# Notes on Morse Theory

by Milnor

Math 5070

Final Project : Morse Theory

Math 5070 | Topology of Manifolds

香港中文大學

# An Introduction to Morse Theory

---

**Fan Yundi**

October 19, 2024

# Preface

This note aims to provide a preliminary understanding of what Morse theory is. I have referenced [3] and have recapitulated most of the conclusions and proofs therein.

Initially, we studied the very basic Morse theory, concerning a finite-dimensional manifold $M$. We seek a Morse function $f$ on $M$; by examining the critical points of $f$ and their indices, we can leverage the Morse lemma to transform abstract geometric problems into a more tangible "plane." Consequently, we can derive the homotopy type of $M$. The specific process is as follows: we want to know the difference in homotopy type between $M^c$ and $M^{(c-\varepsilon)}$, where $c$ denotes a non-degenerate critical value of $f$. To facilitate this, we slightly enlarge $f$ at a designated point to add a "handle." Before the proof, this note has supplemented the requisite topological knowledge, including concepts such as CW complexes.

Morse's original work was accomplished using homology, resulting in the formulation of the Morse inequalities. Thus, in this note, we delve into the basic homology theory, and then deriving the Morse inequalities using it.

Following this, this note introduces some easy concepts from Riemannian geometry. With this knowledge, we can apply Morse theory to path spaces—where we select the energy functional as our "Morse function." In a similar vein, we can define critical points and degenerate points, which aids in the development of Morse theory within the context of path spaces. This part can be applied to the study of Lie groups and symmetric spaces.

From a methodological point of view, Morse theory facilitates the study of the topological properties of manifolds through smooth functions. This phenomenon can be succinctly summarized as follows:

1. We identify some "good objects" that reflect topological properties;

2. These good objects are almost everywhere.

Clearly, Morse functions exemplify such good objects. A similar instance is that of regular values of functions, where we can define the degree of a mapping—an essential homotopy invariant. Regular values are also almost everywhere, thanks to Sard's theorem. Intriguingly, it seems that the second requirement is usually linked to Sard's theorem; for instance, Morse functions are almost everywhere in a manifold. Even when we select a function that is not Morse, we can perturb it to render it so.

In the algebraic topology section of this note, I referenced [1] and [4]. I also referenced [6] and [2] regarding the Whitney Embedding Theorem.

# Contents

# Chapter 1

# Homotopy Type in Terms of Critical Values

## 1.1 Morse's Lemma

**Definition 1.1.** Non-degenerate critical points

Given a function $f \in C^\infty(M)$, we define a point $p \in M$ as a *critical point* if $f$'s differential

$$f_* : T_pM \to T_{f(p)}\mathbb{R}$$

is zero. We call $f(p)$ a *critical value.*

Given a local coordinate system $(U, x^1, \dots, x^n)$ around $p$, $p$ is a critical point if

$$\frac{\partial f}{\partial x^1} = \frac{\partial f}{\partial x^2} = \cdots = 0.$$

We say $p$ is *non-degenerate* if the matrix

$$\left( \frac{\partial^2 f}{\partial x^i \partial x^j} \right)_{i,j}$$

is invertible. This matrix represents a quadratic form.

Intrinsically, we can define the *Hessian of* $f$, denoted $f_{**}$, as a bilinear function on $T_pM$. For $v, w \in T_pM$, extend them to smooth vector fields $\tilde{v}$ and $\tilde{w}$. We define

$$f_{**}(v, w) = \tilde{v}_p(\tilde{w}(f)).$$

**Proposition 1.1.** *$f_{**}$ is well-defined and is symmetric; the matrix of $f_{**}$ under basis $\{\partial/\partial x^i|_p\}$ is $\left( \frac{\partial^2 f}{\partial x^i \partial x^j} \right)_{i,j}$ .*

*Proof.* Note that

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

Here, $[\cdot, \cdot]$ is the Lie bracket, which equals zero because $[\tilde{v}, \tilde{w}]_p$ belongs to $T_pM$. Since $p$ is a critical point, by definition, any point derivation acting on $f$ yields zero.

It is clear that this is well-defined because $\tilde{v}_p(\tilde{w}(f)) = v_p(\tilde{w}(f))$ is independent of the extension of $v$. Due to symmetry, it is also independent of the extension of $w$.

Now compute the matrix. Suppose

$$v = \sum a_i \frac{\partial}{\partial x^i}\bigg|_p, \quad w = \sum b_i \frac{\partial}{\partial x^i}\bigg|_p.$$

We take, in the chart,

$$\tilde{w}_q = \sum b_i \frac{\partial}{\partial x^i}\bigg|_q.$$

Then

$$f_{**}(v,w) = v(\tilde{w}(f))(p) = v\left(\sum b_i \frac{\partial f}{\partial x^i}\right) = \sum_{i,j} a_i b_j \frac{\partial^2 f}{\partial x^i \partial x^j}\bigg|_p.$$

□

**Definition 1.2.** The *index* of a quadratic form $H$ on $V$ is defined as the dimension of the largest subspace $W \subset V$ such that $H$ is negative definite on $W$. Equivalently, it is the number of negative eigenvalues of $H$.

The *nullity* of $H$ is defined as the dimension of the null-space, which is the subspace consisting of all $v$ such that $H(v,w) = 0$ for all $w$.

Now we can introduce the Morse's Lemma. Before proving it, we need the following lemma:

**Lemma 1.2.** *Let $f \in C^\infty(V)$, where $V$ is a convex open subset of $\mathbb{R}^n$ containing $0$, and suppose $f(0) = 0$. Then there exist smooth functions $g_i$ defined on $V$, with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$, such that*

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

*Proof.* Note that

$$f = \int_0^1 \frac{df(tx_1, \ldots, dx_n)}{dt} dt = \int_0^1 \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(tx_1 \ldots, tx_n) dt,$$

so $g_i$ can be

$$\int_0^1 x_i \frac{\partial f}{\partial x_i}(tx_1 \ldots, tx_n) dt.$$

□

**Theorem 1.3.** *Let $p$ be a non-degenerate critical point for $f$. Then there is a local coordinate system $(y^1, \ldots, y^n)$ in a neighborhood $U$ of $p$ with $y^i(p) = 0$ for all $i$ such that the identity*

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

*holds throughout $U$, where $\lambda$ is the index of $f$ at $p$.*

*Proof.* We first show that if there is any such expression for $f$, then $\lambda$ must be the index of $f$ at $p$. For any coordinate system $(z^1, \ldots, z^n)$, if

$$f = f(p) - (z^1)^2 - (z^2)^2 - \ldots - (z^\lambda)^2 + (z^{\lambda+1})^2 + \ldots + (z^n)^2$$

then we have

$$\frac{\partial^2 f}{\partial z^i \partial z^j} = \begin{cases} -2, & \text{if } i = j \leq \lambda, \\ 2, & \text{if } i = j > \lambda, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, with respect to the basis $\{\partial/\partial z^i|_p\}$, $f_{**}$ has matrix form

$$\begin{pmatrix} -2I_\lambda & \\ & 2I_{n-\lambda} \end{pmatrix},$$

hence having index $\lambda$.

We now show such coordinate system $(y^1, \dots, y^n)$ exists. Without loss of generality, we can assume $p = 0 \in \mathbb{R}^n$ and $f(p) = 0$. Then by Lemma 1.2 there exists some $g_i$ such that $g_i(0) = \frac{\partial f}{\partial x_i}(0)$ such that

$$f(x_1, \dots, x_n) = \sum_{j=1}^{n} x_j g_j(x_1, \dots, x_n)$$

in some neighborhood of 0. Since $p = 0$ is a critical point, we have

$$g_j(0) = \frac{\partial f}{\partial x_i}(0) = 0.$$

Now applying Lemma 1.2 to $g_j$ yields there exists some $h_{ij}$ such that

$$g_j(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i h_{ij}(x_1, \dots, x_n)$$

in some smaller neighborhood of 0. It follows that

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

We can assume that $h_{ij} = h_{ji}$, since we can write $\bar{h}_{ij} = \frac{1}{2}(h_{ij} + h_{ji})$, and then have $\bar{h}_{ij} = \bar{h}_{ji}$ and $f = \sum x_i x_j \bar{h}_{ij}$. Additionally, the matrix $(\bar{h}_{ij}(0))$ is non-singular since it is equal to

$$\frac{1}{2}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right).$$

To get the desired expression of $f$, we consider the following steps. Suppose by induction that there exists coordinates $u_1, \dots, u_n$ in a neighborhood $U_1$ of 0 such that

$$f = \sum_{i=1}^{r-1} \pm(u_i)^2 + \sum_{i,j \geq r} u_i u_j H_{ij}(u_1, \dots, u_n),$$

where $(H_{ij})$ is symmetric. We can assume $H_{rr} \neq 0$, since if there is some $H_{ii} \neq 0$, we can reorder it; if there is no $H_{ii} \neq 0$, then there is some $H_{ij} = H_{ji} \neq 0$ (since $h$ is non-singular). By replacing $\tilde{u}_i = u_i + u_j$ and $\tilde{u}_j = u_i - u_j$, since $\tilde{u}_i^2 - \tilde{u}_j^2 = 4u_i u_j$, new $H_{ii}$ and $H_{jj}$ will be nonzero. Now relabeling can make it $H_{rr}$. Let $g(u_1, \dots, u_n) = \sqrt{|H_{rr}|}$ be a smooth and non-zero function in a smaller neighborhood (if needed) $U_2 \subseteq U_1$ containing 0. Now we define new coordinates $v_1, \dots, v_n$ by

$$\begin{cases} v_i = u_i, & \text{if } i \neq r \\ v_r = g(u_1, \dots, u_n) \left( u_r + \sum_{i>r} u_i \dfrac{H_{ir}}{H_{rr}} \right). \end{cases}$$

Notice that $\partial v_i / \partial u_j = \delta_{ij}$ if $i, j \neq r$, so the submatrix of Jacobian matrix of $u$ after deleting $r$-th row and column is $I$. One can verify that $\partial v_r / \partial u_r(0) \neq 0$, so the change of coordinate function from $u$ to $v$ is a local diffeomorphism, due to the inverse function theorem, hence $v$ is a new coordinate system in some smaller neighborhood $U_3 \subseteq U_2$ containing 0.

Assume now, without loss of generality, $H_{rr} > 0$. Then we have

$$\begin{aligned} f &= \sum_{i=1}^{r-1} \pm(u_i)^2 + \sum_{i,j \geq r} u_i u_j H_{ij} \\ &= \sum_{i=1}^{r-1} \pm(u_i)^2 + \left( u_r^2 H_{rr} + 2 \sum_{j>r} u_j u_r H_{rj} + \sum_{i,j>r} u_i u_j H_{ij} \right) \\ &= \sum_{i=1}^{r-1} \pm(u_i)^2 + \left( v_r^2 - \frac{1}{H_{rr}} \left( \sum_{i>r} u_i H_{ir} \right)^2 + \sum_{i,j>r} u_i u_j H_{ij} \right) \\ &= \sum_{i=1}^{r-1} \pm(v_i)^2 + (v_r)^2 + \left( -\frac{1}{H_{rr}} \left( \sum_{i>r} u_i H_{ir} \right)^2 + \sum_{i,j>r} u_i u_j H_{ij} \right) \\ &= \sum_{i=1}^{r-1} \pm(v_i)^2 + (v_r)^2 + \sum_{i,j>r} u_i u_j \tilde{H}_{ij}. \end{aligned}$$

This proves the Morse's lemma. □

**Corollary 1.4.** *Non-degenerate critical points are isolated.*

*Proof.* Let $p$ be a non-degenerate critical point of $f$. Then by Theorem 1.3, there is some chart $(V, \phi) = (V, y^1, \dots, y^n)$ centered at $p$ such that

$$f = f(p) - (y^1)^2 - (y^2)^2 - \dots - (y^\lambda)^2 + (y^{\lambda+1})^2 + \dots + (y^n)^2.$$

Clearly, $\partial f / \partial y^i = 0$ if and only if $y^i = 0$, so $p$ is the only critical point in $V$. □

**Corollary 1.5.** *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 1.4; if $M^a$ is compact then it may only contain finitely many critical points of $f$ in its interior. If we assume $a \in \mathbb{R}$ is a limit point of the set of critical values, then for all $\varepsilon > 0$, there must exist infinite distinct critical values in $(a - \varepsilon, a + \varepsilon)$, which in turn implies the existence of infinite critical points in $M^{a+\varepsilon}$, but this contradicts our result above.

□

## 1.2 1-parameter groups of diffeomorphisms

**Definition 1.3.** A 1-*parameter group of diffeomorphisms* is a smooth group homomorphism $\phi : \mathbb{R} \to \mathrm{Diff}(M)$, where $\mathrm{Diff}(M)$ is the group of diffeomorphisms on $M$.

**Definition 1.4.** A vector field $X$ is said to *generate* the 1-parameter group $\phi$ of diffeomorphisms of $M$, if for each smooth function $f$, we always have

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

**Lemma 1.6.** *A smooth vector field on $M$ which vanishes outside of a compact set $K \subset M$ generates a unique 1-parameter group of diffeomorphisms of $M$.*

*Proof.* Given any smooth curve $t \mapsto c(t) \in M$ with $c(t_0) = p$, it is convenient to define the velocity vector

$$\left.\frac{dc}{dt}\right|_{t_0} \in T_pM$$

by the identity

$$\left.\frac{dc}{dt}\right|_{t_0}(f) = \lim_{h\to 0} \frac{f(c(t_0+h)) - f(c(t_0))}{h}.$$

Now let $\phi$ be a 1-parameter group of diffeomorphisms generated by the vector field $X$. Then for each fixed $q$, the curve $t \mapsto \phi_t(q)$ satisfies the differential equation

$$\left.\frac{d\phi_t(q)}{dt}\right|_{t_0} = X_{\phi_{t_0}(q)},$$

with the initial condition $\phi_0(q) = q$. This is true since

$$\left.\frac{d\phi_t(q)}{dt}\right|_{t_0}(f) = \lim_{h\to 0} \frac{f(\phi_{t_0+h}(q)) - f(\phi_{t_0}(q))}{h} = \lim_{h\to 0} \frac{f(\phi_h(\phi_{t_0}(q))) - f(\phi_{t_0}(q))}{h} = X_{\phi_{t_0}(q)}(f).$$

Thus, for each point in $M$, there exists a $U \subseteq M$ and some $\varepsilon > 0$ such that the differential equation

$$\left.\frac{d\phi_t(q)}{dt}\right|_{t_0} = X_{\phi_{t_0}(q)}, \phi_0(q) = q$$

has unique smooth solution for $q \in U$ and $|t_0| < \varepsilon$.

The compact set $K$ can be covered by a finite number of such neighborhoods $U$. Let $\varepsilon_0 > 0$ denote the smallest of the corresponding numbers $\varepsilon$. Setting $\phi_t(q) = q$ for $q \notin K$, it follows that this differential equation has a unique solution $\phi_t(q)$ for $|t| < \varepsilon_0$ and for all $q \in M$. This solution is smooth as a function of both variables. Furthermore, it is clear that $\phi_{t+s} = \phi_t \circ \phi_s$ provided that $|t|, |s|, |t+s| < \varepsilon_0$. Therefore, each such $\phi_t$ is a diffeomorphism.

It only remains to define $\phi_t$ for $|t| \geq \varepsilon_0$. Any number $t$ can be expressed as a multiple of $\varepsilon_0/2$ plus a remainder $r$ with $|r| < \varepsilon_0/2$. If $t = k(\varepsilon_0/2) + r$ with $k \geq 0$, set

$$\phi_t = \phi_{\varepsilon_0/2} \circ \phi_{\varepsilon_0/2} \circ \cdots \circ \phi_{\varepsilon_0/2} \circ \phi_r,$$

where $\phi_{\varepsilon_0/2}$ is iterated $k$ times. If $k < 0$, it is only necessary to replace $\phi_{\varepsilon_0/2}$ by $\phi_{-\varepsilon_0/2}$ iterated $-k$ times. Thus, $\phi_t$ is defined for all values of $t$. It is not difficult to verify that $\phi_t$ is well defined, smooth, and satisfies the condition $\phi_{t+s} = \phi_t \circ \phi_s$. This completes the proof. $\square$

## 1.3 Topological Preliminaries: Cells, CW-complexes and Homotopy

**Definition 1.5.** $k$-cell is defined to be

$$e^k := \{x \in \mathbb{R}^n \mid \|x\| \leq 1\};\ e^0 = \{\text{pt}\}.$$

We also denote

$$\dot{e}^k := \partial e^k = S^{k-1};\ \dot{e}^0 = \emptyset.$$

**Definition 1.6.** Let $Y$ be a topological space. If $g : \dot{e}^k \to Y$ is continuous, we define $Y$ *with a $k$-cell attached by $g$* to be

$$Y \cup_g e^k := (Y \coprod e^k)/\sim,$$

where $\sim$ is an equivalence relationship defined in $Y \coprod e^k$: $\dot{e}^k \ni x \sim g(x) \in Y$.

**Definition 1.7.** A CW-complex is constructed by taking the union of a sequence of topological spaces

$$\emptyset = X_{-1} \subseteq X_0 \subseteq X_1 \subseteq \cdots,$$

where $X_k$ is obtained by attaching a family of $k$-cells $\{e^k_\alpha\}_\alpha$ to $X_{k-1}$ via mapping $g^k_\alpha : \dot{e}^k_\alpha \to X_{k-1}$, i.e.

$$X_k = X_{k-1} \cup_{g^k_\alpha} e^k_\alpha \cup_{g^k_\beta} e^k_\beta \cdots.$$

Each $X_k$ is called a skeleton of the complex.

The topology of CW-complex $X = \bigcup_k X_k$ is defined as follows. $U \subseteq X$ is defined to be open if $U \cap X_k$ is open in $X_k$ for each $k$.

**Theorem 1.7** (Cellular Approximation Theorem)**.** *We call a CW-complex continuous map $\phi : X \to Y$ cellular, if $\phi$ takes the $n$-skeleton of $X$ to the $n$-skeleton of $Y$ for all $n$, i.e., if*

$$\phi(X^n) \subseteq Y^n \quad \textit{for all } n.$$

*Any continuous map $f : X \to Y$ between CW-complexes $X$ and $Y$ is homotopic to a cellular map, and if $f$ is already cellular on a subcomplex $A$ of $X$, then we can furthermore choose the homotopy to be stationary on $A$.*

**Definition 1.8.** We say continuous maps $f, g : X \to Y$ are *homotopic*, if there is some continuous map $H : X \times [0,1] \to Y$, such that $H(x,0) = f(x)$ and $H(1,x) = g(x)$. We write $f \simeq g$ if they are homotopic.

**Definition 1.9.** We say two topological spaces $X, Y$ to be *homotopy equivalent*, if there is two maps $f : X \to Y$ and $g : Y \to X$ satisfying that $f \circ g$ is homotopic to $1_Y$ and $g \circ f$ is homotopic to $1_X$. We write $X \approx Y$ if they are homotopy equivalent.

**Proposition 1.8.** *If a map $F$ has a left homotopy inverse $L$ and a right homotopy inverse $R$, then $F$ is a homotopy equivalence; and $R$ (or $L$) is a 2-sided homotopy inverse.*

*Proof.* The relations

$$L \circ F \simeq \text{identity}, \quad F \circ R \simeq \text{identity}$$

imply that

$$L \simeq L(F \circ R) = (L \circ F)R \simeq R.$$

Consequently,

$$R \circ F \simeq L \circ F \simeq \text{identity},$$

which proves that $R$ is a 2-sided inverse.

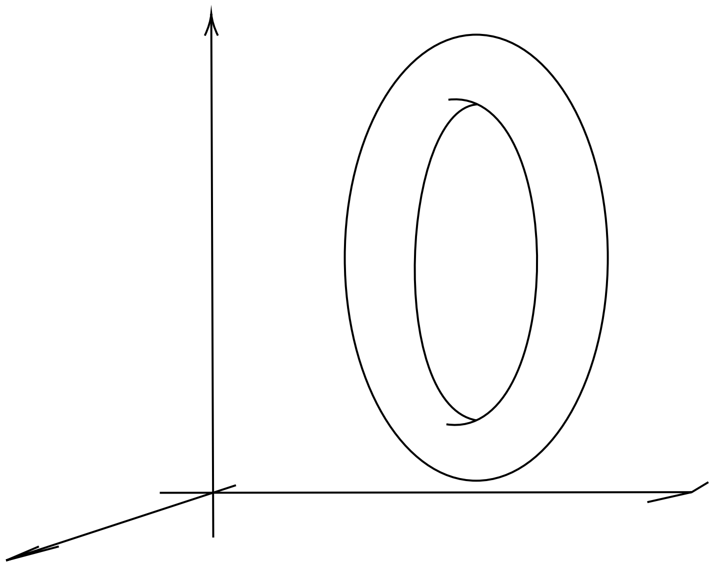
□

**Definition 1.10.** A *deformation retraction* on $X$ is a homotopy between a retraction $r : X \to A \subseteq X$ and the identity map on $X$. In this case, we call the subspace $A$ a *deformation retract* of $X$.

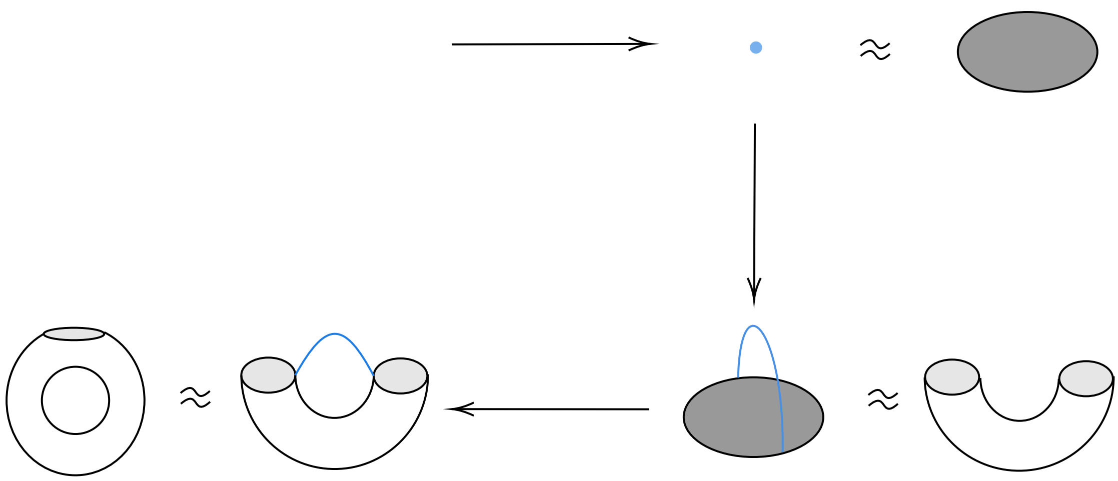
## 1.4 Main Theorems

**Definition 1.11.** We write $M^a = \{p \in M \mid f(p) \leq a\}$ .

When we think about the height function of a torus $M = S^1 \times S^1$ (check the figure below), there will be four critical points.



$M^a$ should remain its homotopy type if the range of $a$ does not contain any critical point. If the range of $a$ contains some critical point, then the homotopy type should change:



Here, the $n$-th arrow represents the change when $a$ passes through the $n$-th lowest critical point. Note that each change corresponds to attaching a 0-cell, a 1-cell, and another 1-cell, respectively. The change of homotopy type when $a$ passes through the highest critical point corresponds to attaching a 2-cell. Indeed, these facts can be generalized, which are the theorems stated below.

**Theorem 1.9.** *Let $f$ be a smooth real-valued function on a manifold $M$. Let $a < b$ and suppose that the set $f^{-1}[a, b]$ is compact and contains no critical points of $f$. Then $M^a$ is diffeomorphic to $M^b$. Furthermore, $M^a$ is a deformation retract of $M^b$, so that the inclusion map $M^a \to M^b$ is a homotopy equivalence.*

**Theorem 1.10.** *Let $f : M \to \mathbb{R}$ be a smooth function, and let $p$ be a non-degenerate critical point with index $\lambda$. Setting $f(p) = c$, suppose that $f^{-1}[c-\varepsilon, c+\varepsilon]$ is compact and contains no critical point of $f$ other than $p$, for some $\varepsilon > 0$. Then, for all sufficiently small $\varepsilon$, the set $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon}$ with a $\lambda$-cell attached.*

**Theorem 1.11.** *If $f$ is a differentiable function on a manifold $M$ with no degenerate critical points, and if each $M^a$ is compact, then $M$ has the homotopy type of a CW-complex, with one cell of dimension $\lambda$ for each critical point of index $\lambda$.*

Due to Theorem 1.11, we have the following corollary.

