YANG-MILLS EQUATIONS ON RIEMANN SURFACES AND MODULI SPACE

BOWEN LIU

Abstract. Reading notes for my graduation thesis.

Contents

Part 1. The classical approach: attaching handles

1. NON-DEGENERATE SMOOTH FUNCTIONS ON A MANIFOLD

1.1. Definitions and Lemmas.

Definition 1.1.1. *Let f be a smooth function on a manifold on a manifold M. A point $p \in M$ is call* ***critical point*** *of f if the induced map $f_* : TM_p \to T\mathbb{R}_{f(p)}$ is zero. If we choose a local coordinate system $(x^1, \ldots, x^n)$ in a neighborhood U of p, this means that:*

$$\frac{\partial f}{\partial x^1}(p) = \cdots = \frac{\partial f}{\partial x^n}(p) = 0$$

The real number $f(p)$ is called a ***critical value*** *of f.*

Definition 1.1.2. *A critical point p is called* ***non-degenerate*** *if and only if the matrix*

$$(\frac{\partial^2 f}{\partial x^i \partial x^j}(p))$$

is nonsingular.

Definition 1.1.3. *If p is a critical point of f, define a symmetric bilinear functional f_{**} on TM_p, called the* ***Hessian*** *of f at p. If $v, w \in TM_p$, then v and w have extensions $\tilde{v}, \tilde{w}$ to vector fields. We let:*

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

Remark 1.1.4. *Hessian of f at p is indeed symmetric since:*

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

and independent of the extension of vector field, since $\tilde{v}_p(\tilde{w}(f))$ just needs the value of $\tilde{v}$ at p, $\tilde{w}_p(\tilde{v}(f))$ just needs the value of $\tilde{w}$ at p

Remark 1.1.5. *If we choose local coordinate $(x^1, \ldots, x^n)$, and suppose:*

$$v = \sum a_i \frac{\partial}{\partial x^i}|_p, \quad w = \sum b_j \frac{\partial}{\partial x^j}|_p$$

then we can extend by taking b_j as constant functions(for convenience: $\frac{\partial b_j}{\partial x^j} = 0$):

$$\tilde{v} = \sum a_i \frac{\partial}{\partial x^i}, \quad \tilde{w} = \sum b_j \frac{\partial}{\partial x^j}$$

then,

$$f_{**}(v, w) = \tilde{v}_p(\tilde{w}(f)) = v_p(\sum b_j \frac{\partial f}{\partial x^j}) = \sum_{i,j} a_i b_j \frac{\partial^2 f}{\partial x^i \partial x^j}(p)$$

so the matrix $(\frac{\partial^2 f}{\partial x^i \partial x^j}(p))$ reprensents the bilinear function f_{**} with respect to the standard basis.

Definition 1.1.6. *The **index** of a bilinear function H, on a vector space V, is defined to be the maximal dimension of a subspace of V on which H is negative definite. The index of f_{**} on TM_p will be referred to simply as the index of f at p.*

Definition 1.1.7. *The **nullity** is the dimension of the nullspace, i.e., the subspace consisting of all $v \in V$ such that $H(v, w) = 0, \forall w \in V$*

Remark 1.1.8. *The point p is obviously a non-degenerate critical point, if and only if f_{**} has nullity equal to 0*

Lemma 1.1.9. *Lef f be a C^∞ function in a convex neighborhood V of 0 in $\mathbb{R}^n$, with $f(0) = 0$. Then:*

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

for some suitable C^∞ functions g_i defined on V, with:

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

Lemma 1.1.10. (*Lemma of Morse*) *Let p be a non-degenerated critical point of f. Then there is a local coordinate system $(y^1, \dots, y^n)$ in a neighborhood U of p with $y^i(p) = 0$ for all i, and such that the identity:*

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

holds throughout U, where λ is the index of f at p.

Remark 1.1.11. *This lemma implies the local properties of f at p is determined totally by the index of f!*

Corollary 1.1.12. *non-degenerate critical points are isolated.*

Definition 1.1.13. *A 1-**parameter group of diffeomorphisms** of a manifold M is a C^∞ map:*

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

such that:

1. *for each $t \in \mathbb{R}$ the map $\varphi_t : M \to M$ defined by $\varphi_t(q) = \varphi(t, q)$ is a diffeomorphism of M onto itself.*
2. *for all $t, s \in \mathbb{R}$ we have $\varphi_{t+s} = \varphi_t \circ \varphi_s$*

Definition 1.1.14. *Given a $1-$parameter group φ of diffeomorphisms of M, we can define a vector field X on M as follows. For every smooth real valued function f let:*

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

*The vector field X is said to **generate** the group φ.*

Lemma 1.1.15. *A vector field on M which vanishes outside a compact set $K \subset M$ generates a unique 1-parameter group of diffeomorphism of M.*

Proof. Now let φ be a 1-parameter group of diffeomorphisms, generated by the vector field X. Then for each fixed q the curve

$$t \to \varphi_t(q)$$

satisfies the differential equation

$$\frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t} = X_{\varphi_t(q)}$$

with initial condition $\varphi_0(q) = q$. This is true since

$$\frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t}(f) = \lim_{h\to 0} \frac{f(\varphi_{t+h})(q) - f(\varphi_t(q))}{h} = \lim_{h\to 0} \frac{\varphi_h(p) - f(p)}{h} = X_p(f)$$

where $p = \varphi_t(q)$. But it is well known that such a differential equation, locally, has a unique solution which depends smoothly on the initial condition.

Thus for each point of M there exists a neighborhood U and a number $\varepsilon > 0$ so that the differential equation

$$\frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t} = X_{\varphi_t(q)}, \quad \varphi_0(q) = q$$

has a unique smooth solution for $q \in U, |t| < \varepsilon$.

The compact set K can be covered a finited number of such neighborhood U. Let $\varepsilon_0 > 0$ denote the smallest of the corresponding numbers ε. Setting $\varphi_t(q) = q$ for $q \notin K$, it follows that this differential equation has a unique solution $\varphi_t(q)$ 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 $\varphi_{s+t} = \varphi_t \circ \varphi_s$ providing that $|t|, |s|, |t+s| < \varepsilon_0$. Therefore each such φ_t is a diffeomorphism.

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

$$\varphi_t = \underbrace{\varphi_{\varepsilon_0/2} \circ \varphi_{\varepsilon_0/2} \circ \cdots \circ \varphi_{\varepsilon_0/2}}_{\text{k times}} \circ \varphi_r$$

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

Remark 1.1.16. *The hypothesis that X vanishes outside of a compact set cannot be omitted.*

1.2. **Homotopy Type in terms of Critical Values.**

Definition 1.2.1. *If f is a real valued function on M, we let:*

$$M^a := f^{-1}(-\infty, a] = \{p \in M \mid f(p) \leq a\}$$

Theorem 1.2.2. *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 point 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.*

The idea of the proof is to push M^b down to M^a along the orthogonal trajectories of the hypersurfaces $f =$ constant.

Definition 1.2.3. *Choose a Riemannian metric on M; and let $\langle X, Y\rangle$ denote the inner product of two tangent vectors, as determined by this metric. The **gradient** of f on M which is characterized by the identity*

$$\langle X, \operatorname{grad} f\rangle = X(f)$$

for any vector field X. This vector field $\operatorname{grad} f$ vanishes precisely at the critical points of f.

If $c : \mathbb{R} \to M$ is a curve with velocity vector $\dfrac{\mathrm{d}c}{\mathrm{d}t}$ note the identity

$$\langle \frac{\mathrm{d}c}{\mathrm{d}t}, \operatorname{grad} f\rangle = \frac{\mathrm{d}(f \circ c)}{\mathrm{d}t}$$

Let $\rho : M \to \mathbb{R}$ be a smooth function which is equal to $1/\langle \operatorname{grad} f, \operatorname{grad} f\rangle$ throughout the compact set $f^{-1}([a, b])$; and which vanishes outside of a compact neighborhood of this set. Then the vector field X, defined by

$$X_q = \rho(q)(\operatorname{grand} f)_q$$

satisfies the condition of lemma 1.1.15. Hencee X generates a 1-parameter group of diffeomorphisms

$$\varphi_t : M \to M$$

For fixed $q \in M$ consider the function $t \to f(\varphi_t(q))$. If $\varphi_t(q)$ lies in the set $f^{-1}[a, b]$, then

$$\frac{\mathrm{d}f(\varphi_t(q))}{\mathrm{d}t} = \langle \frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t}, \operatorname{grad} f\rangle = \langle X, \operatorname{grad} f\rangle = 1$$

Thus the correspondence

$$t \to f(\varphi_t(q))$$

is linear with derivative 1 as long as $f(\varphi_t(q))$ lies in a and b.

Now consider the diffeomorphism φ_{b-a}, clearly this carries M^a diffeomorphically to M^b, since if $q \in M^a$, i.e., $f(q) \leq a$, and

$$f(\varphi_{b-a}(q)) - f(q) = f(\varphi_{b-a}(q)) - f(\varphi_0(q)) = b - a$$

then

$$f(\varphi_{b-a}(q)) \leq b$$

This proves the first part of this theorem.

Define a 1-parameter family of maps

$$r_t : M^b \to M^b$$

by

$$\begin{cases} q, \quad \text{if } f(q) \leq a \\ \varphi_{t(a-f(q))}(q), \quad \text{if } a \leq f(q) \leq b \end{cases}$$

Then r_0 is the identity, and r_1 is a retraction from M^b to M^a. Hence M^a is a deformation retract of M^b. This completes the proof.

Theorem 1.2.4. *Let $f : M \to \mathbb{R}$ be a smooth function, and let p be a non-degenerate critical point with index λ. 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 ε, the set $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon}$ with a λ-cell attached.*

Proof. Choose a coordinate system $u^1, \dots, u^n$ in a neighborhood U 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

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

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 imbedding

   $$(u^1, \dots, u^n) : U \to \mathbb{R}^n$$

   contains the closed ball.

   $$\{(u^1, \dots, u^n) : \sum (u^i)^2 \leq 2\varepsilon\}$$

Now define e^λ to be the set of points in U with

$$(u^1)^2 + \cdots + (u^\lambda)^2 \leq \varepsilon, \quad u^{\lambda+1} = \cdots = u^n = 0$$

Note that $e^\lambda \cap M^{c-\varepsilon}$ is precisely the boundary of $\dot{e}^\lambda$, since

$$c - (u^1)^2 - \cdots - (u^\lambda)^2 \leq c - \varepsilon \implies (u^1)^2 + \cdots + (u^\lambda)^2 \geq \varepsilon$$

so that e^λ is attached to $M^{c-\varepsilon}$ as required. We must show 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^∞ function satisfying the condition

$$\mu(0) > \varepsilon$$

$$u(r) = 0, \quad \text{for } r \geq 2\varepsilon$$

$$-1 < \mu'(r) \leq 0, \quad \text{for all } r$$

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

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

within this coordinate neighborhood. It is easily verified that F is well defined smooth function throughout M.

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 $f = c - \xi + \eta$; so that

$$F(q) = c - \xi(q) + \eta(q) - \mu(\xi(q) + 2\eta(q)), \quad \forall q \in U.$$

Lemma 1.2.5. *The region $F^{-1}(-\infty, c+\varepsilon]$ coincides with the region $M^{c+\varepsilon}$.*

Proof. Outside of the ellipsoid $\xi + 2\eta \leq 2\varepsilon$ the function f and F coincides. Within this ellipsoid we have

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

□

Lemma 1.2.6. *The critical points of F are the same as those of f.*

Proof. 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

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

where the covector $\mathrm{d}\xi$ and $\mathrm{d}\eta$ are simutaneously 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 lemma 1.2.5 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

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

Hence $F^{-1}[c - \varepsilon, c + \varepsilon]$ contains no critical points. together with theorem 1.2.2 this proves the following

Lemma 1.2.7. *The region $F^{-1}(-\infty, c-\varepsilon]$ is a deformation retract of $M^{c+\varepsilon}$*

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 $F^{-1}(-\infty, c-\varepsilon] - M^{c-\varepsilon}$.

