard form. The critical values are 0, 1, 2 and 3. We see that, as illustrated in Figure 3,

- If $c < 0$, S^c is an empty set.

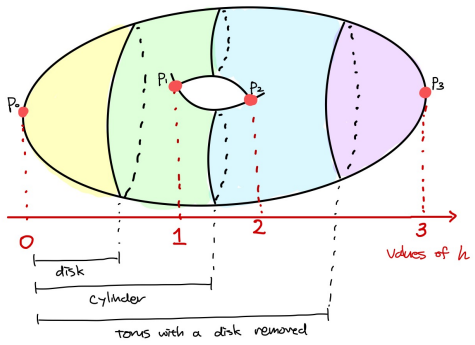

Figure 2: Sublevel sets of 2-torus

- If $0 < c < 1$, S^c is a disk.
- If $1 < c < 2$, S^c is a cylinder.
- If $2 < c < 3$, S^c is a sphere with a puncture (which is a disk).
- If $c > 3$, S^c is a sphere.

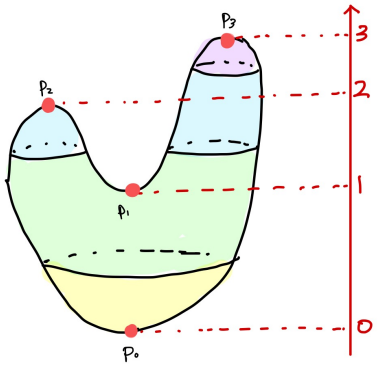

Figure 3: Sublevel sets of Morse heart

What made the difference between two cases? Denote the critical points corresponds to the critical value i by p_i and q_i in torus and Morse heart. One can easily check that p_0 is local minimum, p_1, p_2 are *saddle points*, and p_3 is local maximum. In contrast, in the Morse heart case, q_0 is local minimum, q_1 is saddle point, and q_2, q_3 are local maxima. This local behavior of height functions near critical points make difference between torus and sphere.

Here's the idea: we can see the topology of M by analyzing the behavior of some function f on M near its critical points. This is the starting point of the Morse theory.

1.2 Morse Functions

Let $f : M \to \mathbb{R}$. Recall that the **critical point** of f is a point $p \in M$ such that $d_p f = 0$. In local chart $(x_1, \cdots, x_n)$, we can write this condition as

$$\frac{\partial f}{\partial x_1}(p) = \cdots = \frac{\partial f}{\partial x_n} = 0.$$

We call $f(p)$ **critical value**. At critical point p, $d_p f$ does not contain valuable information of f. Reminiscing the Calculus II course, we would like to see the *second derivatives* at the critical points to extract more information.

Let $\mathrm{Crit}(f)$ be the set of critical points of f. For $p \in \mathrm{Crit}(f)$, define **Hessian** of f at p by $n \times n$-matrix

$$\mathrm{Hess}_p(f) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p) \right)$$

in local chart. In particular, $\mathrm{Hess}_p(f)$ is symmetric and thus defines a symmetric bilinear form on $T_p M$. This matrix describes the local behavior of f near the critical point.

Remark 1.1. Of course, there's a coordinate-free description of Hessian. We would not concern about this. One can look [Mil1] for the definition.

There are standard notions associated to symmetric bilinear form A defined on a real vector space V:

- A can be naturally identified to a map $\tilde{A} : V \to V^*$ by defining $\tilde{A}(v)(w) = A(v, w)$. If $\ker \tilde{A} = 0$, we say A is **nondegenerate**. Otherwise, we say A is **degenerate**. We usually write $\ker A = \ker \tilde{A}$.
- A is **positive definite** if $A(v, v) > 0$ for any $0 \neq v \in V$. Similarly, A is **negative definite** if $A(v, v) < 0$ for any $0 \neq v \in V$.
- **Index** of A is an integer $\mathrm{Ind}(A) := \dim V_-$, where V_- is maximal subspace of V on which A is negative definite.

Let's apply these notions to the Hessian. We call $f : M \to \mathbb{R}$ **Morse function** if it's proper (the inverse image of compact set is compact) and $\mathrm{Hess}_p(f)$ is nondegenerate for any $p \in \mathrm{Crit}(f)$. For Morse function f, define **Morse index** of p by

$$\mathrm{Ind}(p, f) = \mathrm{Ind}(p) := \mathrm{Ind}(\mathrm{Hess}_p(f)).$$

One might check that these definitions do not depend on the choice of local chart which represents Hessian.

Let's see what exactly the nondegeneracy assumption means. The nondegeneracy means, heuristically, the critical points are isolated. Assume that $v \in T_pM$ is degenerate direction of f at p, so that $\mathrm{Hess}_p(v, w) = 0$ for any $w \in T_pM$. For example, let's identify T_pM with $\mathbb{R}^n$ and take $v = e_1$ which is standard basis of $\mathbb{R}^n$, and other basis by e_j. It means that $\partial^2_{1j} f(0) = 0$ for any j, and one can expect that for small $\varepsilon > 0$, $\partial_j f(\pm\varepsilon e_1) = 0$, i.e. $\pm\varepsilon e_1$ is also a critical point of f. The situation is illustrated in Figure 4. We want to avoid such situation, since in this case even $\mathrm{Hess}_p(f)$ cannot catch the behavior of f near p, as the reason we used Hessian instead of $d_p f$.

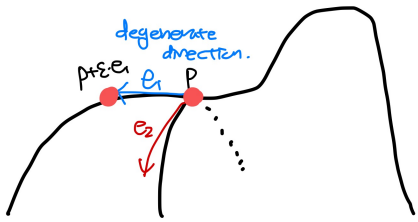

Figure 4: Degenerate critical points

There is an existence theorem for Morse function, so we don't have to concern about the degeneracy of our functions for now.

Theorem 1.2. On any manifold M, there exists a Morse function f. Moreover, Morse functions are generic; the set of Morse functions is open dense subset in $C^\infty(M, \mathbb{R})$ under appropriate topology.

Proof. See Section 6 of [Mil1], or Theorem 2.7 of [Mil2]. □

1.3 Morse Lemma and Cell Decomposition

Here's a basic lemma which describes our observation in the previous sections clearly.

Lemma 1.3 (Morse Lemma)**.** Let $f : M \to \mathbb{R}$ be Morse, and $M^c := f^{-1}(-\infty, c]$.

1. If there's no critical value between a and b, M^a and M^b are diffeomorphic.
2. Let $p \in \mathrm{Crit}(f)$, $\mathrm{Ind}(p, f) = k$. Moreover, assume that there's no critical point other than p in the set $f^{-1}[c - \varepsilon, c + \varepsilon]$. Then $M^{c+\varepsilon}$ has homotopy type of $M^{c-\varepsilon}$ with a k-cell attached.

Sketch of Proof. 1. Let $a < b$. Consider the *gradient vector field* $\mathrm{grad} f$ defined by f at the boundary of M^a, which is $f^{-1}(a)$. Such vector field is defined by an equation

$$\langle \mathrm{grad} f, Y \rangle = df(Y) = Y(f)$$

where $\langle -, - \rangle$ is some Riemannian metric on M. Since f is increasing, grad$f$ is outward at the boundary. Also, grad$f$ is nonvanishing on $f^{-1}[a, b]$, since grad$f(p) = 0$ iff $d_p f = 0$. The integral flow of grad$f$ defines desired diffeomorphism. In fact, this is a deformation retraction.

2. Let $\mathrm{Ind}(p) = k$. We can take chart near p, say $(x_1, \cdots, x_n)$, on which f is represented by

$$f(x_1, \cdots, x_n) = c - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2.$$

This means, if we consider f as a height function, there are k directions which go downward and $(n - k)$ directions which go upward near p. Figure 5 is the illustration of such picture in 2-dimensional case.

We can ignore the upward directions, since they do not touch the manifold $M^{c-\varepsilon}$ and thus can be homotoped out. So we only need to consider the downward direction, which are attached to $M^{c-\varepsilon}$, and has dimension k. Intuitively, the chart described above is a k-cell attached to $M^{c-\varepsilon}$. For detailed proof, see Theorem 3.2 of [Mil1]. □

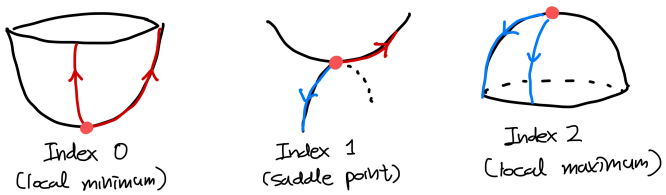

Figure 5: Critical points of index 0,1,2 in dimension 2

This is simple but powerful result, which is at the heart of Morse theory. Note that even if there are many critical points with same critical value, the cell-attaching argument of the above lemma is still valid since the process is local and our critical points are assumed to be discrete. Here's a direct consequence of above results.

Corollary 1.4. Every smooth manifold is homotopy equivalent to some CW complex.

Here's a familiar example which could be computed explicitly.