**Corollary 1.12** (the Weak Version of 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 homotopy equivalent to the $n$-sphere $S^n$.*

This corollary holds because $f$ has precisely one minimum $p$ and one maximum $q$ by the compactness of $M$ and hence $M$ is homotopy equivalent to $e^0 \cup e^n \approx S^n$.

Indeed, we have the following theorem, which can be regarded as a corollary of Theorem 1.9:

**Corollary 1.13** (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 $S^n$.*

*Proof.* We can normalize $f$ such that $f(p) = 0, f(q) = 1$. By Morse's lemma, there are charts $(U_p, \phi_p) = (U_p, x^1, \dots, x^n)$ and $(U_q, \phi_q) = (U_q, y^1, \dots, y^n)$ centered at $p, q$ respectively, such that

$$f = (x^1)^2 + \cdots + (x^n)^2 \text{ in } U_p,$$
$$f = 1 - (y^1)^2 - \cdots - (y^n)^2 \text{ in } U_q.$$

Thus, there is some small enough $\varepsilon > 0$ such that

$$M^\varepsilon = \phi_p^{-1}(\bar{B}_\varepsilon(0)), f^{-1}[1-\varepsilon, 1] = \phi_q^{-1}(\bar{B}_\varepsilon(0).$$

By Theorem 1.9, we can conclude that $M^{1-\varepsilon} \cong M^\varepsilon$ since there is no critical point in $f^{-1}[\varepsilon, 1-\varepsilon]$. We denote this diffeomorphism by $\psi : M^{1-\varepsilon} \to M^\varepsilon$. Recall that a sphere $S^n$ is a union of two closed disks $\bar{B}^n$ after gluing their boundaries $\partial \bar{B}^n$ via the identity map. Thus, we have

$$M^\varepsilon \cup_\rho f^{-1}[1-\varepsilon, 1] \cong S^n,$$

where $\rho = \phi_q^{-1}|_{\partial \bar{B}_\varepsilon(0)} \circ \text{id.} \circ \phi_p|_{\partial M^\varepsilon} : \partial M^\varepsilon \to \partial f^{-1}[1-\varepsilon, 1]$ is a gluing map.

Therefore, it suffices to construct a diffeomorphism between $M^\varepsilon \cup_\rho f^{-1}[1-\varepsilon, 1]$ and $M^{1-\varepsilon} \cup_{\tilde{\rho}} f^{-1}[1-\varepsilon, 1]$, where $\tilde{\rho} = \rho \circ \psi|_{\partial M^{1-\varepsilon}} : \partial M^{1-\varepsilon} \to \partial f^{-1}[1-\varepsilon, 1]$ is a gluing map. We now define

$$g : M^\varepsilon \cup_\rho f^{-1}[1-\varepsilon, 1] \to M^{1-\varepsilon} \cup_{\tilde{\rho}} f^{-1}[1-\varepsilon, 1]$$
$$\text{if } p \in M^\varepsilon,\ [p] \mapsto [\psi(p)],$$
$$\text{if } p \in f^{-1}[1-\varepsilon, 1],\ [p] \mapsto [p].$$

It is obvious that $g$ is a homeomorphism. □

We give the proof of these three main theorems to conclude this chapter.

*proof of Theorem 1.9.* The idea of the proof is to push $M^b$ to $M^a$ along the "orthogonal" trajectories of the hypersurfaces $f =$ constant. To realize it, we introduce the following definitions.

**Definition 1.12.** A Riemannian structure on a manifold $M$ is a smoothly varying $\langle \cdot, \cdot \rangle$ at $T_pM$ at each $p \in M$. The existence of a Riemannian structure is due to the existence of partition of unity.

**Definition 1.13.** The gradient of $f$ is the vector field $\text{grad}f$ on $M$ characterized by

$$\langle X, \text{grad}f \rangle_p = X_p(f).$$

An obvious property of $\text{grad}f$ is that it vanishes precisely at each critical point of $f$.

Now consider a smooth bump function $\tilde{\rho}$ satisfying $\tilde{\rho} \equiv 1$ on the compact $f^{-1}[a, b]$ and is zero outside a compact neighborhood of $f^{-1}[a, b]$ (such neighborhood exists since $M$ is locally Hausdorff). Let $\rho = \tilde{\rho} / \langle \text{grad}f, \text{grad}f \rangle$ and define the vector field $X$ on $M$ by

$$X_p = \rho(p)(\text{grad}f)_p.$$

Then, this vector field $X$ meets the requirement of Lemma 1.6, so there is some unique 1-parameter group $\phi$ of diffeomorphisms generated by it. For fixed $q \in M$ and consider the function $t \mapsto f(\phi_t(q))$. If $\phi_t(q) \in f^{-1}[a,b]$, then by the definition of tangent vector $d\phi_t(q)/dt$, we have

$$\frac{df(\phi_t(q))}{dt} = \left\langle \frac{d\phi_t(q)}{dt}, \operatorname{grad} f \right\rangle = \langle X, \operatorname{grad} f \rangle = 1.$$

Therefore, $t \mapsto f(\phi_t(q))$ is a linear function with slope 1 if $f(\phi_t(q)) \in [a,b]$. It follows that $\phi_{b-a} \in \mathrm{Diff}(M)$ carries $M^a$ to $M^b$ smoothly. Thus $M^a$ and $M^b$ are diffeomorphic.

It suffices to show that $M^a$ is a deformation retract of $M^b$. Define $r_t : M^b \to M^b$ a 1-parameter family of maps by

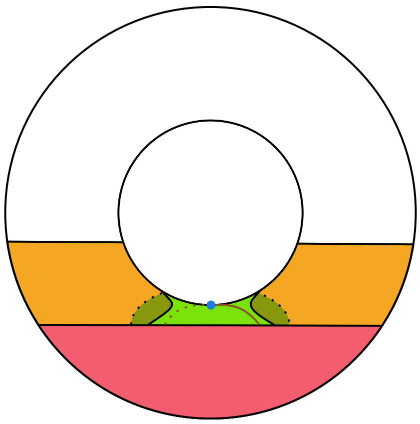
$$r_t(q) = \begin{cases} q & \text{if } f(q) \leq a \\ \phi_{t(a-f(q))}(q) & \text{if } f(q) \in [a,b] \end{cases}$$

Then $r_0 = 1_{M^b}$ and $r_1$ sends $f^{-1}[a,b]$ to $f^{-1}(a)$ and $r_1|_{M^a}$ is the identity, hence a retraction. This completes the proof.

□

**Remark.** *The compactness of $f^{-1}[a,b]$ cannot be omitted.*

*proof of Theorem 1.10.* The idea of this proof comes from the example of torus and the height function (compare the following graph).



Let the blue point be $p$, the red region be $M^{c-\varepsilon}$, and the orange region together with the green region be $f^{-1}[c-\varepsilon, c+\varepsilon]$. Consider a new smooth function $F$ on $M$ which coincides with the height function $f$ except that $F < f$ in a small neighborhood $H$ (the green region) of $p$. Thus, the region $F^{-1}(-\infty, c-\varepsilon]$ will consist of $M^{c-\varepsilon}$ together with $H$.

Choosing a suitable cell $e^\lambda \subset H$ (the brown curve), a direct argument (i.e., pushing in along the horizontal lines) will show that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\varepsilon} \cup H$. Finally, by applying Theorem 1.9 to the function $F$ and the region $F^{-1}[c-\varepsilon, c+\varepsilon]$, we will see that $M^{c-\varepsilon} \cup H$ is a deformation retract of $M^{c+\varepsilon}$. This will complete the proof of the torus case.

Now, back to the general case.

Choose a coordinate system $(U, u^1, \ldots, u^n)$ of $p$ so that the identity

$$f = c - (u^1)^2 - \cdots - (u^\lambda)^2 + (u^{\lambda+1})^2 + \cdots + (u^n)^2$$

holds throughout $U$. Thus, the critical point $p$ will have coordinates such that

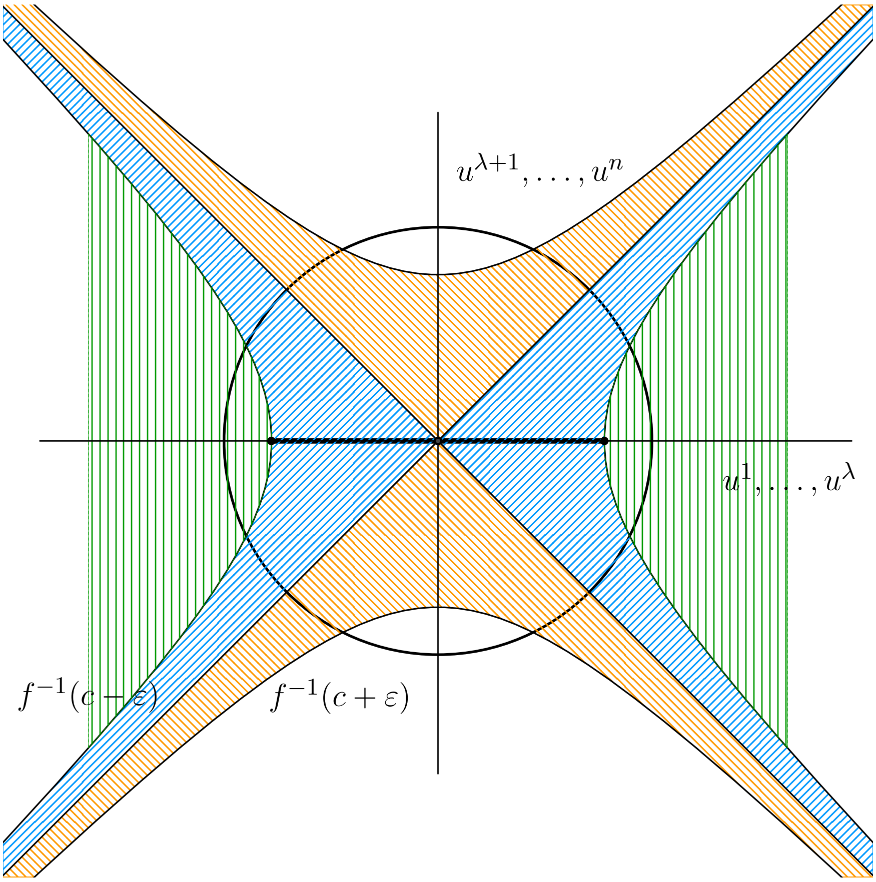
$$u^1(p) = \cdots = u^n(p) = 0.$$

Choose $\varepsilon > 0$ sufficiently small so that

1. The region $f^{-1}[c-\varepsilon, c+\varepsilon]$ is compact and contains no critical points other than $p$.
2. The image of $U$ under the diffeomorphic embedding $(u^1, \ldots, u^n) : U \to \mathbb{R}^n$ contains the closed ball
$$\{(u^1, \ldots, u^n) \mid \sum (u^i)^2 \leq 2\varepsilon\}.$$

Now define $e^\lambda$ to be the set of points in $U$ with $(u^1)^2 + \cdots + (u^\lambda)^2 \leq \varepsilon$ and $u^{\lambda+1} = \cdots = u^n = 0$.

The resulting situation is illustrated schematically in the following figure.




The coordinate lines represent the planes $u^{\lambda+1} = \cdots = u^n = 0$ and $u^1 = \cdots = u^\lambda = 0$ respectively; the circle represents the boundary of the ball with radius $\sqrt{2\varepsilon}$; and the hyperbolas represent the hypersurfaces $f^{-1}(c-\varepsilon)$ and $f^{-1}(c+\varepsilon)$. The region $M^{c-\varepsilon}$ is painted green; the region $f^{-1}[c-\varepsilon, c]$ is painted blue; and the region $f^{-1}[c, c+\varepsilon]$ is painted orange. The horizontal dark line through the origin point $p$ represents the cell $e^\lambda$.

Note that $e^\lambda \cap M^{c-\varepsilon}$ is precisely the boundary $\dot{e}^\lambda$, so that $e^\lambda$ is attached to $M^{c-\varepsilon}$ as required. We must prove that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c+\varepsilon}$.

Construct a new smooth function $F : M \to \mathbb{R}$ as follows. Let $\mu : \mathbb{R} \to \mathbb{R}$ be a $C^\infty$ function satisfying the conditions

$$\begin{aligned}&\mu(0) > \varepsilon,\\ &\mu(r) = 0 \text{ for } r \geq 2\varepsilon,\\ &-1 < \mu'(r) \leq 0 \text{ for all } r.\end{aligned}$$

Now let $F$ coincide with $f$ outside of the coordinate neighborhood $U$, and let

$$F = f - \mu\left((u^1)^2 + \cdots + (u^\lambda)^2 + 2(u^{\lambda+1})^2 + \cdots + 2(u^n)^2\right)$$

within this coordinate neighborhood. It is easily verified that $F$ is a well-defined smooth function throughout $M$ after extending by 0.

It is convenient to define two functions

$$\xi, \eta : U \to [0, \infty)$$

by

$$\xi = (u^1)^2 + \cdots + (u^\lambda)^2, \quad \eta = (u^{\lambda+1})^2 + \cdots + (u^n)^2.$$

Then $f = c - \xi + \eta$; $F(q) = c - \xi(q) + \eta(q) - \mu(\xi(q) + 2\eta(q))$ for all $q \in U$.

Note that outside of the ellipsoid $\xi + 2\eta \leq 2\varepsilon$, the functions $f$ and $F$ coincide. Within this ellipsoid we have

$$F \leq f = c - \xi + \eta \leq c + \frac{1}{2}\xi + \eta \leq c + \varepsilon.$$

Therefore, the region $F^{-1}(-\infty, c+\varepsilon]$ coincides with the region $M^{c+\varepsilon} = f^{-1}(-\infty, c+\varepsilon]$.

Note that

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

Since

$$dF = \frac{\partial F}{\partial \xi} d\xi + \frac{\partial F}{\partial \eta} d\eta,$$

where the forms $d\xi$ and $d\eta$ are simultaneously zero only at the origin, it follows that $F$ has no critical points in $U$ other than the origin. Now consider the region $F^{-1}[c-\varepsilon, c+\varepsilon]$. By the fact that the region $F^{-1}(-\infty, c+\varepsilon]$ coincides with the region $M^{c+\varepsilon} = f^{-1}(-\infty, c+\varepsilon]$, together with the inequality $F \leq f$, we see that

$$F^{-1}[c-\varepsilon, c+\varepsilon] \subset f^{-1}[c-\varepsilon, c+\varepsilon].$$

Therefore, this region is compact. It can contain no critical points of $F$ except possibly $p$. But

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

Hence, $F^{-1}[c-\varepsilon, c+\varepsilon]$ contains no critical points. Therefore, the critical points of $F$ are the same as those of $f$.

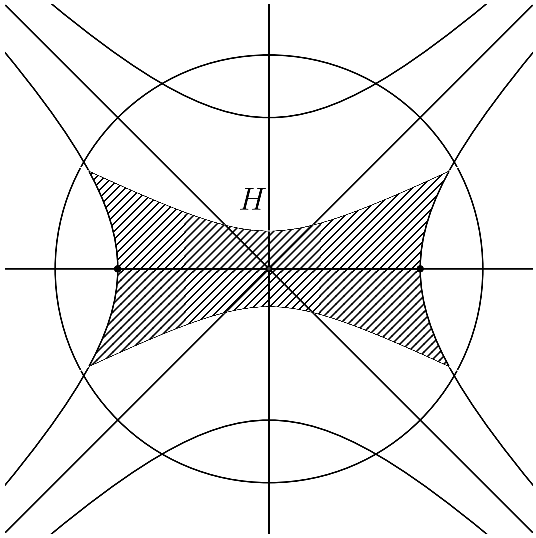
Together with Theorem 1.9, this proves that the region $F^{-1}(-\infty, c-\varepsilon]$ is a deformation retract of $M^{c+\varepsilon}$.

Now it suffices to show that $F^{-1}(-\infty, c-\varepsilon]$ is homotopy equivalent to $M^{c-\varepsilon} \cup e^\lambda$.

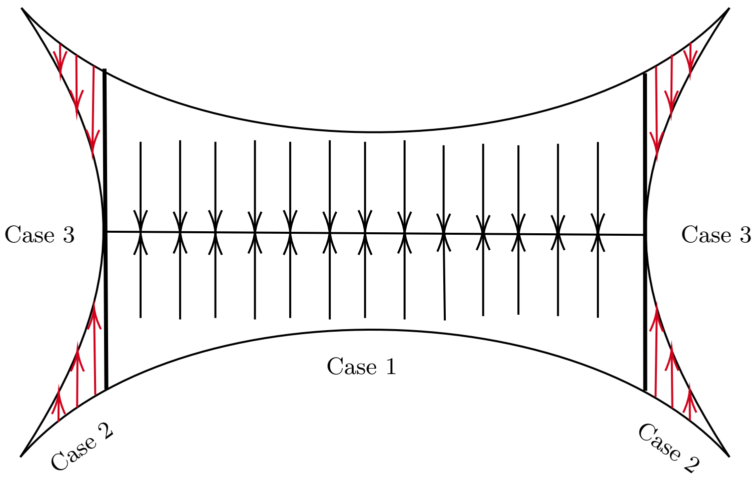
It will be convenient to denote this region $F^{-1}(-\infty, c-\varepsilon]$ by $M^{c-\varepsilon} \cup H$, where $H$ denotes the closure of $F^{-1}(-\infty, c-\varepsilon] \setminus M^{c-\varepsilon}$.

Recall that $e^\lambda = \{q \in U \mid \xi(q) \leq \varepsilon, \eta(q) = 0\}$. Note that $e^\lambda$ is contained in $H$. The reason is that since $\frac{\partial F}{\partial \xi} < 0$, we have $F(q) \leq F(p) < c - \varepsilon$ but $f(q) \geq c - \varepsilon$ for $q \in e^\lambda$.

The present situation is illustrated in the following figure. The region $H$ is shaded.




A deformation retraction $r_t : M^{c-\varepsilon} \cup H \to M^{c-\varepsilon} \cup H$ is indicated schematically by the vertical arrows in the following figure (this figure shows $H$). More precisely, let $r_t$ be the identity outside of $U$; and define $r_t$ within $U$ as follows. It is necessary to distinguish three cases.




**Case 1.** Within the region $\xi \leq \varepsilon$, let $r_t$ correspond to the transformation

$$(u^1, \ldots, u^n) \mapsto (u^1, \ldots, u^\lambda, tu^{\lambda+1}, \ldots, tu^n).$$

Thus, $r_1$ is the identity and $r_0$ maps the entire region into $e^\lambda$. The fact that each $r_t$ maps $F^{-1}(-\infty, c-\varepsilon]$ into itself follows from the inequality $\frac{\partial F}{\partial \eta} > 0$.

**Case 2.** Within the region $\varepsilon \leq \xi \leq \eta + \varepsilon$, let $r_t$ correspond to the transformation

$$(u^1, \ldots, u^n) \mapsto (u^1, \ldots, u^\lambda, s_t u^{\lambda+1}, \ldots, s_t u^n)$$

where the number $s_t \in [0,1]$ is defined by

$$s_t = t + (1-t)\left(\frac{\xi - \varepsilon}{\eta}\right)^{1/2}.$$

Thus, $r_1$ is again the identity, and $r_0$ maps the entire region into the hypersurface $f^{-1}(c-\varepsilon)$. Also note that this definition coincides with that of Case 1 when $\xi = \varepsilon$.

**Case 3.** Within the region $\eta + \varepsilon \leq \xi$ (i.e., within $M^{c-\varepsilon}$), let $r_t$ be the identity. This coincides with the preceding definition when $\xi = \eta + \varepsilon$.

This completes the proof that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $F^{-1}(-\infty, c+\varepsilon]$. □

**Remark.** *More generally, suppose that there are $k$ non-degenerate critical points $p_1, \ldots, p_k$ with indices $\lambda_1, \ldots, \lambda_k$ in $f^{-1}(c)$, then a similar proof shows that $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon} \cup e^{\lambda_1} \cup \cdots \cup e^{\lambda_k}$.*

**Remark.** *A simple modification of the proof of Theorem 1.10 shows that the set $M^c$ is also a deformation retract of $M^{c+\varepsilon}$. In fact, $M^c$ is a deformation retract of $F^{-1}(-\infty, c]$, which is a deformation retract of $M^{c+\varepsilon}$. Combining this fact with Theorem 1.10, we see easily that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^c$.*

*proof of Theorem 1.11.* The proof will be based on two lemmas concerning a topological space $X$ with a cell attached.

**Lemma 1.14** (Whitehead)**.** *Let $\phi_0$ and $\phi_1$ be homotopic maps from the sphere $\dot{e}^\lambda$ to $X$. Then the identity map of $X$ extends to a homotopy equivalence*

$$k : X \cup_{\phi_0} e^\lambda \to X \cup_{\phi_1} e^\lambda.$$

*proof.* (Refer to [5].) Let $\phi_t$ be the homotopy between $\phi_0$ and $\phi_1$. Define

$$k(x) = \begin{cases} x & \text{if } x \in X, \\ 2x & \text{if } e^\lambda \ni x = tu \text{ for some } t \in [0, 1/2], u \in \dot{e}^\lambda, \\ \phi_{2-2t}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in (1/2, 1], u \in \dot{e}^\lambda. \end{cases}$$

Notice that for all $u$ along the boundary of $e^\lambda$, $u$ is mapped to $\phi_0(u)$. From our construction $k(u) = \phi_0(u) = k(\phi_0(u))$, so $k$ is well-defined. Similarly, define $\ell : X \cup_{\phi_1} e^\lambda \to X \cup_{\phi_0} e^\lambda$ by

$$\ell(x) = \begin{cases} x & \text{if } x \in X, \\ 2x & \text{if } e^\lambda \ni x = tu \text{ for some } t \in [0, 1/2], u \in \dot{e}^\lambda, \\ \phi_{2t-1}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in (1/2, 1], u \in \dot{e}^\lambda. \end{cases}$$

We will only verify that $\ell k \simeq 1_{X \cup_{\phi_0} e^\lambda}$. First, we calculate $\ell k$ as follows.

$$\ell k(x) = \begin{cases} x & \text{if } x \in X, \\ 4x & \text{if } e^\lambda \ni x = tu \text{ for some } t \in [0, 1/4], u \in \dot{e}^\lambda, \\ \phi_{4t-1}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in (1/4, 1/2], u \in \dot{e}^\lambda, \\ \phi_{2t-1}\phi_{2-2t}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in (1/2, 1], u \in \dot{e}^\lambda. \end{cases}$$

Here we regard $\phi_{2-2t}(u) \in \dot{e}^\lambda$, hence $\phi_t$ defines a (subset of) 1-parameter group of diffeomorphism on $\dot{e}^\lambda$. Thus, $\phi_{2t-1}\phi_{2-2t} = \phi_1$. Now, A homotopy $H_s$ between $\ell k$ and $1_{X\cup_{\phi_0} e^\lambda}$ can be constructed as follows.

$$H_s(x) = \begin{cases} x & \text{if } x \in X, \\ (4-3s)x & \text{if } e^\lambda \ni x = tu \text{ for some } t \in \left[0, \dfrac{1}{4-3s}\right], u \in \dot{e}^\lambda, \\ (s(t-1)+1)\phi_{(1-s)(4t-1)}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in \left(\dfrac{1}{4-3s}, \dfrac{2-s}{4-3s}\right], u \in \dot{e}^\lambda, \\ (s(t-1)+1)\phi_{1-s}(u) & \text{if } e^\lambda \ni x = tu \text{ for some } t \in \left[\dfrac{2-s}{4-3s}, 1\right], u \in \dot{e}^\lambda. \end{cases}$$

It can be easily verified that $H_0 = \ell k$ and $H_1 = 1_{X\cup_{\phi_0} e^\lambda}$ and is $H$ is continuous. This proves the lemma.