Now consider the cell e^λ consisting of all points q with

$$\xi(q) \leq \varepsilon, \quad \eta(q) = 0$$

Note that e^λ is contained in the "handle" H. In fact, since $\dfrac{\partial F}{\partial \xi} < 0$, we have

$$F(q) \leq F(p) < c - \varepsilon$$

but $f(q) \geq c - \varepsilon$ for $q \in e^\lambda$.

Lemma 1.2.8. *$M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\varepsilon} \cup H$.*

Proof. Let r_t be the identity outside of U; and define r_t within U as follows. It necessary to distinguish three cases.

1. Within the region $\xi \leq \varepsilon$ let r_t correspond to the transformation
$$(u^1, \dots, u^n) \to (u^1, \dots, u^\lambda, tu^{\lambda+1}, \dots, tu^n)$$
Thus r_1 is identity and r_0 maps the entire region into e^λ. The face that each r_t maps $F^{-1}(-\infty, c-\varepsilon]$ into itself, follows from the inequality $\dfrac{\partial F}{\partial \eta} > 0$.
2. Within the region $\varepsilon \leq \xi \leq \eta + \varepsilon$ let r_t correspond to the transformation
$$(u^1, \dots, u^n) \to (u^1, \dots, u^\lambda, s_t u^{\lambda+1}, \dots, s_t u^n)$$
where $s_t \in [0,1]$ is defined by
$$s_t = t + (1-t)((\xi - \varepsilon)/\eta)^{\frac{1}{2}}$$
Thus r_1 is again the identity, and r_0 maps the entire region into the hypersurface $f^{-1}(c-\varepsilon)$
3. Within region $\eta + \varepsilon \leq \xi$ let r_t be identity.

$\square$

This completes the proof that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $F^{-1}(-\infty, c+\varepsilon]$. Together with lemma 1.2.7, it completes the proof of theorem 1.2.4. $\square$

Remark 1.2.9. *More generally suppose that there are k non-degenerate critical points $p_1, \dots, p_k$ with indices $\lambda_1, \dots, \lambda_k$ in $f^{-1}(c)$. Then a similar proof shows that $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon} \cup e^{\lambda_1} \cup \dots \cup e^{\lambda_k}$.*

Theorem 1.2.10. *If f is a differential 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 λ for each critical point of index λ.*

The proof will be based on two lemmas concerning a topological space X with a cell attached.

Lemma 1.2.11. (*Whitehead*) *Let φ_0 and φ_1 be homotopic maps from sphere $\dot{e}^\lambda$ to X. Then the identity map of X extend to a homotopy equivalence*

$$k : X \cup_{\varphi_0} e^\lambda \to X \cup_{\varphi_1} e^\lambda$$

Proof. Define k by the formulas

$$\begin{cases} k(x) = x, \quad \forall x \in X \\ k(tu) = 2tu, \quad \forall 0 \le t \le \frac{1}{2}, u \in \dot{e}^\lambda \\ k(tu) = \varphi_{2-2t}(u), \quad \forall \frac{1}{2} \le t \le 1, u \in \dot{e}^\lambda. \end{cases}$$

Here φ_t denotes the homotopy between φ_0 and φ_1; and tu denotes the product of the scalar t with the unit vector u. A corresponding map

$$l : X \cup_{\varphi_1} e^\lambda \to X \cup_{\varphi_0} e^\lambda$$

is defined by similar formulas. □

Lemma 1.2.12. *Let $\varphi : \dot{e}^\lambda \to X$ be an attaching map. Any homotopy equivalence $f : X \to Y$ extends to a homotopy equivalence*

$$F : X \cup_\varphi e^\lambda \to Y \cup_{f\varphi} e^\lambda$$

Now we give a proof of theorem 1.2.9

Proof. Let $c_1 < c_2 < c_3 < \dots$ be the critical values of $f : M \to \mathbb{R}$. The sequence (c_i) has no cluster point since each M^a is compact. The set M^a is vacuous for $a < c_1$. Suppose $a \neq c_1, c_2, c_3, \dots$ and that M^a is of the homotopy of a CW-complex. Let c be the smallest $c_i > a$. By theorem 1.2.2, 1.2.4 and remark 1.2.9, $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon} \cup_{\varphi_1} e^{\lambda_1} \cup \dots \cup_{\varphi_j} e^{\lambda_{j(c)}}$ for certain maps $\varphi_1, \dots, \varphi_{j(c)}$ when ε is small enough, and there is a homotopy equivalence $h : M^{c-\varepsilon} \to M^a$. We have assume that there is a homotopy equivalence $h' : M^a \to K$, where K is a CW-complex.

Then each $h' \circ h \circ \varphi_j$ is homotopic by cellular approximation to a map

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

Then $K \cup_{\psi_1} e^{\lambda_1} \cup \dots \cup_{\psi_{j(c)}} e^{\lambda_{j(c)}}$ is a CW-complex, and has the homotopy type as $M^{c+\varepsilon}$, by lemmas we mentioned above.

By induction it follows that each M^a has the homotopy type of a CW-complex. If M is compact this completes the proof. If M is not compact, but all critical points lie in one of the compact sets M^a, then a proof similar to that of theorem 1.2.2 shows that the set M^a is a deformation retract of M, so the proof is again complete.

If there are infinitely many critical points then the above construction gives us an infinite sequence of homotopy equivalences

$$\begin{array}{ccccccc} M^{a_1} & \hookrightarrow & M^{a_2} & \hookrightarrow & M^{a_3} & \hookrightarrow & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ K^{a_1} & \hookrightarrow & K^{a_2} & \hookrightarrow & K^{a_3} & \hookrightarrow & \dots \end{array}$$

each extending the previous one. Let K denote the union of the K_i in the direct limit topology, i.e., the finest possible compatible topology, and let $g : M \to K$ be the limit map. Then g induces isomorphism of homotopy groups in all dimensions. We need only apply Theorem 1 of **Combinatorial homotopy I** to conclude g is a homotopy equivalence. This completes the proof. $\square$

1.3. **Examples.**

Theorem 1.3.1. (*Reeb*) *If M is a compact manifold and f is a differentiable function on M with only two critical points, both of which are non-degenerate, then M is homeomorphic to a sphere.*

1.4. The Morse Inequality.

Definition 1.4.1. *Let S be a function from certain pairs of spaces to the integers. S is* ***subadditive*** *if whenever $Z \subset Y \subset X$, we have $S(X,Z) \leq S(X,Y) + S(Y,Z)$. If equality holds, S is called* ***additive****.*

Example 1.4.2. *$R_\lambda(X,Y)$ is λth Betti number of (X,Y) i.e. $\dim_F H_\lambda(X,Y;F)$. Then $R_\lambda(X,Y)$ is subadditive.*

Lemma 1.4.3. *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.*

If we take $S(X,\varnothing) = S(X)$, then take $X_0 = \varnothing$ in lemma 1.4.3, we have:

$$S(X_n) \leq \sum_{i=1}^n S(X_i, X_{i-1}) \tag{1.1}$$

with equality if S is additive.

Let M be a compact manifold and f a differential function on M with isolated, non-degenerate, critical points. Let $a_1 < \cdots < a_k$ be such that M^{a_i} contains exactly i critical points, and $M^{a_k} = M$. Then:

$$H_*(M^{a_i}, M^{a_{i-1}}) = H_*(M^{a_{i-1}} \cup e^{\lambda_i}, M^{a_{i-1}}) \overset{excision}{=} H_*(e^{\lambda_i}, \dot{e}^{\lambda_i})$$

Applying (1.1) to $\varnothing = M^{a_0} \subset \cdots \subset M^{a_k} = M$ with $S = R_\lambda$ we have:

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

where C_λ denotes the number of critical points of index λ. Applying this formula to the case $X = \chi$ we have:

$$\chi(M) = \sum_{i=1}^k \chi(M^{a_i}, M^{a_{i-1}}) = C_0 - C_1 + C_2 - \cdots + (-1)^n C_n$$

where χ is Euler characteristic.

Thus we have proven:

Theorem 1.4.4. (*Weak Morse Inequality*) *If C_λ denotes the number of critical points of index λ on the compact manifold M, then:*

$$R_\lambda(M) \leq C_\lambda \tag{1.2}$$

$$\sum (-1)^\lambda R_\lambda(M) = \sum (-1)^\lambda C_\lambda \tag{1.3}$$

Slightly sharper inequalities can be proven by the following argument:

Lemma 1.4.5. *The function S_λ is subadditive, where*

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

Applying this subadditive function S_λ to the spaces $\varnothing = M^{a_0} \subset \cdots \subset M^{a_k} = M$ with $S = R_\lambda$ we have:

Theorem 1.4.6. (*Morse Inequality*)

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots + (-1)^\lambda R_0(M) \leq C_\lambda - C_{\lambda-1} + \cdots + (-1)^\lambda C_0 \tag{1.4}$$

Corollary 1.4.7. *If* $C_{\lambda+1} = C_{\lambda-1} = 0$*, then* $R_\lambda = C_\lambda$ *and* $R_{\lambda+1} = R_{\lambda-1} = 0$

1.5. Manifolds in Euclidean Space: The existence of Non-Degenerate Functions.

Definition 1.5.1. *Let $M \subset \mathbb{R}^n$ be a manifold of dimension $k < n$, differentiably embedded in $\mathbb{R}^n$. Let $N \subset M \times \mathbb{R}^n$ be defined by:*

$$N = \{(q, v) : q \in M, v \text{ is perpendicular to } M \text{ at } q\}$$

and let $E : N \to \mathbb{R}^n$ be $E(q, v) = q + v$.

Remark 1.5.2. *It is not difficult to show that N is a n-dimensional manifold differentiably embedded in $\mathbb{R}^{2n}$. In fact, N is the total space of the normal vector bundle of M. E is the "end point" map.*

Definition 1.5.3. *$e \in \mathbb{R}^n$ is a **focal point** of (M, q) with multiplicity μ if $e = q + v$ where $(q, v) \in N$ and the Jacobian of E at (q, v) has nullity $\mu > 0$. The point e will be called a focal point of M if e is a focal point of (M, q) for some $q \in M$.*

Theorem 1.5.4. (*Sard*) *If M_1, M_2 are differentiable manifolds having countable basis, of the same dimension, and $f : M_1 \to M_2$ is of class C^1, then the image of the set of critical points has measure 0 in M_2.*

Corollary 1.5.5. *For almost all $x \in \mathbb{R}^n$, the point x is not a focal point of M.*

Proof. We have just seen that N is a n-dimensional manifold. The point x is a focal point iff x is in the image of critical points of $E : N \to \mathbb{R}^n$. Therefore the set of focal points has measure 0. □

For better understanding of the concept of focal point, it's convenient to introduce the "second fundamental form" of a manifold in Euclidean space. We will make use of a fixed local coordinate system.

Let $u^1, \dots, u^k$ be coordinates for a region of the manifold $M \subset \mathbb{R}^n$. then the inclusion map from $M \to \mathbb{R}^n$ determines n smooth functions:

$$x^1(u^1, \dots, u^k), \dots, x^n(u^1, \dots, u^k)$$

This functions will be written briefly as $\overrightarrow{x} = (x_1, \dots, x_n)$.

Definition 1.5.6. *The **first fundamental form** associated with the coordinate system is defined to be the symmetric matrix of real valued functions:*

$$(g_{ij}) = (\frac{\partial \overrightarrow{x}}{\partial u^i} \cdot \frac{\partial \overrightarrow{x}}{\partial u^j})$$

Definition 1.5.7. *The **second fundamental form** on the other hand, is a symmetric matrix $(\overrightarrow{l_{ij}})$ of vector valued functions. It is defined as follows. The vector $\dfrac{\partial^2 \overrightarrow{x}}{\partial u^i \partial u^j}$ at a point of M can be expressed as the sum of a vector tangent to M and a vector normal to M. Define $\overrightarrow{l_{ij}}$ to be the normal* *component of it. Given any unit vector $\overrightarrow{v}$ which is normal to M at $\overrightarrow{q}$ the matrix:*

$$(\overrightarrow{v}\cdot\frac{\partial^2\overrightarrow{x}}{\partial u^i\partial u^j})=(\overrightarrow{v}\cdot\overrightarrow{l_{ij}})$$

can be called the second fundamental form of M at $\overrightarrow{q}$ in the direction $\overrightarrow{v}$.