Example 1.5 (Morse Function of Complex Projective Space)**.** Let $M = \mathbb{CP}^n$, a n-dimensional complex projective space. Our convention is $\mathbb{CP}^n = S^{2n+1}/\mathbb{C}^*$ where $S^{2n+1} \subset \mathbb{C}^{n+1}$, so we can use all coordinates normalized.

$$\mathbb{CP}^n = \left\{ [z_0, \cdots, z_n] : \sum |z_i|^2 = 1 \right\}.$$

Now define Morse function f on $\mathbb{CP}^n$ by

$$f([z_0, \cdots, z_n]) = \sum i|z_i|^2.$$

This is well-defined, since the other choices of $[z_0, \cdots, z_n]$ are $[cz_0, \cdots, cz_n]$ where $c \in S^1 \subset \mathbb{C}$. Use the standard coordinate chart

$$U_j = \{[z] \in \mathbb{CP}^n : z_j \neq 0\} \simeq \mathbb{C}^n = \{(w_0, \cdots, 0, \cdots, w_n)\} \subset \mathbb{C}^{n+1},$$
$$[z_0, \cdots, z_n] \mapsto \left(\frac{z_0}{z_j}, \cdots, \frac{\hat{z}_j}{z_j}, \cdots, \frac{z_n}{z_j}\right)$$

Here $\hat{z}$ means the coordinate is zero. In this chart, we can find $|z_j|^2$ since $1 + \sum_{i \neq j} |w_i|^2 = (\sum |z_i|^2)/|z_j|^2 = 1/|z_j|^2$, i.e. $|z_j|^2 = 1/(1 + \sum_{i \neq j} |w_i|^2)$. Now our f has the form

$$\begin{aligned} f(w_0, \cdots, w_n) = \sum i|z_i|^2 &= \left(j + \sum_{i \neq j} i|w_i|^2\right) |z_j|^2 \\ &= \frac{j + \sum_{i \neq j} i|w_i|^2}{1 + \sum_{i \neq j} |w_i|^2}. \end{aligned}$$

Using $\partial/\partial w_k$, we can find that

$$\begin{aligned} \frac{\partial f}{\partial w_k} &= \frac{k\bar{w}_k(1 + \sum |w_i|^2) - (j + \sum i|w_i|^2)\bar{w}_k}{(1 + \sum |w_i|^2)^2} \\ &= \frac{\bar{w}_k((k-j) + \sum (k-i)|w_i|^2)}{(1 + \sum |w_i|^2)^2} \end{aligned}$$

and similar for $\partial f/\partial \bar{w}_k$. Now it's straightforward that the origin is unique critical point of f in U_j.

By taking Hessian of the function and put the origin in it, it turns out that every term is zero except

$$\frac{\partial^2 f}{\partial w_k \partial \bar{w}_k} = k - j.$$

It means that, the Hessian is negative definite on the subspace of $T_0 U_j$ generated by $\partial/\partial w_k, \partial/\partial \bar{w}_k$ if $k < j$. Thus we find that $\mathrm{Ind}(0, f) = 2j$ where $0 \in U_j$.

We have coordinate charts $U_0, \cdots, U_n$ for $\mathbb{CP}^n$. Denote the corresponding points of $\mathbb{CP}^n$ to the origin of U_j by $p_0, \cdots, p_n$. Such p_j's are only critical points of f, and $\mathrm{Ind}(p_j) = 2j$. Now we can see that there are one $2j$-cell for each $j = 0, \cdots, n$, which corresponds to the well-known CW structure of $\mathbb{CP}^n$; one cell in each even dimension.

1.4 Morse Inequalities

We introduce a direct consequence of our analysis on the Morse functions. First, it's a well-known fact that we can make our Morse function **self-indexing** by perturbing a bit, so that for each $p \in \mathrm{Crit}(f)$, $f(p) = \mathrm{Ind}(p, f)$. See Lemma 2.8. of [Mil2] for some discussion. If we take self-indexing Morse function, it's convenient to write out its cell structure. For example, let $k \in \mathbb{Z}_{\geq 0}$ and consider self-indexing Morse function $f : M \to \mathbb{R}$. Again, denote $M^c = f^{-1}(-\infty, c]$.

Then our Lemma 1.3. tells us that $M^{k+1/2}$ is homotopy equivalent to $M^{k-1/2}$ with k-cells attached, and the number of k-cells is equal to the number of critical points of index k. If we denote

$$C_k = C_k(f, M) := \#\{p \in \mathrm{Crit}(f) : \mathrm{Ind}(p, f) = k\},$$

then we have following.

$$H_k(M^{k+1/2}, M^{k-1/2}) \simeq H_k(\cup_{C_k} e^k, \cup_{C_k} \dot{e}^k; \mathbb{Z}) \simeq \mathbb{Z}^{\oplus C_k}.$$

The first isomorphism is induced by excision of homology.

$M^{k+1/2}$ is k-skeleton of M, and thus $H_k(M^{k+1/2}, M^{k-1/2})$ are the chain groups of cellular homology of M, and we know that it is isomorphic to the singular homology. Let $b_k(M) := \mathrm{rank} H_k(M)$ be the **k-th Betti number of M**. Then $\chi(M) := \sum(-1)^k b_k(M)$ is **Euler characteristic of M**. From the above, we have following.

Proposition 1.6 (Weak Morse Inequality)**.**

$$b_k(M) \leq C_k(f, M), \quad \sum(-1)^k C_k(f, M) = \chi(M).$$

Notice that the number C_k depends on the choice of a Morse function f. It means that the topology of M gives the lower bound of the number of critical points of f, *assumed that f is a Morse function.* (If f is not Morse, this bound might not hold.)

For stronger result, we need some algebraic notions. We call function S assigns a pair of topological spaces (X, Y) (that is, $Y \subset X$) to $\mathbb{Z}$ is **subadditive** if for $Z \subset Y \subset X$,

$$S(X, Z) \leq S(X, Y) + S(Y, Z).$$

If the equality holds, we say S is **additive**. For example, our Betti number $b_k(X, Y) = \mathrm{rank} H_k(X, Y)$ is subadditive and $\chi(X, Y) = \sum(-1)^k b_k(X, Y)$ is additive. One can easily check by induction that if a function S is subadditive and $X_0 \subset \cdots \subset X_n$, then $S(X_n, X_0) \leq \sum S(X_i, X_{i-1})$.

Lemma 1.7. Define S_k by following, then S_k is subadditive.

$$S_k(X, Y) = \sum(-1)^i b_{k-i}(X, Y) = b_k(X, Y) - b_{k-1}(X, Y) + \cdots \pm b_0(X, Y).$$

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

$$\cdots \longrightarrow H_{k+1}(X, Z) \longrightarrow H_{k+1}(X, Y) \xrightarrow{\partial} H_k(Y, Z) \longrightarrow \cdots \longrightarrow 0.$$

By exactness, we have

$$\mathrm{rank}\partial = b_k(Y, Z) - b_k(X, Z) + b_k(X, Y) - b_{k-1}(Y, Z) + \cdots \pm b_0(X, Y) \geq 0.$$

Collecting terms by pair, we have desired result. □

Now we apply this lemma to get the Morse inequality.

Theorem 1.8 (Morse Inequality)**.** For any k, we have

$$b_k(M) - b_{k-1}(M) + \cdots \pm b_0(M) \leq C_k(f) - C_{k-1}(f) + \cdots \pm C_0(f).$$

Proof. From S_k we've defined above, we have

$$S_k(M) \leq \sum_{i=1}^{k} S_k(M^{i+1/2}, M^{i-1/2}) = C_k - C_{k-1} + \cdots \pm C_0.$$

The result follows. □

From the above inequality, we can easily deduce the following.

Corollary 1.9. Let $f : M \to \mathbb{R}$ be a Morse function.

1. If $C_k = 0$, then $b_k = 0$.
2. If $C_{k+1} = C_{k-1} = 0$, then $C_k = b_k$.

Finally, we note the most trivial inequality which can be deduced directly from the fact that $\mathrm{Crit}(f)$ corresponds to the k-cells in M.

Theorem 1.10. $\mathrm{Crit}(f) \geq \sum \mathrm{rank} H_k(M)$.

This is rather easy to see, but it generalizes to the famous conjecture of Arnold.

Conjecture 1 (Arnold, now is a theorem)**.** Let M be a compact symplectic manifold and $H : M \times \mathbb{R} \to \mathbb{R}$ be a time-dependent Hamiltonian, whose period 1 orbits are all nondegenerate. Denote $\mathcal{O}(H)$ for such orbits, then

$$\#\mathcal{O}(H) \geq \sum \mathrm{rank} H_k(M).$$

Let's just ignore the complicated notions in the statement for now. The Hamiltonian orbits, which is a periodic trajectory of some vector field X_H generated by H, can be viewed as a critical point of *action functional* defined on the loop space of M. If H is C^2*-small* Hamiltonian, then we can show that there is no time-dependent periodic orbit of X_H. Time-independent orbit of X_H corresponds to the critical point of H, so it means that $\mathcal{O}(H) = \mathrm{Crit}(H)$. Then the problem degenerates into the last inequality. However, we cannot solve the problem if H is large for now, even if we understand every notion in the statement clearly. This is the starting point of *Floer theory*.

2 Morse Homology

In this section, we introduce the construction of Morse complex with $\mathbb{Z}/2\mathbb{Z}$-coefficient, and define the Morse homology. The theory might be rather easy to understand, but the underlying idea is very important and useful. The main idea is following.

1. Generate the chain groups with critical points of f.
2. Assign Morse index for each chain, so that CM_* becomes a graded module.
3. Define ∂ by counting *Morse trajectories* connecting two critical points.
4. Show that $\partial^2 = 0$, and get homology.

The Morse homology can be seen as finite dimensional prototype for many complicated homology theory, for example *Floer homology*. The main reference for this section is Section 2,3 of [AD].

2.1 Stable and Unstable Manifolds

From now, we assume that M is closed. Let $f : M \to \mathbb{R}$ be a Morse function and $\mathrm{Fl}^t_{-\mathrm{grad} f}(p) = \varphi^t(p)$ be the negative gradient flowline. (This convention of sign is traditional, I guess.) We define **stable manifolds** and **unstable manifolds** as follows. Let $p \in \mathrm{Crit}(f)$.

$$W^s(p) = \left\{ q \in M : \lim_{t\to\infty} \varphi^t(q) = p \right\},$$
$$W^u(p) = \left\{ q \in M : \lim_{t\to-\infty} \varphi^t(q) = p \right\}.$$

If $q \in W^s(p)$, then the flowline through q tends to go into p, while if $q \in W^u(p)$, the flowline tends to escape from p. This explains the names, stable and unstable. It's natural to expect that for any q, $\varphi^t(q)$ tends to a critical point, at least when M is closed. See Proposition 2.1.6 of [AD] for detailed proof. Figure 6 illustrates the stable and unstable manifolds of Morse heart.