($\Box$)

**Lemma 1.15.** *Let $\phi : \dot{e}_\lambda \to X$ be an attaching map. Any homotopy equivalence $f : X \to Y$ extends to a homotopy equivalence*

$$F : X \cup_\phi e_\lambda \to X \cup_{f\circ\phi} e_\lambda.$$

*proof.* Define $F$ by the conditions:

$$F|_X = f, \quad F|_{e^\lambda} = \text{identity}.$$

Let $g : Y \to X$ be a homotopy inverse to $f$ and define

$$G : Y \cup_{f\phi} e^\lambda \to Y \cup_{gf\phi} e^\lambda$$

by the corresponding conditions:

$$G|_Y = g, \quad G|_{e^\lambda} = \text{identity}.$$

Since $g \circ f \circ \phi$ is homotopic to $\phi$, it follows from Lemma 1.14 that there is a homotopy equivalence

$$k : X \cup_{gf\phi} e^\lambda \to X \cup_\phi e^\lambda.$$

We will first prove that the composition

$$k \circ G \circ F : X \cup_\phi e^\lambda \to X \cup_\phi e^\lambda$$

is homotopic to the identity map.

Let $h_t$ be a homotopy between $g \circ f$ and the identity. Using the specific definitions of $k$, $G$, and $F$, note that

$$\begin{cases} k \circ G \circ F(x) = gf(x) & \text{for } x \in X, \\ k \circ G \circ F(tu) = 2tu & \text{for } 0 \le t \le \dfrac{1}{2}, u \in \dot{e}^\lambda, \\ k \circ G \circ F(tu) = h_{2-2t}(u) & \text{for } \dfrac{1}{2} < t \le 1, u \in \dot{e}^\lambda. \end{cases}$$

The required homotopy

$$\begin{cases} q_\tau(x) = h_\tau(x) & \text{for } x \in X \\ q_\tau(tu) = \dfrac{2}{1+\tau}tu & \text{for } 0 \le t \le \dfrac{1+\tau}{2},\ u \in \dot{e}^\lambda \\ q_\tau(tu) = h_{2-2t+\tau}(u) & \text{for } \dfrac{1+\tau}{2} \le t \le 1,\ u \in \dot{e}^\lambda \end{cases}$$

Therefore, $F$ has a left homotopy inverse.

Similarly, we can find a homotopy equivalence $\tilde{k} : X \cup_{fg\phi} e^\lambda \to X \cup_\phi e^\lambda$ and $\tilde{k} \circ F \circ G \simeq 1_{X \cup_\phi e^\lambda}$, hence $G$ also has a left homotopy inverse. Now we use Proposition 1.8. Since $k(G \circ F) \simeq$ identity, and $k$ is known to have a left inverse, it follows that $(G \circ F)k \simeq$ identity. Since $G(F \circ k) \simeq$ identity, and $G$ is known to have a left inverse, it follows that $(F \circ k)G \simeq$ identity. Since $F(k \circ G) \simeq$ identity, and $F$ has $k \circ G$ as left inverse also, it follows that $F$ is a homotopy equivalence. This completes the proof. (□)

Now, let $c_1 < c_2 < c_3 < \cdots$ be the critical values of $f : M \to \mathbb{R}$. The sequence $\{c_i\}$ has no cluster point due to Corollary 1.5. The set $M^a$ is vacuous for $a < c_1$. Suppose $a \neq c_1, c_2, c_3, \ldots$ and that $M^a$ is of the homotopy type of a CW-complex. Let $c$ be the smallest $c_i > a$. By Theorem 1.9 and Theorem 1.10 and its corresponding remark, $M^{c+\varepsilon}$ has the homotopy type of

$$M^{c-\varepsilon} \cup_{\phi_1} e^{\lambda_1} \cup \cdots \cup_{\phi_n} e^{\lambda_n}$$

for certain $\phi_1, \ldots, \phi_n$ maps when $\varepsilon$ is small enough, and there is a homotopy equivalence $h : M^{c-\varepsilon} \to M^a$. We have assumed that there is a homotopy equivalence $h' : M^a \to K$, where $K$ is a CW-complex.

Then each $h' \circ h \circ \phi_j : \dot{e}^{\lambda_j} \to K$ is homotopic by Cellular Approximation Theorem to a map

$$\psi_j : \dot{e}^{\lambda_j} \to K^{\lambda_j - 1},$$

since $\dot{e}^{\lambda_j} = e^0 \cup e^{\lambda_j - 1}$.

Therefore,

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

is a CW-complex. Note that $h' \circ h$ is a homotopy equivalence from $M^{c-\varepsilon}$ to $K$, so by Lemma 1.15, for each $\phi_j$, we have

$$M^{c-\varepsilon} \cup_{\phi_j} e^{\lambda_j} \approx K \cup_{h' \circ h \circ \phi_j} e^{\lambda_j}.$$

Since $h' \circ h \circ \phi_j \simeq \psi_j$ and by Lemma 1.14, we conclude that

$$K \cup_{h' \circ h \circ \phi_j} e^{\lambda_j} \approx K \cup_{\psi_j} e^{\lambda_j}.$$

Therefore,

$$M^{c+\varepsilon} \approx K \cup_{\psi_1} e^{\lambda_1} \cup \cdots \cup_{\psi_n} e^{\lambda_n}.$$

has the homotopy type of a CW-complex. By induction, for each $a < c_k$ for some integer $k$, $M^a$ is homotopy equivalent to a CW-complex. Thus, if $M$ is compact, Theorem 1.11 is proved.

The proof for non-compact case seems to be related with *combinatorial homotopy*. Given that the author has not extensively explored these topics, we will skip this part. □

# Chapter 2

# The Morse Inequalities

## 2.1 Topological Preliminaries: Singular Homology

### Singular Homology on Topological Spaces

**Definition 2.1.** A *standard simplex* in $\mathbb{R}^n$ is

$$\Delta^n := \left\{ (x_i) \mid \sum_{1 \le i \le n} x_i = 1 \right\}.$$

**Definition 2.2.** A *simplex* $[p_0, \dots, p_n] \subseteq \mathbb{R}^n$ is defined as

$$\left\{ \sum t_i p_i \mid t_i \in [0,1] \right\},$$

where $\{p_0, \dots, p_n\}$ is affine independent, or equivalently $\{p_1 - p_0, \dots, p_n - p_0\}$ is linearly independent.

**Definition 2.3.** An *orientation* on a simplex $[p_0, \dots, p_n]$ is a linear order of $\{p_0, \dots, p_n\}$. Viewing it as a permutation of $\{p_0, \dots, p_n\}$, there are precisely two orientations on $[p_0, \dots, p_n]$, one with the positive sign and one with the negative sign.

**Definition 2.4.** The *i-th face map* $\varepsilon_i^n : \Delta^{n-1} \to \Delta^n$ taking $[e_0, \dots, e_{n-1}]$ to $[e_0, \dots, \hat{e}_i, \dots, e_n]$ maps $(t_0, \dots, t_{n-1})$ to $(t_0, \dots, t_{i-1}, 0, t_i, \dots, t_{n=1})$.

**Definition 2.5.** A *singular n-simplex* in a topological space $X$ is a continuous map $\sigma : \Delta^n \to X$.

**Definition 2.6.** For a topological space $X$, we define $S_n(X)$ to be the free abelian group generated by all singular $n$-simplexes in $X$ and define $S_{-1}(X) = 0$.

**Definition 2.7.** The *boundary map* $\partial_n$ is defined by linearity a homomorphism mapping $S_n$ to $S_{n-1}$, with

$$\partial_n \sigma = \sum_{i=0}^{n} (-1)^i \sigma \varepsilon_i^n.$$

**Definition 2.8.** The complex

$$\cdots \longrightarrow S_n \xrightarrow{\partial} S_{n-1} \xrightarrow{\partial} \cdots \longrightarrow S_1 \xrightarrow{\partial} 0$$

is called the *singular complex* of $X$, denoted by $S_*(X)$. It is a cochain complex in abelian group, since $\partial^2 = 0$. The homology group of this complex, $H_n(X) := B_n(X)/Z_n(X)$ where $B_n(X) = \operatorname{im}\partial_n$ is called the *n-boundaries* and $Z_n(X) = \ker\partial_n$ is called the *n-cycles.*

**Proposition 2.1.** *$H_n$ is a functor* $\texttt{Top} \to \texttt{Ab}$.

*proof.* For $f \in \operatorname{Hom}_{\texttt{Top}}(X, Y)$, define $f_\# \in \operatorname{Hom}_{\texttt{Ab}}(S_n(X), S_n(Y))$ by

$$f_\# \sum m_\sigma \sigma = \sum m_\sigma f\sigma, m_\sigma \in \mathbb{Z}.$$

Now, define $f_* \in \operatorname{Hom}_{\texttt{Ab}}(H_n(X), H_n(Y))$ by

$$f_*[z] = [f_\# z].$$

Since we can verify easily that $\partial f_\# = f_\#\partial$ and $f_\#$ preserves $B_n$ and $Z_n$, $f_*$ is well-defined. Note that $H_n(f) = f_*$. This shows the functoriality of $H_n$.

□

**Definition 2.9.** The *n-th Betti number* of $X$ is

$$R_n(X) := \operatorname{rank} H_n(X).$$

**Theorem 2.2** (Dimension Axiom)**.** *If $X$ is a one point space, then $H_n(X) = 0$ for all $n > 0$.*

**Definition 2.10.** A space $X$ is called *acyclic* if $H_n(X) = 0$ for all $n \geq 1$.

**Theorem 2.3.** *If $\{X_\lambda\}$ is the set of all path components of $X$, then*

$$H_n(X) = \bigoplus_\lambda H_n(X_\lambda).$$

**Proposition 2.4.** *If $X$ is a nonempty path connected space, then $H_0(X) = \mathbb{Z}$.*

*proof.* Consider

$$\cdots \longrightarrow S_1(X) \xrightarrow{\partial} S_0(X) \xrightarrow{\partial} 0$$

Then we have

$$Z_0(X) = \ker\partial = S_0(X),$$

and we claim

$$B_0(X) = \left\{\sum m_x x \mid x \in X, \sum m_x = 0\right\}.$$

If so, then we can define a homomorphsm

$$\theta : Z_0(X) \to \mathbb{Z}, \sum m_x x \mapsto \sum m_x.$$

By the first isomorphism theorem we have

$$Z_0(X)/B_0(X) \cong \mathbb{Z}.$$

Now we prove the claim. Let $\gamma = \sum_{i=0}^k m_i x_i \in S_0(X)$ with $\sum m_i = 0$. For each $x \in X$, there is a path connecting $x$ and $x_i$. This path is a simplex $\sigma_i : \Delta^1 = [0, 1] \to X$. Note that we have

$$\partial\sigma_i = \sigma_i(1) - \sigma_i(0) = x_i - x,$$

so

$$\partial \sum m_i\sigma_i = \sum m_i\partial\sigma_i = \sum m_i x_i - x\sum m_i = \gamma,$$

which gives that $\gamma \in B_0(X)$.

On the other hand, if $\gamma \in B_0(X)$, we have $\gamma = \partial \sum n_i\tau_i$ where $n_i \in \mathbb{Z}$ and $\tau_i$ are simplexes. Hence

$$\gamma = \sum n_i(\tau_i(1) - \tau_i(0)).$$

Note that each coefficient $n_i$ occurs twice and with opposite sign, thus the sum of the coefficients is zero.

□

**Remark.** *Every point in $X$ is in the same homology class and each generates the whole group $\mathbb{Z}$.*

**Proposition 2.5.** *Let $A$ be a subspace of $X$ with inclusion $j : A \to X$. Then $j_\# : S_n(A) \to S_n(X)$ is an injection for every $n \geq 0$.*

**Definition 2.11.** The support of $\zeta = \sum m_i\sigma_i \in S_n$ is defined by

$$\operatorname{supp}\zeta := \bigcup \sigma_i(\Delta^n).$$

**Proposition 2.6.** *If $[\zeta] \in H_n(X)$, then there exists a compact subspace $A$ of $X$ such that $[\zeta]$ is in $\operatorname{im} j_*$, where $j : A \to X$ is the inclusion.*

*proof.* Note that we can just take $A = \operatorname{supp}\zeta$. □

## The Homotopy Axiom

**Definition 2.12.** For chain maps $f, g : (C, d) \to (D, d)$, we say they are *chain homotopic* via $h : C_n \to D_{n+1}$ if

$$\pm(dh \pm hd) = f - g.$$

**Lemma 2.7.** *If $f \simeq g : C \to D$, then the induced homomorphism between homologies $f_*, g_* : H_n(C) \to H_n(D)$ are the same.*

**Lemma 2.8.** *If $X$ is contractible, then $H_n(X) \cong \delta_0^n\mathbb{Z}$.*

*proof.* Since $X$ is contractible, there is a homotopy $k : X \times [0, 1] \to X$ with $k(x, 0) = 1_X$ and $k(x, 1) = \star$ for some $\star \in X$. Consider a chain map $\varepsilon : S_*(X) \to S_*(X)$ with

$$\varepsilon(\sigma) = \delta_0^n\star$$

where $\sigma : \Delta^n \to X$ is a simplex. Note that since $\varepsilon_* : H_n(X) \to H_n(\star)$, it suffices to show $1 \simeq \varepsilon$.

Define $h(\sigma) : \Delta^{n+1} \to X$ for $\sigma : \Delta^n \to X$ by

$$\begin{aligned} (t_0, \dots, t_{n+1}) &\mapsto k\left(\sigma\left(\frac{t_1}{1-t_0}, \dots, \frac{t_{n+1}}{1-t_0}\right), t_0\right) \text{ for } 0 \leq t_0 < 1, \\ (1, 0, \dots, 0) &\mapsto \star, \end{aligned}$$

then this is a chain homotopy. □

**Lemma 2.9.** *If $\lambda_i : X \to X \times [0,1], x \mapsto (x,i)$ satisfies $(\lambda_0)_* = (\lambda_1)_* : H_n(X) \to H_n(X \times [0,1])$, then if $f \simeq g : X \to Y$, we have $f_* = g_*$.*

*proof.* Denote the homotopy between $f \simeq g$ to be $k : X \times [0,1] \to Y$. Then we have

$$f_* = (k\lambda_0)_* = k_*(\lambda_0)_* = k_*(\lambda_1)_* = g_*.$$

□

**Theorem 2.10** (The Homotopy Axiom)**.** *If $f \simeq g$, then $f_* = g_*$.*

*proof.* Due to Lemma 2.9, it suffices to show

$$(\lambda_0)_* = (\lambda_1)_*.$$

By Lemma 2.7, it suffice to show the existence of chain homotopy. We claim further that there is a natural chain homotopy $h : (\lambda_0)_\# \simeq (\lambda_1)_\#$. Here, naturality means that for a continuous map $f : X \to Y$, the following diagram commutes:

$$\begin{tikzcd} S_n(X) \arrow[r, "h"] \arrow[d, "f_\#"'] & S_{n+1}(X \times [0,1]) \arrow[d, "(f\times 1)_\#"] \\ S_n(Y) \arrow[r, "h"'] & S_{n+1}(Y \times [0,1]) \end{tikzcd}$$

We shall construct $h_n$ simultaneously for all topological spaces $X$ by induction on $n$. For $n = 0$, choose an affine map $\iota : \Delta^1 \to [0,1]$ which sends $(0,1) \mapsto 0$, $(1,0) \mapsto 1$. Then for a singular 0-simplex in $X$, which can be identified with a point $x \in X$, let

$$h(x) = \bullet \times \iota,$$

where $\bullet \in X$.

Now suppose $h_{n-1}$ is defined and natural. We will first define $h(\kappa_n)$ where $\kappa_n = \mathrm{Id} : \Delta^n \to \Delta^n$. Then for any singular $n$-simplex $\sigma : \Delta^n \to X$ for any topological space $X$, we have

$$\sigma = \sigma_\#(\kappa_n).$$

Thus, to satisfy naturality, we can (and must) put

$$h(\sigma) = \sigma_\#(h(\kappa_n)).$$

Now to construct $\nu := h(\kappa_n)$, we must solve the equation

$$\partial\nu = (\lambda_0)_\#\kappa_n - (\lambda_1)_\#\kappa_n - h(d\kappa_n).$$

Let us verify that the right-hand side is a cycle. We have

$$\begin{aligned} & \partial\left((\lambda_0)_\#\kappa_n - (\lambda_1)_\#\kappa_n - h(\partial\kappa_n)\right) \\ =& (\lambda_0)_\#\partial\kappa_n - (\lambda_1)_\#\partial\kappa_n - \partial h(\partial\kappa_n) && \text{(by } f_\#\partial = \partial f_\#) \\ =& \partial h(\partial\kappa_n) - h(\partial\partial\kappa_n) - \partial h(\partial\kappa_n) && \text{(by induction hypothesis)} \\ =& 0. \end{aligned}$$

Since $\Delta^n \times [0,1]$ is contractible, by Lemma 2.8, a cycle is a boundary if $n \geq 1$. Thus, such $\nu$ exists. The induction completes. □

## Relative Homology Groups

**Definition 2.13.** Given a chain complex $(C, d)$. We say $(D, d')$ is a *subcomplex* of $C$ if $D_n \leq C_n$ and $d' = d|_{D_*}$.

**Definition 2.14.** If $(D, d')$ is a subcomplex of $(C, d)$, then we define the *quotient complex* to be

$$(C_n/D_n, \bar{d}),\ \bar{d}(c + D_n) := dc + D_{n+1}.$$

It is well-defined since $d(D_n) \subseteq D_{n+1}$.

**Definition 2.15** (Relative Homology Group)**.** Let $A \subseteq X$. We define the *relative homology group*

$$H_n(X, A) := H_n(S_*(X)/S_*(A)).$$

**Definition 2.16.** The *n-th Betti number* of $(X, A)$ is

$$R_n(X, A) := \operatorname{rank} H_n(X, A).$$

**Theorem 2.11** (Zig-zag Lemma)**.** *It is well known that a short exact sequence of complexes* $(C_i, D)$

$$0 \longrightarrow C_1 \xrightarrow{\alpha} C_2 \xrightarrow{\beta} C_3 \longrightarrow 0$$

*gives rise to a long exact sequence*

$$\begin{tikzcd}
 & & \cdots \arrow[dll] \\
H_n(C_1) \arrow[r, "\alpha_*"] & H_n(C_2) \arrow[r, "\beta_*"] & H_n(C_3) \arrow[dll, "d"'] \\
H_{n-1}(C_1) \arrow[r, "\alpha_*"] & H_{n-1}(C_2) \arrow[r, "\beta_*"] & H_{n-1}(C_3) \arrow[dll] \\
\cdots & &
\end{tikzcd}$$

*where $\alpha D \beta d[a] = [a]$ is the connecting map, illustrating in the following diagram:*

$$\begin{tikzcd}
 & {} \arrow[d] & {} \arrow[d] & {} \arrow[d] & \\
0 \arrow[r] & (C_1)_n \arrow[r, "\alpha"] \arrow[d, "D"] & (C_2)_n \arrow[r, "\beta"] \arrow[d, "D"] & (C_3)_n \arrow[r] \arrow[d, "D"] \arrow[dll, dashed, bend left] & 0 \\
0 \arrow[r] & (C_1)_{n-1} \arrow[r, "\alpha"] \arrow[d] & (C_2)_{n-1} \arrow[r, "\beta"] \arrow[d] & (C_3)_{n-1} \arrow[r] \arrow[d] & 0 \\
 & {} & {} & {} &
\end{tikzcd}$$

$$\begin{tikzcd}
* & * & \exists b \arrow[d, mapsto] & a \text{ is a cycle} \arrow[l, mapsto, "\beta \text{ is onto}"'] \arrow[d, mapsto, dashed] & * \\
* & \exists c & Db \in \ker \beta \arrow[l, mapsto, "\text{exactness}"'] \arrow[r, mapsto, dashed] & 0 & *
\end{tikzcd}$$

**Theorem 2.12** (Naturality of the Connecting Map)**.** *The commutative diagram of complexes with exact rows*

$$\begin{tikzcd} 0 \arrow[r] & M \arrow[r, "\alpha"] \arrow[d, "f"] & N \arrow[r, "\beta"] \arrow[d, "g"] & L \arrow[r] \arrow[d, "h"] & 0 \\ 0 \arrow[r] & M' \arrow[r, "\gamma"] & N' \arrow[r, "\delta"] & L' \arrow[r] & 0 \end{tikzcd}$$

*gives rise to the following commutative diagram of complexes with exact rows*

$$\begin{tikzcd} \cdots \arrow[r] & H_n(M) \arrow[r] \arrow[d, "f_*"] & H_n(N) \arrow[r] \arrow[d, "g_*"] & H_n(L) \arrow[r, "d"] \arrow[d, "h_*"] & H_{n-1}(M) \arrow[r] \arrow[d, "f_*"] & \cdots \\ \cdots \arrow[r] & H_n(M') \arrow[r] & H_n(N') \arrow[r] & H_n(L') \arrow[r, "d"] & H_{n-1}(M') \arrow[r] & \cdots \end{tikzcd}$$

*proof.* The exactness follows from Theorem 2.11. The commutativity of the first two squares come from the functorality of $H_n$. It suffices to show the third square is commutative. We denote $L = (L, D)$ and $L' = (L', D')$. Write $[\ell] \in H_n(L)$ and suppose $[\ell] = [\beta n]$. Then we have

$$\begin{aligned} f_* d[\ell] &= f_* d[\beta n] = f_*[\alpha^{-1} Dn] \\ &= [f\alpha^{-1} Dn] = [\gamma^{-1} gDn] \\ &= [\gamma^{-1} D' gn] \\ &= d[\delta gn] = d[h\beta n] \\ &= dh_*[\ell]. \end{aligned}$$

Hence $f_* d = dh_*$.

□

**Theorem 2.13** (Relative Homotopy Version of Theorem 2.11 and Theorem 2.12)**.** *If $A \subseteq X$, then we have the following long exact sequence*

$$\cdots \longrightarrow H_n(A) \xrightarrow{i_*} H_n(X) \xrightarrow{\pi_*} H_n(X, A) \xrightarrow{d} H_{n-1}(A) \longrightarrow \cdots$$

*If $f : X \to Y$ is continuous and $f(A) \subseteq B \subseteq Y$ (if so, we write $f : (X, A) \to (Y, B)$), then we have a commutative diagram with exact rows*

$$\begin{tikzcd} \cdots \arrow[r] & H_n(A) \arrow[r] \arrow[d, "f_* i_*"] & H_n(X) \arrow[r] \arrow[d, "f_*"] & H_n(X, A) \arrow[r, "d"] \arrow[d] & H_{n-1}(A) \arrow[r] \arrow[d] & \cdots \\ \cdots \arrow[r] & H_n(B) \arrow[r] & H_n(Y) \arrow[r] & H_n(Y, B) \arrow[r, "d"] & H_{n-1}(B) \arrow[r] & \cdots \end{tikzcd}$$

*proof.* They come from the short exact sequence

$$0 \longrightarrow S_*(A) \xrightarrow{i_\#} S_*(X) \xrightarrow{\pi} S_*(X)/S_*(A) \longrightarrow 0$$

□

**Theorem 2.14.** *If we have $A' \subseteq A \subseteq X$, then there is a long exact sequence*

$$\cdots \longrightarrow H_n(A, A') \xrightarrow{i_*} H_n(X, A') \xrightarrow{\pi_*} H_n(X, A) \xrightarrow{d} H_{n-1}(A, A') \longrightarrow \cdots$$

*Moreover, a commutative diagram of pairs of spaces*

$$\begin{tikzcd}
(A, A') \arrow[r] \arrow[d] & (X, A') \arrow[r] \arrow[d, "f"] & (X, A) \arrow[d] \\
(B, B') \arrow[r] & (Y, B') \arrow[r] & (Y, B)
\end{tikzcd}$$

*give rise to a commutative diagram with exact rows*

$$\begin{tikzcd}
\cdots \arrow[r] & H_n(A, A') \arrow[r] \arrow[d] & H_n(X, A') \arrow[r] \arrow[d, "f_*"] & H_n(X, A) \arrow[r, "d"] \arrow[d] & H_{n-1}(A) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & H_n(B, B') \arrow[r] & H_n(Y, B') \arrow[r] & H_n(Y, B) \arrow[r, "d"] & H_{n-1}(B) \arrow[r] & \cdots
\end{tikzcd}$$

*proof.* They come from the third isomorphism theorem.

□

**Remark.** *Inspired by Theorem 2.14, we note that $H_n(X) = H_n(X, \emptyset)$.*

Noticing that the definition of $H_n(X, A)$ is too abstract, we try to give a more concrete definition.