Definition 1.5.8. *It will simplify the discussion to assume that the coordinates have been chosen to that g_{ij}, evaluated at $\overrightarrow{q}$, is the identity matrix. Then the eigenvalue of the matrix $(\overrightarrow{v}\cdot\overrightarrow{l_{ij}})$ are called the* ***principal curvature*** *$K_1,\dots,K_n$ of M at $\overrightarrow{q}$ in the normal direction $\overrightarrow{v}$. $K_1^{-1},\dots,K_n^{-1}$ are called the* ***principal radii of curvature.***

Now consider the normal line l consisting of all $\overrightarrow{q}+t\overrightarrow{v}$, which $\overrightarrow{v}$ is a fixed unit vector orthogonal to M at $\overrightarrow{q}$

Lemma 1.5.9. *The focal point of $(M,\overrightarrow{q})$ along l are precisely the points $\overrightarrow{q}+K_i^{-1}\overrightarrow{v}$, where $1\le i\le k, K_i\neq 0$. Thus there are at most, k focal points of $M,\overrightarrow{q})$ along l, each being counted with its proper multiplicity.*

Proof. Choose $n-k$ vector fields $\overrightarrow{W}_1(u^1,\dots,u^k),\dots,\overrightarrow{W}_{n-k}(u^1,\dots,u^k)$ along the manifold so that $\overrightarrow{W}_1,\dots,\overrightarrow{W}_{n-k}$ are unit vectors which are orthogonal to each other and to M. We can introduce coordinates $(u^1,\dots,u^k,t^1,\dots,t^{n-k})$ on manifold $N\subset M\times\mathbb{R}^n$ as follows. Let $(u^1,\dots,u^k,t^1,\dots,t^{n-k})$ correspond to the point

$$(\overrightarrow{x}(u^1,\dots,u^k),\sum_{\alpha=1}^{n-k}t^\alpha\overrightarrow{W}_\alpha(u^1,\dots,u^k))\in N$$

Then the function

$$E:N\to\mathbb{R}^n$$

gives rise to the correspondence

$$(u^1,\dots,u^k,t^1,\dots,t^{n-k})\xrightarrow{\overrightarrow{e}}\overrightarrow{x}(u^1,\dots,u^k)+\sum_{\alpha=1}^{n-k}t^\alpha\overrightarrow{W}_\alpha(u^1,\dots,u^k)$$

with partial derivatives

$$\begin{cases}\dfrac{\partial\overrightarrow{e}}{\partial u^i}=\dfrac{\partial\overrightarrow{x}}{\partial u^i}+\sum_\alpha t^\alpha\dfrac{\partial\overrightarrow{W}_\alpha}{\partial u^i}\\ \dfrac{\partial\overrightarrow{e}}{\partial t^\beta}=\overrightarrow{W}_\beta\end{cases}$$

Taking the inner products of these n-vectors with the linear independent vector $\frac{\partial\overrightarrow{x}}{\partial u^1},\dots,\frac{\partial\overrightarrow{x}}{\partial u^k},\overrightarrow{W}_1,\dots,\overrightarrow{W}_{n-k}$ we will obtain an $n\times n$ matrix whose rank equals the rank of the Jacobian of E at the corresponding point.

This $n\times n$ matrix clearly has the following form

$$\begin{pmatrix}(\frac{\partial\overrightarrow{x}}{\partial u^i}\cdot\frac{\partial\overrightarrow{x}}{\partial u^j}+\sum_\alpha t^\alpha\frac{\partial\overrightarrow{W}_\alpha}{\partial u^i}\cdot\frac{\partial\overrightarrow{x}}{\partial u^j}) & (\sum_\alpha t^\alpha\frac{\partial\overrightarrow{W}_\alpha}{\partial u^i}\cdot\overrightarrow{W}_\beta)\\ 0 & \text{identity matrix}\end{pmatrix}$$

Thus the nullity is equal to the nullity of the upper left hand block. Using the identity

$$0 = \frac{\partial}{\partial u^i}(\overrightarrow{W}_\alpha \cdot \frac{\partial \overrightarrow{x}}{\partial u^j}) = \frac{\partial \overrightarrow{W}_\alpha}{\partial u^i} \cdot \frac{\partial \overrightarrow{x}}{\partial u^j} + \overrightarrow{W}_\alpha \cdot \frac{\partial^2 \overrightarrow{x}}{\partial u^i \partial u^j}$$

We see that this upper left hand block is just the matrix

$$(g_{ij} - \sum_\alpha t^\alpha \overrightarrow{W}_\alpha \cdot \overrightarrow{l}_{ij})$$

Thus $\overrightarrow{q} + t\overrightarrow{v}$ is a focal point of $(M, \overrightarrow{q})$ with multiplicity μ if and only if the matrix

$$(g_{ij} - t\overrightarrow{v} \cdot \overrightarrow{l}_{ij})$$

is singular, with nullity μ.

Now suppose that (g_{ij}) is identity matrix. Then the matrix above is singular if and only if $\frac{1}{t}$ is an eigenvalue of the matrix $(\overrightarrow{v} \cdot \overrightarrow{l}_{ij})$. Furthermore the multiplicity μ is equal to the multiplicity of $\frac{1}{t}$ as eigenvalue. This completes the proof. □

Now for fixed point $\overrightarrow{p} \in \mathbb{R}^n$ let us study the function:

$$L_{\overrightarrow{p}} = f : M \to \mathbb{R}$$

where:

$$f(\overrightarrow{x}(u^1, \dots, u^k)) = \|\overrightarrow{x}(u^1, \dots, u^k) - \overrightarrow{p}\|^2 = \overrightarrow{x} \cdot \overrightarrow{x} - 2\overrightarrow{x} \cdot \overrightarrow{p} + \overrightarrow{p} \cdot \overrightarrow{p}$$

we have:

$$\frac{\partial f}{\partial u^i} = 2\frac{\partial \overrightarrow{x}}{\partial u^i} = 2\frac{\partial \overrightarrow{x}}{\partial u^i} \cdot (\overrightarrow{x} - \overrightarrow{p})$$

Thus f has a critical point at $\overrightarrow{p}$ if and only if $\overrightarrow{q} - \overrightarrow{p}$ is normal to M at $\overrightarrow{q}$.

the second partial derivatives at a critical point are given by:

$$\frac{\partial^2 f}{\partial u^i \partial u^j} = 2(\frac{\partial \overrightarrow{x}}{\partial u^i} \cdot \frac{\partial \overrightarrow{x}}{\partial u^j} + \frac{\partial^2 \overrightarrow{x}}{\partial u^i \partial u^j} \cdot (\overrightarrow{x} - \overrightarrow{p}))$$

Setting $\overrightarrow{p} = \overrightarrow{x} + t\overrightarrow{v}$, it becomes:

$$\frac{\partial^2 f}{\partial u^i \partial u^j} = 2(g_{ij} - t\overrightarrow{v} \cdot \overrightarrow{l_{ij}})$$

Therefore:

Lemma 1.5.10. *The point $\overrightarrow{q} \in M$ is a degenerate critical point of $f = L_{\overrightarrow{p}}$ if and only if $\overrightarrow{p}$ is a focal point of $(M, \overrightarrow{q})$. The nullity of $\overrightarrow{q}$ as critical point is equal to the multiplicity of $\overrightarrow{p}$ as focal point.*

Theorem 1.5.11. *For almost all $\overrightarrow{p} \in \mathbb{R}^n$ the function $L_{\overrightarrow{p}}$ has no degenerate critical points.*

Corollary 1.5.12. *On any manifold M there exists a differentiable function, with no degenerate critical points, for which each M^a is compact.*

Corollary 1.5.13. *A differentiable manifold has the homotopy type of a CW-complex.*

Corollary 1.5.14. (*Poincaré-Hopf theorem*) *On a compact manifold M there is a vector field X such that the sum of the indices of the critical points of X equals $\chi(M)$, the Euler characteristic of M.*

Proof. For any differentiable function f on M we have

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

where C_λ is the number of critical points with index λ. But $(-1)^\lambda$ is the index of the vector field $\operatorname{grad} f$ at a point where f has index λ. □

It follows that the sum of the indeces of any vector field on M is equal to $\chi(M)$ because this sum is a topological invariant.

The preceding corollary can be sharpend as follows: Let $k \geq 0$ be an integer and let $K \subset M$ be a compact set:

Corollary 1.5.15. *Any bounded smooth function $f : M \to \mathbb{R}$ can be uniformly approximated by a smooth function g which has no degenerate critical points. Furthermore g can be chosen so that i-th derivatives of g on the compact set K uniformly approximate the corresponding derivatives of f, for $i \leq k$*

Lemma 1.5.16. (*Index theorem for L_p*) *The index of L_p at a non-degenerate critical point $q \in M$ is equal to the numberof focal point of (M, q) which lie on the segment from q to p each focal point being counted with its multiplicity.*

1.6. The Lefschetz Theorem on Hyperplane Sections.

Theorem 1.6.1. *If $M \subset \mathbb{C}^n$ is a non-singular affine algebraic variety in complex n-space with real dimension $2k$, then:*

$$H_i(M; \mathbb{Z}) = 0, \quad i > k$$

This is a consequence of the stronger:

Theorem 1.6.2. *A complex analytic manifold M of complex dimension k, bianalytically embedded as a closed subset of $\mathbb{C}^n$ has the homotopy type of a k-dimensional CW-complex.*

2. A RAPID COURSE IN RIEMANNIAN GEOMETRY

2.1. Covariant Differentiation.

Let M be a smooth manifold.

Definition 2.1.1. *An **affine connection** at a point $p \in M$ is a function which assigns to each tangent vector $X_p \in TM_p$ and to each vector field Y a new tangent vector:*

$$X_p \vdash Y \in TM_p$$

*called the **covariant derivative*** of Y in the direction X_p. This is required to be bilinear as a function of X_p and Y. Furthermore, if*

$$f : M \to \mathbb{R}$$

is a real valued function. and if fY denotes the vector field

$$(fY)_q = f(q)Y_q$$

then $\vdash$ is required to satisfy the identity:

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

Definition 2.1.2. *A **Global affine connection** on M is a function which assigns to each $p \in M$ an affine connection $\vdash_p$ at p, satisfying the following smoothness condition:*

1. *If X and Y are smooth vector fields on M then the vector field $X \vdash Y$, defined by the identity*
$$(X \vdash Y)_p = X_p \vdash_p Y$$
is also be smooth.
2. *$X \vdash Y$ is bilinear as a function of X and Y.*
3. *$(fX) \vdash Y = f(X \vdash Y)$*
4. *$X \vdash (fY) = (Xf)Y + f(X \vdash Y)$*

In terms of local coordinates $u^1, \ldots, u^n$ defined on a coordinate neighborhood $U \subset M$, the connection $\vdash$ is determined by n^3 smooth real valued functions r_{ij}^k on U as follows. Let ∂_k denote the vector field $\dfrac{\partial}{\partial u^k}$ on U. Then any vector field X on U can be expressed as

$$X = \sum_k x^k \partial_k$$

where x_k are real valued functions on U. In particular the vector field $\partial_i \vdash \partial_j$ can be expressed as:

$$\partial_i \vdash \partial_j = \sum_k r_{ij}^k \partial_k$$

This functions determine the connection completely on U. In fact, if $X = \sum x^i \partial_i, Y = \sum y^j \partial_j$, one can expand $X \vdash Y$ as follows:

$$X \vdash Y = \sum_k (\sum_i x^i y_i^k) \partial_k$$

*Note that our $X \vdash Y$ coincides with $\nabla_X Y$

where

$$y_i^k = \partial_i y^k + \sum_j r_{ij}^k y^j$$

Using the connection one can define the covariant derivative of a vector field along a curve in M.