Remark 2.1. If M is not closed, there are two cases come in mind such that $\varphi^t(q)$ does not converge. First, consider the height function on a half-infinite cylinder with end homeomorphic to a hemisphere. Then $\varphi^t(q)$ can escape to infinity. Secondly, consider the manifold with boundary, for example a torus with a disk cut out. Then $\varphi^t(q)$ cannot be extended to $t > T$, for some large T. So it's better to assume that M is closed.

In fact, these are the *cells* which we attached.

Proposition 2.2. $W^u(p)$, $W^s(p)$ are submanifolds of M diffeomorphic to open disks. Moreover, we have

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

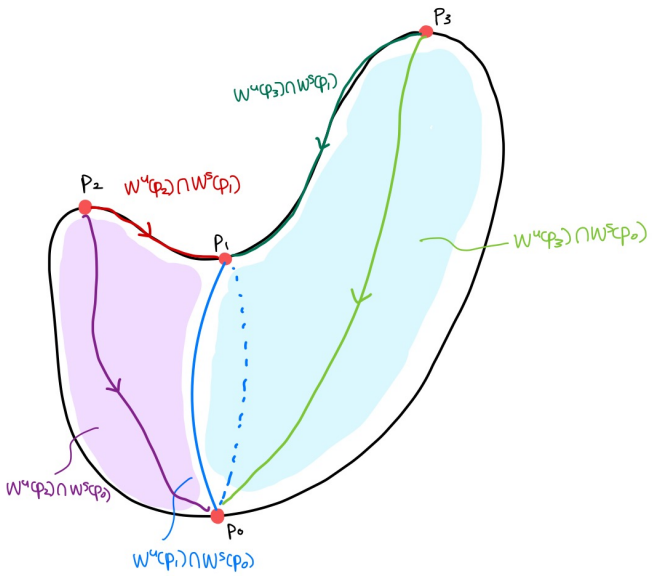

Figure 6: (Intersections of) Stable and unstable manifolds of Morse heart

Sketch of Proof. Take the chart near p which we've taken in the proof of Lemma 1.3, so that

$$f(x_1, \cdots, x_n) = c - x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2.$$

Then we can see that if $q = (x_1, \cdots, x_k, 0, \cdots, 0)$, then the *negative* gradient flowline will tend to p as $t \to -\infty$. Conversely, if $q = (0, \cdots, 0, x_{k+1}, \cdots, x_n)$, then the negative gradient flowline will tend to p as $t \to \infty$. See Proposition 2.1.5 of [AD] for details. □

2.2 Space of Trajectories $\mathcal{M}(p_+, p_-)$

We're interested in the intersections of stable and unstable manifolds. Setwise, we have

$$W^u(p_+) \cap W^s(p_-) = \left\{ p \in M : \lim_{t \to \pm\infty} \varphi^s(p) = p_\pm \right\}.$$

As in the Morse heart figure, we can see this as an image of trajectories *connecting* p_+ and p_-. Recall the properties of $\mathrm{grad} f$. That is,

1. $\mathrm{grad} f(p) = 0$ iff $p \in \mathrm{Crit}(f)$.
2. f decreases along the flowline of $-\mathrm{grad} f$, i.e. for $\mathrm{Fl}^t_{-\mathrm{grad} f} = \varphi^t$,

$$\frac{d}{dt}(f \circ \varphi^t(p)) = -||\mathrm{grad} f(\varphi^t(p))||^2 < 0.$$

Here, $\mathrm{Fl}_X^t(p)$ is the integral curve of X starting at p, i.e. it solves a differential equation

$$\frac{d\mathrm{Fl}_X^t}{dt} = X(\mathrm{Fl}_X^t(p)),\ \mathrm{Fl}_X^0(p) = p.$$

The negative gradient flowlines are enough for previous results, but they are not from now. We want the intersection $W^u(p_+) \cap W^s(p_-)$ to be a smooth manifold, and it's sufficient to assume that $W^u(p_+)$ and $W^s(p_-)$ intersect transversally. But might not be the case for the gradient flowlines.

To avoid this problem, we define **Smale condition** on a vector field. A vector field X satiesfies Smale condition for f if

- X is **pseudo-gradient**, i.e. X satisfies two properties of negative gradient vector field stated above and $X = -\mathrm{grad} f$ near the critical points.
- Define W_X^u and W_X^s by replacing φ^t to Fl_X^t. Then $W_X^u(p_+) \pitchfork W_X^s(p_-)$ for any $p_\pm \in \mathrm{Crit}(f)$.

We have an existence theorem for Smale vector field.

Theorem 2.3 (Smale)**.** For any Morse function f, there exists a Smale vector field X of f which is C^1-close to $-\mathrm{grad} f$.

Proof. See Theorem 2.2.5 of [AD]. □

By definition, we have following observation, which would be critical to define the differential of Morse complex.

Proposition 2.4. If X is Smale, $W_X^u(p_+) \cap W_X^s(p_-)$ is a smooth submanifold of M of dimension $\mathrm{Ind}(p_+) - \mathrm{Ind}(p_-)$.

Proof. If N_1, N_2 are transversely intersecting submanifolds of M, we have a standard formula

$$\mathrm{codim} N_1 \cap N_2 = \mathrm{codim} N_1 + \mathrm{codim} N_2.$$

Thus we have

$$\mathrm{codim} W_X^u(p_+) \cap W_X^s(p_-) = \mathrm{codim} W_X^u(p_+) + \mathrm{codim} W_X^s(p_-).$$

Since $\dim W_X^u(p_+) = \mathrm{Ind}(p_+, f)$ and $\mathrm{codim} W_X^s(p_-) = \mathrm{Ind}(p_-, f)$, we get the dimension formula. □

Now assume that X is Smale, and consider $W_X^u(p_+) \cap W_X^s(p_-)$. We have natural $\mathbb{R}$-action on this space, defined by

$$t \cdot q = \varphi^t(q).$$

One can easily check that this is free action (unless $p_+ = p_-$), and therefore the quotient space of intersection by $\mathbb{R}$-action is also a manifold. We define the **moduli space of trajectories** by

$$\mathcal{M}(p_+, p_-) = W_X^u(p_+) \cap W_X^s(p_-)/\mathbb{R}.$$

If $[p] = [q]$ in $\mathcal{M}(p_+, p_-)$, it means that p, q lie on the same trajectory. Thus we can understand $\mathcal{M}(p_+, p_-)$ as a space of *trajectories*, instead of a quotient of submanifold of M.

Corollary 2.5. For $p_\pm \in \mathrm{Crit}(f)$, we have

- $\dim \mathcal{M}(p_+, p_-) = \mathrm{Ind}(p_+) - \mathrm{Ind}(p_-) - 1$.
- In particular, there's no trajectory from p_+ to p_- if $\mathrm{Ind}(p_+) < \mathrm{Ind}(p_-)$. It means that the Morse index decreases along the Morse trajectory.
- If $\mathrm{Ind}(p_+) = \mathrm{Ind}(p_-)$, $\mathcal{M}(p_+, p_-) = \emptyset$ unless $p_+ = p_-$. In this case, $\mathcal{M}(p, p) = \{p\}$ where p is the constant trajectory.

Remark 2.6. In particular, we can see that the negative gradient of height function on $\mathbb{T}^2$ which we've given at the start of note is *not* Smale, since there is a flowline between different critical points of index 1!

2.3 Compactness of $\mathcal{M}(p_+, p_-)$

Since $\mathcal{M}(p_+, p_-)$ is a quotient of submanifold of M, there exists a natural topology equipped on $\mathcal{M}(p_+, p_-)$. Intuitively, this could be seen as a topology on the space of trajectories, which is a space of functions and rather hard to deal with. We'll not go into this topology deeply. Instead, we will discuss the compactification of $\mathcal{M}(p_+, p_-)$ intuitively.

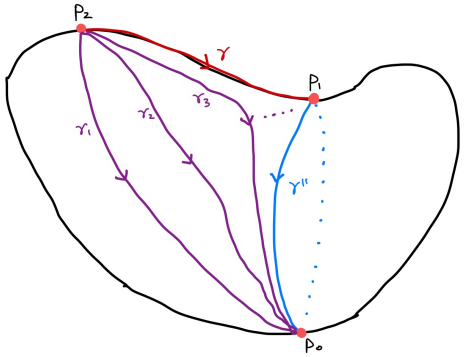

Figure 7: Sequence of Morse trajectories converging to a broken trajectory

Let's see that Figure 7. This figure describes a behavior of a sequence of trajectories (γ_n) connecting p_2 and p_0. Let's denote $\gamma = \lim_n \gamma_n$ for now. There is a sequence of real numbers, say (t_n), such that $\lim_n \gamma_n(t_n) = p_1$. So one might, and should expect that p_1 is in the image of γ. But it cannot happen; once γ touches p_1, then it would eternally stay at p_1. Thus one can see, at least visually, that the limit γ is *not* in $\mathcal{M}(p_2, p_0)$.

To solve this problem, we introduce the notion of **broken trajectory**. Let $\gamma' \in \mathcal{M}(p_2, p_1)$, $\gamma'' \in \mathcal{M}(p_1, p_0)$ be the trajectories in each space, marked in the figure. We call a pair $(\gamma', \gamma'') \in \mathcal{M}(p_2, p_1) \times \mathcal{M}(p_1, p_0)$ a broken trajectory, and consider this as a limit of sequence (γ_n).

In general, let (γ_n) be a sequence in $\mathcal{M}(p_+, p_-)$. Even if γ_n converge to the broken trajectory, we have a rule that index decreases along the Morse trajectory. Our candidate for the compactification of $\mathcal{M}(p_+, p_-)$ is following:

$$\overline{\mathcal{M}}(p, q) = \bigcup_{\mathrm{Ind}(p_i) > \mathrm{Ind}(p_{i+1})} \mathcal{M}(p_+, p_1) \times \cdots \times \mathcal{M}(p_l, p_-).$$

Here, the union goes through all pairs of critical points $(p_1, \cdots, p_l)$ (with l not fixed) such that $\mathrm{Ind}(p_+) > \mathrm{Ind}(p_1) > \cdots > \mathrm{Ind}(p_l) > \mathrm{Ind}(p_-)$. Of course, the empty pair gives original $\mathcal{M}(p_+, p_-)$. We say (γ_n) in $\mathcal{M}(p_+, p_-)$ converges to $(\gamma^{(0)}, \cdots, \gamma^{(r)})$ if there exists a real sequence $(t_n^{(i)})$ such that $\gamma_n(t + t_n^{(i)})$ converges to $\gamma^{(i)}(t)$ for each i. We have following results on the compactness of $\mathcal{M}(p_+, p_-)$.