**Definition 2.17.** The group of *relative $n$-cycles* mod $A$ is

$$Z_n(X, A) := \{\gamma \in S_n(X) \mid \partial\gamma \in S_{n-1}(A)\}.$$

The group of *relative $n$-boundaries* mod $A$ is

$$B_n(X, A) := \{\gamma \in S_n(X) \mid \exists \gamma' \in S_n(A) \text{ such that } \gamma - \gamma' \in B_n(X)\} = B_n(X) + S_n(A).$$

It can be easily verified that $S_n(A) \subseteq Z_n(X, A) \subseteq B_n(X, A) \subseteq S_n(X)$.

**Theorem 2.15** (Equivalent Definition of Relative Homologies)**.**

$$H_n(X, A) \cong \frac{Z_n(X, A)}{B_n(X, A)}.$$

*proof.* We prove it by just writing down the quotient complex.

$$\cdots \longrightarrow \tfrac{S_n(X)}{S_n(A)} \xrightarrow{\bar{\partial}} \tfrac{S_{n-1}(X)}{S_{n-1}(A)} \longrightarrow \cdots$$

Recall

$$\bar{\partial} : \gamma + S_n(A) \mapsto \partial\gamma + S_{n-1}(A).$$

Note that

$$\operatorname{im} \bar{\partial} = \{\gamma + S_{n-1}(A) \mid \gamma \in \operatorname{im} \partial\} = \frac{B_n(X, A)}{S_n(A)},$$
$$\ker \bar{\partial} = \{\gamma + S_n(A) \mid \partial\gamma \in S_{n-1}(A)\} = \frac{Z_{n-1}(X, A)}{S_{n-1}(A)}.$$

Then the theorem follows due to the third isomorphism theorem.

□

**The Excision Axiom**

In this part, we only state the excision axiom (there are two equivalent versions) and some of its corollaries without the proof of it.

**Theorem 2.16** (The Excision Axiom I). *Assume $U \subseteq A \subseteq X$ and $\bar{U} \subseteq A^\circ$, then the inclusion $i : (X \setminus U, A \setminus U) \to (X, A)$ induces an isomorphism*

$$i_* : H_n(X \setminus U, A \setminus U) \to H_n(X, A).$$

**Theorem 2.17** (The Excision Axiom II). *Assume $X_1, X_2 \subseteq X$ and $X = X_1^\circ \cup X_2^\circ$, then the inclusion $j : (X_1, X_1 \cap X_2) \to (X, X_2)$ induces an isomorphism*

$$j_* : H_n(X_1, X_1 \cap X_2) \to H_n(X, X_2).$$

**Lemma 2.18** (Barratt-Whitehead). *Consider a commutative diagram with exact rows*

$$\begin{tikzcd}
\cdots \arrow[r] & M_n \arrow[r, "\alpha"] \arrow[d, "f"] & N_n \arrow[r, "\beta"] \arrow[d, "g"] & L_n \arrow[r, "d"] \arrow[d, "h"] & M_{n-1} \arrow[r] \arrow[d, "f"] & \cdots \\
\cdots \arrow[r] & M'_n \arrow[r, "\gamma"] & N'_n \arrow[r, "\delta"] & L'_n \arrow[r, "d'"] & M'_{n-1} \arrow[r] & \cdots
\end{tikzcd}$$

*where $h$ is an isomorphism. It give rise to a long exact sequence*

$$\cdots \longrightarrow M_n \xrightarrow{(\alpha,f)} N_n \oplus M'_n \xrightarrow{g-\gamma} N'_n \xrightarrow{dh^{-1}\delta} M_{n-1} \longrightarrow \cdots$$

*where $(\alpha, f)(a) = (\alpha(a), f(a))$ and $(g - \gamma)(b, a') = g(b) - \gamma(a')$.*

**Theorem 2.19** (Mayer-Vietoris). *If $X_1, X_2$ are subspaces of $X$ and $X = X_1^\circ \cup X_2^\circ$, then we have the following commutative diagram with exact rows*

$$\cdots \longrightarrow H_n(X_1 \cap X_2) \xrightarrow{(i_{1*},i_{2*})} H_n(X_1) \oplus H_n(X_2) \xrightarrow{g_*-j_*} H_n(X) \xrightarrow{D} H_{n-1}(X_1 \cap X_2) \longrightarrow \cdots$$

*where $i_k : X_1 \cup X_2 \to X_k$, $g : X_1 \to X$, and $j : X_2 \to X$ are the inclusion maps, $D = dh_*^{-1}q_*$ where $h : X_1 \to X$ and $q : (X, \emptyset) \to (X, X_2)$ are inclusion, and $d : H_n(X_1, X_1 \cap X_2) \to H_{n-1}(X_1 \cap X_2)$ is the connecting homomorphism.*

*proof.* Note that we have the following commutative diagram of maps

$$\begin{tikzcd}
(X_1 \cap X_2, \emptyset) \arrow[r, "i_1"] \arrow[d, "i_2"'] & (X_1, \emptyset) \arrow[r, "p"] \arrow[d, "g"'] & (X_1, X_1 \cap X_2) \arrow[d, "h"] \\
(X_2, \emptyset) \arrow[r, "j"] & (X, \emptyset) \arrow[r, "q"] & (X, X_2)
\end{tikzcd}$$

By Theorem 2.14, we have the following commutative diagram with exact rows

$$\begin{tikzcd}
\cdots \arrow[r] & H_n(X_1 \cap X_2) \arrow[r, "i_{1*}"] \arrow[d, "i_{2*}"] & H_n(X_1) \arrow[r, "p_*"] \arrow[d, "g_*"] & H_n(X, X_1 \cap X_2) \arrow[r, "d"] \arrow[d, "h_*"] & H_{n-1}(X_1 \cap X_2) \arrow[r] \arrow[d, "i_{2*}"] & \cdots \\
\cdots \arrow[r] & H_n(X_2) \arrow[r, "j_*"] & H_n(X) \arrow[r, "q_*"] & H_n(X, X_2) \arrow[r, "d"] & H_{n-1}(X_2) \arrow[r] & \cdots
\end{tikzcd}$$

Theorem 2.17 gives that $h_*$ is an isomorphism. Thus, applying Lemma 2.18, we complete the proof.

$\square$

## Reduced Homology Groups

**Definition 2.18.** We define

$$\tilde{S}_*(X): \cdots \longrightarrow S_1(X) \xrightarrow{\partial_1} S_0(X) \xrightarrow{\tilde{\partial}_0} S_{-1}(X) \longrightarrow 0$$

to be the reduced complex of $X$, where $S_{-1}(X) = \langle [\,] \rangle \cong \mathbb{Z}$ and $\tilde{\partial}_0 : \sum m_i x_i \mapsto \sum m_i[\,]$. We define the *reduced homology groups* of $X$ to be $\tilde{H}_n(X) := H_n(\tilde{S}_*(X))$ for all $n \geq 0$.

**Theorem 2.20.** *Theorem 2.13 becomes*

$$\cdots \longrightarrow \tilde{H}_n(A) \longrightarrow \tilde{H}_n(X) \longrightarrow H_n(X,A) \longrightarrow \tilde{H}_{n-1}(A) \longrightarrow \cdots$$

*which ends*

$$\cdots \longrightarrow \tilde{H}_0(A) \longrightarrow \tilde{H}_0(X) \longrightarrow H_0(X,A) \longrightarrow 0$$

## Homology Group of Spheres

**Theorem 2.21.** *We have*

$$\tilde{H}_k(S^n) = \begin{cases} \mathbb{Z} & k = n, \\ 0 & k \neq n. \end{cases}$$

*proof.* Let $X_i = S^n \setminus \{x_i\}$ where $\{x_1, x_2\}$ are two distinct points on $S^n$, then by Theorem 2.19, we have the following long exact sequence

$$\tilde{H}_p(X_1) \oplus \tilde{H}_p(X_2) \longrightarrow \tilde{H}_p(S^n) \longrightarrow \tilde{H}_{p-1}(X_1 \cap X_2) \longrightarrow \tilde{H}_{p-1}(X_1) \oplus \tilde{H}_{p-1}(X_2)$$

Since $X_i$ are contractible, the first and the fourth term are zero. Thus

$$\tilde{H}_p(S^n) \cong \tilde{H}_{p-1}(X_1 \cap X_2) \cong \tilde{H}_{p-1}(S^{n-1})$$

Use induction, we see the theorem follows.

□

**Theorem 2.22.** *We have*

$$H_k(e^n, \dot{e}^n) = \begin{cases} \mathbb{Z} & k = n, \\ 0 & k \neq n. \end{cases}$$

*proof.* By Theorem 2.20, we have the following long exact sequence

$$\cdots \longrightarrow \tilde{H}_k(e^n) \longrightarrow H_k(e^n, \dot{e}^n) \longrightarrow \tilde{H}_{k-1}(\dot{e}^n) \longrightarrow \tilde{H}_{k-1}(e^n) \longrightarrow \cdots$$

The first and the third term is zero, so

$$H_k(e^n, \dot{e}^n) \cong \tilde{H}_{k-1}(\dot{e}^n).$$

Done.

□

## 2.2 The Morse Inequalities

**Definition 2.19.** Let $S$ be a function from certain pairs of spaces to the integers. $S$ is *subadditive* if whenever $X \supseteq Y \supseteq Z$, we have

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

If equality holds, $S$ is called *additive.*

**Definition 2.20.** The *Euler characteristic* of the pair $(X,Y)$ is defined as

$$\chi(X,Y) = \sum_{\lambda=0}^{\infty} (-1)^{\lambda} R_{\lambda}(X,Y).$$

**Proposition 2.23.** *The Betti number $R_\lambda$ is subadditive, whereas the Euler characteristic is additive. More precisely, if $Z \subseteq Y \subseteq X$, then*

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

*and*

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

*proof.* Consider, by Theorem 2.14, the long exact sequence

$$\cdots \longrightarrow H_\lambda(Y,Z) \xrightarrow{\alpha} H_\lambda(X,Z) \xrightarrow{\beta} H_\lambda(X,Y) \xrightarrow{d} H_{\lambda-1}(Y,Z) \longrightarrow \cdots$$

It gives rise to the short exact sequence

$$0 \longrightarrow \operatorname{im}\alpha \longrightarrow H_\lambda(X,Z) \longrightarrow \ker d \longrightarrow 0$$

Thus, $\operatorname{rank} H_\lambda(X,Z) = \operatorname{rank}\operatorname{im}\alpha + \operatorname{rank}\ker d \leq \operatorname{rank} H_\lambda(Y,Z) + \operatorname{rank} H_\lambda(X,Y)$.

Also note that we have

$$0 \longrightarrow \ker\beta \longrightarrow H_\lambda(X,Z) \longrightarrow \operatorname{im}\beta \longrightarrow 0$$

Thus,

$$\operatorname{rank} H_\lambda(X,Z) = \operatorname{rank}\ker\beta + \operatorname{rank}\operatorname{im}\beta,$$

so we have

$$\begin{aligned}
&\sum(-1)^\lambda R_\lambda(X,Z) - \sum(-1)^\lambda R_\lambda(Y,Z) - \sum(-1)^\lambda R_\lambda(X,Y) \\
=&\sum(-1)^\lambda \operatorname{rank}\ker\beta_\lambda + \sum(-1)^\lambda \operatorname{rank}\operatorname{im}\beta_\lambda - \\
&\sum(-1)^\lambda \operatorname{rank}\ker\alpha_\lambda - \sum(-1)^\lambda \operatorname{rank}\operatorname{im}\alpha_\lambda - \\
&\sum(-1)^\lambda \operatorname{rank}\ker d_\lambda - \sum(-1)^\lambda \operatorname{rank}\operatorname{im} d_\lambda \\
=&\ 0,
\end{aligned}$$

due to the exactness. $\square$

**Lemma 2.24.** *Let $S$ be subadditive and let $X_0 \subset \cdots \subset X_n$. Then*

$$S(X_n, X_0) \leq \sum_{i=1}^{n} S(X_i, X_{i-1}).$$

*If $S$ is additive, then equality holds.*

*proof.* Induction on $n$. For $n = 1$, equality holds trivially, and the case $n = 2$ is the definition of (sub)additivity.

Assume the result is true for $n - 1$. Then

$$S(X_{n-1}, X_0) \leq \sum_{i=1}^{n-1} S(X_i, X_{i-1}).$$

Thus, we have

$$S(X_n, X_0) \leq S(X_{n-1}, X_0) + S(X_n, X_{n-1}) \leq \sum_{i=1}^{n} S(X_i, X_{i-1}),$$

and the result is true for $n$. □

**Remark.** *Let $S(X, \emptyset) = S(X)$. Taking $X_0 = \emptyset$ in Lemma 2.24, we have*

$$S(X_n) \leq \sum_{i=1}^{n} S(X_i, X_{i-1}),$$

*with equality if $S$ is additive.*

**Theorem 2.25** (Morse's Weak Inequalities)**.** *Let $M$ be a compact manifold and $f$ a smooth function on $M$ with isolated, non-degenerate, critical points. If $C_\lambda$ denotes the number of critical points of $f$ with index $\lambda$, then*

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

*and*

$$\chi(M) = \sum (-1)^\lambda C_\lambda.$$

*proof.* Let $a_1 < \cdots < a_k$ be such that $M^{a_i}$ contains exactly $i$ critical points, and $M^{a_k} = M$. Then by Theorem 1.10, we have

$$\begin{aligned} H_n(M^{a_i}, M^{a_{i-1}}) &= H_n(M^{a_{i-1}} \cup e^{\lambda_i}, M^{a_{i-1}}) \\ &= H_n(e^{\lambda_i}, \dot{e}^{\lambda_i}) && \text{(Excision Axiom)} \\ &= \begin{cases} \mathbb{Z} & n = \lambda_i \\ 0 & \text{otherwise} \end{cases} && \text{(Theorem 2.22)} \end{aligned}$$

Thus by Lemma 2.24, we have

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

□

**Lemma 2.26.** *The function $S_\lambda$ is subadditive, where*

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

*proof.* Consider the long exact sequence

$$\cdots \longrightarrow H_\lambda(Y,Z) \xrightarrow{\alpha} H_\lambda(X,Z) \xrightarrow{\beta} H_\lambda(X,Y) \xrightarrow{d} H_{\lambda-1}(Y,Z) \longrightarrow \cdots$$

Then we have

$$\begin{aligned}
\operatorname{rank} d_{\lambda+1} &= R_\lambda(Y,Z) - \operatorname{rank} \alpha_\lambda \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + \operatorname{rank} \beta_\lambda \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + R_\lambda(X,Y) - \operatorname{rank} d_\lambda \\
&= R_\lambda(Y,Z) - R_\lambda(X,Z) + R_\lambda(X,Y) - \cdots \pm R_0(X,Y) \\
&\geq 0.
\end{aligned}$$

Collecting the like terms and rearranging yields

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

completing the proof.

□

**Theorem 2.27** (Morse's Inequalities)**.** *Use notations and assumptions as above, we have*

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

*proof.* From Lemma 2.26 we get the subadditivity of $S_\lambda$, and apply Lemma 2.24 on

$$\emptyset \subseteq M^{a_1} \subseteq \cdots \subseteq M^{a_k} = M,$$

we have

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

The right hand side, by Theorem 2.25, is $C_\lambda - C_{\lambda-1} + \cdots \pm C_0$, so

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

□

**Corollary 2.28.** *If $C_{\lambda+1} = C_{\lambda-1} = 0$, then $R_\lambda(M) = C_\lambda$.*

*proof.* By Theorem 2.25, if $C_{\lambda+1} = C_{\lambda-1} = 0$, then $R_{\lambda+1}(M) = R_{\lambda-1}(M) = 0$. By Theorem 2.27 on $\lambda+1$, we get

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

Thus,

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

Applying Theorem 2.27 on $\lambda$, we see that

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

Thus we must have

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

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

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

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

# Chapter 3

# Existence of Morse Functions

The theory we have developed relies on functions with nondegenerate critical points, which we now define as *Morse functions.* Specifically, if $f : M \to \mathbb{R}$ is a smooth function with all nondegenerate critical points, it is a Morse function. It is crucial to establish that such functions exist on our manifold; otherwise, Morse theory would be nonsense. This chapter shows that most smooth functions are Morse, and those that are not can be adjusted to form new Morse functions.

## 3.1 The Euclidean Case

**Theorem 3.1.** *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 zero.*

**Definition 3.1.** Let $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. Given any smooth real-valued function $f$ on $M$, we define a new function

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

**Lemma 3.2.** *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 $a = (a_1, \dots, a_n)$ and define

$$g = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right).$$

We write the Hessian matrix of $f$ at $x$ to be $H(f)(x)$, then

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

Also note that

$$df_a = g + a,$$

so firstly we have $H(f_a)(x) = dg(x)$ and secondly we have the set of critical points of $f_a$ is the set of points satisfying $g = -a$. Sard's Theorem tells us that almost every $-a$ is a regular value of $g$, so for almost every $a \in \mathbb{R}$, we have $dg(x) = H(f_a)(x)$ is nonsingular, where $x$ is a critical point of $f_a$.

□

## 3.2 A Weak Whitney Embedding Theorem on Possibly Non-compact Manifolds

**Definition 3.2.** A real-valued continuous function $f$ on $M$ is called an *exhaustion function* for $M$ if for any $c \in \mathbb{R}$, $M^c$ is compact.

**Lemma 3.3.** *There exists a positive smooth exhaustion function on any smooth manifold.*

*proof.* Since $M$ is a manifold, it admits a countable open cover $\{V_j\}_{j\in\mathbb{N}}$ with $V_j$ precompact. Let $\{\rho_j\}$ be a smooth partition of unity subordinate to it. Define

$$f(x) = \sum_j j\rho_j.$$

It is positive since $f \geq \sum \rho_j = 1$.

Now, let $c \in \mathbb{R}$ and and choose some positive integer $N > c$. If $p \notin \bigcup_{j=1}^N \bar{V}_j$, then $\rho_j(p) = 0$ for $0 \leq j \leq N$. Thus,

$$f(p) = \sum_{j=N+1}^{\infty} j\rho_j(p) \geq \sum_{j=N+1}^{\infty} N\rho_j(p) = N > c.$$

Therefore, equivalently, if $f(p) \leq c$, we must have $p \in \bigcup_{j=1}^N \bar{V}_j$, so $M^c \subseteq \bigcup_{j=1}^N \bar{V}_j$. Since a closed subset of a compact set is compact, we conclude that $M^c$ is compact.

□

**Lemma 3.4.** *If $M$ is a manifold which can be covered by finite charts, then $M$ admits an injective immersion into $\mathbb{R}^K$ for sufficiently large $K$.*

*proof.* Let $M$ be covered by charts $\{(U_i, \phi_i)\}_{i=1}^k$ and $\{\rho_i\}$ be a partition of unity subordinate to this cover. Define

$$\iota : M \to \mathbb{R}^{k(m+1)}, p \mapsto (\rho_1(p)\phi_1(p), \dots, \rho_k(p)\phi_k(p), \rho_1(p), \dots, \rho_k(p)),$$

where $m = \dim M$ and each $\rho_i\phi_i$ is defined on the whole $M$ by extending to 0. It is obvious that this map is injective. It is an immersion because for all $X_p \in T_pM$, we have

$$\begin{aligned}(d\iota)_p(X_p) = (&X_p(\rho_1)\phi_1(p) + \rho_1(p)(d\phi_1)_p(X_p), \dots, X_p(\rho_k)\phi_k(p) + \rho_k(p)(d\phi_k)_p(X_p),\\ &X_p(\rho_1), \dots, X_p(\rho_k))\end{aligned}$$

If $(d\iota)_pX_p$ is zero, $X_p(\rho_i)$ must be zero, so $(d\phi_i)_pX_p = 0$ if $\rho_i(p) \neq 0$. Since $\{\rho_i\}$ is a partition of unity, there is an $i$ such that $\rho_i(p) \neq 0$. Since $d\phi_i$ is an isomorphism, we must have $X_p$ is zero. Thus, $\iota$ is an immersion. □

**Theorem 3.5.** *If $M$ is an $m$ dimensional manifold which can be covered by finite charts, then $M$ admits an injective immersion into $\mathbb{R}^{2m+1}$.*

*proof.* By Lemma 3.4, there is an injective immersion $\iota : M \to \mathbb{R}^K$ with $K > 2m + 1$. Let $[v] \in \mathbb{R}P^{K-1}$ and define

$$P_{[v]} = \{w \in \mathbb{R}^K \mid v \cdot w = 0\} \cong \mathbb{R}^K.$$

Let $\pi_{[v]} : \mathbb{R}^K \to P_{[v]}$ be the canonical projection. Denote $\iota_{[v]} := \pi_{[v]} \circ \iota$. We claim that for almost every $[v]$, $\iota_{[v]}$ is an injective immersion. Equivalently, the set of $[v]$ making $\iota_{[v]}$ not injective or not immersion has measure zero.

First, suppose $[v]$ makes $\iota_{[v]}$ not injective, then one can pick $p_1 \neq p_2$ in $M$ such that $[\iota(p_1) - \iota(p_2)] = [v]$. Define

$$\alpha : M \times M \setminus \{(p,p) \mid p \in M\} \to \mathbb{R}P^{K-1}, (p_1, p_2) \mapsto [\iota(p_1) - \iota(p_2)].$$

It is well defined since $\iota$ is injective. Since

$$\dim(M \times M \setminus \{(p,p) \mid p \in M)\} = \dim M \times M = 2m < K - 1 = \dim \mathbb{R}P^{K-1},$$

the image of $\alpha$ is composed of critical values, so by Sard's Theorem, $\operatorname{im} \alpha$ is of measure zero.

Second, suppose $[v]$ makes $\iota_{[v]}$ not an immersion, then there is some $p \in M$ and a nonzero $X_p \in T_pM$ such that $(d\iota_{[v]})_p X_p = 0$. Thus,

$$d\pi_{[v]} \circ d\iota X_p = 0.$$

Since $\pi_{[v]}$ is a linear map, $\pi_{[v]} = \pi_{[v]}$, so

$$\pi_{[v]} \circ d\iota X_p = 0.$$

Equivalently, we have

$$[d\iota X_p] = [v].$$

Define

$$\beta : TM \to \mathbb{R}P^{K-1}, (p, X) \mapsto [d\iota X_p].$$

Again, since

$$\dim TM = 2m < K - 1 = \dim \mathbb{R}P^{K-1},$$

$\operatorname{im} \beta$ is of measure zero.

□

**Theorem 3.6.** *Any smooth manifold $M$ admits an injective immersion into $\mathbb{R}^K$ for sufficiently large $K$.*

*proof.* By Lemma 3.3, there exists some positive exhaustion function $f$ on $M$. For each $i \in \mathbb{N}^*$, we define

$$M_i = f^{-1}[i, i+1].$$

Since each $M_i$ is compact, it is contained in a finite cover of charts $U_1, \ldots, U_k$. Let

$$N_i = \left(\bigcup_{i=1}^{k} U_i\right) f^{-1} \cap f(i - \varepsilon, i + 1 + \varepsilon)$$

for some positive $\varepsilon < 1/2$. Then $N_i$ is an open submanifold of $M$ containing $M_i$, and $N_i \cap N_j = \emptyset$ if $|i - j| \geq 2$. Note that $N_i$ is a manifold which can be covered by finite charts, so by Theorem 3.5, $N_i$ admits an injective immersion $\phi_i : N_i \to \mathbb{R}^{2m+1}$, where $m = \dim M$.

Now define a smooth bump function $\rho_i$ such that $\rho_i|_{M_i} \equiv 1$ and $\operatorname{supp}(\rho_i) \subseteq N_i$. Define

$$\iota : M \to \mathbb{R}^{4m+3}, p \mapsto \left( \sum_{i \text{ odd}} \rho_i(p)\phi_i(p), \sum_{i \text{ even}} \rho_i(p)\phi_i(p), f(p) \right).$$

First we argue that $\iota$ is injective. If $\iota(p_1) = \iota(p_2)$, then $\exists i \in \mathbb{N}$ such that $f(p_1) = f(p_2) \in [i, i+1]$. Therefore, $p_1, p_2 \in M_i \subset N_i$ and $\phi_i(p_1) = \phi_i(p_2)$. Since $\phi_i$ is injective, we have $p_1 = p_2$.