Definition 2.1.3. *A* ***parametrized curve*** *in M is a smooth function c from the real number to M. A* ***vector field*** *V* ***along*** *the curve c is a function which assigns to each $t \in \mathbb{R}$ a tangent vector.*

$$V_t \in TM_{c(t)}$$

This is required to be smooth in the following sense: For any smooth function f on M the correspondence.

$$t \to V_t f$$

should define a smooth function on $\mathbb{R}$.

Example 2.1.4. ***velocity*** *vector field* $\dfrac{\mathrm{d}c}{\mathrm{d}t}$ *of the curve is the vector field along c which is defined by the rule:*

$$\frac{\mathrm{d}c}{\mathrm{d}t} = c_* \frac{\mathrm{d}}{\mathrm{d}t}$$

Here $\frac{\mathrm{d}}{\mathrm{d}t}$ *denotes the standard vector field on the real number, and:*

$$c_* : T\mathbb{R}_t \to TM_{c(t)}$$

denotes the homomorphism of tangent spaces induced by the map c.

Now suppose that M is provided with an affine connection.

Definition 2.1.5. *Any vector field V along c determines a new vector field* $\frac{\mathrm{D}V}{\mathrm{d}t}$ *along c called the* ***covariant derivative*** *of V. The operation is characterized by the following three axioms:*

1. $\frac{\mathrm{D}(V+W)}{\mathrm{d}t} = \frac{\mathrm{D}V}{\mathrm{d}t} + \frac{\mathrm{D}W}{\mathrm{d}t}$
2. *If f is a smooth real valued function on $\mathbb{R}$ then* $\frac{\mathrm{D}(fV)}{\mathrm{d}t} = \frac{\mathrm{d}f}{\mathrm{d}t}V + f\frac{\mathrm{D}V}{\mathrm{d}t}$
3. *If V is induced by a vector field Y on M, that is if $V_t = Y_{c(t)}$ for each t, then* $\frac{\mathrm{D}V}{\mathrm{d}t}$ *is qual to* $\frac{\mathrm{d}c}{\mathrm{d}t} \vdash Y$

Lemma 2.1.6. *There is one and only one operation* $V \to \frac{\mathrm{D}V}{\mathrm{d}t}$ *which satisfies these conditions.*

Proof. Choose a local coordinate system for M, and let $u^1(t), \dots, u^n(t)$ denote the coordinates of the point $c(t)$. The vector field V can be expressed uniquely in the form

$$V = \sum v^j \partial_j$$

where $v^1, \dots, v^n$ are real valued functions on $\mathbb{R}$, and $\partial_1, \dots, \partial_n$ are the standard vector field on the coordinate neighborhood. It follows from $1, 2$ and 3 that

$$\frac{\mathrm{D}V}{\mathrm{d}t} = \sum_j \left(\frac{\mathrm{d}v^j}{\mathrm{d}t}\partial_j + v^j \frac{\mathrm{d}c}{\mathrm{d}t} \vdash \partial_j\right) = \sum_k \left(\frac{\mathrm{d}v^j}{\mathrm{d}t} + \sum_{i,j} \frac{\mathrm{d}u^i}{\mathrm{d}t} r^k_{ij} v^j\right)\partial_k$$

Conversely, defining $\frac{\mathrm{D}V}{\mathrm{d}t}$ by this formula, it is not difficult to verify that $1, 2$and 3 are satisfied. ☐

Definition 2.1.7. *A vector field V along c is said to be **parallel vector field** if the covariant derivative $\frac{\mathrm{D}V}{\mathrm{d}t}$ is zero.*

Lemma 2.1.8. *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

$$\frac{\mathrm{d}v^k}{\mathrm{d}t} + \sum_{i,j} \frac{\mathrm{d}u^i}{\mathrm{d}t} r^k_{ij} v^i = 0$$

has 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 relevant values of t. ☐

Definition 2.1.9. *The vector V_t is said to be obtained from V_0 by **parallel translation** along c.*

Definition 2.1.10. *A connection $\vdash$ on M is **compatible** with the Riemannian metric 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.*

Lemma 2.1.11. *Suppose that the connection is compatible with the metric. Let V, W be any two vector fields along c. Then:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle V, W\rangle = \langle \frac{\mathrm{D}V}{\mathrm{d}t}, W\rangle + \langle V, \frac{\mathrm{D}W}{\mathrm{d}t}\rangle$$

Proof. Choose parallel vector fields $P_1, \dots, P_n$ along c which are orthogonal at one point of c and hence at every point of c. Then the given fields V and W can be expressed as $\sum v^i P_i$ and $\sum w^j P_j$ respectively. If follows that $\langle V, W\rangle = \sum v^i w^j$ and that

$$\frac{\mathrm{D}V}{\mathrm{d}t} = \sum \frac{\mathrm{d}v^i}{\mathrm{d}t} P_i, \quad \frac{\mathrm{D}W}{\mathrm{d}t} = \sum \frac{\mathrm{d}w^j}{\mathrm{d}t} P_j$$

Therefore

$$\langle \frac{\mathrm{D}V}{\mathrm{d}t}, W\rangle + \langle V, \frac{\mathrm{D}W}{\mathrm{d}t} = \sum \left(\frac{\mathrm{d}v^i}{\mathrm{d}t} w^i + v^i \frac{\mathrm{d}w^i}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\langle V, W\rangle$$

which completes the proof. ☐

Corollary 2.1.12. *Given any vector field Y, Y' on M and any vector $X_p \in TM_p$, then:*

$$X_p\langle Y, Y'\rangle = \langle X_p \vdash Y, Y'\rangle + \langle Y, X_p \vdash Y'\rangle$$

Proof. Choose a curve c whoose velocity vector at $t = 0$ is X_p. □

Definition 2.1.13. *A connection $\vdash$ is called* ***symmetric***[†]*if it satisfies the identity*

$$(X \vdash Y) - (Y \vdash X) = [X, Y]$$

Remark 2.1.14. *Applying this identity to the case $X = \partial_i, Y = \partial_j$, then since $[\partial_i, \partial_j] = 0$, one obtains:*

$$r^k_{ij} = r^k_{ji}$$

Conversely, if this relation holds, the connection $\vdash$ is also symmetric.

Lemma 2.1.15. (*Fundamental lemma of Riemannian geometry*)*A Riemannian manifold possesses one and only one symmetric connection which is compatible with its metric.*

Proof. Applying corollary 2.1.12 to the vector fields $\partial_i, \partial_j, \partial_k$ and setting $\langle\partial_j, \partial_k\rangle = g_{jk}$ one obtains the identity

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

Permuting i, j and k this gives three linear equations relating the three quantities

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

There are only three such quantities since $\partial_i \vdash \partial_j = \partial_j \vdash \partial_i$. These equations can be solved uniquely; yielding the **first Christoffel identity**

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

The left hand side of this identity is equal to $\sum_l r^l_{ij} g_{lk}$. Multiplying by the inverse (g^{kl}) of the matrix (g_{lk}) this yields the **second Christoffel identity**

$$r^l_{ij} = \sum_k \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) g^{kl}$$

Thus the connection is uniquely determined by the metric.

Conversely, defining r^l_{ij} by this formula, one can verify that the resulting connection is symmetric and compatible with the metric. This completes the proof. □

[†]The following reformulation may seem more intuitive: Define **covariant second derivative** of a real valued function f along two vectors X_p, Y_p to be expression:

$$X_p(Yf) - (X_p \vdash Y)f$$

where Y denotes any vector field extending Y_p. Then the connection is symmetric if this second derivative is symmtric as a function of X_p, Y_p.

An alternative characterization of symmetry will be very useful later. Consider a "parametrized surface" in M: that is a smooth function:

$$s : \mathbb{R}^2 \to M$$

By a vector field V along s is meant a function which assigns to each $(x, y) \in \mathbb{R}^2$ a tangent vector $V_{(x,y)} \in TM_{s(x,y)}$. As an example, the two standard vector fields $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$ give rise to vector fields $s_* \frac{\partial}{\partial x}, s_* \frac{\partial}{\partial y}$ along s. These will be denoted briefly by $\frac{\partial s}{\partial x}, \frac{\partial s}{\partial y}$

For any smooth vector field V along s the **covariant derivatives** $\frac{\mathrm{D}V}{\mathrm{d}x}, \frac{\mathrm{D}V}{\mathrm{d}y}$ are constructed as follows. For each fixed y_0, restricting V to the curve

$$x \to s(x, y_0)$$

one obtains a vector field along this curve. Its covariant derivative with respect to x is defined to be $(\frac{\mathrm{D}V}{\partial x})_{(x,y_0)}$. This defines $\frac{\mathrm{D}V}{\partial x}$ along s.

Lemma 2.1.16. *If the connection is symmetric then* $\frac{\mathrm{D}}{\partial x}\frac{\partial s}{\partial y} = \frac{\mathrm{D}}{\partial y}\frac{\partial s}{\partial x}$

2.2. **The curvature Tensor.**

Definition 2.2.1. *Given vector fields X, Y, Z define a new vector field $R(X,Y)Z$ by the identity*

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

called it ***curvature****.*

Lemma 2.2.2. *The value of $R(X,Y)Z$ at a point $p \in M$ depends only on the vectors X_p, Y_p, Z_p at this point and not on their values at nearby points. Furthermore the correspondence:*

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

is tri-linear. Briefly, this lemma can be expressed by saying that R is a "tensor".

Proof. Clearly $R(X,Y)Z$ is a tri-linear function of X, Y and Z. If X is replaced by a multiple fX then the three terms $-X \vdash (Y \vdash Z), Y \vdash (X \vdash Z), [X,Y] \vdash Z$ are replaced respectively by $-fX \vdash (Y \vdash Z), (Yf)(X \vdash Z) + fY \vdash (X \vdash Z), -(Yf)(X \vdash Z) + f[X,Y] \vdash Z$.

Adding thest three terms one obtains the identity

$$R(fX,Y)Z = fR(X,Y)Z$$

Corresponding identities for Y and Z are easily obtained by similar computaions.

Now suppose that $X = \sum x^i \partial_i, Y = \sum y^j \partial_j, Z = \sum z^k \partial_k$. Then

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

Evaluating this expression at p one obtains the formula

$$(R(X,Y)Z)_p = \sum x^i(p) y^j(p) z^k(p) (R(\partial_i, \partial_j)\partial_k)_p$$

which depends only on the values of the function x^i, y^j, z^k at p, and not on their values at nearby points. This completes the proof. $\square$

Now consider a parametrized surface

$$s : \mathbb{R}^2 \to M$$

Given any vector field V along s, one can apply the two covariant differentiation operators $\frac{\mathrm{D}}{\partial x}, \frac{\mathrm{D}}{\partial y}$ to V. In general these operators will not commute with each other.

Lemma 2.2.3.

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

Proof. Express both sides in terms of a local coordinate system, and compute, making use of the identity

$$\partial_j \vdash (\partial_i \vdash \partial_k) - \partial_i(\partial_j \vdash \partial_k) = R(\partial_i, \partial_j)\partial_k$$

$\square$

It's interesting to ask whether one can construct a vector field P along s which is parallel, in the sense that

$$\frac{\mathrm{D}P}{\partial x} = \frac{\mathrm{D}P}{\partial y} = 0$$

In general no such vector field exists. However, if the curvature tensor happens to be zero then P can be constructed as follows. Let $P_{(x,0)}$ be a parallel vector field along x-axis, satisfying the given initial condition. Clearly $\frac{\mathrm{D}P}{\partial y}$ is identically zero, and $\frac{\mathrm{D}V}{\partial x}$ is zero along x-axis. Now the identity

$$\frac{\mathrm{D}}{\partial y}\frac{\mathrm{D}P}{\partial x} - \frac{\mathrm{D}}{\partial x}\frac{\mathrm{D}P}{\partial y} = R(\frac{\partial s}{\partial x}, \frac{\partial s}{\partial y})P = 0$$

implies $\frac{\mathrm{D}}{\partial y}\frac{\mathrm{D}P}{\partial x} = 0$. In other words, the vector field $\frac{\mathrm{D}P}{\partial x}$ is parallel along the curve

$$y \to s_{(x_0,y)}$$

Since $(\frac{\mathrm{D}P}{\partial x})_{(x_0,0)} = 0$, implies $\frac{\mathrm{D}P}{\partial x}$ is identically zero! and completes the proof that P is parallel along s.