Theorem 2.7. Let $p_+, p_- \in \mathrm{Crit}(f)$.

1. $\overline{\mathcal{M}}(p_+, p_-)$ is compact.
2. If $\mathrm{Ind}(p_+) - \mathrm{Ind}(p_-) = 1$, $\mathcal{M}(p_+, p_-)$ is compact 0-manifold, i.e. a finite set.
3. If $\mathrm{Ind}(p_+) - \mathrm{Ind}(p_-) = 2$, $\mathcal{M}(p_+, p_-)$ is compact 1-manifold with boundary, i.e. a disjoint union of circles and closed intervals. In particular, we have
$$\partial\overline{\mathcal{M}}(p_+, p_-) = \bigcup_{\mathrm{Ind}(p) = \mathrm{Ind}(p_+) - 1} \mathcal{M}(p_+, p) \times \mathcal{M}(p, p_-).$$

For the proof, see Section 3.2 of [AD].

2.4 The Morse Complex $(CM_*(f), \partial_X)$

We are now ready to define the **Morse complex**. Let $CM_*(f)$ be $\mathbb{Z}/2\mathbb{Z}$-module generated by $\mathrm{Crit}(f)$. Assign each $p \in \mathrm{Crit}(f)$ an index $\mathrm{Ind}(p)$ as degree, so that $CM_*(f)$ becomes a graded $\mathbb{Z}/2\mathbb{Z}$-module.

Note 2.8. We might assign orientation for each critical points and make CM_* into $\mathbb{Z}$-module. See Section 3.3 of [AD] for the discussion.

Let $m_X(p_+, p_-) = \#\mathcal{M}(p_+, p_-)$ if $\mathrm{Ind}(p_+) - \mathrm{Ind}(p_-) = 1$. Since $\mathcal{M}(p_+, p_-)$ is a finite set in this case, we can count the number of elements. Define differential for each $p_+ \in \mathrm{Crit}(f)$ as follows, and extend to whole module.

$$\partial_X(p_+) = \sum_{\mathrm{Ind}(p_-) = \mathrm{Ind}(p_+) - 1} m_X(p_+, p_-) p_-.$$

We're assuming that M is closed, so there exists only finitely many critical points. Hence the sum in the definition is finite, so well-defined. Also, by definition, we see that ∂_X is degree -1 map, i.e.

$$\partial_X : CM_*(f) \to CM_{*-1}(f).$$

Theorem 2.9. $\partial_X^2 = 0$, i.e. $(CM_*(f), \partial_X)$ is a chain complex.

Proof. By definition, we have

$$\begin{aligned}\partial_X^2(p) &= \partial_X \left(\sum_{\mathrm{Ind}(q)=\mathrm{Ind}(p)-1} m_X(p,q)q \right) \\ &= \sum_{\mathrm{Ind}(q)=\mathrm{Ind}(p)-1} \left(\sum_{\mathrm{Ind}(r)=\mathrm{Ind}(p)-1} m_X(p,q)m_X(q,r)r \right)\end{aligned}$$

$m_X(p,q)m_X(q,r)$ is cardinality of $\mathcal{M}(p,q) \times \mathcal{M}(q,r)$. For fixed r, from the result follows from 3. of Theorem 2.5. we have that

$$\bigcup_q \mathcal{M}(p,q) \times \mathcal{M}(q,r) = \partial \mathcal{M}(p,r).$$

Moreover, $\mathcal{M}(p,r)$ is compact 1-manifold with boundary, so its boundary is even points. It means that $\sum_q m_X(p,q)m_X(q,r) = 0$ modulo 2. Hence we get the result. □

Hence we can define **Morse homology** by

$$HM_*(M; f, X) = \ker \partial_X / \mathrm{im} \partial_X.$$

Note 2.10. In fact, $HM_*(M; f, X)$ is isomorphic to the singular homology of M. The idea comes from Section 1 of this note. A critical point of index k correspond to k-cell, and the trajectory corresponds to the attaching map of k-cell to $(k-1)$-cells. So it could be identified with cellular homology, which is isomorphic to the singular homology. See Section 4.9 of [AD] for the discussion.

Example 2.11. Let's see that Morse heart picture Figure 6 again. We have $\mathrm{Ind}(p_3) = \mathrm{Ind}(p_2) = 2$, $\mathrm{Ind}(p_1) = 1$ and $\mathrm{Ind}(p_0) = 0$. Thus $CM_2 = \mathbb{Z}/2\mathbb{Z}(p_3, p_2)$, $CM_1 = \mathbb{Z}/2\mathbb{Z}(p_1)$ and $CM_0 = \mathbb{Z}/2\mathbb{Z}(p_0)$. Let's see the flowlines between the critical points with index difference 1. We see that there is 2 from p_1 to p_0 and 1 from p_3, p_2 to p_1 each. And there is no other one. It means that, our chain complex is

$$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \xrightarrow{(-)+(-)} \mathbb{Z}/2\mathbb{Z} \xrightarrow{2(-)} \mathbb{Z}/2\mathbb{Z}$$

One can easily see that $HM_2(M) = \mathbb{Z}/2\mathbb{Z}$, $HM_1(M) = 0$, $HM_0(M) = \mathbb{Z}/2\mathbb{Z}$.

2.5 Independence of HM_*

In this subsection, we will see that HM_* is independent of the choice of a Morse function f and the choice of Smale vector field X. The idea is originally from Floer, and detailed proof is in Section 3.4 of [AD].

Let's call the pair (f, X) of a Morse function f and a Smale vector field X a *Morse pair* (which is not standard notion). Let (f_0, X_0) and (f_1, X_1) be Morse pairs.

Theorem 2.12. $HM_*(f_0, X_0)$ and $HM_*(f_1, X_1)$ are isomorphic.

For this, we have several steps.

1. Choose a function $F : M \times [0,1] \to \mathbb{R}$ such that $F(-,t) = f_0$ for $0 \leq t < \varepsilon$, and $F(-,1) = f_1$ for $1-\varepsilon < t \leq 1$. It's known that such F could be chosen as a Morse function on $M \times [0,1]$. We call this **homotopy** from f_0 to f_1.

2. Define a chain map with such F,

$$\Phi^F : (CM_*(f_0), X_0) \to (CM_*(f_1), X_1)$$

   so that Φ^F induces a map on homology level.

3. For $\tilde{f} : M \times [0,1] \to \mathbb{R}$ such that $\tilde{f}(x,t) = f(x)$ for any t, $\Phi^{\tilde{f}} = \mathrm{Id}_{CM_*(f)}$.

4. For homotopy G from f_1 to f_2 and H from f_0 to f_2, we have

$$(\Phi^G \circ \Phi^F)_* = \Phi^H_*$$

   Here, Φ^F_* is induced map on homology.

If these are shown, then the result easily follows. As F^{-1} defined by $F^{-1}(-,t) = F(-,1-t)$ is homotopy from f_1 to f_0, and $\tilde{f}$ is homotopy from f_0 to f_1, we have from 3 and 4 that

$$\Phi^{F^{-1}}_* \circ \Phi^F_* = \Phi^{\tilde{f}_0}_* = \mathrm{Id}$$

and similar result in the opposite direction. It means that Φ^F_* is isomorphism on homology, which gives us desired result.

Sketch of Proof. (See Section 3.4. of [AD].) First, we can add some function $g : \mathbb{R} \to \mathbb{R}$ which has 0 as local maximum and 1 as local minimum to the extended $F : M \times [-1,2] \to \mathbb{R}$, so that for $F = F + g$,

$$\mathrm{Crit}(F) = \mathrm{Crit}(f_0) \times \{0\} \cup \mathrm{Crit}(f_1) \times \{1\}.$$

Here, g controls the ill-bahavior of F between 0 and 1, which cannot be predicted. We build a Morse complex with $M \times [-1,2]$ and F. We can see that $\mathrm{Ind}((p,0),F) = \mathrm{Ind}(p,f_0) + 1$ and $\mathrm{Ind}((p,1),F) = \mathrm{Ind}(p,f_1)$. It means that

$$CM_{k+1}(F) = CM_k(f_0) \oplus CM_{k+1}(f_1).$$

Choose nice X on $M \times [-1,2]$ so that (F,X) becomes a Morse pair. Our Morse differential becomes

$$\partial_X : CM_k(f_0) \oplus CM_{k+1}(f_1) \to CM_{k-1}(f_0) \oplus CM_k(f_1).$$

In matrix form, we have

$$\partial_X = \begin{pmatrix} \partial_{X_0} & 0 \\ \Phi^F & \partial_{X_1} \end{pmatrix}$$

The (1,2)-term is zero because of index difference, and Φ^F is given by

$$\begin{aligned} \Phi^F : CM_k(f_0) &\to CM_k(f_1) \\ p &\mapsto \sum_{q \in CM_k(f_1)} m_X(p,q) q \end{aligned}$$

where $m_X(p,q) = \#\bar{\mathcal{M}}(p,q)$ is number of trajectories in $M \times [-1,2]$ connecting p and q. Since $\partial_X^2 = 0$, we have

$$\Phi^F \circ \partial_{X_0} + \partial_{X_1} \circ \Phi^F = 0,$$

i.e. Φ^F is a chain map. With this definition, one can easily show that $\Phi^{\tilde{f}} = \mathrm{Id}$.