Second we argue that $\iota$ is an immersion. Suppose $p \in M_i$. Without loss of generality, we assume $i$ is odd. Then for any $0 \neq X_p \in T_pM$,

$$d\iota_p(X_p) = ((d\phi_i)_p(X_p), *, *).$$

Since $\phi_i$ is an immersion on $U_i \ni p$, we have $(d\phi_i)_p(X_p) \neq 0$. Thus, $d\iota_p(X_p) \neq 0$. $\square$

Unfortunatly, since we do not require $M$ to be compact, an injective immersion may not be an embedding. We need the following technical lemma to prove that $M$ might be embedded into a larger Euclidean space.

**Lemma 3.7.** *If $f : M \to N$ is a proper injective immersion, then $f$ is an embedding.*

*proof.* $f$ is proper means $f$ is open. Since $f$ is injective, the left inverse of $f$ from its image can be define: $f^{-1} : f(M) \to M$. Note that $(f^{-1})^{-1}(U) = f(U)$ for any open $U \subseteq M$. Since $f$ is open, $f(U)$ is open in $N$. Thus, $f^{-1}$ is continuous, so $M$ is homeomorphic to $f(M)$ via $f$. $\square$

Note that an exhaustion function is a proper map, so we have

**Theorem 3.8.** *Any smooth manifold $M$ admits an embedding into $\mathbb{R}^K$ for sufficiently large $K$.*

## 3.3 Existence of Morse Function on Smooth Manifolds

Translating the result from Euclidean spaces to manifolds will be a bit more tricky; we will need to establish some new concepts and lemmas before we can tackle it. Consider $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^\ell$. Let $c \in \mathbb{R}^k$, and we denote the vertical slice as $V_c = \{c\} \times \mathbb{R}^\ell$. Let $A$ be a subset of $V_c$; then $A = \{c\} \times U$ for some subset $U$ of $\mathbb{R}^\ell$. We say $A$ has measure zero in $V_c$ if $U$ has measure zero in $\mathbb{R}^\ell$.

The following two lemmas will be important in the proof of Fubini's theorem on measure zero.

**Lemma 3.9.** *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*

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

*proof.* Since every open set may be expressed as a countable union of open intervals, we may express each $S_i$ as

$$S_i = \bigcup_{k=1}^{\infty} (a_{ik}, b_{ik}).$$

Then, the union of all the open intervals of $S_1, \dots, S_n$ 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 $j = 1, 2, \dots, m$. Now let

$$\varepsilon = b - a.$$

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 $Q_1 = (a - \varepsilon_1, b_1)$ where $\varepsilon_1 < \varepsilon$ and $a - \varepsilon_1 \in (a_1, b_1)$. From our indexing, $b_1 \in (a_2, b_2)$; then we may choose $\varepsilon_2$ so that $b_1 - \varepsilon_2 \in (a_2, b_2)$ and $\varepsilon_2 < \varepsilon$. Therefore, let $Q_2 = (b_1 - \varepsilon_2, b_2)$. Repeat the construction for $j = 3, 4, \dots, m - 1$. For $Q_m$, choose $\varepsilon_m$ such that $\varepsilon_m < \varepsilon$ and both $b + \varepsilon_m$ and $b_{m-1} - \varepsilon_m$ are contained in $(a_m, b_m)$. Finally, let $Q_m = (b_{m-1} - \varepsilon_m/2, b + \varepsilon_m/2)$.

Now, we compute the total length:

$$\begin{aligned}\sum_{j=1}^{m} \text{length}(Q_j) =&(b_1 - a + \varepsilon_1) + (b_2 - b_1 + \varepsilon_2) + \cdots \\ &+ (b_{m-1} - b_{m-2} + \varepsilon_{m-2}) + \left(b + \frac{\varepsilon_m}{2} - b_{m-1} + \frac{\varepsilon_m}{2}\right) \\ =&(b - a) + (\varepsilon_1 + \cdots + \varepsilon_m) \\ <&(b - a) + m\varepsilon \\ =&2(b - a).\end{aligned}$$

Therefore, $Q_1, \dots, Q_m$ is our desired cover.

□

**Lemma 3.10.** *Let $A$ be a compact subset of $\mathbb{R}^n$ and let $c \in \mathbb{R}$. Suppose $A \cap V_c \subset \{c\} \times U$ for $U$ open in $\mathbb{R}^{n-1}$; then for a sufficiently small interval $I$ containing $c$, $A \cap V_I \subset I \times U$.*

*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 $x_i \notin U$.

Since $A$ is compact, 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.

□

**Theorem 3.11** (Fubini's Theorem on Measure Zero)**.** *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 $\ell = n-1$. For $k > 1$, the argument follows by induction. Since $A$ is closed, we may assume $A$ is made up of countably many compact sets (just take $A_j = [-j, j] \cap A$). Therefore, it suffices to show that a compact set $A$ with the above property has measure 0.

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

By Lemma 3.10, there exists an open interval $J_c$ such that

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

This holds for all $c \in I$ and clearly $J_c$ is an open cover of $I$. By compactness of $I$ and Lemma 3.9, 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 $J'(c_i) \subset J_c$ for some $c$, therefore

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

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

□

**Theorem 3.12.** *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 first want to 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 $(x_1, \ldots, x_N)$ be the usual coordinates on $\mathbb{R}^N$. Consider the dual space of $\mathbb{R}^N$ with its usual basis $\phi_1, \ldots, \phi_N$, that is

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

Let $V = T_pM$ be the tangent space of $M$ at $p$. Then the set $\phi_1|_V, \ldots, \phi_N|_V$ spans the dual space of $V$ but is not linearly independent as $\dim T_xM = k$. There will however exist a subset $\phi_{i_1}|_V, \ldots, \phi_{i_k}|_V$ 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 $\phi_i$ to the tangent space $T_pM$. For the prior indexed sub-collection $i_1, \ldots, i_k$, the function $f : M \to \mathbb{R}^k$ by

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

has the following derivative

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

But $\phi_{i_1}|_V, \ldots, \phi_{i_k}|_V$ forms a basis on $T_xM$ so $df$ is an isomorphism. By the Inverse Function Theorem, $(x_{i_1}, \ldots, x_{i_k})$ 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 $x_{i_1}, \ldots, x_{i_k}$ mapping $U_x$ to an open set in $\mathbb{R}^k$. Applying the second countability of $\mathbb{R}^N$, we may assume there are countably many $U_i$ that cover $M$. Fix a $U_i$ and for convenience assume $x_1, \ldots, x_k$ 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

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

Since $U_i$ is homeomorphic to an open set in $\mathbb{R}^k$ and non-degeneracy is preserved under diffeomorphic composition, Lemma 3.2 tells us for almost all $b \in \mathbb{R}^k$, $f_{(b,c)}$ is Morse on $U_i$. Thus, we have shown that $S_i \cap V_c$ has measure 0 for all $c \in \mathbb{R}^{N-k}$. By Theorem 3.11, $S_i$ 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 $U_i$; hence the set of $a \in \mathbb{R}^N$ in which $f_a$ is not Morse is equal to the union of $S_i$ and must also have measure 0.

□

# Chapter 4

# Basic Knowledge in Riemannian Geometry

This chapter may offer essential groundwork for later chapters.

## 4.1 Covariant Differentiation

**Definition 4.1** (Affine Connection)**.** An *affine connection* at $p \in M$ is a map assigning to $X_p \in T_pM$ and to a smooth vector field $Y$ a vector in $T_pM$, where we denote

$$X_p \vdash Y \in T_pM.$$

Some also denotes it by $\nabla_X Y$.

It should satisfy some properties. First, it is a bilinear map of $X_p$ and $Y$. Second, it satisfies

$$X_p \vdash (fY) = (X_p f)Y_p + fX_p \vdash Y.$$

**Definition 4.2** (Global Connection)**.** A *connection* assigns to each point $p \in M$ an affine connection $\vdash_p$. It should satisfy four conditions:

1. If we define a vector field by

   $$X \vdash Y := X_p \vdash Y,$$

   where $X, Y$ are smooth vector fields, then this new vector field is smooth.

2. $\vdash$ is bilinear in $X, Y$.

3. $(fX) \vdash Y = f(X \vdash Y)$.

4. $X \vdash (fY) = (Xf)Y + f(X \vdash Y)$.

In a local chart $(U, u^1, \ldots, u^n)$, we can determine $\vdash$ locally by setting $n^3$ functions $\Gamma^k_{ij}$. If we write $\partial_i := \partial/\partial u^i$, we express

$$\partial_i \vdash \partial_j = \sum_k \Gamma^k_{ij} \partial_k.$$

This can determine $\vdash$. Particularly, if we pick two vector fields $X, Y$ in $U$, then we can express $X = \sum x^i \partial_i$ and $Y = \sum y^i \partial_i$ for some smooth functions $x_i, y_i$ in $U$. Then

$$
\begin{aligned}
X \vdash Y &= \sum_i x^i \sum_j \partial_i \vdash y^j \partial_j \\
&= \sum_i x^i \left( \sum_j (\partial_i y^j) \partial_j + y^j (\partial_i \vdash \partial_j) \right) \\
&= \sum_i x^i \left( \sum_j (\partial_i y^j) \partial_j + \sum_j y^j \sum_k \Gamma_{ij}^k \partial_k \right) \\
&= \sum_i x^i \left( \sum_k \left( \partial_i y^k + \sum_j y^j \Gamma_{ij}^k \right) \partial_k \right) \\
&= \sum_k \left( \sum_i x^i y^k_{,i} \partial_k \right),
\end{aligned}
$$

where we write for simplicity

$$
y^k_{,i} := \partial_i y^k + \sum_j y^j \Gamma_{ij}^k
$$

**Definition 4.3.** A vector field along the parametrized curve $c : \mathbb{R} \to M$ is a function assigning to each $t \in \mathbb{R}$ a vector $V_t$ in $T_{c(t)}M$ satisfying the smooth condition: for any smooth function $f$ on $M$, the map $t \mapsto V_t f$ is smooth.

Recall that $dc/dt := c_* d/dt$ is a vector field along $c$, since $dc/dt(t)f = d/dt f c(t)$ is smooth. We now define a derivative of any vector field along $c$ called the covariant derivative.

**Definition 4.4** (Covariant Derivative)**.** Let $M$ be a smooth manifold with a global connection $\vdash$. Let $V$ be a vector field along $c$. We define $DV/dt|_t \in T_{c(t)}M$ the covariant derivative of $V$ at $t$. It should satisfy three axioms:

1. $D(V + W)/dt = DV/dt + DW/dt$.
2. $D(fV)/dt = df/dt\, V + f DV/dt$, where $f \in C^\infty(\mathbb{R})$ and $(fV)_t = f(t)V_t$.
3. If there is a vector field $Y$ such that $Y_{c(t)} = V_t$, then we have
$$
\frac{DV}{dt} = \frac{dc}{dt} \vdash Y.
$$

We can write down, locally, how $DV/dt$ looks like. Suppose $(U, u^1, \ldots, u^n)$ is a local chart which intersects $c$, and we write $\tilde{u}^i := u^i \circ c$ smooth functions on $t$ and $\partial_i = \partial/\partial u^i$. Then if we suppose

$$
V = \sum_i v^i \partial_i,
$$

we have

$$\begin{aligned}
\frac{DV}{dt} &= \sum_i \left( \frac{dv^i}{dt}\partial_i + v^i \frac{D(\tilde{\partial}_i)_t}{dt} \right) \qquad \text{where } (\tilde{\partial}_i)_t = (\partial_i)_{c(t)} \\
&= \sum_i \left( \frac{dv^i}{dt}\partial_i + v^i \frac{dc}{dt} \vdash \partial_i \right) \\
&= \sum_i \left( \frac{dv^i}{dt}\partial_i + v^i \sum_j \frac{d\tilde{u}^j}{dt}\partial_j \vdash \partial_i \right) \\
&= \sum_i \left( \frac{dv^i}{dt}\partial_i + v^i \sum_j \frac{d\tilde{u}^j}{dt} \sum_k \Gamma_{ij}^k \partial_k \right) \\
&= \sum_k \left( \frac{dv^k}{dt} + \sum_{i,j} \frac{d\tilde{u}^j}{dt}\Gamma_{ij}^k v^i \right) \partial_k.
\end{aligned}$$

One can check, if we define $DV/dt$ by the formula above, it will automatically satisfies the three axioms. Thus,

**Proposition 4.1.** *There is exactly one operation $V \mapsto DV/dt$.*

**Definition 4.5.** A vector field $V$ along $c$ is said to be a parallel vector field if the covariant derivative $DV/dt$ is identically zero.

**Proposition 4.2.** *Given a curve $c$ and a tangent vector $V_0$ at the point $c(0)$, there is one and only one parallel vector field $V$ along $c$ which extends $V_0$.*

*proof.* The differential equations

$$\frac{dv^k}{dt} + \sum_{i,j} \frac{d\tilde{u}^j}{dt}\Gamma_{ij}^k v^i = 0$$

have solutions $v_k(t)$ which are uniquely determined by the initial values $v_k(0)$. Since these equations are linear, the solutions can be defined for all relevant values of $t$.

□

**Definition 4.6.** The vector $V_t$ is said to be obtained from $V_0$ by *parallel translation* along $c$.

**Definition 4.7.** A connection $\vdash$ on $M$ is *compatible with the Riemannian structure* $\langle \cdot, \cdot \rangle$ if parallel translation preserves inner products. In other words, for any parametrized curve $c$ and any pair $P, P'$ of parallel vector fields along $c$, the inner product $\langle P, P' \rangle$ should be constant.

**Proposition 4.3.** *Suppose $\vdash$ and $\langle \cdot, \cdot \rangle$ are compatible on $M$ and $V, W$ are vector fields along curve $c$, then*

$$\frac{d}{dt}\langle V, W \rangle = \left\langle \frac{DV}{dt}, W \right\rangle + \left\langle V, \frac{DW}{dt} \right\rangle$$

*proof.* Choose parallel vector fields $P_1, \ldots, P_n$ along $c$ which are orthonormal at one point in $c$ and hence at any point in $c$. Then we can write

$$V = \sum v^i P_i, W = \sum w^i P_i$$

and hence

$$\langle V, W \rangle = \sum v^i w^i,$$

so

$$\frac{d}{dt}\langle V, W \rangle = \sum \frac{dv^i}{dt} w^i + v^i \frac{dw^i}{dt}.$$

Note that

$$\frac{DV}{dt} = \sum \frac{dv^i}{dt} P_i + v^i \frac{DP_i}{dt} = \sum \frac{dv^i}{dt} P_i,$$

and similarly we have

$$\frac{DW}{dt} = \sum \frac{dw^i}{dt} P_i.$$

Thus the right hand side equals

$$\sum \frac{dv^i}{dt} w^i + \frac{dw^i}{dt} v^i.$$

This completes the proof. □

**Corollary 4.4.** *For any smooth vector fields $Y, Y'$ on $M$ and $X_p \in T_pM$, we have*

$$X_p\langle Y, Y' \rangle = \langle X_p \vdash Y, Y'_p \rangle + \langle Y_p, X_p \vdash Y' \rangle.$$

*proof.* Let $c$ be a curve such that $c(0) = p$ and $dc/dt(0) = X_p$. Then it suffices to show

$$dc/dt(0)\langle Y, Y' \rangle = \langle dc/dt(0) \vdash Y, Y'_p \rangle + \langle Y_p, dc/dt(0) \vdash Y' \rangle.$$

The left hand side:

$$\begin{aligned} dc/dt(0)\langle Y, Y' \rangle &= c_* d/dt(0)\langle Y, Y' \rangle \\ &= d/dt(0)(\langle Y, Y' \rangle \circ c) \\ &= d/dt(0)\langle Y \circ c, Y' \circ c \rangle; \end{aligned}$$

The right hand side:

$$\langle dc/dt(0) \vdash Y, Y'_p \rangle + \langle Y_p, dc/dt(0) \vdash Y' \rangle = \langle D(Y \circ c)/dt(0), Y'_p \rangle + \langle Y_p, D(Y' \circ c)/dt(0) \rangle$$

Then by Proposition 4.3, the corollary follows. □

**Definition 4.8.** A connection $\vdash$ is called *symmetric* if

$$X \vdash Y - Y \vdash X = [X, Y].$$

Since $[\partial_i, \partial_j] = 0$, if we suppose $X = \partial_i$ and $Y = \partial_j$, we may get

$$\partial_i \vdash \partial_j = \partial_j \vdash \partial_i,$$

so $\Gamma^k_{ij} = \Gamma^k_{ji}$. One can show that if $\Gamma^k_{ij} = \Gamma^k_{ji}$, $\vdash$ must be symmetric.

**Theorem 4.5** (Fundamental Lemma of Riemannian Geometry)**.** *A Riemannian manifold possesses one and only one symmetric connection which is compatible with its metric.*

*proof.* Suppose there exists such a symmetric connection. By Corollary 4.4 and by setting $g_{ij} := \langle \partial_i, \partial_j \rangle$, we have

$$\partial_i g_{jk} = \langle \partial_i \vdash \partial_j, \partial_k \rangle + \langle \partial_j, \partial_i \vdash \partial_k \rangle.$$

Permuting $i, j, k$ we get three equations about

$$\langle \partial_i \vdash \partial_j, \partial_k \rangle, \langle \partial_j \vdash \partial_k, \partial_i \rangle, \langle \partial_k \vdash \partial_i, \partial_j \rangle,$$

and thus we can calculate that

$$\langle \partial_i \vdash \partial_j, \partial_k \rangle = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}).$$

This formula is called *the first Christoffel identity.*

The left hand side is

$$\langle \partial_i \vdash \partial_j, \partial_k \rangle = \left\langle \sum_\ell \Gamma^\ell_{ij} \partial_\ell, \partial_k \right\rangle = \sum_\ell \Gamma^\ell_{ij} g_{\ell k}.$$

Multiplying by $(g^{k\ell})$, the inverse of $(g_{\ell k})$, we get

$$\Gamma^\ell_{ij} = \sum_k \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) g^{k\ell}.$$

This formula is called *the second Christoffel identity.*

From the second Christoffel identity, we know that the connection is uniquely determined by the Riemannian metric. Conversely, if we define $\Gamma^k_{ij}$ using the second Christoffel identity, we get the connection we need. □

An alternative characterization of symmetry is useful. Consider a parametrized surface $s : \mathbb{R}^2 \to M$, we define that a vector field along $s$ is a map assigning to each point $(x, y)$ a vector

$$V_{(x,y)} \in T_{s(x,y)} M$$

satisfying $(x, y) \mapsto V_{(x,y)} f$ is smooth for any smooth function $f$.

We denote $\partial s/\partial x := s_* \partial/\partial x$ and $\partial s/\partial y := s_* \partial/\partial y$. Note that both are vector fields along $s$. Let $V$ be a vector field along $s$, we define $(DV/dx)_{(x,y_0)}$ to be the covariant derivative by considering curve $x \mapsto s(x, y_0)$. Hence $DV/dx$ is defined anywhere along $s$, and similarly we can define $DV/dy$. We can characterize symmetry using these new concepts:

**Proposition 4.6.** *If* $\vdash$ *is symmetric, then*

$$\frac{D}{dx}\frac{\partial s}{\partial y} = \frac{D}{dy}\frac{\partial s}{\partial x}.$$

*proof.* Pick $(U, u^1, \dots, u^n)$ a local chart intersecting $s(\mathbb{R}^2)$. Define $\tilde{u}^i = u^i \circ s$ and $(\tilde{\partial}_i)_x = (\partial_i)_{s(x,y_0)}$ at some $y_0$. Then

$$\begin{aligned}
\frac{D}{dx}\frac{\partial s}{\partial y} &= \sum_i \frac{D}{dx}\frac{\partial \tilde{u}^i}{\partial y}\partial_i \\
&= \sum_i \frac{d}{dx}\frac{\partial \tilde{u}^i}{\partial y}\partial_i + \frac{\partial \tilde{u}^i}{\partial y}\frac{D(\tilde{\partial}_i)_x}{dx} \\
&= \sum_i \frac{d}{dx}\frac{\partial \tilde{u}^i}{\partial y}\partial_i + \frac{\partial \tilde{u}^i}{\partial y}\frac{ds^{y_0}(x)}{dx} \vdash \partial_i \\
&= \sum_i \frac{d}{dx}\frac{\partial \tilde{u}^i}{\partial y}\partial_i + \frac{\partial \tilde{u}^i}{\partial y}\left(\sum_j \frac{\partial \tilde{u}^j}{\partial x}\partial_j\right) \vdash \partial_i \\
&= \sum_i \frac{d}{dx}\frac{\partial \tilde{u}^i}{\partial y}\partial_i + \frac{\partial \tilde{u}^i}{\partial y}\sum_j \frac{\partial \tilde{u}^j}{\partial x}\sum_k \Gamma^k_{ij}\partial_k \\
&= \sum_k \left(\frac{\partial^2 \tilde{u}^k}{\partial x \partial y} + \sum_{i,j}\frac{\partial \tilde{u}^i}{\partial y}\frac{\partial \tilde{u}^j}{\partial x}\Gamma^k_{ij}\right)\partial_k.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\frac{D}{dx}\frac{\partial s}{\partial y} &= \sum_k \left(\frac{\partial^2 \tilde{u}^k}{\partial x \partial y} + \sum_{i,j}\frac{\partial \tilde{u}^i}{\partial x}\frac{\partial \tilde{u}^j}{\partial y}\Gamma^k_{ij}\right)\partial_k \\
&= \sum_k \left(\frac{\partial^2 \tilde{u}^k}{\partial x \partial y} + \sum_{i,j}\frac{\partial \tilde{u}^i}{\partial y}\frac{\partial \tilde{u}^j}{\partial x}\Gamma^k_{ji}\right)\partial_k \\
&= \frac{D}{dx}\frac{\partial s}{\partial y}.
\end{aligned}$$

□

## 4.2 The Curvature Tensor

**Definition 4.9.** The *curvature tensor* $R$ of connection $\vdash$ is defined as follows.

$$R(X,Y)Z := -X \vdash (Y \vdash Z) + Y \vdash (X \vdash Z) + [X,Y] \vdash Z,$$

where $X, Y, Z$ are smooth vector fields on $M$. Equivalently, if we use $\nabla$ symbol,

$$R(X,Y) = \nabla_{[X,Y]} - [\nabla_X, \nabla_Y].$$

**Proposition 4.7.** *For smooth functions $f, g, h$, we have*

$$R(fX, gY)hZ = fghR(X,Y)Z.$$

*proof.* Note that

$$\begin{aligned}
&R(fX,Y)Z \\
=& -fX \vdash (Y \vdash Z) + (Yf)(X \vdash Z) + fY \vdash (X \vdash Z) - (Yf)(X \vdash Z) + f[X,Y] \vdash Z \\
=& fR(X,Y)Z
\end{aligned}$$

and similarly we have $R(X, gY)Z = gR(X, Y)Z$. Finally,

$$\begin{aligned}
&R(X,Y)hZ\\
=&-(X(Yh))Z-(Yh)X\vdash Z-(Xh)Y\vdash Z-hX\vdash(Y\vdash Z)\\
&+(Y(Xh))Z+(Xh)Y\vdash Z+(Yh)X\vdash Z+hY\vdash(X\vdash Z)\\
&+([X,Y]h)Z+h[X,Y]\vdash Z\\
&=-hX\vdash(Y\vdash Z)+hY\vdash(X\vdash Z)+h[X,Y]\vdash Z\\
&=hR(X,Y)Z.
\end{aligned}$$

□

**Proposition 4.8.** *$(R(X, Y)Z)_p$ depends only on $p$ and is denoted by $R(X_p, Y_p)Z_p$. Furthermore, the correspondence*

$$(X_p, Y_p, Z_p) \mapsto R(X_p, Y_p)Z_p$$

*is tri-linear.*

*proof.* Write $X = \sum x^i\partial_i, Y = \sum y^i\partial_i, Z = \sum z^i\partial_i$, we have

$$R(X,Y)Z = \sum_{i,j,k} x^i y^j z^k R(\partial_i, \partial_j)\partial_k.$$

Thus, $(R(X, Y)Z)_p$ is determined by the values of $x^i, y^i, z^i$ at $p$ only. The tri-linear-ness comes from Proposition 4.7.

□

Now we consider a parametrized surface $s : \mathbb{R}^2 \to M$.