Henceforth we will assume that M is a Riemannian manifold, provided with the unique symmetric connection which is compatible with its metric.

Lemma 2.2.4. *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. $\langle R(X,Y)Z, W\rangle + \langle R(Z,W)X, Y\rangle$

Proof. The skew-symmetry relaion 1. follows immediately from the definition of R.

Since all three terms of 2. are tensors, it is sufficient to prove 2. when the bracket product $[X,Y],[X,Z],[Y,Z]$ are all zero. Under this hypothesis we must verify the identity

$$\begin{aligned} &- X \vdash (Y \vdash Z) + Y \vdash (X \vdash Z) \\ &- Y \vdash (Z \vdash X) + Z \vdash (Y \vdash X) \\ &- Z \vdash (X \vdash Y) + X \vdash (Z \vdash Y) = 0 \end{aligned}$$

But the symmetry of the connection implies that

$$Y \vdash Z - Z \vdash Y = [Y,Z] = 0$$

Thus the upper left term cancels the low right term. Similarly the remaining terms cancel in pairs. This proves 2.

To prove 3. we must show that the expression $\langle R(X,Y)Z, W\rangle$ is skew-symmetry in Z and W. This is clear equivalent to the assertion that

$$\langle R(X,Y)Z, Z\rangle = 0$$

for all X, Y, Z. Again we may assume $[X, Y] = 0$, so that $\langle R(X,Y)Z, Z\rangle = 0$ is equal to

$$\langle -X \vdash (Y \vdash Z) + Y \vdash (X \vdash Z), Z\rangle$$

In other words we must prove that the expression

$$\langle Y \vdash (X \vdash Z), Z\rangle$$

is symmetric in X and Y.

Since $[X, Y] = 0$ the expression $YX\langle Z, Z\rangle$ is symmetric in X and Y. Since the connection is compatible with the metric, we have

$$X\langle Z, Z\rangle = 2\langle X \vdash Z, Z\rangle$$

Hence

$$YX\langle Z, Z\rangle = 2\langle Y \vdash (X \vdash Z), Z\rangle + 2\langle X \vdash Z, Y \vdash Z\rangle$$

But the right hand term is clearly symmetric in X and Y. Therefore $\langle Y \vdash (X \vdash Z), Z\rangle$ is symmetric in X and Y; which completes the proof.

Property 4. may be proved from 1., 2. and 3. $\square$

2.3. Geodesics and Completeness.

Definition 2.3.1. *A parametrized path*

$$\gamma : I \to M$$

where I denotes any interval of real numbers, is called a ***geodesic*** *if the acceleration vector field $\frac{\mathrm{D}}{\mathrm{d}t}\frac{\mathrm{d}\gamma}{\mathrm{d}t}$ is identically zero. Thus the velocity vector field $\frac{\mathrm{d}\gamma}{\mathrm{d}t}$ must be parallel along γ.*

Remark 2.3.2. *If γ is a geodesic, then the identity*

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\frac{\mathrm{d}\gamma}{\mathrm{d}t},\frac{\mathrm{d}\gamma}{\mathrm{d}t}\rangle = 2\langle\frac{\mathrm{D}}{\mathrm{d}t}\frac{\mathrm{d}\gamma}{\mathrm{d}t},\frac{\mathrm{d}\gamma}{\mathrm{d}t}\rangle = 0$$

shows that the length $\|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\|$ of the velocity vector is constant along γ.

Definition 2.3.3. *Introducing the* ***arc-length*** *function*

$$s(t) = \int \|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\|\mathrm{d}t + \text{constant}$$

Remark 2.3.4. *Remark* 2.3.2 *can be rephrased as follows: The parameter t along a geodesic is a linear function of the arc-length. The parameter t is actually equal to the arc-length if and only if $\|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\| = 1$*

In terms of a local coordinate system with coordinates $u^1,\ldots,u^n$ a curve $t \to \gamma(t) \in M$ determines n smooth function $u^1(t),\ldots,u^n(t)$. The equation $\frac{\mathrm{D}}{\mathrm{d}t}\frac{\mathrm{d}\gamma}{\mathrm{d}t}$ for a geodesic then take the form

$$\frac{\mathrm{d}^2u^k}{\mathrm{d}t^2} + \sum_{i,j=1}^{n} r^k_{ij}(u^1,\ldots,u^n)\frac{\mathrm{d}u^i}{\mathrm{d}t}\frac{\mathrm{d}u^j}{\mathrm{d}t} = 0, \quad k = 1,\ldots,n$$

The existence of geodesic depends, therefore, on the solutions of a certain system of second order differential equations.

More generally consider any system of equations of the form

$$\frac{\mathrm{d}^2\overrightarrow{u}}{\mathrm{d}t^2} = \overrightarrow{F}(\overrightarrow{u},\frac{\mathrm{d}\overrightarrow{u}}{\mathrm{d}t})$$

Here $\overrightarrow{u}$ stands for $(u^1,\ldots,u^n)$ and $\overrightarrow{F}$ stands for an n-tuple of C^∞ functions, all defined throughout some neighborhoond U of a point

$$(\overrightarrow{u}_1,\overrightarrow{v}_1) \in \mathbb{R}^{2n}$$

Theorem 2.3.5. (*Existence and Uniqueness theorem*) *There exists a neighborhood W of the point $(\overrightarrow{u}_1,\overrightarrow{v}_1)$ and a number $\epsilon > 0$ so that, for each $(\overrightarrow{u}_0,\overrightarrow{v}_0) \in W$ the differential equation*

$$\frac{\mathrm{d}^2\overrightarrow{u}}{\mathrm{d}t^2} = \overrightarrow{F}(\overrightarrow{u},\frac{\mathrm{d}\overrightarrow{u}}{\mathrm{d}t})$$

has a unique solution $t \to \overrightarrow{u}(t)$ which is defined for $|t| < \epsilon$, and satisfies the initial condition

$$\overrightarrow{u}(0) = \overrightarrow{u}_0, \quad \frac{\mathrm{d}\overrightarrow{u}}{\mathrm{d}t}(0) = \overrightarrow{v}_0$$

Furthermore, the solution denpends smoothly on the initial conditions.

Lemma 2.3.6. *For every point p_0 on a Riemannian manifold M there exists a neighborhood U of p_0 and a number $\epsilon > 0$ so that: for each $p \in U$ and each tangent vector $v \in TM_p$ with length $< \varepsilon$ there is a unique geodesic*

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

satisfying the conditions

$$\gamma_v(0) = p, \quad \frac{\mathrm{d}\gamma_v}{\mathrm{d}t}(0) = v$$

Proof. If we were willing to replace the interval $(-2, 2)$ by an arbitrary small interval, then this statement would follow immediately from above theorem. To be more precise, there exists a neighborhood U of p_0 and number $\varepsilon_1, \varepsilon_2 > 0$ so that: for each $p \in U$ and each $v \in TM_p$ 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 $t \to \gamma(t)$ is a geodesic, then the parametrized curve $t \to \gamma(ct)$ will also be a geodesic.

Now suppose that ε is smaller than $\varepsilon_1\varepsilon_2$. Then if $\|v\| < \varepsilon$ and $|t| < 2$ not that

$$\|v/\varepsilon_2\| < \varepsilon_1 \quad \text{and} \quad |\varepsilon_2 t| < 2\varepsilon_2$$

Hence we can define $\gamma_v(t)$ to be $\gamma_{v/\varepsilon_2}(\varepsilon_2 t)$. This completes the proof. □

This following notation will be convenient.

Definition 2.3.7. *Let $v \in TM_q$ be a tangent vector, and suppose that there exits a geodesic*

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

satisfying the conditions

$$\gamma(0) = q, \quad \frac{\mathrm{d}\gamma}{\mathrm{d}t}(0) = v$$

*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 γ can thus be described by the formula*

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

Lemma 2.3.6 says that $\exp_q(v)$ is defined providing 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 2.3.8. *The manifold M is **geodesic complete** if $\exp_q(v)$ is defined for all $q \in M$ and all vector $v \in TM_q$.*

Remark 2.3.9. *The above is clearly equivalent to the following requirement: For every geodesic segment $\gamma_0 : [a, b] \to M$ it should be possible to extend γ_0 to an infinite geodesic*

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

Let TM be the tangent manifold of M, consisting of all pairs (p, v) with $p \in M, v \in TM_p$. We give TM the following C^∞ structure: If $(u^1, \dots, u^n)$ is a coordinate system in an open set $U \subset M$ then every tangent vector at $q \in U$ can be expressed uniquely as $t^1\partial_1 + \cdots + t^n\partial_n$, where $\partial_i = \frac{\partial}{\partial u^i}|_q$. Then the functions $u^1, \dots, u^n, t^1, \dots, t^n$ constitute a coordinate system on the open set $TU \subset TM$.

Lemma 2.3.6 says that for each $p \in M$ the map

$$(q, v) \to \exp_q(v)$$

is defined throught a neighborhood V of the point $(p, 0) \in TM$. Furthermore this map is differentiable throughout V.

Now consider the smooth function $F : V \to M \times M$ defined by $F(q, v) = (q, \exp_q(v))$. We claim that the Jacobian of F at the point $(p, 0)$ is non-singular. In fact, denoting the induced coordinates on $U \times U \subset M \times M$ by $(u_1^1, \dots, u_1^n, u_2^1, \dots, u_2^n)$, we have

$$F_*(\frac{\partial}{\partial u^i}) = \frac{\partial}{\partial u_1^i} + \frac{\partial}{\partial u_2^i}$$

$$F_*(\frac{\partial}{\partial t^j}) = \frac{\partial}{\partial u_2^j}$$

Thus the Jacobian matrix of F at $(p, 0)$ is non-singular.

It follows from the implicit function theorem that F maps some neighborhood V' of $(p, 0) \in TM$ diffeomorphically onto some neighborhood of $(p, p) \in M \times M$. 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 $W \times W \subset F(V')$. Then we have proven the following.

Lemma 2.3.10. *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.*
3. *For each $q \in W$ the map $\exp_q$ maps the open ε-ball in TM_q diffeomorphically onto an open set $W \subset U_q$.*

Remark 2.3.11. *With more care it would be possible to choose W so that the geodesic joining any two of its points lies completely within W, i.e. W is a convex neighborhood of p.*

Theorem 2.3.12. *Let W and ε be as in Lemma* 2.3.10. *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

$$\int_0^1 \|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\|\mathrm{d}t \leq \int_0^1 \|\frac{\mathrm{d}\omega}{\mathrm{d}t}\|\mathrm{d}t$$

where equality can hold only if the points set $\omega([0,1])$ coincides with $\gamma([0,1])$. Thus γ is the shortest path joining its end points.

Proof. The proof will be based on two lemmas. Let $q = \gamma(0)$ and let U_q be as in lemma 2.3.10.

Lemma 2.3.13. *In U_q, the geodesics through q are the orthogonal trajectories of hypersurfaces*

$$\{\exp_q(v) : v \in TM_q, \|v\| = \textit{constant}\}$$

Proof. Let $t \to v(t)$ defines any curve in TM_q with $\|v(t\| = 1$. We must show that the corresponding curves

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

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

$$r \to \exp_q(rv(t_0))$$

In terms of the parametrized sufface f given by

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

we must prove

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

for all (r,t).