Similarly, we can make $K : M \times [-1,2] \times [-1,2] \to \mathbb{R}$ which connects F, G and H, and

$$\begin{aligned} \mathrm{Crit}(K) =& \mathrm{Crit}(f_0) \times \{(0,0)\} \cup \mathrm{Crit}(f_1) \times \{(1,0)\} \\ & \cup \mathrm{Crit}(f_2) \times \{(0,1)\} \cup \mathrm{Crit}(f_2) \times \{(1,1)\} \end{aligned}$$

where the indices are given by

$$\begin{aligned} \mathrm{Ind}((p,0,0),K) &= \mathrm{Ind}(p,f_0) + 2, \\ \mathrm{Ind}((p,1,0),K) &= \mathrm{Ind}(p,f_1) + 1, \\ \mathrm{Ind}((p,0,1),K) &= \mathrm{Ind}(p,f_2) + 1, \\ \mathrm{Ind}((p,1,1),K) &= \mathrm{Ind}(p,f_2). \end{aligned}$$

It means $CM_{k+1}(K) = CM_{k-1}(f_0) \oplus CM_k(f_1) \oplus CM_k(f_2) \oplus CM_{k+1}(f_2)$, and with appropriate choice of X we have

$$\partial_X = \begin{pmatrix} \partial_{X_0} & 0 & 0 & 0 \\ \Phi^F & \partial_{X_1} & 0 & 0 \\ \Phi^H & 0 & \partial_{X_2} & 0 \\ S & \Phi^G & \mathrm{Id} & \partial_{X_2} \end{pmatrix}$$

Again, by $\partial_X^2 = 0$, we have that

$$\Phi^G \circ \Phi^F + \Phi^H + S \circ \partial_{X_0} + \partial_{X_2} \circ S = 0.$$

Thus, the map S is a chain homotopy between $\Phi^G \circ \Phi^F$ and Φ^H. □

Considering F as a Morse function on $M \times [0,1]$ is the main point of the proof. With this point of view, the Morse differential gives us chain map with desired properties. This idea apply almost similarly with some difference in technical details in the parallel statement for Floer homology.

3 The Energy Functional on Path Spaces

In this section, we introduce the infinite dimensional analogy of Morse theory. The manifold M will be replaced by its path space $\mathcal{P}M$ or loop space ΩM, and the Morse function would be given by *energy functional*. Even though there are many technical difficulties, the idea of Morse theory still can be applied. The main reference for this section is Part III of [Mil1].

Remark. Honestly, I realized that this section is nothing more than a summary of Part III of [Mil1] after the writing is done. Please understand me, since John Milnor is the best writer who I know in mathematics.

3.1 Differentiation on the Path Spaces

Let $p, q \in M$. Define the **path space** of M connecting p, q by

$$\mathcal{P}M(p,q) = \mathcal{P}M := \{\gamma : [0,1] \to M \,|\, \gamma(0) = p, \gamma(1) = q, \gamma \text{ is piecewise smooth}\}$$

The **loop space** based on p is $\Omega M(p) = \Omega M := \mathcal{P}M(p,p)$. One might mind defining such space with *piecewise* smooth curves, but we have enough reason to do so. We can give smooth structure on $\mathcal{P}M$, but we'll not go into it. Instead, we just see how to *differentiate* the functions on $\mathcal{P}M$.

First, let $f : M \to \mathbb{R}$. Recall that, on smooth manifolds, we defined *directional derivative* at $p \in M$ by taking a path $\alpha : (-\varepsilon, \varepsilon) \to M$ such that $\alpha(0) = p$, $\dot{\alpha}(0) = v$ and define

$$d_p f(v) = \left.\frac{d}{dt}\right|_0 f \circ \alpha(t).$$

Here, v is tangent vector and it could be defined as an equivalence class (germ) of paths at p.

The same thing works in infinite dimension. Let $\alpha : (-\varepsilon, \varepsilon) \to \mathcal{P}M$ be a path such that $\alpha(0) = \gamma$, $d\alpha/ds(0) = X$. Here, α could also be seen as a function

$$\begin{aligned} \alpha : (-\varepsilon, \varepsilon) \times [0,1] &\to M \\ (s,t) &\mapsto \alpha(s)(t) \end{aligned}$$

where $\alpha(0,t) = \gamma(t)$ and $\alpha(s,0) = p$, $\alpha(s,1) = q$ for any s. We call such α **variation** of γ. Our tangent vector $X = X_t$ is vector field along γ which vanishes at the endpoints, since we have that

$$\left.\frac{d}{ds}\right|_0 \alpha(s,t) \in T_{\gamma(t)}M.$$

Hence, we might define $T_\gamma \mathcal{P}M$ as a space of piecewise smooth vector fields along γ. Figure 8 explains that why this definition makes sense.

Now let $F : \mathcal{P}M \to \mathbb{R}$ be a functional. (Functional is nothing but a function on the infinite dimensional space.) Then we define the differential by

$$d_p F(X) = \left.\frac{d}{ds}\right|_0 F \circ \alpha(0,s).$$

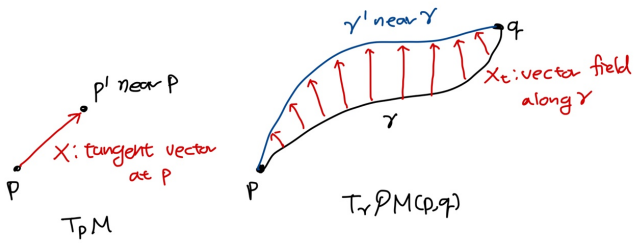

Figure 8: Analogy of tangent spaces of M and $\mathcal{P}M$

As before, we call γ is **critical path** of F if $d_pF(X) = 0$ for any $X \in T_\gamma \mathcal{P}M$. Of course, we cannot say $d_pF(X)$ is defined for every F, but the functional we will use is differentiable.

3.2 Critical Points of Energy Functional E : Geodesics

From now, we assume that M is a closed n-manifold. Equip a Riemannian metric $\langle -, - \rangle$ on M. Then we can define **energy functional** for $\gamma \in \mathcal{P}M$

$$E(\gamma) = \frac{1}{2} \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle dt.$$

We can easily check that $E : \mathcal{P}M \to \mathbb{R}$, $E(\gamma) \geq 0$ and $E(\gamma) = 0$ iff γ is a constant path. The critical points of E would be interesting for us, as we'll see. The factor $1/2$ is traditional; it came from the notion of *kinetic energy* in the physics. It also makes our formulas look better, except the next one.

Define **length functional** $L : \mathcal{P}M \to \mathbb{R}$ by

$$L(\gamma) = \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} dt.$$

By Schwarz inequality, we have

$$L(\gamma)^2 \leq \int_0^1 dt \cdot \int_0^1 \langle \dot{\gamma}, \dot{\gamma} \rangle dt = 2E(\gamma)$$

and the equality holds iff $\langle \dot{\gamma}, \dot{\gamma} \rangle$ is constant, i.e. γ has constant speed.

We say γ is a **geodesic** if it is smooth and has constant speed. We can show that if γ is a geodesic, then it's locally length-minimizing; for small $\varepsilon > 0$, we have $L(\gamma|_{[0,\varepsilon]}) = d(\gamma(0), \gamma(\varepsilon))$. If γ is globally length-minimizing, we call γ **minimizing geodesic**. Now assume that γ' is a minimizing geodesic from p to q. Then we have

$$2E(\gamma') = L(\gamma')^2 \leq L(\gamma)^2 \leq 2E(\gamma).$$

We can easily see that $L(\gamma') = L(\gamma)$ iff γ is also a (re-parametrized) minimizing geodesic and $L(\gamma)^2 = 2E(\gamma)$ iff γ has a constant speed. Thus we have following.

Lemma 3.1. The minimizing geodesic γ from p to q is a minimum point of E on $\mathcal{P}M(p,q)$. The corresponding minimum value is $d(p,q)^2/2$.

We can expect that the minimizing geodesics are actually a critical paths of E, and it is actually true. In fact, not only minimizing geodesics but also any geodesics are the critical paths of E.

Theorem 3.2 (First Variation Formula)**.** Let $X \in T_\gamma \mathcal{P}M$, and define for non-smooth point t of γ, $\Delta_t\dot{\gamma} = \dot{\gamma}(t+) - \dot{\gamma}(t-)$. Then,

$$d_\gamma E(X) = -\sum_t \langle X_t, \Delta_t\dot{\gamma}\rangle - \int_0^1 \langle X_t, \ddot{\gamma}(t)\rangle dt.$$

One can (and might have been) find similar statement in elementary differential geometry books.

Proof. Let α be a variation related to X. Then,

$$\begin{aligned}
\frac{dE(\alpha(s,t))}{ds} &= \frac{d}{ds}\left(\frac{1}{2}\int_0^1 \left\langle \frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial t}\right\rangle dt\right) \\
&= \int_0^1 \left\langle \frac{\partial}{\partial s}\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial t}\right\rangle dt \\
&= \int_0^1 \left\langle \frac{\partial}{\partial t}\frac{\partial\alpha}{\partial s}, \frac{\partial\alpha}{\partial t}\right\rangle dt \\
&= \int_0^1 \frac{\partial}{\partial t}\left\langle \frac{\partial\alpha}{\partial s}, \frac{\partial\alpha}{\partial t}\right\rangle dt - \int_0^1 \left\langle \frac{\partial\alpha}{\partial s}, \frac{\partial}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle dt \\
&= \sum_i \left\langle \frac{\partial\alpha}{\partial s}, \frac{\partial\alpha}{\partial t}\right\rangle_{t_{i+1}+}^{t_i-} - \int_0^1 \left\langle \frac{\partial\alpha}{\partial s}, \frac{\partial}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle \\
&= -\sum_i \left\langle \frac{\partial\alpha}{\partial s}, \Delta_{t_i}\frac{\partial\alpha}{\partial t}\right\rangle - \int_0^1 \left\langle \frac{\partial\alpha}{\partial u}, \frac{\partial}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle dt.
\end{aligned}$$

where $t_1, \cdots, t_k$ are the non-smooth points of γ. Put $s = 0$ to get the formula in the statement. □

This formula characterizes the critical points of E completely. Recall that γ is critical path of E iff $d_\gamma E(X) = 0$ for any X. If γ is a critical path, one must have $\Delta_t\dot{\gamma} = 0$ for any t. It means that $\dot{\gamma}$ has no discontinuity, hence γ is smooth. Also, one must have $\ddot{\gamma} = 0$, i.e. γ has constant speed. In short, we have following.