**Proposition 4.9.**

$$\frac{D}{dy}\frac{D}{dx}V - \frac{D}{dx}\frac{D}{dy}V = R\left(\frac{\partial s}{\partial x}, \frac{\partial s}{\partial y}\right)V$$

*for each $V$ along $s$.*

*proof.* As usual, we work in $(U, u^i)$ with $\tilde{u}^i = u^i \circ s$. First we calculate the right hand side:

$$\begin{aligned}
R\left(\frac{\partial s}{\partial x}, \frac{\partial s}{\partial x}\right)V =&R\left(\sum_i \frac{\partial \tilde{u}^i}{\partial x}\partial_i, \sum_j \frac{\partial \tilde{u}^j}{\partial y}\partial_j\right)\sum_k v^k\partial_k\\
=&\sum_{i,j,k}\frac{\partial \tilde{u}^i}{\partial x}\frac{\partial \tilde{u}^j}{\partial y}v^k R(\partial_i,\partial_j)\partial_k\\
=&\sum_{i,j,k}\frac{\partial \tilde{u}^i}{\partial x}\frac{\partial \tilde{u}^j}{\partial y}v^k(\partial_j\vdash(\partial_i\vdash\partial_k)-\partial_i\vdash(\partial_j\vdash\partial_k)),
\end{aligned}$$

since $[\partial_i, \partial_j] = 0$.

Then we calculate

$$\begin{aligned}
\frac{D}{dy}\frac{D}{dx}V =& \frac{D}{dy}\frac{D}{dx}\sum_i v^i\partial_i \\
=& \frac{D}{dy}\sum_i \frac{\partial v^i}{\partial x}\partial_i + v^i\frac{\partial s}{\partial x}\vdash\partial_i \\
=& \frac{D}{dy}\sum_{i,j} \frac{\partial v^i}{\partial x}\partial_i + v^i\frac{\partial \tilde{u}^j}{\partial x}\partial_j\vdash\partial_i \\
=& \sum_i \left(\frac{\partial^2 v^i}{\partial x\partial y}\partial_i + \frac{\partial v^i}{\partial x}\frac{D\partial_i}{dy}\right) + \\
& \sum_{i,j}\left(\frac{\partial}{\partial y}\left(v^i\frac{\partial \tilde{u}^j}{\partial x}\right)\partial_j\vdash\partial_i + v^i\frac{\partial \tilde{u}^j}{\partial x}\frac{D\partial_j\vdash\partial_i}{dy}\right) \\
=& \sum_i \left(\frac{\partial^2 v^i}{\partial x\partial y}\partial_i + \frac{\partial v^i}{\partial x}\sum_j\frac{\partial \tilde{u}^j}{\partial y}\partial_j\vdash\partial_i\right) + \\
& \sum_{i,j}\left(\frac{\partial}{\partial y}\left(v^i\frac{\partial \tilde{u}^j}{\partial x}\right)\partial_j\vdash\partial_i + v^i\frac{\partial \tilde{u}^j}{\partial x}\sum_k\frac{\partial \tilde{u}^k}{\partial y}\partial_k\vdash(\partial_j\vdash\partial_i)\right) \\
=& \sum_i \frac{\partial^2 v^i}{\partial x\partial y}\partial_i + \sum_{i,j}\left(\frac{\partial v^i}{\partial x}\frac{\partial \tilde{u}^j}{\partial y} + \frac{\partial v^i}{\partial y}\frac{\partial \tilde{u}^j}{\partial x} + v^i\frac{\partial^2 \tilde{u}^j}{\partial x\partial y}\right)\partial_j\vdash\partial_i + \\
& \sum_{i,j,k} v^i\frac{\partial \tilde{u}^j}{\partial x}\frac{\partial \tilde{u}^k}{\partial y}\partial_k\vdash(\partial_j\vdash\partial_i).
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{D}{dy}\frac{D}{dx}V - \frac{D}{dx}\frac{D}{dy}V =& \sum_{i,j,k}\left(v^i\frac{\partial \tilde{u}^j}{\partial x}\frac{\partial \tilde{u}^k}{\partial y} - v^i\frac{\partial \tilde{u}^j}{\partial y}\frac{\partial \tilde{u}^k}{\partial x}\right)\partial_k\vdash(\partial_j\vdash\partial_i) \\
=& \sum_{i,j,k} v^i\frac{\partial \tilde{u}^j}{\partial x}\frac{\partial \tilde{u}^k}{\partial y}(\partial_k\vdash(\partial_j\vdash\partial_i) - \partial_j\vdash(\partial_k\vdash\partial_i)).
\end{aligned}$$

The proposition follows.

□

**Remark.** *In general, one cannot construct a parallel vector field $P$ along $s$, where parallel means $DP/dx = DP/dy = 0$. However, when the curvature tensor is zero, one can always construct such a vector field. Given any $P_{(0,0)} := P_0$, we construct $P_{(x,0)}$ the parallel vector field along $s(x,0)$. For each fixed $x_0$, define $P_{(x_0,y)}$ be a parallel vector field along the curve $s(x_0,y)$ with the initial value $P_{(x_0,0)}$ defined before. Clearly, $DP/dy = 0$, and $DP/dx$ is zero along $s(x,0)$. Now since*

$$\frac{D}{dy}\frac{D}{dx}P - \frac{D}{dx}\frac{D}{dy}P = R\left(\frac{\partial s}{\partial x},\frac{\partial s}{\partial y}\right)P = 0,$$

*we have $\frac{D}{dy}\frac{D}{dx}P = 0$. Thus, the vector space $DP/dx$ is parallel along $s(x_0,y)$. Since the initial value $DP/dx|_{y=0} = 0$, we conclude that $DP/dx = 0$.*

Henceforth we will assume that $M$ is a Riemannian manifold, provided with the unique symmetric connection which is compatible with its metric. Then we have the following properties of $R$.

**Proposition 4.10.** *The curvature tensor of a Riemannian manifold satisfies:*

1. $R(X,Y)Z + R(Y,X)Z = 0$,
2. $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$,
3. $\langle R(X,Y)Z, W\rangle + \langle R(X,Y)W, Z\rangle = 0$,
4. $\langle R(X,Y)Z, W\rangle + \langle R(Z,W)X, Y\rangle = 0$.

*proof.* Just use the definition. □

## 4.3 Geodesics and Completeness

Let $M$ be a connected Riemannian manifold. In this section, we first introduce the definition of geodesics, and then prove that locally, the geodesic is the path with least arc length. Finally, we will develop that geodesically complete is equivalent to complete when regarding $M$ as a metric space. Now we start with the definition.

**Definition 4.10.** A parametrized path $\gamma : I \to M$ where $I$ denotes any interval of real numbers, is called a *geodesic* if the acceleration vector field $\frac{D}{dt}\frac{d\gamma}{dt}$ is identically zero.

**Proposition 4.11.** *The length of the velocity vector $d\gamma/dt$ is a constant.*

*proof.* It is because

$$\frac{d}{dt}\left\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right\rangle = 2\left\langle \frac{D}{dt}\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right\rangle = 0.$$

□

In a local chart $(U, u^i)$, a path $\gamma$ is geodesic if and only if for each $k$,

$$\frac{d^2\tilde{u}^k}{dt^2} + \sum_{i,j}\frac{d\tilde{u}^i}{dt}\frac{d\tilde{u}^j}{dt}\Gamma^k_{ij} = 0,$$

where $\tilde{u}^i := u^i \circ \gamma$.

Note that $\Gamma^k_{ij}$ is a function of point $p \in M$, hence in $U$, it is a function of $(u^1, \ldots, u^n)$. Since it is restricted in the path, we can regard $u$ as $\tilde{u}$. Thus, we omit tilde symbol and rewrite the second order differential equations as

$$\frac{d^2u^k}{dt^2} + \sum_{i,j}\frac{du^i}{dt}\frac{du^j}{dt}\Gamma^k_{ij}(u) = 0,$$

where $u^i$ are real-valued smooth functions of $t$.

More generally, we can consider the following differential equations

$$\frac{d^2\mathbf{u}}{dt^2} = \mathbf{F}\left(\mathbf{u}, \frac{d\mathbf{u}}{dt}\right),$$

where $\mathbf{u}$ stands for $(u^1, \ldots, u^n)$ and $\mathbf{F}$ stands for an $n$-tuple of $C^\infty$ functions, all defined throughout some neighborhood $U$ of a point $(\mathbf{u}_1, \mathbf{v}_1) \in \mathbb{R}^{2n}$.

**Theorem 4.12.** *There exists a neighborhood $W$ of $(\mathbf{u}_1, \mathbf{v}_1)$ and $\varepsilon > 0$ so that for any $(\mathbf{u}_0, \mathbf{v}_0) \in W$, the initial value problem*

$$\begin{cases} \frac{d^2\mathbf{u}}{dt^2} = \mathbf{F}\left(\mathbf{u}, \frac{d\mathbf{u}}{dt}\right) \\ \mathbf{u}(0) = \mathbf{u}_0 \\ \frac{d\mathbf{u}}{dt}(0) = \mathbf{v}_0 \end{cases}$$

*has a unique solution $(-\varepsilon, \varepsilon) \to \mathbb{R}^n$, $t \mapsto \mathbf{u}(t)$.*

*Furthermore, the solution depends smoothly on the initial conditions. In other words, the correspondence*

$$(\mathbf{u}_0, \mathbf{v}_0, t) \mapsto \mathbf{u}(t)$$

*is smooth as a map from $W \times (-\varepsilon, \varepsilon)$ to $\mathbb{R}^n$.*

*proof.* Define $v^i = du^i/dt$ and $\mathbf{v} = (v^1, \dots, v^n)$, then the initial value problem becomes

$$\begin{cases} \frac{d\mathbf{u}}{dt} = \mathbf{v} \\ \frac{d\mathbf{v}}{dt} = \mathbf{F}(\mathbf{u}, \mathbf{v}) \\ \mathbf{u}(0) = \mathbf{u}_0 \\ \mathbf{v}(0) = \mathbf{v}_0 \end{cases}$$

Write $\mathbf{x} = (\mathbf{u}, \mathbf{v})$, then the equations become

$$\begin{cases} \frac{d\mathbf{x}}{dt} = (\mathbf{v}, \mathbf{F}(\mathbf{u}, \mathbf{v})) \\ \mathbf{x}(0) = (\mathbf{u}_0, \mathbf{v}_0) \end{cases}$$

Then the theorem follows due to Picard's Theorem. □

Applying this theorem to geodesics, we get the following theorem.

**Lemma 4.13.** *For every point $p_0$ on a Riemannian manifold $M$ there exists a neighborhood $U$ of $p_0$ and a number $\varepsilon > 0$ so that: for each $p \in U$ and each tangent vector $v \in T_pM$ with length $< \varepsilon$ there is a unique geodesic*

$$\gamma_v : (-2, 2) \to M$$

*satisfying the condition*

$$\gamma_v(0) = p, \ \frac{d\gamma_v}{dt}(0) = v.$$

*proof.* Theorem 4.12 implies that there exists a neighborhood $U$ of $p_0$ and numbers $\varepsilon_1, \varepsilon_2 > 0$ so that: for each $p \in U$ and each $v \in T_pM$ with $\|v\| < \varepsilon_1$, there is a unique geodesic $\gamma_v : (-2\varepsilon_2, 2\varepsilon_2) \to M$ satisfying the required initial conditions. To obtain the sharper statement, it is only necessary to observe that the differential equation for geodesics has the following homogeneity property. Let $c$ be any constant. If the parametrized curve

$$t \mapsto \gamma(t)$$

is a geodesic, then the parametrized curve

$$t \mapsto \gamma(ct)$$

is also a geodesic. Now suppose that $\varepsilon$ is smaller than $\varepsilon_1\varepsilon_2$. Then if $\|v\| < \varepsilon$ and $|t| < 2$, note that

$$\left\|\frac{v}{\varepsilon_2}\right\| < \varepsilon_1 \quad \text{and} \quad |\varepsilon_2 t| < 2\varepsilon_2.$$

Hence we can define $\gamma_v(t)$ to be $\gamma_{\frac{v}{\varepsilon_2}}(\varepsilon_2 t)$. This proves the lemma. □

**Definition 4.11** (Riemannian Exponential)**.** Let $V \in T_qM$ be a tangent vector, and suppose that there exists a geodesic

$$\gamma : [0,1] \to M$$

satisfying the conditions

$$\gamma(0) = q, \quad \frac{d\gamma}{dt}(0) = v.$$

Then the point $\gamma(1) \in M$ will be denoted by $\exp_q(v)$ and called the *exponential* of the tangent vector $v$. The geodesic $\gamma$ can thus be described by the formula

$$\gamma(t) = \exp_q(tv).$$

Lemma 4.13 says that $\exp_q(v)$ is defined provided that $\|v\|$ is small enough. In general, $\exp_q(v)$ is not defined for large vectors $v$. However, if defined at all, $\exp_q(v)$ is always uniquely defined.

**Definition 4.12.** The manifold $M$ is *geodesically complete* if $\exp_q(v)$ is defined for all $q \in M$ and all vectors $v \in T_qM$.

This is clearly equivalent to the following requirement: For every geodesic segment $\gamma_0 : [a,b] \to M$, it should be possible to extend $\gamma_0$ to an infinite geodesic

$$\gamma : \mathbb{R} \to M.$$

Now we consider $F : TM \to M \times M$ by $F(p,v) = \exp_p(v)$. We give $TM$ the local coordinates as follows. If $p \in U$ where $(U, u^1, \ldots, u^n)$ is a local chart, $v = t^1\partial_1 + \cdots + t^n\partial_n \in T_pM$, then we give $(p,v)$ the coordinate $(u^1, \ldots, u^n, t^1, \ldots, t^n)$. If $(U, u^i)$ is a local chart in $M$, then we denote $(U \times U, u_1^i, u_2^i)$ a local chart in $M \times M$.

Therefore we can calculate the Jacobian matrix of $F$. Note that

$$\left.\frac{\partial(u_k^i \circ F)}{\partial u^j}\right|_{(p,v)} = \frac{\partial u^i}{\partial u^j} = \delta_i^j,$$

and

$$(d\exp_p)_0 v = \left.\frac{d}{dt}\right|_{t=0} \exp_p(tv) = \gamma_v'(0) = v,$$

where

$$v \in T_pM = T_0(T_pM).$$

Thus the Jacobian of $F$ is non-singular, so $F$ is a local diffeomorphism. Hence we may assume that the first neighborhood $V'$ consists of all pairs $(q,v)$ such that $q$ belongs to a given neighborhood $U'$ of $p$ and such that $\|v\| < \varepsilon$. Choose a smaller neighborhood $W$ of $p$ so that $F(V') \supset W \times W$. Then we have proven the following.

**Lemma 4.14.** *For each $p \in M$ there exists a neighborhood $W$ and a number $\varepsilon > 0$ so that:*

1. *Any two points of $W$ are joined by a unique geodesic in $M$ of length $< \varepsilon$.*
2. *This geodesic depends smoothly upon the two points. (i.e., if $t \mapsto \exp_{q_1}(tv)$, $0 \le t \le 1$, is the geodesic joining $q_1$ and $q_2$, then the pair $(q_1, v) \in TM$ depends smoothly on $(q_1, q_2)$.*
3. *For each $q \in W$ the map $\exp_q$ maps the open $\varepsilon$-ball in $T_qM$ diffeomorphically onto an open set $U_q \supset W$.*

Let $q = y(0)$ and let $U_q$ be as in Lemma 4.14.

**Lemma 4.15.** *In $U_q$, the geodesics through $q$ are the orthogonal trajectories of hypersurfaces $\exp_q(v)$ where $v \in T_qM$, $\|v\| =$ constant.*

*proof.* Let $t \mapsto v(t)$ denote any curve in $T_qM$ with $\|v(t)\| = 1$. We must show that the corresponding curves

$$t \mapsto \exp_q(r_0 v(t))$$

in $U_q$, where $0 < r_0 < \varepsilon$, are orthogonal to the radial geodesics

$$r \mapsto \exp_q(rv(t_0)).$$

In terms of the parametrized surface $f$ given by

$$f(r,t) = \exp_q(rv(t)), \quad 0 \le r < \varepsilon,$$

we must prove that

$$\left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = 0$$

for all $(r,t)$. Now,

$$\frac{\partial}{\partial r}\left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle = \left\langle \frac{D}{\partial r}\frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle + \left\langle \frac{\partial f}{\partial r}, \frac{D}{\partial r}\frac{\partial f}{\partial t} \right\rangle.$$

The first expression on the right is zero since the curves $r \mapsto f(r,t)$ are geodesics. The second expression is equal to

$$\left\langle \frac{\partial f}{\partial r}, \frac{D}{\partial t}\frac{\partial f}{\partial t} \right\rangle = \frac{1}{2}\frac{\partial}{\partial t}\left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial r} \right\rangle = 0,$$

due to symmetric-ness and since $\left\|\frac{\partial f}{\partial r}\right\| = \|v(t)\| = 1$. Therefore the quantity $\left\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \right\rangle$ is independent of $r$. But for $r = 0$ we have

$$f(0,t) = \exp_q(0) = q.$$

hence

$$\frac{\partial f}{\partial t}(0,t) = 0,$$

so

$$\left\langle \frac{\partial f}{\partial t}, \frac{\partial f}{\partial r} \right\rangle = 0$$

everywhere.

□

**Lemma 4.16.** *Now consider any piecewise smooth curve $\omega : [a, b] \to U_q \backslash \{q\}$. Each point $\omega(t)$ can be expressed uniquely in the form $\exp_q(r(t)v(t))$ with $0 < r(t) < \varepsilon$, and $\|v(t)\| = 1$, $v(t) \in T_qM$. Then the length $\int_a^b \left\| \frac{d\omega}{dt} \right\| dt$ is greater than or equal to $|r(b) - r(a)|$, where equality holds only if the function $r(t)$ is monotone, and the function $v(t)$ is constant.*

*Thus the shortest path joining two concentric spherical shells around $q$ is a radial geodesic.*

*proof.* Let $f(r, t) = \exp_q(rv(t))$, so that $\omega(t) = f(r(t), t)$. Then

$$\frac{d\omega}{dt} = \frac{\partial f}{\partial r} r'(t) + \frac{\partial f}{\partial t}.$$

Since the two vectors on the right are mutually orthogonal (by Lemma 4.15), and since $\left\| \frac{\partial f}{\partial r} \right\| = 1$, this gives

$$\left\| \frac{d\omega}{dt} \right\|^2 = |r'(t)|^2 + \left\| \frac{\partial f}{\partial t} \right\|^2 \geq |r'(t)|^2$$

where equality holds only if $\frac{\partial f}{\partial t} = 0$; hence only if $\frac{\partial v}{\partial t} = 0$. Thus

$$\int_a^b \left\| \frac{d\omega}{dt} \right\| dt \geq \int_a^b |r'(t)| dt \geq |r(b) - r(a)|$$

where equality holds only if $r(t)$ is monotone and $v(t)$ is constant. This completes the proof. □

With these two lemmas, we can prove the following theorem.

**Theorem 4.17.** *Let $W$ and $\varepsilon$ be as in Lemma 4.14. Let $\gamma : [0, 1] \to M$ be the geodesic of length $< \varepsilon$ joining two points of $W$, and let $\omega : [0, 1] \to M$ be any other piecewise smooth path joining the same two points. Then,*

$$\int_0^1 \left\| \frac{d\gamma}{dt} \right\| dt \leq \int_0^1 \left\| \frac{d\omega}{dt} \right\| dt,$$

*where equality can hold only if the point set $\omega([0, 1])$ coincides with $\gamma([0, 1])$.*

*proof.* Consider any piecewise smooth path $\omega$ from $q$ to a point $q' = \exp_q(rv) \in U_q$, where $0 < r < \varepsilon$ and $\|v\| = 1$. Then for any $\delta > 0$, the path $\omega$ must contain a segment joining the spherical shell of radius $\delta$ to the spherical shell of radius $r$, and lying between these two shells. The length of this segment will be $\geq r - \delta$ by Lemma 4.16; hence letting $\delta$ tend to 0, the length of $\omega$ will be $\geq r$. If $\omega([0, 1])$ does not coincide with $\gamma([0, 1])$, then we easily obtain a strict inequality. This completes the proof. □

**Corollary 4.18.** *Suppose that a path $\omega : [0, \ell] \to M$, parametrized by arc-length, has length less than or equal to the length of any other path from $\omega(0)$ to $\omega(\ell)$. Then $\omega$ is a geodesic.*

*proof.* Consider any segment of $\omega$ lying within an open set $W$, as above, and having length $< \varepsilon$. This segment must be a geodesic by Theorem 4.17. Hence the entire path $\omega$ is a geodesic. □

**Definition 4.13.** A geodesic $\gamma : [a, b] \to M$ will be called *minimal* if its length is less than or equal to the length of any other piecewise smooth path joining its endpoints.

Theorem 4.17 asserts that any sufficiently small segment of a geodesic is minimal. On the other hand, a long geodesic may not be minimal. For example, we will see shortly that a great circle arc on the unit sphere is a geodesic. If such an arc has length greater than $\pi$, it is certainly not minimal.

**Definition 4.14.** Define the distance $\rho(p, q)$ between two points $p, q \in M$ to be the greatest lower bound for the arc-lengths of piecewise smooth paths joining these points. This clearly makes $M$ into a metric space. It follows easily from Theorem 4.17 that this metric is compatible with the usual topology of $M$ (since small open metric ball is diffeomorphic to an open ball in $T_pM$ (by Lemma 4.14 and Theorem 4.17), hence diffeomorphic to an open neighborhood in $M$).

**Corollary 4.19.** *Given a compact set $K \subset M$, there exists a number $\delta > 0$ so that any two points of $K$ with distance less than $\delta$ are joined by a unique geodesic of length less than $\delta$. Furthermore, this geodesic is minimal and depends differentiably on its endpoints.*

*proof.* Cover $K$ by open sets $W_\alpha$, as in Lemma 4.14, and let $\delta$ be small enough so that any two points in $K$ with distance less than $\delta$ lie in a common $W_\alpha$. This completes the proof. $\square$

Recall that the manifold $M$ is geodesically complete if every geodesic segment can be extended infinitely.

**Theorem 4.20** (Hopf and Rinow)**.** *If $M$ is geodesically complete, then any two points can be joined by a minimal geodesic.*

*proof.* Given $p, q \in M$ with distance $r > 0$, choose a neighborhood $U_p$ as in Lemma 4.14. Let $S \subseteq U_p$ denote a spherical shell of radius $\delta < \varepsilon$ about $p$, i.e., $S := \{\exp_p(\delta x) \mid x \in T_pM, \|x\| = 1\}$. Since $S$ is compact, there exists a point

$$p_0 = \exp_p(\delta v), \quad \|v\| = 1,$$

on $S$ for which the distance to $q$ is minimized. We will prove that

$$\exp_p(rv) = q.$$

This implies that the geodesic segment $t \mapsto \gamma(t) = \exp_p(tv)$, $0 \le t \le r$, is actually a minimal geodesic from $p$ to $q$.

The proof will amount to showing that a point which moves along the geodesic $\gamma := \gamma_v$ must get closer and closer to $q$. In fact for each $t \in [\delta, r]$ we will prove that

$$\rho(\gamma(t), q) = r - t. \tag{$1_t$}$$

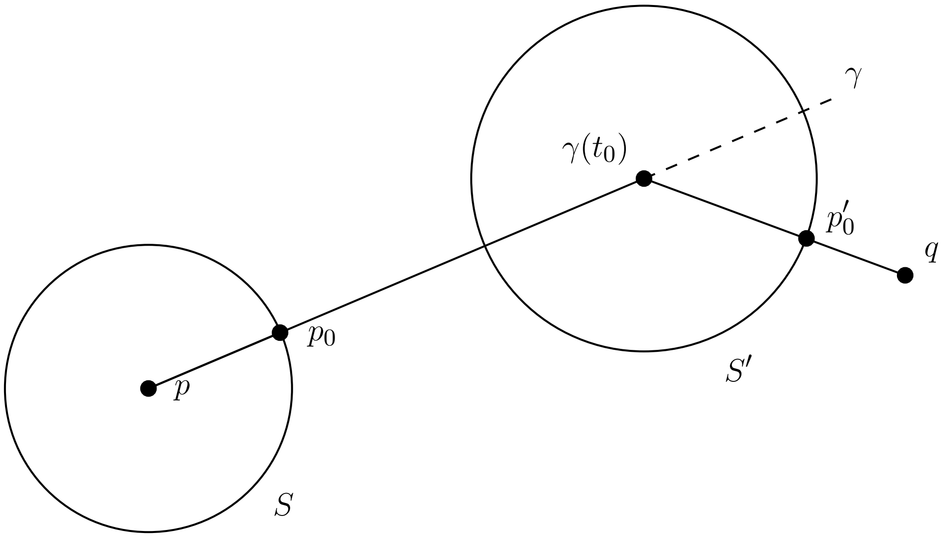
This identity, for $t = r$, will complete the proof.