Now

$$\frac{\partial}{\partial r}\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle = \langle \frac{\mathrm{D}}{\partial r}\frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle + \langle \frac{\partial f}{\partial r}, \frac{\mathrm{D}}{\partial r}\frac{\partial f}{\partial t} \rangle$$

The first expression on the right is zero since the curves

$$r \to f(r,t)$$

are geodesics. Tbe second expression is equal to

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

since $\|\frac{\partial f}{\partial r}\| = \|v(t)\| = 1$. Therefore the quantity $\langle \frac{\partial f}{\partial r}, \frac{\partial f}{\partial t} \rangle$ is independent of r. But for $r = 0$ we have

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

hence $\frac{\partial f}{\partial t}(0,t)=0$. Therefore $\langle\frac{\partial f}{\partial r},\frac{\partial f}{\partial t}\rangle$ is identically zero, which completes the proof. $\square$

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 $\|v(t)\|=1, v(t)\in TM_q$.

Lemma 2.3.14. *The length $\int_a^b\|\frac{\mathrm{d}\omega}{\mathrm{d}t}\|$ 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

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

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

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

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

$$\int_a^n\|\frac{\mathrm{d}\omega}{\mathrm{d}t}\|\mathrm{d}t\geq\int_a^b|r'(t)|\mathrm{d}t\geq|r(b)-r(a)|$$

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

The proof of theorem 2.3.12 is now straightforward. Consider any piecewise smooth path ω from q to a point

$$q'=\exp_q(rv)\in U_q$$

where $0<r<\varepsilon,\|v\|=1$. Then for any $\delta>0$ the path ω must contains a segment joining the spherical shell of radius δ to the spherical shell of radius r, and lying between these two shells. The length of this segment will be $\geq r-s$; hence letting δ tend to 0 the length of ω 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. $\square$

Corollary 2.3.15. *Suppose that a path $\omega:[0,l]\to M$, parametrized by arc-length, has length less than or equal to the length of any other path from $\omega(0)$ to $\omega(l)$. Then ω is a geodesic.*

Proof. Consider any segment of ω lying within an open set W, as above, and having length $<\varepsilon$. This segment must be a geodesic by theorem 2.3.12. Hence the entire path ω is a geodesic. $\square$

Definition 2.3.16. *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.*

Remark 2.3.17. *Theorem* 2.3.12 *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 π, it is certainly not minimal.*

And in general, minimal geodesics are not unique. For example two antipodal points on a unit sphere are joined by infinitely many minimal geodesic.

Definition 2.3.18. *Define the* ***distance*** *$\rho(p, q)$ between two points $p, q \in M$ to be the greatest lower bound for the arc-length of piecewise smooth paths joining these points.*

Corollary 2.3.19. *Given a compact set $K \subset M$ there exists a number $\delta > 0$ so that any two points of K with distance less than δ are joined by a unique geodesic of length less than δ. Furthermore this geodesic is minimal and depends differentiably on its endpoints.*

Proof. Cover K by open set W_α as in lemma 2.3.10, and let δ be small enough so that any two points of K with distance less than δ lie in a common W_α. This completes the proof. ☐

Theorem 2.3.20. (*Hopf and Rinow*) *If M is geodesically complete, then any two points can be joined by a minimal geodesic.*

Corollary 2.3.21. *If M is geodesically complete then every bounded subset of M has compact closure. Consequently M is complete as a metric space.*

Remark 2.3.22. *Conversely, if M is complete as a metric space, then it is not difficult to prove M is geodesically complete. Henceforth we will not distinguish between geodesic completeness and metric completeness, but will refer simply to a* ***complete Riemannian manifold****.*

Now we will show some familiar examples of geodesics

Example 2.3.23. *In Euclidean n-space, $\mathbb{R}^n$, with the usual coordinate system $x_1, \ldots, x_n$ and the usual Riemannian metric $\mathrm{d}x_1 \otimes \mathrm{d}x_1 + \cdots + \mathrm{d}x_n \otimes \mathrm{d}x_n$ we have $r_{ij}^k = 0$ and the equations for a geodesic γ, given by $t \to (x_1(t), \ldots, x_n(t))$ becomes*

$$\frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} = 0$$

whose solutions are the straight lines. This could also have been seen as follows: it is easy to show that the formula for arc length

$$\int (\sum_{i=1}^{n} (\frac{\mathrm{d}x_i}{\mathrm{d}t})^2)^{\frac{1}{2}} \mathrm{d}t$$

coincides with the usual definitions of arc length as the least upper bound of the lengths of inscribed polygons; from this definition it is clear that straight lines have minimal length, and are therefore geodesics.

Example 2.3.24. *The geodesics on* S^n *are precisely the greatest circles, that is, the intersections of* S^n *with the planes through the center of* S^n*. Reflection through a plane* E^2 *is an isometry* $I : S^n \to S^n$ *whose fixed point set is* $C = S^n \cap E^2$*. Let* x *and* y *be two points of* C *with a unique geodesic* C' *of minimal length between them. Then, since* I *is an isometry, the curve* $I(C')$ *is a geodesic of the same length as* C' *between* $I(x) = x, I(y) = y$*. Therefore* $C' = I(C')$*. This implies that* $C' \in C$

Finaly, since there is a great circle through any point of S^n *in any given direction, these are all the geodesics.*

Antipodal points on the sphere have a continium of geodesics of minimal length between them. All other pairs of points have a unique geodesic of minimal length between them, but an infinite family of non-minimal geodesics, depending on how many times the geodesic goes around the sphere and in which direction it starts.

By the same reasoning every meridian line on a surface of revolution is a geodesic.

3. THE CALCULUS OF VARIATIONS APPILED TO GEODESICS

3.1. **The Path Space of a Smooth Manifold.**

Definition 3.1.1. *Let M be a smooth manifold and let p and q be two points of M. By a* ***piecewise smooth path*** *from p to q will be meant a map $\omega : [0,1] \to M$ such that:*

1. *there exists a subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$ of $[0,1]$ so that each $\omega|_{[t_{i-1},t_i]}$ is differentiable of class C^∞;*
2. $\omega(0) = p, \omega(1) = q$.

The set of all piecewise smooth paths from p to q in M will be denoted by $\Omega(M; p, q)$, or briefly by $\Omega(M)$ or Ω.

Remark 3.1.2. *Later Ω will be given the structure of a topological space, but for the moment this will not be necessary. We will think of Ω as being something like an "infinite dimensional manifold".*

Definition 3.1.3. *By the* ***tangent space*** *of Ω at a path ω will be meant the vector space consisting of all piecewise smooth vector fields W along ω for which $W(0) = 0, W(1) = 0$. The notation $T\Omega_\omega$ will be used for this vector space.*

If F is a real valued function on Ω it is natural to ask what

$$F_* : T\Omega_\omega \to T\mathbb{R}_{F(\omega)}$$

the induced map on the tangent space, should mean. When F is a function which is smooth in the usual sense. Given $X \in TM_p$ choose a smooth path $u \to \alpha(u)$ in M, which is defined for $-\varepsilon < u < \varepsilon$, so that

$$\alpha(0) = p, \quad \frac{\mathrm{d}\alpha}{\mathrm{d}u}(0) = X$$

Then $F_*(X)$ is equal to $\dfrac{\mathrm{d}(F(\alpha(u)))}{\mathrm{d}u}|_{u=0}$, multiplied by the basis vector $(\dfrac{\mathrm{d}}{\mathrm{d}t})|_{F(p)} \in T\mathbb{R}_{F(p)}$

In order to carry out an analogous construction for $F : \Omega \to \mathbb{R}$, the following concept is needed.

Definition 3.1.4. *A* ***variation*** *of ω 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 $\alpha(u,t) = \bar{\alpha}(u)(t)$ is C^∞ on each strip $(-\varepsilon, \varepsilon) \times (t_{i-1}, t_i], i = 1, \ldots, k$.

Since each $\bar{\alpha}(u)$ belongs to $\Omega(M;p,q)$, note that:

$$\alpha(u,0)=p,\quad \alpha(u,1)=q,\quad \forall u\in(-\varepsilon,\varepsilon)$$

More generally, in the above definition, $(-\varepsilon,\varepsilon)$ is replaced by a neighborhood of 0 in $\mathbb{R}^n$, then α or $\bar{\alpha}$ is called an ***n-parameter variation*** *of ω.*

Definition 3.1.5. *Now $\bar{\alpha}$ may be considered as a "smooth path" in Ω. Its "velocity vector" $\dfrac{\mathrm{d}\bar{\alpha}}{\mathrm{d}u}\in T\Omega_\omega$ is defined to be the vector field W along ω given by*

$$W_t=\frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}u}(0)_t=\frac{\partial\alpha}{\partial u}(0,t)$$

This vector field W is also called the ***variation vector field*** *associated with variation α.*

Given any $W\in T\Omega_\omega$ note that there exists a variation $\bar{\alpha}:(-\varepsilon,\varepsilon)\to\Omega$ which satisfies the condition $\bar{\alpha}(0)=\omega,\dfrac{\mathrm{d}\bar{\alpha}}{\mathrm{d}u}(0)=W$.

By analogy with the definition given above, if F is a real valued function on Ω, we attempt to define

$$F_*:T\Omega_\omega\to T\mathbb{R}_{F(\omega)}$$

as follows.

Definition 3.1.6. *Given $W\in T\Omega_\omega$ choose a variation $\bar{\alpha}:(-\varepsilon,\varepsilon)\to\Omega$ with*

$$\bar{\alpha}(0)=\omega,\frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}u}(0)=W$$

and set $F_*(W)$ equal to $\dfrac{\mathrm{d}(F(\bar{\alpha}(u)))}{\mathrm{d}u}|_{u=0}$ multiplied by the tangent vector $(\dfrac{\mathrm{d}}{\mathrm{d}t})_{F(\omega)}$

Definition 3.1.7. *A path ω is a* ***critical path*** *for a function $F:\Omega\to\mathbb{R}$ if and only if $\dfrac{\mathrm{d}F(\bar{\alpha}(u))}{\mathrm{d}u}|_{u=0}$ is zero for every variation $\bar{\alpha}$ of ω.*

3.2. **The Energy of a Path.**

Definition 3.2.1. *For $\omega \in \Omega$ define the* ***energy*** *of ω from a to b as*

$$E_a^b(\omega) = \int_a^b \|\frac{\mathrm{d}\omega}{\mathrm{d}t}\|^2 \mathrm{d}t$$

we will use E to denote E_0^1.

Remark 3.2.2. *This can be compared with the arc-length from a to b as follows. Applying Schwarz's inequality*

$$(\int_a^b fg)^2 \leq (\int_a^b f^2)(\int_a^b g^2)$$

with $f = 1, g = \|\frac{\mathrm{d}\omega}{\mathrm{d}t}\|$, we see that

$$(L_a^b)^2 \leq (b-a)E_a^b$$

where equality holds if and only if g is constant; that is if and only if the parameter t is proportional to arc-length.

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

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

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

Lemma 3.2.3. *Let M be a complete Riemannian manifold and let $p, q \in M$ have distance d. Then the energy function takes on its minimum d^2 precisely on the set of minimal geodesic from p to q.*

We will now see which paths $\omega \in \Omega$ are critical paths for the energy function E.

Let $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ be a variation of ω, and let $W_t = \dfrac{\partial \alpha}{\partial u}(0,t)$ be the associated variation vector field. Furthermore, let

$$V_t = \frac{\mathrm{d}\omega}{\mathrm{d}t}, \quad A_t = \frac{\mathrm{D}}{\mathrm{d}t}\frac{\mathrm{d}\omega}{\mathrm{d}t}$$

and $\Delta_t V = V_{t_+} - V_{t_-}$ be the discontinuity in the velocity vector at t, where $0 < t < 1$.