Corollary 3.3. $\mathrm{Crit}(E)$ is a set of geodesics.

Note 3.4. Such method is often called *calculus of variation*. It's very useful, for example one can deduce the celebrated *Euler-Lagrange equation*

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{x}} = \frac{\partial\mathcal{L}}{\partial x}$$

exactly in the same way.

Remark 3.5. In the proof, we didn't use the exact terminology of differentiation on manifolds. That is, we need to use covariant derivative (which is usually denoted by D/ds, etc.) to differentiate the things on manifolds. What we wrote down is rather a local form. I hope this would be enough to understand the idea.

3.3 Hessian of E : Second Variation Formula

Since we've found the critical points of E, we have to see the Hessian of E at the geodesics γ to apply our Morse theory on $(\mathcal{P}M, E)$. As before, we can define the second derivative of E on $\mathcal{P}M$ by using *two-dimensional variation*. Let $X, Y \in T_\gamma \mathcal{P}M$, i.e. vector fields along γ, and $\alpha : (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon) \times [0,1] \to \mathcal{P}M$ be a function such that

$$\alpha(0,0,t) = \gamma(t), \ \frac{\partial \alpha}{\partial s_1}(0,0,t) = X_t, \ \frac{\partial \alpha}{\partial s_2}(0,0,t) = Y_t,$$

and preseves the endpoints, i.e. $\alpha(s_1, s_2, 0) = p$, $\alpha(s_1, s_2, 1) = q$ for any $(s_1, s_2) \in (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon)$. Then we can define the **Hessian** of E at γ by a matrix

$$\mathrm{Hess}_\gamma(E)(X, Y) = \frac{\partial^2 (E \circ \alpha)}{\partial s_1 \partial s_2}(0,0).$$

As before, we also have nice formula for $\mathrm{Hess}_\gamma(E)$.

Theorem 3.6 (Second Variation Formula)**.** For $X, Y \in T_\gamma \mathcal{P}M$, we have

$$\mathrm{Hess}_\gamma(E)(X, Y) = -\sum_t \left\langle Y_t, \Delta_t \frac{dX_t}{dt} \right\rangle - \int_0^1 \left\langle Y_t, \frac{d^2 X_t}{dt^2} + R(\dot\gamma(t), X_t)\dot\gamma(t) \right\rangle dt.$$

Here, $R(-,-)$ is Riemann curvature tensor.

Proof. Let α be defined as above. Then the first variation formula gives us

$$\frac{\partial (E \circ \alpha)}{\partial s_2} = -\sum_t \left\langle \frac{\partial \alpha}{\partial s_2}, \Delta_t \frac{\partial \alpha}{\partial t} \right\rangle - \int_0^1 \left\langle \frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial t} \frac{\partial \alpha}{\partial t} \right\rangle dt.$$

Taking $\partial / \partial s_1$, we get

$$\begin{aligned} \frac{\partial^2 (E \circ \alpha)}{\partial s_1 \partial s_2} = &-\sum_t \left\langle \frac{\partial}{\partial s_1} \frac{\partial \alpha}{\partial s_2}, \Delta_t \frac{\partial \alpha}{\partial t} \right\rangle - \sum_t \left\langle \frac{\partial}{\partial s_1} \frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial s_1} \Delta_t \frac{\partial \alpha}{\partial t} \right\rangle \\ &- \int_0^1 \left\langle \frac{\partial}{\partial s_1} \frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial t} \frac{\partial \alpha}{\partial t} \right\rangle dt - \int_0^1 \left\langle \frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial s_1} \frac{\partial}{\partial t} \frac{\partial \alpha}{\partial t} \right\rangle dt. \end{aligned}$$

Let's evaluate the terms at $(s_1, s_2) = (0,0)$. We're looking at the geodesic γ, so that $\partial \alpha / \partial t$ and $\partial^2 \alpha / \partial t^2$ are zero. Hence the first term and the third term

vanishes. For the second term, we can interchange the order of partial derivative for function α so after evaluation we have

$$-\sum_t \left\langle \frac{\partial}{\partial s_1}\frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial s_1}\Delta_t \frac{\partial \alpha}{\partial t} \right\rangle \Bigg|_{(0,0)} = -\sum_t \left\langle Y_t, \Delta_t \frac{dX_t}{dt} \right\rangle.$$

For the fourth term, we first have evaluation

$$-\int_0^1 \left\langle \frac{\partial \alpha}{\partial s_2}, \frac{\partial}{\partial s_1}\frac{\partial}{\partial t}\frac{\partial \alpha}{\partial t} \right\rangle dt \Bigg|_{(0,0)} = -\int_0^1 \left\langle Y_t, \frac{\partial}{\partial s_1}\frac{\partial}{\partial t}\dot{\gamma} \right\rangle dt$$

Note that

$$\frac{\partial}{\partial s_1}\dot{\gamma} = \frac{\partial}{\partial s_1}\frac{\partial \alpha}{\partial t} = \frac{\partial}{\partial t}\frac{\partial \alpha}{\partial s_1} = \frac{\partial X_t}{\partial t}.$$

For the second derivatives of vector fields, the order of differentiation matters; we have curvature-involving formula

$$\frac{\partial}{\partial s_1}\frac{\partial}{\partial t}\dot{\gamma} - \frac{\partial}{\partial t}\frac{\partial}{\partial s_1}\dot{\gamma} = R\left(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial s_1}\right)\dot{\gamma} = R(\dot{\gamma}, X)\dot{\gamma}.$$

(See Remark 3.4. of prior subsection. We recommend to check with precise differential geometry. [S] would help a lot.) Sum up these equations, we get the desired terms. □

3.4 Kernel of $\mathrm{Hess}_\gamma(E)$: Jacobi Fields and Conjuate Points

As the first variation formula reveals the critical points of E, the second variation formula reveals the properties of Hessian of E. Note that $\mathrm{Hess}_\gamma(E)$ is symmetric, which is straightforward by definition using variation. As before, we define the **index** $\mathrm{Ind}(\gamma)$ by the maximal negative definite dimension of $\mathrm{Hess}_\gamma(E)$, and say that γ is **nondegenerate** if $\mathrm{Hess}_\gamma(E)$ is nondegenerate, or equivalently, $\ker \mathrm{Hess}_\gamma(E) = 0$.

Unlike in the Morse case, we cannot assume that E is nondegenerate; we fixed specific E. The appropriate choice of metric can be used to avoid this problem, but we should throw away the symmetries of our space. Instead, we directly analyze $\ker \mathrm{Hess}_\gamma(E)$ first since itself contains a rich geometry.

We call a vector field J along γ **Jacobi field** if

$$\frac{d^2 J}{dt^2} + R(\dot{\gamma}, J)\dot{\gamma} = 0.$$

We call this equation **Jacobi equation**. We do not assume J to be vanish at $\gamma(0)$ and $\gamma(1)$ in general, just for the convenience later. (This is not very important point.) If we assume that $J(0) = 0 = J(1)$, then $J \in T_\gamma \mathcal{P}M$.

Also, note that the Jacobi equation is second order ODE, and thus the solution J is always smooth.

Proposition 3.7. $J \in T_\gamma \mathcal{P}M$ is in $\ker \mathrm{Hess}_\gamma(E)$ iff it's a Jacobi field.

Proof. The second variation formula explains that if J is a Jacobi field, then $\mathrm{Hess}_\gamma(E)(J,Y) = 0$ for any Y. The converse is similar. Let $\mathrm{Hess}_\gamma(E)(J,Y) = 0$. Take appropriate Y to show that J satisfies Jacobi equation locally, and then take another Y to show that J cannot have discontinuity. □

As our observation on the degenerate critical points in the finite dimensional case, we might expect that if we move our critical path γ along Jacobi field J a bit, then we again have geodesics $\mathrm{Fl}_J^t(\gamma)$ for small t. For this, we define the **variation of geodesics** by a variation α such that

$$\alpha(0,t) = \gamma(t),\ \alpha(s,-) \text{ is a geodesic.}$$

Proposition 3.8. If α is variation of geodesics, $J = \partial\alpha/\partial s$ is a Jacobi field. Conversely, any Jacobi field can be obtained by such variations. If we assume that $J(0) = 0 = J(1)$, then J can be obtained by a variation α satisfying $\alpha(s,0) = p$ and $\alpha(s,1) = q$.

Proof. The proof of Lemma 3.10. of this note provides an idea. For the full proof, see Lemma 14.3. and Lemma 14.4. of [Mil1]. □

Figure 9 illustrates the Jacobi field on S^2 with standard metric. Let p, q be the north pole and south pole of S^2. Then it's elementary to see that every great half-circle is geodesic which connects p and q. For example, γ_1 and γ_2 in the picture are geodesics. Then there exists a Jacobi field J connecting two geodesics, which could be obtained by taking derivative of appropriate variation. This example could be extended to S^n for any $n \in \mathbb{Z}_{\geq 2}$.

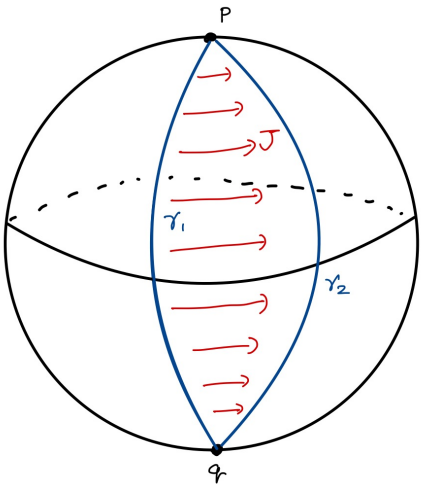

Figure 9: Jacobi field J on S^2, connecting γ_1 and γ_2

Now let's put this property of Jacobi fields to describe the dimension of $\ker \mathrm{Hess}_\gamma(E)$. As mentioned, the Jacobi equation is second order ODE and

the solution is determined by $J(0)$ and $dJ/dt(0)$. Moreover, $J \in \ker \mathrm{Hess}_\gamma(E)$ means that $J(0) = 0$. Hence the degree of freedom is at most n, and it follows that the dimension of Jacobi field is less or equal then n. In particular, it is finite dimensional.