First we will show that the equality $(1_\delta)$ is true. Since every path from $p$ to $q$ must pass through $S$, we have

$$\rho(p, q) = \inf_{s \in S}(\rho(p, s) + \rho(s, q)) = \delta + \rho(p_0, q).$$

Therefore $\rho(p_0, q) = r - \delta$. Since $p_0 = \gamma(\delta)$, this proves $(1_\delta)$.

Let $t_0 \in [\delta, r]$ denote the supremum of those numbers $t$ for which $1_t$ is true. Then by continuity the equality $(1_{t_0})$ is true also. If $t_0 < r$ we will obtain a contradiction. Let $S'$ denote a small spherical shell of radius $\delta'$ about the point $\gamma(t_0)$; and let $p_0' \in S'$ be a point of $S'$ with minimum distance from $q$. Check the figure below.




Then

$$\rho(\gamma(t_0), q) = \inf_{s \in S'} (\rho(\gamma(t_0), s) + \rho(s, q)) = \delta' + \rho(p_0', q),$$

hence

$$\rho(p_0', q) = (r - t_0) - \delta'.$$

We claim that $p_0'$ is equal to $\gamma(t_0 + \delta')$. In fact, the triangle inequality states that

$$\rho(p, p_0') \geq \rho(p, q) - \rho(p_0', q) = t_0 + \delta'.$$

But a path of length precisely $t_0 + \delta'$ from $p$ to $p_0'$ can be obtained by following $\gamma$ from $p$ to $\gamma(t_0)$, and then following a minimal geodesic from $\gamma(t_0)$ to $p_0'$, which is illustrated in the figure. Since this broken geodesic has minimal length, it follows from Corollary 4.18 that it is an (unbroken) geodesic, and hence coincides with $\gamma$.

Thus $\gamma(t_0 + \delta') = p_0'$. Now the equality becomes

$$\rho(\gamma(t_0 + \delta'), q) = r - (t_0 + \delta'). \qquad (1_{t_0+\delta'})$$

This contradicts the definition of $t_0$, and completes the proof. □

**Corollary 4.21.** *If $M$ is geodesically complete, then every bounded subset of $M$ has compact closure. Consequently, $M$ is complete as a metric space (i.e., every Cauchy sequence converges).*

*proof.* If $X \subset M$ has diameter $d$, then for any $p \in X$, the map $\exp_p : T_pM \to M$ maps the disk of radius $d$ in $T_pM$ onto a compact subset of $M$ which (making use of Theorem 4.20) contains $X$. Hence the closure of $X$ is compact.

It is a well-known theorem that this means completeness, but we still write the proof here. Pick any Cauchy sequence $\{x_n\}$ and since it is in a metric space, it is bounded, so its closure contained in a closed ball, hence compact. In metric space, compactness is sequential compactness, so $\{x_n\}$ admits at least one limit point, hence it has exactly one limit since it is Cauchy. $\square$

Conversely, if $M$ is complete as a metric space, then we can use Lemma 4.14 to prove that $M$ is geodesically complete. Therefore, we have the following theorem.

**Theorem 4.22.** *A Riemannian manifold $M$ is geodesically complete if and only if it is complete as a metric space. Henceforth, we will not distinguish these two concepts, and just call such a manifold complete.*

# Chapter 5

# Morse Theory on Path Spaces

In order to understand the homotopy of a manifold, we do Morse theory on its path space.We realize it by considering the energy functional.

We will define, in analogy with Chapter 1, the differential and Hessian of energy functional $E$ and will see that the "critical points" of $E$ is geodesics and "degenerate critical points" are Jacobi fields.

In this chapter, we assume $M$ is a manifold with Riemannian structure $\langle \cdot, \cdot \rangle$ (or Riemannian metric $g$) and the symmetric connection compatible with its Riemannian structure.

## 5.1 The Path Space of a Smooth Manifold

**Definition 5.1** (Piecewise Smooth Path)**.** , We call $\omega : [0,1] \to M$ a *piecewise smooth path* if there exists a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of $[0,1]$ such that $\omega_{[t_i, t_{i+1}]}$ is $C^\infty$.

We define $\Omega(M; p, q)$ to be the set of all piecewise smooth path from $p$ to $q$ in $M$. For simplicity, we often denote it by $\Omega(M)$ or $\Omega$.

**Definition 5.2** (Tangent Space)**.** We define $T_\omega\Omega$ to be the *tangent space* at $\omega$ of $\Omega$. It will be meant the vector space of all vector fields $W$ along $\omega$ with $W_0 = W_1 = 0$.

**Definition 5.3** (Variation)**.** A *variation* of $\omega$ keeping endpoints fixed is a function

$$\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$$

for some $\varepsilon > 0$, such that

1. $\bar{\alpha}(0) = \omega$,
2. There is a subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$ of $[0,1]$ so that the map

   $$\alpha : (-\varepsilon, \varepsilon) \times [0,1] \to M$$

   defined by $\alpha(u,t) = \overline{\alpha}(u)(t)$ is $C^\infty$ on each strip $(-\varepsilon, \varepsilon) \times [t_{i-1}, t_i]$, $i = 1, \ldots, k$.

We will use either $\alpha$ or $\bar{\alpha}$ to refer to the variation. More generally, if, in the above definition, $(-\varepsilon, \varepsilon)$ is replaced by a neighborhood $U$ of 0 in $\mathbb{R}^n$, then $\alpha$ (or $\bar{\alpha}$) is called an $n$-parameter variation of $\omega$.

**Definition 5.4** (Velocity Vector). $\bar{\alpha}$ may be viewed as a smooth path in $\Omega$. Its *velocity vector* at 0 is defined as a vector field $W$ along $\omega$ given by

$$\left(\frac{d\bar{\alpha}}{du}(0)\right)_t = W_t := \frac{\partial \alpha}{\partial u}(0,t).$$

Clearly we have $W \in T_\omega \Omega$.

Given any tangent vector $W \in T_\omega\Omega$, set

$$\bar{\alpha}(u)(t) = \exp_{\omega(t)}(uW_t), u \in (-\varepsilon, \varepsilon),$$

then we have

$$\begin{aligned} \bar{\alpha}(0)(t) &= \exp_{\omega(t)}(0) = \omega(t), \\ \left(\frac{d\bar{\alpha}}{du}(0)\right)_t &= \frac{\partial\alpha}{\partial u}(0,t) = \left(\frac{\partial}{\partial u}\exp_{\omega(t)}(uW_t)\right)(0,t) = W_t. \end{aligned}$$

Thus $\bar{\alpha}$ is a curve satisfying $\bar{\alpha}(0) = \omega$ and $\bar{\alpha}'(0) = W$. Now, let $F : \Omega \to \mathbb{R}$, we can define its differential $F_* \in \mathrm{Hom}(T_\omega\Omega, T_{F(\omega)}\mathbb{R})$ as follows.

**Definition 5.5.** Given any $W \in T_\omega\Omega$, define

$$F_*W := \frac{d(F\circ\bar{\alpha})}{du}(0)\cdot \left.\frac{d}{dt}\right|_{F(\omega)},$$

where $\bar{\alpha}(0) = \omega$ and $\bar{\alpha}'(0) = W$.

**Definition 5.6** (Critical Path). A path $\omega$ is a *critical path* for function $F : \Omega \to \mathbb{R}$ if for each variation $\bar{\alpha}$ of $\omega$, we have

$$\frac{d(F\circ\bar{\alpha})}{du}(0) = 0.$$

For example, if $F$ takes on its minimum at $\omega_0$ and if $\frac{d(F\circ\bar{\alpha})}{du}$ is well-defined for each $\bar{\alpha}$, then $\omega_0$ is a critical path of $F$.

## 5.2 The Energy of a Path

We introduce a very important functional called energy in this section and show that its critical paths are precisely geodesics.

**Definition 5.7** (Energy). For $\omega \in \Omega$, we define its *energy* from $a$ to $b$ where $0 \le a < b \le 1$ to be

$$E_a^b(\omega) := \int_a^b \left\|\frac{d\omega}{dt}\right\|^2 dt.$$

We will write $E$ for $E_0^1$.

**Proposition 5.1.** *Let $L_a^b(\omega)$ the arc length of $\omega$ from $a$ to $b$, then*

$$(L_a^b)^2 \le (b-a)E_a^b,$$

*where equality holds if and only if $\|d\omega/dt\|$ is constant.*

*proof.* Let $f = 1$, $g(t) = \|d\omega/dt|_t\|$, then by Cauchy-Swartz inequality

$$\left(\int_{[a,b]} fg\right)^2 \leq \left(\int_{[a,b]} f^2\right)\left(\int_{[a,b]} g^2\right),$$

the proposition follows. □

Now suppose that there exists a minimal geodesic $\gamma$ from $p = \omega(0)$ to $q = \omega(1)$. Then

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

Here the equality $L(\gamma)^2 = L(\omega)^2$ can hold only if $\omega$ is also a minimal geodesic, possibly reparametrized. On the other hand, the equality $L(\omega)^2 = E(\omega)$ can hold only if the parameter is proportional to arc-length along $\omega$. This proves that $E(\gamma) < E(\omega)$ unless $\omega$ is also a minimal geodesic. Thus,

**Proposition 5.2.** *Let $M$ be a complete Riemannian manifold and let $p, q \in M$ have distance $d$. Then the energy function*

$$E : \Omega(M; p, q) \to \mathbb{R}$$

*takes on its minimum $d^2$ precisely on the set of minimal geodesics from $p$ to $q$.*

Let $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ be a variation of $\omega$ and $W_t = \frac{\partial \alpha}{\partial u}(0, t)$ be the associated variation vector field. Furthermore, we define (if exist) $V = d\omega/dt$, $A = \frac{D}{dt}\frac{d\omega}{dt}$, and $\Delta_t V = V_{t+} - V_{t-}$. Of course $\Delta_t V \neq 0$ and $V_t$ or $A_t$ cannot be defined for only finite number of values of $t$. We let $V_t$ or $A_t$ be zero if they are not defined.

**Theorem 5.3** (First Variation Formula)**.**

$$\frac{1}{2}\frac{d(E \circ \bar{\alpha})}{du}(0) = -\sum_t \langle W_t, \Delta_t V\rangle - \int_0^1 \langle W_t, A_t\rangle dt.$$

*proof.* Since

$$\frac{\partial}{\partial u}\left\langle \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\right\rangle = 2\left\langle \frac{D}{du}\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\right\rangle$$

and

$$\frac{D}{du}\frac{\partial}{\partial t} = \frac{D}{dt}\frac{\partial}{\partial u},$$

we have

$$\frac{dE\bar{\alpha}}{du} = \frac{d}{du}\int_0^1 \left\langle \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial t}\right\rangle dt = 2\int_0^1 \left\langle \frac{D}{dt}\frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right\rangle dt.$$

Now we choose a partition $\{t_i\}_0^k$ of $[0, 1]$ such that $\alpha$ is smooth in each strip $(-\varepsilon, \varepsilon) \times [t_{i-1}, t_i]$. Since

$$\frac{\partial}{\partial t}\left\langle \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right\rangle = \left\langle \frac{D}{dt}\frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right\rangle + \left\langle \frac{\partial \alpha}{\partial u}, \frac{D}{dt}\frac{\partial \alpha}{\partial t}\right\rangle,$$

we have

$$\int_{t_{i-1}}^{t_i} \left\langle \frac{D}{dt}\frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right\rangle dt = \left\langle \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right\rangle\Bigg|_{t_{i-1}+}^{t_i-} - \int_{t_{i-1}}^{t_i} \left\langle \frac{\partial \alpha}{\partial u}, \frac{D}{dt}\frac{\partial \alpha}{\partial t}\right\rangle dt.$$

Evaluating the above formula at $(0,t)$ and adding up it for $i=1,\dots,k$ gives

$$\frac{1}{2}\frac{d(E\circ\bar{\alpha})}{du}(0)=-\sum_t\langle W_t,\Delta_t V\rangle-\int_0^1\langle W_t,A_t\rangle dt.$$

□

**Corollary 5.4.** *The path $\omega$ is a critical point for the function $E$ if and only if $\omega$ is a geodesic.*

*proof.* Clearly, a geodesic is a critical point, since $\Delta_t V=0$ and $A_t=0$. Let $\omega$ be a critical point. There is a variation of $\omega$ with $W_t=f(t)A_t$ where $f(t)$ is positive except that it vanishes at the $t_i$. Then

$$\frac{1}{2}\frac{dE\bar{\alpha}}{du}(0)=-\int_0^1 f(t)\langle A(t),A(t)\rangle\,dt.$$

This is zero if and only if $A_t=0$ for all $t$. Hence each $\omega|_{[t_i,t_{i+1}]}$ is a geodesic.

Now pick a variation such that $W_{t_i}=\Delta t_i V$. Then

$$\frac{1}{2}\frac{dE\bar{\alpha}}{du}(0)=-\sum\langle\Delta_{t_i}V,\Delta_{t_i}V\rangle.$$

If this is zero, then all $\Delta_t V$ are 0, and $\omega$ is differentiable of class $C^1$, even at the points $t_i$. Now it follows from the uniqueness theorem for differential equations that $\omega$ is $C^\infty$ everywhere: thus $\omega$ is an unbroken geodesic. □

## 5.3 The Hessian of the Energy Function at a Critical Path

Recall that we define the Hessian of $f\in C^\infty(M)$ as

$$f_{**}:T_pM\times T_pM\to\mathbb{R}.$$

Explicitly, given $X_1,X_2\in T_pM$, choose $\alpha:U\subseteq\mathbb{R}^2\to M$,where $(0,0)\in U$ and

$$\alpha(0,0)=p,\quad\frac{\partial\alpha}{\partial u_1}(0,0)=X_1,\quad\frac{\partial\alpha}{\partial u_2}(0,0)=X_2,$$

then

$$f_{**}(X_1,X_2)=\frac{\partial^2(f\circ\alpha)}{\partial u_1\partial u_2}(0,0).$$

It is equivalent to the definition that

$$f_{**}(X_1,X_2)=\tilde{X}_1(\tilde{X}_2(f)).$$

This suggest defining $E_{**}$ as follows.

**Definition 5.8.** Let $\gamma$ be a geodesic. Let $W_1,W_2\in T_\gamma\Omega$, choose a 2-parameter variation $\bar{\alpha}:U\subseteq\mathbb{R}\to\Omega$ of $\gamma$, such that

$$\bar{\alpha}(0,0)=\gamma,\quad\frac{\partial\alpha}{\partial u_1}(0,0,t)=W_1(t),\quad\frac{\partial\alpha}{\partial u_2}(0,0,t)=W_2(t).$$

We define

$$E_{**}(W_1,W_2):=\frac{\partial^2(E\circ\bar{\alpha})}{\partial u_1\partial u_2}(0,0).$$

The following theorem is needed to prove $E_{**}$ is well-defined.

**Theorem 5.5** (Second Variation Formula). *With notations above, we have*

$$\frac{1}{2}\frac{\partial^2(E\circ\bar{\alpha})}{\partial u_1\partial u_2} = -\sum_t \left\langle W_2(t), \Delta_t \frac{DW_1}{dt}\right\rangle - \int_0^1 \left\langle W_2(t), \frac{D^2W_1}{dt^2}(t) + (R(V,W_1)V)_t\right\rangle dt,$$

*where* $V = \frac{d\gamma}{dt}$ *is the velocity vector field, and*

$$\Delta_t \frac{DW_1}{dt} := \frac{DW_1}{dt}(t+) - \frac{DW_1}{dt}(t-).$$

*proof.* By Theorem 5.3,

$$\frac{1}{2}\frac{\partial(E\circ\bar{\alpha})}{\partial u_2}(0) = -\sum_t \left\langle \frac{\partial\alpha}{\partial u_2}, \Delta_t\frac{\partial\alpha}{\partial t}\right\rangle - \int_0^1 \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{dt}\frac{\partial\alpha}{\partial t}\right\rangle dt.$$

Therefore,

$$\begin{aligned}\frac{1}{2}\frac{\partial^2(E\circ\bar{\alpha})}{\partial u_1\partial u_2} = &-\sum_t \left\langle \frac{D}{du_1}\frac{\partial\alpha}{\partial u_2}, \Delta_t\frac{\partial\alpha}{\partial t}\right\rangle - \sum_t \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{du_1}\Delta_t\frac{\partial\alpha}{\partial t}\right\rangle \\ &- \int_0^1 \left\langle \frac{D}{du_1}\frac{\partial\alpha}{\partial u_2}, \frac{D}{dt}\frac{\partial\alpha}{\partial t}\right\rangle dt - \int_0^1 \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{du_1}\frac{D}{dt}\frac{\partial\alpha}{\partial t}\right\rangle dt.\end{aligned}$$

Let us evaluate this expression for $(u_1, u_2) = (0, 0)$. Since $\gamma = \bar{\alpha}(0, 0)$ is an unbroken geodesic, we have

$$\Delta_t\frac{\partial\alpha}{\partial t} = 0, \quad \frac{D}{dt}\frac{\partial\alpha}{\partial t} = 0,$$

so that the first and third terms are zero. Rearranging the second term, we obtain

$$\frac{1}{2}\frac{\partial^2 E}{\partial u_1\partial u_2}(0,0) = -\sum_t \left\langle W_2, \Delta_t\frac{D}{dt}W_1\right\rangle - \int_0^1 \left\langle W_2, \frac{D}{du_1}\frac{D}{dt}V\right\rangle dt.$$

Now, note that

$$\frac{D}{du_1}\frac{D}{dt}V - \frac{D}{dt}\frac{D}{du_1}V = R\left(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial u_1}\right)V = R(V, W_1)V$$

and

$$\frac{D}{dt}\frac{D}{du_1}V = \frac{D}{dt}\frac{D}{dt}\frac{\partial\alpha}{\partial u_1} = \frac{D^2}{dt^2}W_1,$$

so

$$\frac{1}{2}\frac{\partial^2 E}{\partial u_1\partial u_2}(0,0) = -\sum_t \left\langle W_2, \Delta_t\frac{DW_1}{dt}\right\rangle - \int_0^1 \left\langle W_2, \frac{D^2W_1}{dt^2} + R(V,W_1)V\right\rangle dt.$$

□

**Corollary 5.6.** *The expression* $E_{**}(W_1, W_2) = \dfrac{\partial^2 E\bar{\alpha}}{\partial u_1\partial u_2}(0,0)$ *is a well-defined symmetric and bilinear function.*

*proof.* The second variation formula shows that $\dfrac{\partial^2 E\bar{\alpha}}{\partial u_1 \partial u_2}(0,0)$ depends only on the variation vector fields $W_1$ and $W_2$, so that $E_{**}(W_1, W_2)$ is well-defined. This formula also shows that $E^{**}$ is bilinear. The symmetry property $E_{**}(W_1, W_2) = E_{**}(W_2, W_1)$ is not at all obvious from the second variation formula; but does follow immediately from the symmetry property $\dfrac{\partial^2 E\bar{\alpha}}{\partial u_1 \partial u_2} = \dfrac{\partial^2 E\bar{\alpha}}{\partial u_2 \partial u_1}$. □

Simple calculation shows that

$$E_{**}(W) = \frac{d^2 E(\overline{\alpha}(u))}{du^2}(0),$$

where $\alpha : (-\varepsilon, \varepsilon) \to \Omega$ is a variation of a geodesic $\gamma$ with $d\alpha/du(0) = W$.

**Proposition 5.7.** *If $\gamma$ is a minimal geodesic from $p$ to $q$, then the bilinear pairing $E_{**}$ is positive semi-definite. Hence the index $\lambda$ of $E_{**}$ is zero.*

*proof.* The inequality $E(\bar{\alpha}(u)) \geq E(\gamma) = E(\bar{\alpha}(0))$ implies that $\dfrac{d^2 E(\overline{\alpha}(u))}{du^2}$, evaluated at $u = 0$, is $\geq 0$. Hence $E_{**}(W, W) \geq 0$ for all $W$. □

## 5.4 Jacobi Fields: The Null Space of $E_{**}$

In this section we show the space of all Jacobi fields which vanish at two endpoints is the null space of $E_{**}$. Furthermore, we show that Jacobi fields along a geodesic are exactly the vector fields produced by moving geodesics around.

**Definition 5.9** (Jacobi Fields). A vector field $\mathbf{J}$ along a geodesic $\gamma$ is called a *Jacobi field* if it satisfies

$$\frac{D^2 \mathbf{J}}{dt^2} + R(V, \mathbf{J})V = 0,$$

where $V = d\gamma/dt$ is the velocity vector field. Choosing $n$ orthonormal parallel vector fields $P_i$ along $\gamma$ and write $\mathbf{J} = \sum f^i P_i$, we turn the differential equation to $n$ second order linear ODEs

$$\frac{d^2 f^i}{dt^2} + \sum_i a^i_j(t) f^i(t) = 0,$$

where

$$a^i_j = \langle R(V, P_j)V, P_j \rangle.$$

Thus the Jacobi equation has $2n$ linearly independent solutions, each of which can be defined throughout $\gamma$.

A given Jacobi field $\mathbf{J}$ is determined by initial value $\mathbf{J}(0)$ and $D\mathbf{J}/dt(0)$

**Definition 5.10.** Let $p = \gamma(a)$ and $q = \gamma(b)$ be two points on the geodesic $\gamma$, with $a \neq b$. $p$ and $q$ are *conjugate* along $\gamma$ if there exists a non-zero Jacobi field $\mathbf{J}$ along $\gamma$ which vanishes for $t = a$ and $t = b$.

The *multiplicity* of $p$ and $q$ as conjugate points is equal to the dimension of the vector space consisting of all such Jacobi fields.

Geometrically, the multiplicity of two conjugate points $p, q$ indicates "how many" geodesics with the same endpoints $p, q$ exist in a small neighborhood of a geodesic. Corollary 5.10 gives an example to illustrate it.

Now let $\gamma$ be a geodesic in $\Omega = \Omega(M; p, q)$. Recall that the null space of the Hessian

$$E_{**} : T_\gamma\Omega \times T_\gamma\Omega \to \mathbb{R}$$

is the vector space consisting of those $W_1 \in T_\gamma\Omega$ such that $E_{**}(W_1, W_2) = 0$ for all $W_2$. The nullity $\nu$ of $E_{**}$ is equal to the dimension of this null space. $E_{**}$ is degenerate if $\nu > 0$.

**Theorem 5.8.** *A vector field $W_1 \in T_\gamma\Omega$ belongs to the null space of $E_{**}$ if and only if $W_1$ is a Jacobi field. Hence $E_{**}$ is degenerate if and only if the endpoints $p$ and $q$ are conjugate along $\gamma$ and the nullity of $E_{**}$ is equal to the multiplicity of $p$ and $q$ as conjugate points.*

*proof.* If $\mathbf{J}$ is a Jacobi field which vanishes at $p$ and $q$, then $\mathbf{J}$ certainly belongs to $T_\gamma\Omega$. The second variation formula (Theorem 5.5) states that

$$-\frac{1}{2}E_{**}(\mathbf{J}, W_2) = \sum_t \langle W_2(t), 0\rangle + \int_0^1 \langle W_2, 0\rangle dt = 0.$$

Hence $\mathbf{J}$ belongs to the null space.

Conversely, suppose that $W_1$ belongs to the null space of $E_{**}$. Choose a subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$ of $[0,1]$ so that $W_1|_{[t_{i-1},t_i]}$ is smooth for each $i$. Let $f : [0,1] \to [0,1]$ be a smooth function which vanishes for the parameter values $t_0, t_1, \ldots, t_k$ and is positive otherwise; and let

$$W_2(t) := f(t)\left(\frac{D^2W_1}{dt^2} + R(V, W_1)V\right)_t,$$

then

$$-\frac{1}{2}E_{**}(W_1, W_2) = \sum_t 0 + \int_0^1 f(t)\left\|\left(\frac{D^2W_1}{dt^2} + R(V, W_1)V\right)_t\right\|^2 dt = 0,$$

so

$$\left(\frac{D^2W_1}{dt^2} + R(V, W_1)V\right)_t = 0$$

when $t \neq t_i$, which means $W_1|_{[t_{i-1},t_i]}$ is Jacobi field.