Theorem 3.2.4. (*Fisrt variation formula*)

$$\frac{1}{2}\frac{\mathrm{d}E(\bar{\alpha}(u))}{\mathrm{d}u}|_{u=0} = -\sum_t \langle W_t, \Delta_t V\rangle - \int_0^1 \langle W_t, A_t\rangle \mathrm{d}t$$

Proof. Since we have
$$\frac{\partial}{\partial u}\langle\frac{\partial\alpha}{\partial t},\frac{\partial u}{\partial t}\rangle=2\langle\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial t},\frac{\partial\alpha}{\partial t}\rangle$$
Therefore
$$\frac{\mathrm{d}E(\bar{\alpha}(u))}{\mathrm{d}u}=\frac{\mathrm{d}}{\mathrm{d}u}\int_0^1\langle\frac{\partial\alpha}{\partial t},\frac{\partial\alpha}{\partial t}\rangle=2\int_0^1\langle\frac{\mathrm{D}}{\partial u}\frac{\partial\alpha}{\partial t},\frac{\partial\alpha}{\partial t}\rangle$$
and we can Substitute $\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial u}$ for $\frac{\mathrm{D}}{\partial u}\frac{\partial\alpha}{\partial t}$ in this last formula.

Choose $0=t_0<t_1<\cdots<t_k=1$ so that α is differentiable on each strip $(-\varepsilon,\varepsilon)\times[t_{i-1},t_i]$. Then we can integrate by parts on $[t_{i-1},t_i]$ as follows. The identity
$$\frac{\partial}{\partial t}\langle\frac{\partial\alpha}{\partial u},\frac{\partial u}{\partial t}\rangle=\langle\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial u},\frac{\partial\alpha}{\partial t}\rangle+\langle\frac{\partial\alpha}{\partial u},\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial t}\rangle$$
implies that
$$\int_{t_{i-1}}^{t_i}\langle\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial u},\frac{\partial\alpha}{\partial t}\rangle\mathrm{d}t=\langle\frac{\partial\alpha}{\partial u},\frac{\partial\alpha}{\partial t}\rangle|_{(t_{i-1})_+}^{(t_i)_-}-\int_{t_{i-1}}^{t_i}\langle\frac{\partial\alpha}{\partial u},\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial t}\rangle\mathrm{d}t$$
Adding up the corresponding formula for $i=1,\dots,k$; and using the fact that $\frac{\partial\alpha}{\partial u}=0$ for $t=0,1$, this gives
$$\frac{1}{2}\frac{\mathrm{d}E(\bar{\alpha}(u))}{\mathrm{d}u}=-\sum_{i=1}^{k-1}\langle\frac{\partial\alpha}{\partial u},\Delta_{t_i}\frac{\partial\alpha}{\partial t}\rangle-\int_0^1\langle\frac{\partial\alpha}{\partial u},\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial t}\rangle\mathrm{d}t$$
Setting $u=0$, we now obtain the required formula. This completes the proof. □

Corollary 3.2.5. *The path ω is a critical path for the function E if and only if ω is a geodesic.*

Proof. Clearly a geodesic is a critical point. Let ω be a critical path. There is a variation of ω with $W(t)=f(t)A(t)$ where $f(t)$ is positive except that it vanishes at the t_i. Then
$$\frac{1}{2}\frac{\mathrm{d}E}{\mathrm{d}u}(0)=-\int_0^1 f(t)\langle A(t),A(t)\rangle\mathrm{d}t$$
This is zero if and only if $A(t)=0$ for all t. Hence each $\omega|_{[t_{i-1},t_i]}$ is a geodesic.

Now pick a variation such that $W(t_i)=\Delta_{t_i}V$. Then
$$\frac{1}{2}\frac{\mathrm{d}E}{\mathrm{d}u}(0)=-\langle\Delta_{t_i}V,\Delta_{t_i}V\rangle$$
If this is zero then all Δ_{t_i} is zero and ω is differentiable of class C^1, even at the points t_i. Now it follows from the uniqueness theorem for differential equations that ω is C^∞ everywhere: thus ω is an unbroken geodesic. □

3.3. The Hessian of the Energy Function at a Critical Path.

Definition 3.3.1. *Define a bilinear functional*

$$E_{**} : T\Omega_\gamma \times T\Omega_\gamma \to \mathbb{R}$$

*when γ is a critical path of the function E, i.e., a geodesic. This bilinear functional will be called the **Hessian** of E at γ.*

We define it as follows. Givene vector fields $W_1, W_2 \in T\Omega_\gamma$, choose a 2-parameter variation

$$\alpha : U \times [0,1] \to M$$

where U is a neighborhood of $(0,0)$ in $\mathbb{R}^2$, so that

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

The Hessian $E_{**}(W_1, W_2)$ will be defined to be the second partial derivative

$$\frac{\partial^2 E(\bar{\alpha}(u_1,u_2))}{\partial u_1 \partial u_2}|_{(0,0)}$$

This second derivative will be written briefly as $\dfrac{\partial^2 E}{\partial u_1 \partial u_2}(0,0)$

Theorem 3.3.2. (*Second variation formula*) *Let $\bar{\alpha} : U \to \Omega$ be a 2-parameter variation of the geodesic γ with variation vector fields*

$$W_i = \frac{\partial \bar{\alpha}}{\partial u_i}(0,0) \in T\Omega_\gamma, \quad i = 1,2$$

Then the second derivative of the energy function is satisfied

$$\frac{1}{2}\frac{\partial^2 E}{\partial u_1 \partial u_2}(0,0) = -\sum_t \langle W_2(t), \Delta_t \frac{\mathrm{D}W_1}{\mathrm{d}t}\rangle - \int_0^1 \langle W_2, \frac{\mathrm{D}^2 W_1}{\mathrm{d}t^2} + R(V, W_1)V\rangle \mathrm{d}t$$

Proof. According to first variation formula, we have

$$\frac{1}{2}\frac{\partial E}{\partial u_2} = -\sum_t \langle \frac{\partial \alpha}{\partial u_2}, \Delta_t \frac{\partial \alpha}{\partial t}\rangle - \int_0^1 \langle \frac{\partial \alpha}{\partial u_2}, \frac{\mathrm{D}}{\partial t}\frac{\partial \alpha}{\partial t}\rangle \mathrm{d}t$$

Therefore

$$\begin{aligned}\frac{1}{2}\frac{\partial^2 E}{\partial u_1 \partial u_2} = &- \sum_t \langle \frac{\mathrm{D}}{\partial u_1}\frac{\partial \alpha}{\partial u_2}, \Delta_t \frac{\partial \alpha}{\partial t}\rangle - \sum_t \langle \frac{\partial \alpha}{\partial u_2}, \frac{\mathrm{D}}{\partial u_1}\Delta_t \frac{\partial \alpha}{\partial t}\rangle \\ &- \int_0^1 \langle \frac{\mathrm{D}}{\partial u_1}\frac{\partial \alpha}{\partial u_2}, \frac{\mathrm{D}}{\partial t}\frac{\partial \alpha}{\partial t}\rangle \mathrm{d}t - \int_0^1 \langle \frac{\partial \alpha}{\partial u_2}, \frac{\mathrm{D}}{\partial u_1}\frac{\mathrm{D}}{\partial t}\frac{\partial \alpha}{\partial t}\rangle \mathrm{d}t\end{aligned}$$

Let us evaluate this expression for $(u_1, u_2) = (0, 8)$. Since $\gamma = \bar{\alpha}(0,0)$ is an unbroken geodesic, we have

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

so that the first and third terms are zero.

Rearranging the second term, we obtain
$$\frac{1}{2}\frac{\partial^2 E}{\partial u_1 \partial u_2}(0,0) = -\sum \langle W_2, \Delta_t \frac{\mathrm{D}}{\mathrm{d}t} W_1 \rangle - \int_0^1 \langle W_2, \frac{\mathrm{D}}{\partial u_1}\frac{\mathrm{D}}{\partial t} V \rangle \mathrm{d}t$$
In order to interchange the two operators $\frac{\mathrm{D}}{\partial u_1}$ and $\frac{\mathrm{D}}{\partial t}$, we need to bring in the curvature formula
$$\frac{\mathrm{D}}{\partial u_1}\frac{\mathrm{D}}{\partial t} V - \frac{\mathrm{D}}{\partial t}\frac{\mathrm{D}}{\partial u_1} V = R(\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial u_1})V = R(V, W_1)V$$
together with the identity $\frac{\mathrm{D}}{\partial u_1} V = \frac{\mathrm{D}}{\partial t}\frac{\partial \alpha}{\partial u_1} = \frac{\mathrm{D}}{\mathrm{d}t} W_1$, this yields
$$\frac{\mathrm{D}}{\partial u_1}\frac{\mathrm{D}}{\partial t} V = \frac{\mathrm{D}^2 W_1}{\mathrm{d}t^2} + R(V, W_1)V$$
Substituting this expression into above, this completes the proof. □

Corollary 3.3.3. *The expression $E_{**}(W_1, W_2)$ is a well defined symmetric and bilinear function of W_1, W_2.*

Proof. The second variation formula shows that $\frac{\partial^2 E}{\partial u_1 \partial u_2}(0,0)$ denpends only on the variation vector field 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 $\frac{\partial^2 E}{\partial u_1, \partial u_2} = \frac{\partial^2 E}{\partial u_1, \partial u_2}$. □

Remark 3.3.4. *The diagonal terms $E_{**}(W, W)$ of the bilinear pairing E_{**} can be described in terms of a 1-parameter variation of γ. In fact*
$$E_{**}(W, W) = \frac{\mathrm{d}^2 E \circ \bar{\alpha}}{\mathrm{d}^2 u}(0)$$
where $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ denotes any variation of γ with variation vector fields $\dfrac{\mathrm{d}\bar{\alpha}}{\mathrm{d}u}(0) = W$.

Lemma 3.3.5. *If γ is a minimal geodesic from p to q then the bilinear pairing E_{**} is positive semi-definite, Hence the index of E_{**} is zero.*

Proof. The inequality $E(\bar{\alpha}(u)) \geq E(\gamma) = E(\bar{\alpha}(0))$ implies that $\frac{\mathrm{d}^2 E(\bar{\alpha}(u))}{\mathrm{d}u^2}$, evaluated at $u = 0$ is ≥ 0. Hence $E_{**}(W, W) \geq 0$ for all W. □

3.4. Jacobi Fields: The Null Space of E_{**}.

Definition 3.4.1. *A vector field J along a geodesic γ is called a* ***Jacobi field*** *if it satisfies the Jacobi differential equation*

$$\frac{\mathrm{D}^2 J}{\mathrm{d}t^2} + R(V, J)V = 0$$

where $V = \dfrac{\mathrm{d}\gamma}{\mathrm{d}t}$. This is a linear, second order differential equation.

Remark 3.4.2. *It can be put in a more familiar form by choosing orthogonal parallel vector fields $P_1, \ldots, P_n$ along γ. Then setting $J(t) = \sum f^i(t)P_i(t)$, the equation becomes*

$$\frac{\mathrm{d}^2 f^i}{\mathrm{d}t^2} + \sum_{j=1}^{n} a^i_j(t) f^j(t) = 0, \quad i = 1, \ldots, n$$

where $a^i_j = \langle R(V, P_j)V, P_i\rangle$. Thus the Jacobi equation has 2n linearly independent solutions, each of which can be defined throughout γ. The solutions are all C^∞-differentiable. A given Jacobi field J is completely determined by its initial conditions

$$J(0), \frac{\mathrm{D}J}{\mathrm{d}t}(0) \in TM_{\gamma(0)}$$

Definition 3.4.3. *Let $p = \gamma(a), q = \gamma(b)$ be two points on the geodesic γ, with $a \neq b$. p and q are* ***conjugate*** *along γ if there exists a non-zero Jacobi field J along γ which vanishes for $t = a, t = b$. The* ***multiplicity*** *of p and q as conjugate points is equal to the dimension of the vector space consisting of all such Jacobi fields.*

Theorem 3.4.4. *A vector field $W_1 \in T\Omega_\gamma$ belongs to the null space of E_{**} if and only if W_1 is a Jacobi field. Hencee E_{**} is degenerate if and only if the end poitns p and q are conjugate along γ. Thu nullity of E_{**} is equal to the multiplicity of p and q as conjugate points.*

Proof. If J is a Jacobi field which vanishes at p and q, then J certainly belongs to $T\Omega_\gamma$. The second variation formula states that

$$-\frac{1}{2}E_{**}(J, W_2) = \sum_t \langle W_2(t), 0\rangle + \int_0^1 \langle W_2, 0\rangle \mathrm{d}t = 0$$

Hence 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 $W_i|_{[t_{i-1}, t_i]}$ 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

$$W_2(t) = f(t)(\frac{\mathrm{D}^2}{\mathrm{d}t^2} + R(V, W_1)V)_t$$

Then

$$-\frac{1}{2}E_{**}(W_1,W_2)=\sum 0+\int_0^1 f(t)\|\frac{\mathrm{D}^2W_1}{\mathrm{d}t^2}+R(V,W_1)V\|^2\mathrm{d}t$$

Since this is zero, it follows that $W_1|_{[t_{i-1},t_i]}$ is a Jacobi field for each i.