We say $p, q \in M$ are **conjugate** along γ if there exists a nonzero Jacobi field $J \in T_\gamma \mathcal{P}M(p, q)$. Define the **multiplicity** $\nu_{p,q}$ of p, q by the dimension of such Jacobi fields, then we can see that $\nu_{p,q}$ equals to the dimension of $\ker \mathrm{Hess}_\gamma(E)$ for any geodesic γ connects p, q. For example, consider the north pole p and south pole q of S^n again. One can easily check that p, q are conjugate with $\nu_{p,q} = n - 1$.

3.5 Index of $\mathrm{Hess}_\gamma(E)$: Broken Jacobi Fields

Now we have to describe $\mathrm{Ind}(\gamma)$, which is maximal negative definite dimension of $\mathrm{Hess}_\gamma(E)$. Let's begin with basic observation.

Lemma 3.9. Let $\gamma \in \mathrm{Crit}(E)$ be *minimizing* geodesic connecting p and q. Then $\mathrm{Hess}_\gamma(E)$ is positive semi-definite, so $\mathrm{Ind}(\gamma) = 0$.

Such γ corresponds to the local minimum point in finite dimensional case. Note that we still have the dimension of kernel left i.e. it's positive *semi*-definite.

Proof. Let $\alpha : (-\varepsilon, \varepsilon) \to \mathcal{P}M$ be the variation of γ. Then as we've seen in Lemma 3.1., we have $E(\alpha(s)) \geq E(\gamma) = E(\alpha(0))$ for any s. It follows that

$$\left.\frac{d^2(E \circ \alpha)}{ds^2}\right|_0 \geq 0,$$

which is elementary result from the calculus. Thus, considering the definition of $\mathrm{Hess}_\gamma(E)(X, Y)$ with $X = Y$, we see that $\mathrm{Hess}_\gamma(E)(X, X) \geq 0$ for any X. □

At first glance, the index of general geodesic γ does not seem to be easy to describe, even to show that it is finite. But it turns out that there is a elegant result for this, thanks to Morse. We will devote the rest of this subsection on the proof of following theorem.

Theorem 3.10 (Morse)**.** For any $\gamma \in \mathrm{Crit}(E)$,

$$\mathrm{Ind}(\gamma) = \#\left\{\gamma(t) : \gamma(t) \text{ is conjugate to } \gamma(0)\right\}.$$

The conjugate points $\gamma(t)$ are counted with multiplicity. In particular, $\mathrm{Ind}(\gamma)$ is always finite.

Note that, under the completeness assumption for M, for any $p \in M$ there exists a contractible neighborhood U of p, so that any two points in U can be connected by a unique minimizing geodesic. It means that, any two *close* points has a unique minimizing geodesic between them. Moreover, we have following.

Lemma 3.11. Let U be a small neighborhood, $p, q \in U$ and $\gamma : [0, \delta] \to U$ be the unique minimizing geodesic connecting p and q. Then a Jacobi field J along γ is completely characterized by its boundary condition $J(0)$ and $J(\delta)$.

Proof. Let $J(0) \in T_{\gamma(0)}M$ and $J(\delta) \in T_{\gamma(\delta)}M$ be given by curves c_0 and c_δ, defined on some small interval $(-\varepsilon, \varepsilon)$. Define a geodesic variation $\alpha : (-\varepsilon, \varepsilon) \to M$ by $\alpha(s)$ to be the unique minimizing geodesic from $c_0(s)$ to $c_\delta(s)$. Then we can see that α gives a Jacobi field with boundary conditions $J(0)$ and $J(\delta)$.

Notice that there exists $2n$ degree of freedom to choose $J(0)$ and $J(\delta)$, which equals to the dimension of $T_{\gamma(0)}M \times T_{\gamma(\delta)}M$. However, we've seen that the space of solutions of Jacobi equation is at less or equal then $2n$, where the number $2n$ was from the choice of $J(0)$ and $dJ/dt(0)$. The previous argument shows that in this case the solution of Jacobi equation *is equal to* $2n$, and it's isomorphic to $T_{\gamma(0)}M \times T_{\gamma(\delta)}M$ under the map $J \mapsto (J(0), J(\delta))$. □

Let's fix a geodesic $\gamma \in \mathcal{P}M(p, q)$, and $0 = t_0 < t_1 < \cdots < t_k = 1$ be the partition of $[0, 1]$ so that $\gamma[t_i, t_{i+1}]$ lies in such neighborhood. Define a subspace of **broken Jacobi fields** in $T_\gamma \mathcal{P}M$,

$$T(t_0, \cdots, t_k) = T := \left\{ X \in T_\gamma \mathcal{P}M : X|_{[t_i, t_{i+1}]} \text{ is a Jacobi field along } \gamma|_{[t_i, t_{i+1}]}. \right\}.$$

(Recall that we've defined $T_\gamma \mathcal{P}M$ to be the *broken* vector fields along γ.) Since the dimension of a space of Jacobi field is always finite dimensional, $T(t_0, \cdots, t_k)$ is finite dimensional. In particular, Lemma 3.10. provides us an isomorphism

$$T(t_0, \cdots, t_k) \simeq T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_{k-1})}M.$$

Define another subspace S by following.

$$S(t_0, \cdots, t_k) = S := \{ X \in T_\gamma \mathcal{P}M : X(t_i) = 0 \text{ for any } i \}.$$

We can easily see that $T \cap S = 0$, since the Jacobi field on small interval is determined by its value on endpoints and if $J \in T \cap S$, we can and must take $J = 0$ to get the boundary condition $J(t_i) = 0$.

Lemma 3.12. We have orthogonal decomposition

$$T_\gamma \mathcal{P}M = T(t_0, \cdots, t_k) \oplus S(t_0, \cdots, t_k).$$

The orthogonality means, under bilinear form $\mathrm{Hess}_\gamma(E)$. Moreover, S is positive definite subspace for $\mathrm{Hess}_\gamma(E)$.

Proof. Let's see the decomposition first. Let $X \in T_\gamma \mathcal{P}M$, and X_T be the broken Jacobi field in T such that $X_T(t_i) = X(t_i)$ for each i. There exists unique such $X_T \in T$, and it's clear by definition that $X_S := X - X_T \in S$. As we've seen, $T \cap S = 0$. Hence this gives a direct sum decomposition.

Now let's see the orthogonality. Let $X \in T$ and $Y \in S$ and put this into the second variation formula,

$$\mathrm{Hess}_\gamma(E)(X, Y) = -\sum_t \left\langle Y_t, \Delta_t \frac{dX_t}{dt} \right\rangle - \int_0^1 \left\langle Y_t, \frac{d^2 X_t}{dt^2} + R(\dot\gamma(t), X_t)\dot\gamma(t) \right\rangle dt.$$

Since X satisfies Jacobi equation, $\ddot{X} + R(\dot{\gamma}, X)\dot{\gamma} = 0$. Hence the second term vanishes. The discontinuity of X which occurs in the first term are $t_0, \cdots, t_k$, but we've chosen $Y \in S$ so $Y(t_i) = 0$ for any i. Hence the first term also vanishes, and we see that T, S are orthogonal under $\mathrm{Hess}_\gamma(E)$.

For the positive-definiteness, let α be the variation associated to $X \in S$ Then we must have $\alpha(t_i) = \gamma(t_i)$ for each i, since $Y(t_i) = 0$. Since γ provides a minimizing geodesic in each segments $[t_i, t_{i+1}]$, as in Lemma 3.8., we can see that $\mathrm{Hess}_\gamma(E)(X, X) \geq 0$. If $X \in S$ and $\mathrm{Hess}_\gamma(E)(X, X) = 0$, one can easily check (even algebraically) X lies in the kernel of $\mathrm{Hess}_\gamma(E)$. That is, X is a Jacobi field. But it's impossible unless $X = 0$. □

Since $\mathrm{Hess}_\gamma(E)$ is positive definite on S, we might restrict $\mathrm{Hess}_\gamma(E)$ to the *finite dimensional space* $T(t_0, \cdots, t_k)$ to find its index, and even the dimension of its kernel.

Now we go into the proof of main theorem. For $0 \leq \tau \leq 1$, denote $\gamma_\tau = \gamma|_{[0,\tau]}$, $H_\tau = \mathrm{Hess}_{\gamma_\tau}(E_\tau)$ where

$$E_\tau(\gamma_\tau) = \frac{1}{2}\int_0^\tau \langle \dot{\gamma}, \dot{\gamma} \rangle dt$$

and $\mathrm{Ind}(\tau)$ to be the index of H_τ. Then $\mathrm{Ind}(1) = \mathrm{Ind}(\gamma)$, and we'll analyze the properties of $\mathrm{Ind}(\tau)$. The graph of $\mathrm{Ind}(\tau)$, which could be deduced from the following lemma, is illustrated in Figure 10.

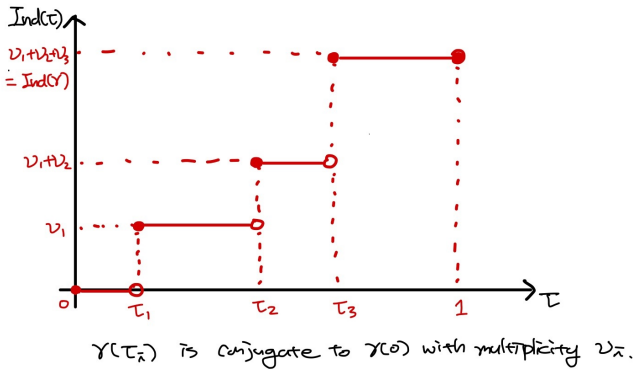

Figure 10: The graph of $\mathrm{Ind}(\tau)$ in simple case

Lemma 3.13. 1. $\mathrm{Ind}(\tau)$ is monotone increasing function on τ.