Now let $W_2' = \Delta_{t_i}\dfrac{DW_1}{dt}$, then

$$-\frac{1}{2}E_{**}(W_1, W_2') = \sum_t \left\|\Delta_t\frac{DW_1}{dt}\right\|^2 + 0 = 0,$$

hence $\frac{DW_1}{dt}$ has no jumps. But a solution $W_1$ of the Jacobi equation is completely determined by the vectors $W_1(t_i)$ and $\dfrac{DW_1}{dt}(t_i)$. Thus it follows that the $k$ Jacobi fields $W_1|_{[t_{i-1},t_i]}$, for $i = 1, \ldots, k$, fit together to give a Jacobi field $W_1$ which is $C^\infty$-differentiable throughout the entire unit interval. This completes the proof. $\square$

**Remark.** *The proof of Theorem 5.8 is similar to the proof of Corollary 5.4, both constructing two vector fields with zero or nonzero value at connections and using variation formula.*

It follows that the nullity $\nu$ of $E_{**}$ is always finite, for there are only finitely many linearly independent Jacobi fields along $\gamma$. Actually, $\nu < n$. $\nu \leq n$ follows from that fact that if $\mathbf{J}(0) = 0$ is determined, there are only $n$ linearly independent Jacobi fields along $\gamma$. We will construct one example of a Jacobi field which vanishes for $t = 0$, but not for $t = 1$. This will imply that $\nu < n$. In fact, let $\mathbf{J}_t = tV_t$ where $V = \frac{d\gamma}{dt}$ denotes the velocity vector field. Then

$$\frac{D\mathbf{J}}{dt} = V + t\frac{DV}{dt} = V$$

hence

$$\frac{D^2\mathbf{J}}{dt^2} = 0.$$

Furthermore, $R(V, \mathbf{J})V = tR(V, V)V = 0$ since $R$ is skew symmetric in the first two variables. Thus $\mathbf{J}$ satisfies the Jacobi equation. Since $\mathbf{J}(0) = 0$ and $\mathbf{J}(1) \neq 0$, this completes the proof.

**Proposition 5.9.** *Let $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ be a variation (not necessarily keeping the end-points fixed) of geodesic $\gamma$ such that each $\bar{\alpha}(u)$ is a geodesic. Then the vector field*

$$W_t := \frac{\partial \alpha}{\partial u}(0, t)$$

*is a Jacobi field.*

*proof.* It suffices to show that

$$\frac{D^2 W_t}{dt} + R(V, W_t)V = 0.$$

Note that

$$\begin{aligned}
\frac{D}{dt}\frac{D}{dt}\frac{\partial \alpha}{\partial u} &= \frac{D}{dt}\frac{D}{du}\frac{\partial \alpha}{\partial t} \\
&= \frac{D}{du}\frac{D}{dt}\frac{\partial \alpha}{\partial t} - R\left(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial u}\right)\frac{\partial \alpha}{\partial t} \\
&= -R\left(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial u}\right)\frac{\partial \alpha}{\partial t} & \frac{D}{dt}\frac{\partial \alpha}{\partial t} = 0 \text{ since } \bar{\alpha}(u) \text{ is geodesic} \\
&= -R(V, W_t)V.
\end{aligned}$$

This completes the proof. ∎

Thus one way of obtaining Jacobi fields is to move geodesics around.

**Corollary 5.10.** *Suppose that $p$ and $q$ are antipodal points on the unit sphere $S^n$, and let $\gamma$ be a great circle arc from $p$ to $q$. Then we will see that $p$ and $q$ are conjugate with multiplicity $n-1$. Thus the nullity $\nu$ of $E_{**}$ takes its largest possible value.*

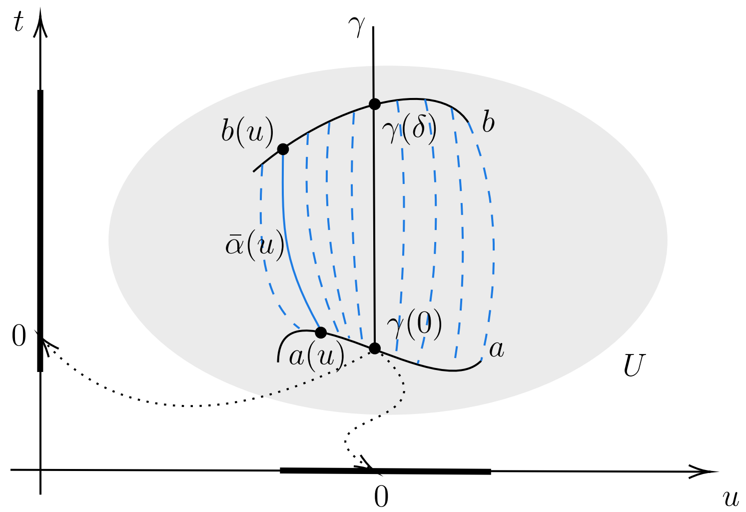
*proof.* Rotating the sphere, keeping $p$ and $q$ fixed, the variation vector field along the geodesic $\gamma$ will be a Jacobi field vanishing at $p$ and $q$. Rotating in $n-1$ different directions, one obtains $n-1$ linearly independent Jacobi fields. Thus $p$ and $q$ are conjugate along $\gamma$ with multiplicity $n-1$. ∎

**Proposition 5.11.** *Every Jacobi field along a geodesic $\gamma : [0, 1] \to M$ may be obtained by a variation of $\gamma$ through geodesics.*

*proof.* Choose a neighborhood $U$ of $\gamma(0)$ so that any two points of $U$ are joined by a unique minimal geodesic which depends differentiably on the endpoints. Suppose that $\gamma(t) \in U$ for $0 \leq t \leq \delta$. We will first construct a Jacobi field $W$ along $\gamma|_{[0,\delta]}$ with arbitrarily prescribed values at $t = 0$ and $t = \delta$. Choose a curve $a : (-\varepsilon, \varepsilon) \to U$ so that $a(0) = \gamma(0)$ and so that $\dfrac{da}{du}(0)$ is any prescribed vector in $T_{\gamma(0)}M$. Similarly, choose $b : (-\varepsilon, \varepsilon) \to U$ with $b(0) = \gamma(\delta)$ and $\dfrac{db}{du}(0)$ arbitrary. Now define the variation

$$\alpha : (-\varepsilon, \varepsilon) \times [0, \delta] \to M$$

by letting $\bar{\alpha}(u) : [0, \delta] \to M$ be the unique minimal geodesic from $a(u)$ to $b(u)$. (Compare the figure below.)




Then the formula

$$t \mapsto \frac{\partial \alpha}{\partial u}(0, t)$$

defines a Jacobi field with the given end conditions, by Proposition 5.9.

Any Jacobi field along $\gamma|_{[0,\delta]}$ can be obtained in this way: if $\mathcal{J}(\gamma)$ denotes the vector space of all Jacobi fields $W$ along $\gamma$, then the formula

$$W \mapsto (W_0, W_\delta)$$

defines a linear map

$$\ell : \mathcal{J}(\gamma) \to T_{\gamma(0)}M \times T_{\gamma(\delta)}M.$$

We have just shown that $\ell$ is onto. Since both vector spaces have the same dimension $2n$, it follows that $\ell$ is an isomorphism. That is, a Jacobi field is determined by its values at $\gamma(0)$ and $\gamma(\delta)$. (More generally, a Jacobi field is determined by its values at any two non-conjugate points.) Therefore, the above construction yields all possible Jacobi fields along $\gamma|_{[0,\delta]}$.

The restriction of $\bar{\alpha}(u)$ to the interval $[0, \delta]$ is not essential. If $u$ is sufficiently small, then, using the compactness of $[0, 1]$, $\overline{\alpha}(u)$ can be extended to a geodesic defined over the entire unit interval $[0, 1]$. This yields a variation through geodesics:

$$\alpha' : (-\varepsilon, \varepsilon) \times [0, 1] \to M$$

with any given Jacobi field as variation vector. □

**Remark.** *This argument shows that in any such neighborhood $U$, the Jacobi fields along a geodesic segment in $U$ are uniquely determined by their values at the endpoints of the geodesic.*

**Remark.** *The proof also shows that there is a neighborhood $(-\delta, \delta)$ of $0$ such that if $t \in (-\delta, \delta)$, then $\gamma(t)$ is not conjugate to $\gamma(0)$ along $\gamma$.*

## 5.5 The Index Theorem

Let $\gamma$ be a geodesic. The index $\lambda$ of the Hessian

$$E_{**} : T_\gamma\Omega \times T_\gamma\Omega \to \mathbb{R}$$

is defined to be the maximum dimension of a subspace of $T_\gamma\Omega$ on which $E_{**}$ is negative definite. This section will be entirely dedicated to proving the following theorem.

**Theorem 5.12** (Morse's Index Theorem)**.** *The index $\lambda$ of $E_{**}$ is equal to the number of points $\gamma(t)$, with $0 < t < 1$, such that $\gamma(t)$ is conjugate to $\gamma(0)$ along $\gamma$; each such conjugate point being counted with its multiplicity. This index $\lambda$ is always finite.*

**Idea of Proof.** *We can split $T_\gamma\Omega$ into two orthogonal direct factors, and the index of $E_{**}$ is the index of its restriction in one of it. Then we consider the restriction of $(E_0^\tau)_{**}$ in such a direct factor of $T_{\gamma|_{[0,\tau]}}\Omega$ and examine the property of index function $\lambda(\tau)$. We will see $\lambda(\tau)$ is increasing and left-continuous; it increases by the nullity of $E_0^\tau$ at $\tau+$, so finally $\lambda = \lambda(1)$ is the nullity of each $E_0^\tau$ where $\gamma(\tau)$ is conjugate to $\gamma(0)$.*

*proof.* Each point $\gamma(t)$ is contained in an open set $U$ such that any two points of $U$ are joined by a unique minimal geodesic which depends differentiably on the endpoints. Choose a subdivision $0 = t_0 < t_1 < \ldots < t_k = 1$ of the unit interval which is sufficiently fine so that each segment $\gamma[t_{i-1}, t_i]$ lies within such an open set $U$; and so that each $\gamma|_{[t_{i-1},t_i]}$ is minimal.

Let $T_\gamma\Omega(t_0, t_1, t_2, \ldots, t_k) \subset T_\gamma\Omega$ be the vector space consisting of all vector fields $W$ along $\gamma$ such that

(a) $W|_{[t_{i-1},t_i]}$ is a Jacobi field along $\gamma|_{[t_{i-1},t_i]}$ for each $i$;

(b) $W$ vanishes at the endpoints $t = 0$ and $t = 1$.

Thus $T_\gamma\Omega(t_0, t_1, t_2, \ldots, t_k)$ is a finite-dimensional vector space consisting of broken Jacobi fields along $\gamma$. Let $T' \subset T_\gamma\Omega$ be the vector space consisting of all vector fields $W \in T_\gamma\Omega$ for which $W(t_0) = 0$, $W(t_1) = 0$, $W(t_2) = 0$, $\ldots$, $W(t_k) = 0$.

**Lemma 5.13.** *The vector space $T_\gamma\Omega$ splits as the direct sum $T_\gamma\Omega(t_0, t_1, t_2, \ldots, t_k) \oplus T'$. These two subspaces are mutually perpendicular with respect to the inner product $E_{**}$. Furthermore, $E_{**}$ restricted to $T'$ is positive definite.*

*proof of Lemma 5.13.* Given any vector field $W \in T_\gamma\Omega$ let $W_1$ denote the unique "broken Jacobi field" in $T_\gamma\Omega(t_0, \ldots, t_k)$ such that $W_1(t_i) = W(t_i)$ for $i = 0, 1, \ldots, k$. It follows from the proof of Proposition 5.11 that $W_1$ exists and is unique. Clearly $W - W_1$ belongs to $T'$. Thus the two subspaces, $T_\gamma\Omega(t_0, \ldots, t_k)$ and $T'$ generate $T_\gamma\Omega$, and have only zero vector field in common, since $\gamma|_{[t_i, t_{i-1}]}$ is the only geodesic with these two endpoints in a neighborhood $U$.

If $W_1$ belongs to $T_\gamma\Omega(t_0, t_1, t_2, \ldots, t_k)$ and $W_2$ belongs to $T'$, then the second variation formula (Theorem 5.5) takes the form

$$\frac{1}{2}E_{**}(W_1, W_2) = -\sum_t \left\langle W_2(t), \Delta_t \frac{DW_1}{dt} \right\rangle - \int_0^1 \langle W_2, 0\rangle \, dt = 0.$$

Thus the two subspaces are mutually perpendicular with respect to $E_{**}$.

For any $W \in T_\gamma\Omega$ the Hessian $E_{**}(W, W)$ can be interpreted as the second derivative $d^2(E \circ \bar{\alpha})/du^2(0)$; where $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ is any variation of $\gamma$ with variation vector field $d\bar{\alpha}/dt(0)$ equal to $W$. If $W$ belongs to $T'$ then we may assume that $\bar{\alpha}$ is chosen so as to leave the points $\gamma(t_0), \gamma(t_1), \ldots, \gamma(t_k)$ fixed. In other words we may assume that $\bar{\alpha}(u)(t_i) = \gamma(t_i)$ for $i = 0, 1, \ldots, k$.

Proof that $E_{**}(W, W) \geq 0$ for $W \in T'$. Each $\bar{\alpha}(u) \in \Omega$ is a piecewise smooth path from $\gamma(0)$ to $\gamma(t_1)$ to $\gamma(t_2)$ to $\ldots$ to $\gamma(1)$. But each $\gamma|_{[t_{i-1}, t_i]}$ is a minimal geodesic, and therefore has smaller energy than any other path between its endpoints. This proves that

$$E(\bar{\alpha}(u)) \geq E(\gamma) = E(\bar{\alpha}(0)).$$

Therefore the second derivative, evaluated at $u = 0$, must be $\geq 0$.

Proof that $E_{**}(W, W) > 0$ for $W \in T'$, $W \neq 0$. Suppose that $E_{**}(W, W)$ were equal to 0. Then $W$ would lie in the null space of $E_{**}$. In fact for any $W_1 \in T_\gamma\Omega(t_0, t_1, \ldots, t_k)$ we have already seen that $E_{**}(W_1, W) = 0$. For any $W_2 \in T'$ the inequality

$$0 \leq E_{**}(W + cW_2, W + cW_2) = 2cE_{**}(W_2, W) + c^2 E_{**}(W_2, W_2)$$

for all values of $c$ implies that $E_{**}(W_2, W) = 0$. Thus $W$ lies in the null space. But the null space of $E_{**}$ consists of Jacobi fields. Since $T'$ contains no Jacobi fields other than zero, this implies that $W = 0$.

Thus the quadratic form $E_{**}$ is positive definite on $T'$. This completes the proof. $(\square)$

An immediate consequence is the following:

**Lemma 5.14.** *The index (or the nullity) of $E_{**}$ is equal to the index (or nullity) of $E_{**}$ restricted to the space $T_\gamma\Omega(t_0, t_1, \ldots, t_k)$ of broken Jacobi fields. In particular (since $T_\gamma\Omega(t_0, t_1, \ldots, t_k)$ is a finite dimensional vector space) the index $\lambda$ is always finite.*

The proof is straightforward.

Therefore, now, it suffices to show that the index of $E_{**}$ restricted to $T_\gamma\Omega(t_0, \ldots, t_k)$ is equal to the number of points $\gamma(t)$ with $0 < t < 1$ such that $\gamma(t)$ is conjugate to $\gamma(0)$ along $\gamma$.

Let $\gamma_\tau$ denote the restriction of $\gamma$ to the interval $[0, \tau]$. Thus $\gamma_\tau : [0, \tau] \to M$ is a geodesic from $\gamma(0)$ to $\gamma(\tau)$. Let $\lambda(\tau)$ denote the index of the Hessian $(E_0^\tau)_{**}$ which is associated with this geodesic. Thus $\lambda(1)$ is the index which we are actually trying to compute. First note that:

**Assertion 1.** *$\lambda(\tau)$ is a monotone function of $\tau$.*

*proof of Assertion 1.* For if $\tau < \tau'$, then there exists a $\lambda(\tau)$ dimensional space $V$ of vector fields along $\gamma_\tau$ which vanish at $\gamma(0)$ and $\gamma(\tau)$ such that the Hessian $(E_0^\tau)_{**}$ is negative definite on this vector space. Each vector field in $V$ extends to a vector field along $\gamma_{\tau'}$ which vanishes identically between $\gamma(\tau)$ and $\gamma(\tau')$. Thus we obtain a $\lambda(\tau)$ dimensional vector space of fields along $\gamma_{\tau'}$, on which $(E_{\tau'})_{**}$ is negative definite. Hence $\lambda(\tau) \leq \lambda(\tau')$. (□)

**Assertion 2.** *$\lambda(\tau) = 0$ for small values of $\tau$.*

*proof of Assertion 2.* For if $\tau$ is sufficiently small then $\gamma_\tau$ is a minimal geodesic, hence $\lambda(\tau) = 0$. (□)

Now let us examine the discontinuities of the function $\lambda(\tau)$. First note that $\lambda(\tau)$ is continuous from the left:

**Assertion 3.** *For all sufficiently small $\varepsilon > 0$ we have $\lambda(\tau - \varepsilon) = \lambda(\tau)$.*

*proof of Assertion 3.* According to Lemma 5.14 the number $\lambda(1)$ can be interpreted as the index of a quadratic form on a finite dimensional vector space $T_\gamma\Omega(t_0, \ldots, t_k)$. We may assume that the subdivision is chosen so that say $t_i < \tau < t_{i+1}$. Then the index $\lambda(\tau)$ can be interpreted as the index of a corresponding quadratic form $H_\tau$ on a corresponding vector space of broken Jacobi fields along $\gamma_\tau$. This vector space is to be constructed using the subdivision $0 = t_0 < t_1 < t_2 < \ldots < t_i < \tau$ of $[0, \tau]$. Since a broken Jacobi field is uniquely determined by its values at the break points $\gamma(t_i)$ by the proof of Proposition 5.11, this vector space is isomorphic to the direct sum

$$\Sigma = \bigoplus_{j=1}^{i} T_{\gamma(t_j)} M.$$

Note that this vector space $\Sigma$ is independent of $\tau$. Evidently the quadratic form $H_\tau$ on $\Sigma$ varies continuously with $\tau$.

Now $H_\tau$ is negative definite on a subspace $V \subset \Sigma$ of dimension $\lambda(\tau)$. For all $\tau'$ sufficiently close to $\tau$ it follows that $H_{\tau'}$ is negative definite on $V$. Therefore $\lambda(\tau') \geq \lambda(\tau)$. But if $\tau' = \tau - \varepsilon < \tau$ then we also have $\lambda(\tau - \varepsilon) \leq \lambda(\tau)$ by Assertion 1. Hence $\lambda(\tau - \varepsilon) = \lambda(\tau)$. (□)

**Assertion 4.** *Let $\nu$ be the nullity of the Hessian $(E_0^\tau)_{**}$. Then for all sufficiently small $\varepsilon > 0$ we have*

$$\lambda(\tau + \varepsilon) = \lambda(\tau) + \nu.$$

Thus the function $\lambda(t)$ jumps by $\nu$ when the variable $t$ passes a conjugate point of multiplicity $\nu$; and is continuous otherwise. Clearly this assertion will complete the proof of the Index Theorem (Theorem 5.12).

*proof of Assertion 4.* Proof that $\lambda(\tau + \varepsilon) \leq \lambda(\tau) + \nu$. Let $H_\tau$ and $\Sigma$ be as in the proof of Assertion 3. Since $\dim \Sigma = ni$ we see that $H_\tau$ is positive definite on some subspace $V' \subset \Sigma$ of dimension $ni - \lambda(\tau) - \nu$. For all $\tau'$ sufficiently close to $\tau$, it follows that $H_{\tau'}$ is positive definite on $V'$. Hence

$$\lambda(\tau') \leq \dim \Sigma - \dim V' = \lambda(\tau) + \nu.$$

Proof that $\lambda(\tau+\varepsilon) \geq \lambda(\tau)+\nu$. Let $W_1, \dots, W_{\lambda(\tau)}$ be $\lambda(\tau)$ vector fields along $\gamma_\tau$, vanishing at the endpoints, such that the matrix

$$\Big((E_0^\tau)_{**}(W_i, W_j)\Big)_{ij}$$

is negative definite. Let $\mathbf{J}_1, \dots, \mathbf{J}_\nu$ be $\nu$ linearly independent Jacobi fields along $\gamma_\tau$, also vanishing at the endpoints. Note that the $\nu$ vectors

$$\frac{D\mathbf{J}_h}{dt}(\tau) \in T_{\gamma(\tau)}M$$

are linearly independent. Hence it is possible to choose $\nu$ vector fields $X_1, \dots, X_\nu$ along $\gamma_{\tau+\varepsilon}$, vanishing at the endpoints of $\gamma_{\tau+\varepsilon}$, so that

$$\left(\left\langle \frac{D\mathbf{J}_h}{dt}(\tau), X_k(\tau)\right\rangle\right)_{hk}$$

is equal to the $\nu \times \nu$ identity matrix (by choosing $X_k(\tau) \perp \delta_{hk}\dfrac{D\mathbf{J}_h}{dt}(\tau)$ and resizing). Extend the vector fields $W_i$ and $\mathbf{J}_h$ over $\gamma_{\tau+\varepsilon}$ by setting these fields equal to 0 for $\tau \leq t \leq \tau+\varepsilon$.

Using the second variation formula we see easily that

$$(E_0^{\tau+\varepsilon})_{**}(\mathbf{J}_h, W_i) = -0 - 0 = 0,$$

$$(E_0^{\tau+\varepsilon})_{**}(\mathbf{J}_h, X_k) = -2\sum_{t=\tau}\left(0 - \left\langle \frac{D\mathbf{J}_h}{dt}(t), X_k(t)\right\rangle\right) - 0 = 2\delta_{hk}.$$

Now let $c$ be a small number, and consider the $\lambda(\tau)+\nu$ vector fields

$$W_1, \dots, W_{\lambda(\tau)}, c^{-1}\mathbf{J}_1 - cX_1, \dots, c^{-1}\mathbf{J}_\nu - cX_\nu$$

along $\gamma_{\tau+\varepsilon}$. We claim that these vector fields span a vector space of dimension $\lambda(\tau)+\nu$ on which the quadratic form $E_0^{\tau+\varepsilon **}$ is negative definite. In fact, the matrix of $E_0^{\tau+\varepsilon **}$ with respect to this basis is

$$\begin{pmatrix} \Big((E_0^\tau)_{**}(W_i, W_j)\Big)_{ij} & cA \\ cA^T & -4I + c^2B \end{pmatrix}$$

where

$$A = \Big((E_0^{\tau+\varepsilon})_{**}(W_i, X_j)\Big)_{ij},$$

$$B = \Big((E_0^{\tau+\varepsilon})_{**}(X_\ell, X_k)\Big)_{\ell k}.$$

are fixed matrices. If $c$ is sufficiently small, this compound matrix is certainly negative definite. This proves Assertion 4. ($\square$)

By Assertion 4, the function $\lambda(t)$ jumps by $\nu$ when the variable $t$ passes a conjugate point of multiplicity $\nu$; and is continuous otherwise. Clearly this assertion complete the proof of the Index Theorem (Theorem 5.12).

$\square$

## 5.6 A Finite Dimensional Approximation to $\Omega^c$

# References


[1] Allen Hatcher, *Algebraic topology*, Cambridge University Press, 2002.

[2] John M. Lee, *Introduction to smooth manifolds*, Springer, 2003.

[3] John Willard Milnor, *Morse theory*, vol. 51, Princeton university press, 1963.

[4] Joseph J. Rotman, *An introduction to algebraic topology*, vol. 119, Springer Science & Business Media, 2013.

[5] John Henry Constantine Whitehead, *On simply connected, 4-dimensional polyhedra*, Commentarii Mathematici Helvetici **22** (1949), 48–92.

[6] Wang Zuoqin, *Smooth manifolds*, `http://staff.ustc.edu.cn/~wangzuoq/Courses/18F-Manifolds/index.html`.