2. $\mathrm{Ind}(\tau) = 0$ for small $\tau > 0$.

3. For small $\varepsilon > 0$, $\mathrm{Ind}(\tau - \varepsilon) = \mathrm{Ind}(\tau)$.

4. Let $\nu = \dim \ker H_\tau > 0$, i.e. $\gamma(\tau)$ is conjuate point of $\gamma(0)$ of multiplicity ν. Then for small $\varepsilon > 0$, $\mathrm{Ind}(\tau + \varepsilon) = \mathrm{Ind}(\tau) + \nu$.

Proof. 1. Let $\tau < \tau'$ and $H_\tau(X_\tau, X_\tau) < 0$. Then we can extend X_τ to $[0, \tau']$ by defining $X_{\tau'}$ to be equal to X_τ on $[0, \tau]$ and equal to 0 on $[\tau, \tau']$. (Since X_τ vanishes at the endpoint, this definition works well.) Then it's clear that $X_{\tau'}$ gives negative definitive dimension for $H_{\tau'}$, i.e. $\mathrm{Ind}(\tau) \leq \mathrm{Ind}(\tau')$.

2. If τ is small enough so that $\gamma(0)$ and $\gamma(\tau)$ can be connected by minimizing geodesic, then by Lemma 3.8. $\mathrm{Ind}(\tau) = 0$.

3. For fixed τ, we can choose the partition $(t_0, \cdots, t_k)$ so that $t_i < \tau < t_{i+1}$. By Lemma 3.11., we can compute $\mathrm{Ind}(\tau)$ by resticting the domain of H_τ to the subspace
$$T_i \simeq T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_i)}M.$$
So T_i does not depend on τ (locally), while H_τ depends on τ continuously. Note that the negative definiteness is *open condition*, i.e. if $H_\tau(X, X) < 0$ for some X, then $H_{\tau'}(X, X) < 0$ for τ' close to τ, assumed that H depends continuously on τ. Thus we have $\mathrm{Ind}(\tau - \varepsilon) \geq \mathrm{Ind}(\tau)$ for small $\varepsilon > 0$. However, by monotonicity, we have $\mathrm{Ind}(\tau - \varepsilon) = \mathrm{Ind}(\tau)$.

4. Let's use that same notations as above. With assumptions in the statement, $\dim \ker H_\tau = \nu$ and the dimension of negative definite subspace is $\mathrm{Ind}(\tau)$. It means that H_τ is positive definite on the (orthogonal) complement of those spaces, which has dimension $ni - \nu - \mathrm{Ind}(\tau)$. As mentioned, positive definiteness is also an open condition. So we get one side of equality, $\mathrm{Ind}(\tau + \varepsilon) \leq \mathrm{Ind}(\tau) + \nu$.

   For the converse, we need to show that $\ker H_\tau$ can be included to the negative definite subspace of $H_{\tau'}$. Denote $\mathrm{Ind}(\tau) = \mu$ for a while, for convenience. Take the basis of negative definite subspace of H_τ, say $X_1, \cdots, X_\mu \in T_{\gamma_\tau}\mathcal{P}M$. Likewise, take the basis of kernel of H_τ, which are the Jacobi fields $J_1, \cdots, J_\nu \in T_{\gamma_\tau}\mathcal{P}M$. In particular, $dJ_i/dt(\tau) \in T_{\gamma(\tau)}M$ are linearly independent, since they should define the independent solutions for ODE. Take vector fields $Y_1, \cdots, Y_\nu \in T_{\gamma_{\tau+\varepsilon}}\mathcal{P}M$ such that
$$\left\langle \frac{dJ_i}{dt}(\tau), Y_j \right\rangle = \delta_{ij}.$$
   Extend X_i, J_i to the interval $[0, \tau + \varepsilon]$ by extension by zero. We have, from the second variation formula,
$$\begin{aligned} H_{\tau+\varepsilon}(J_i, X_j) &= 0, \\ H_{\tau+\varepsilon}(J_i, Y_j) &= \delta_{ij}. \end{aligned}$$
   Consider the vector fields along $\gamma_{\tau+\varepsilon}$,
$$X_1, \cdots, X_\mu, c^{-1}J_1 - cY_1, \cdots, c^{-1}J_\nu - cY_\nu.$$
   One can check that, on the subspace generated by these vector fields,
$$H_{\tau+\varepsilon} = \begin{pmatrix} H_\tau|_X & cA \\ cA^t & -2\mathrm{Id} + c^2 B \end{pmatrix}$$

where $H_\tau|_X$ is H_τ restricted to the subspace generated by $X_1, \cdots, X_\mu$, which is negative defnite, and A, B are some fixed matrices. For small c, we can easily deduce that this matrix is negative definite, i.e. the other side of equality, $\mathrm{Ind}(\tau + \varepsilon) \geq \mathrm{Ind}(\tau) + \nu$.

□

Combining these, the proof of Theorem 3.9. follows. We have an unexpected consequence from the theorem.

Corollary 3.14. On a geodesic $\gamma : [0, 1] \to M$, there exists only finitely many t such that $\gamma(t)$ is conjugate to $\gamma(0)$ along γ.

3.6 Cell Structure of Loop Space ΩS^n

We've seen the ways to compute the index of geodesic γ with respect to energy functional E. We expect that this result would reveal the topology of $\mathcal{P}M$. There is a theorem we wanted, which is analogous to the Morse lemma.

Theorem 3.15. Let M be complete Riemannian manifold and $p, q \in M$ be points which are not conjugate along any γ. Then $\mathcal{P}M(p, q)$ has a homotopy type of a countable CW complex, which contains one $\mathrm{Ind}(\gamma)$-cell for each geodesic γ.

The proof is analogous to the one of Morse lemma, except the point that we need finite dimensional approximation. See Section 16 and 17 of [Mil1] for the discussion.

Again, we see that the geodesics (critical points) and the topology of path spaces (manifolds) have a close relation. Even more, we might expect that the homology of path space can give lower bound of the number of geodesics.

There is a simple example to apply this theorem. Consider the based loop space of n-sphere, ΩS^n. For fixed point p say the north pole of S^n, we have a countable geodesics, say $\{\gamma_k : k = 0, 1, 2, \cdots\}$ which goes around the great circle of S^n k-times. The antipodal point q, which is the south pole, is conjugate to p with multiplicity $(n - 1)$, as we've seen in Section 3.4. In the geodesic γ_k, q appears k-times, so that we have

$$\mathrm{Ind}(\gamma_k) = k(n - 1).$$

It follows that ΩS^n has a cell structure with one cell in dimension $k(n - 1)$ for each $k \in \mathbb{Z}_{\geq 0}$.

Note that $\pi_k(\Omega S^n) \simeq \pi_{k+1}(S^n)$, which could be shown by using suspension. If $n = 1$, the above result shows that ΩS^1 is a countable discrete set. We can see this from the fact that $\pi_0(\Omega S^1) \simeq \pi_1(S^1) \simeq \mathbb{Z}$, and $\pi_k(S^1) = 0$ in any other dimensions. If $n > 2$, we directly have the homology of ΩS^n, which is given by

$$H_*(\Omega S^n; \mathbb{Z}) \simeq \begin{cases} \mathbb{Z} & \text{if } * = k(n-1) \\ 0 & \text{otherwise} \end{cases}$$

In particular, we have that $H_*(\Omega S^3) \simeq H_*(\mathbb{CP}^\infty)$. So one might ask that whether $\mathbb{CP}^\infty$ is homotopy equivalent to ΩS^3 or not. The answer is false, since $\mathbb{CP}^\infty$ is Eilenberg-MacLane space $K(\mathbb{Z}, 2)$ so that $\pi_2(\mathbb{CP}^\infty) \simeq \mathbb{Z}$ and for any other k, $\pi_k(\mathbb{CP}^\infty) = 0$. But $\pi_k(\Omega S^3) \simeq \pi_{k+1}(S^3)$ does not vanish at high dimensions.

There is also a beautiful application of this theory on the *symmetric spaces*, on which there are symmetries which preserve geodesics. For example, Lie groups are symmetric spaces. On such spaces, computing the index of geodesics becomes much easier than on the general spaces. One can prove the celebrated *Bott's periodicity theorem* with this tool. The Part IV of [Mil1] is devoted to this topic.

Even though we get a nice result for now, there is a missing point; we can count the cells, but there is no information about the attaching maps. For example, we cannot say anything about the homology of ΩS^2 (except the point that it would have rank ≤ 1 in each dimension) with previous results. The differential of Morse homology contains such information, so to get more precise information on the topology of a space, we need to construct *Morse homology in infinite dimension.*

Of course, we can use another nice function instead of E, which contains another information than E that we're interested in. For example, *Floer theory* uses a functional called *Hamiltonian action functional* $\mathcal{A}_H$ instead of E, which has Hamiltonian orbits as its critical points. Then the Floer homology is analogy of Morse homology constructed with $\mathcal{A}_H$ on the loop space of a symplectic manifold M. There is a differential given by trajectories between the orbits, which corresponds to the Morse trajectories. Our next topic will be the introduction to the Floer homology.

References

[AD] M.Audin, M.Damian: *Morse Theory and Floer Homology*, Springer (2010)

[Mil1] J.Milnor: *Morse Theory*, Princeton University Press (1963)

[Mil2] J.Milnor: *Lectures on the H-Cobordism Theorem*, Princeton Legacy Library (1965)

[Mat] Y.Mastumoto: *An Introduction to Morse Theory*, Iwanami Series in Modern Mathematics (2002)

[S] M. Spivak: *A Comprehensive Introduction to Differential Geometry, Volume II*, Second Edition, Publish or Perish (1979)