Now let $W_2'\in T\Omega_\gamma$ be a field such that $W_2'(t_i)=\Delta_{t_i}\frac{\mathrm{D}W_1}{\mathrm{d}t}$ for $i=1,2,\dots,k-1$. Then

$$-\frac{1}{2}E_{**}(W_1,W_2')=\sum_{i=1}^{k-1}\|\Delta_{t_i}\frac{\mathrm{D}W_1}{\mathrm{d}t}\|^2+\int_0^1 0\mathrm{d}t=0$$

Hence $\frac{\mathrm{D}W_1}{\mathrm{d}t}$ has no jumps. But a solution W_1 of the Jacobi equation is completely determined by the vectors $W_1(t_i)$ and $\frac{\mathrm{D}W_1}{\mathrm{d}t}(t_i)$. Thus it follows that the k Jacobi fields $W_i|_{[t_{i-1},t_i]}, i=1,\dots,k$, fit together to give a Jacobi field W_1 which is C^∞-differentiable throughout the entire unit interval. This completes the proof. □

Remark 3.4.5. *Actually the nullity ν satisfies $0\leq\nu<n$. Since the space of Jacobi fields which vanish for $t=0$ has dimension precise n, it is clear that $\nu\leq n$. We will construct one example of Jacobi field which vanishes for $t=0$, but not for $t=1$. This will imply that $\nu<n$. In fact let $J_t=tV_t$ where $V_t=\frac{\mathrm{d}\gamma}{\mathrm{d}t}$ denotes the velocity vector field. Then*

$$\frac{\mathrm{D}V}{\mathrm{d}t}=1\cdot V+t\frac{\mathrm{D}V}{\mathrm{d}t}=V\implies\frac{\mathrm{D}^2V}{\mathrm{d}t^2}=0$$

Furthermore $R(V,J)V=tR(V,V)V=0$ since R is skew symmetric in the first two variables. Thus J satisfies the Jacobi equation. Since $J_0=0,J_1\neq 0$, this completes the proof

Example 3.4.6. *Suppose M is "flat" in the sense that the curvature tensor is identically zero. Then the Jacobi equation becomes $\frac{\mathrm{D}^2V}{\mathrm{d}t^2}=0$. Setting $J(t)=\sum f^i(t)P_i(t)$ where P_i are parallel, this becomes $\frac{\mathrm{d}^2f^i}{\mathrm{d}t^2}=0$. Evidently a Jacobi field along γ can have at most one zero. Thus there are no conjugate points, and E_{**} is non-degenerate.*

Example 3.4.7. *Suppose that p and q are antipodal points on the unit sphere S^n, and let γ be a great circle arc from p to q. Then we will see that p and q are conjugate with multiplicity n-1. Thus in this example the nullity ν of E_{**} takes its largest possible value. The proof will depend on the following discussion.*

Let α be a 1-parameter variation of γ. not necessarily keeping the end-points fixed, such that each $\bar{\alpha}(u)$ is a geodesic.

Lemma 3.4.8. *If α is such a variation of γ throughout geodesics, then the variation vector field $W(t)=\frac{\partial\alpha}{\partial u}(0,t)$ is a Jacobi field along γ.*

Proof. If α is a variation of γ through geodesics, then $\frac{\mathrm{D}}{\mathrm{d}t}\frac{\partial\alpha}{\partial t}$ is identically zero. Hence

$$\begin{aligned}0 =&\frac{\mathrm{D}}{\partial u}\frac{\mathrm{D}}{\partial t}\frac{\partial\alpha}{\partial t} = \frac{\mathrm{D}}{\partial t}\frac{\mathrm{D}}{\partial u}\frac{\partial\alpha}{\partial t} + R(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial u})\frac{\partial\alpha}{\partial t} \\ =&\frac{\mathrm{D}^2}{\partial t^2}\frac{\partial\alpha}{\partial u} + R(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial u})\frac{\partial\alpha}{\partial t}\end{aligned}$$

Therefore the variation vector field $\frac{\partial\alpha}{\partial u}$ is a Jacobi field. □

Thus one way of obtaining Jacobi fields is to move geodesics around. Now let us return to the example of two antipodal points on a unit n-sphere. Rotating the sphere, keeping p and q fixed, the variation vector field along the geodesic γ will be a Jacobi field vanishing at p and q. Rotating in $n-1$ different direcitons one obtains $n-1$ independent Jacobi fields. Thus p and q are conjugate along γ with multiplicity $n-1$.

Lemma 3.4.9. *Every Jacobi field along a geodesic γ may be obtained by a variation of γ throughout 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 endpoitns. Suppose that $\gamma(t) \in U$ for $0 \le t \le \delta$. 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 $\frac{\mathrm{d}a}{\mathrm{d}u}(0)$ is any prescribed vector in $TM_{\gamma(0)}$. Similarly choose $b : (-\varepsilon, \varepsilon) \to U$ with $b(0) = \gamma(\delta)$ and $\frac{\mathrm{d}b}{\mathrm{d}t}(0)$ 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 $b(u)$. Then the formula $t \to \frac{\partial\alpha}{\partial u}(0, t)$ defines a Jacobi field with the given end conditions.

Any Jacobi field along $\gamma|_{[0,\delta]}$ can be obtained in this way: If $\mathscr{J}(\gamma)$ denotes the vector space of all Jacobi fields W along γ, then the formula $W \to (W(0), W(\delta))$ defines a linear map

$$\ell : \mathscr{J}(\gamma) \to TM_{\gamma(0)} \times TM_{\gamma(\delta)}$$

we have just shown that ℓ is onto. Since both vector spaces have the same dimension $2n$ it follows that ℓ is an isomorphism. I.e., a Jacobi field is determined by its values at $\gamma(0)$ and $\gamma(\delta)$. (More generally a Jacobi field is determined by its varlues 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 $[0, 1]$, $\bar{\alpha}(u)$ can be extended to a geodesic defined over the entire unit interval $[0, 1]$. This yields a variation through geodesics:

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

with any given Jacobi as variation vector. □

Remark 3.4.10. *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 3.4.11. *The proof shows also, that there is a neighborhood $(-\delta, \delta)$ of 0 so that if $t \in (-\delta, \delta)$ then $\gamma(t)$ is not conjugate to $\gamma(0)$ along γ.*

3.5. **The Index Theorem.**

Definition 3.5.1. *The **index** λ of the Hessian*

$$E_{**} : T\Omega_\gamma \times T\Omega_\gamma \to \mathbb{R}$$

is defined to be the maximum dimension of a subspace of $T\Omega_\gamma$ on which E_{**} is negative definite.

Theorem 3.5.2. (*Morse*) *The index λ 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 γ; each such conjugate point being counted with its multiplicity. This index λ is always finite.*

As an immediate consequence one has:

Corollary 3.5.3. *A geodesic segment $\gamma : [0,1] \to M$ can contain only finitely many points which are conjugate to $\gamma(0)$ along γ.*

Proof. In order to prove theorem 3.5.2 we will first make an estimate for λ by splitting the vector space $T\Omega_\gamma$ into two mutually orthogonal subspaces, on one of which E_{**} is positive definite.

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 endpoitns. Choose a subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$ 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 $T\Omega_\gamma(t_0, t_1, \ldots, t_k) \subset T\Omega_\gamma$ be the vector space consisting of all vector fields W along γ such that

1. $W|_{[t_{i-1},t_i]}$ is a Jacobi field along $\gamma|_{[t_{i-1},t_i]}$ for each i;
2. W vanishes at the endpoints $t = 0, t = 1$.

Thus $T\Omega_\gamma(t_0, t_1, \ldots, t_k)$ is a finite dimensional vector space consisting of broken Jacobi fields along γ.

Let $T' \subset T\Omega_\gamma$ be the vector space consisting of all vector fields $W \in T\Omega_\gamma$ for which $W(t_0) = W(t_1) = \cdots = W(t_k) = 0$.

Lemma 3.5.4. *The vector space $T\Omega_\gamma$ splits as the direct sum $T\Omega_\gamma(t_0, t_1, \ldots, t_k) \oplus T'$. These two subspace are mutually perpendicular with respect to the inner product E_{**}. Furthermore, E_{**} restricted to T' is positive definite.*

Proof. Given any vector field $W \in T\Omega_\gamma$ let W_1 denote the unique broken Jacobi fields in $T\Omega_\gamma(t_0, t_1, \ldots, t_k)$ such that $W_1(t_i) = W(t_i)$ for $i = 0, 1, \ldots, k$. It follows from remark 3.4.10 that W_1 exists and is unique.

Clearly $W - W_1$ belongs to T'. Thus the two subspace generate the whole space, and have only the zero vector field in common.

If W_1 belongs to $T\Omega_\gamma(t_0, t_1, \ldots, t_k)$ and W_2 belongs to T', thus the second variation formula takes the form

$$\frac{1}{2}E_{**}(W_1, W_2) = -\sum_t \langle W_2(t), \Delta_t \frac{\mathrm{D}W_1}{\mathrm{d}t}\rangle - \int_0^1 \langle W_2, 0\rangle \mathrm{d}t = 0$$

Thus the two subspaces are mutually perpendicular with respect to E_{**}.

For any $W \in T\Omega_\gamma$ the Hessian $E_{**}(W,W)$ can be interpreted as the second derivative $\frac{\mathrm{d}^2 E\circ\bar{\alpha}}{\mathrm{d}u^2}(0)$, where $\bar{\alpha}:(-\varepsilon,\varepsilon)\to\Omega$ is any variation of γ with variation vector field $\frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}t}=W$. If W belongs to T', then we may assume that $\bar{\alpha}$ is chosen so as to leave the points $\gamma(t_0),\dots,\gamma(t_k)$ fixed. In other words we may assume that $\bar{\alpha}(u)(t_i)=\gamma(t_i)$ for $i=0,1,\dots,k$.

Proof that $E_{**}(W,W)\geq 0$ for $W\in T'$. Each $\bar{\alpha}\in\Omega$ is a piecewise smooth path from $\gamma(0)$ to $\gamma(t_1)$ to $\gamma(t_2)$ to ... 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 ≥ 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 $W_1\in T\Omega_\gamma(t_0,t_1,\dots,t_k)$ we have already seen that $E_{**}(W_1,W)=0$. For any $W_2\in T'$ the inequality

$$0\leq E_{**}(W+cW_2,W+cW_2)=2cE_{**}(W_2,W)+c^2E_{**}(W_2,W_2)$$

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 $W=0$.

Thus the quadratic form E_{**} is positive definite on T'. This completes the proof. □

An immediate consequence is the following:

Lemma 3.5.5. *The index or the nullity of E_{**} is equal to the index or nullity of E_{**} restricted to the space $T\Omega_\gamma(t_0,t_1,\dots,t_k)$ of broken Jacobi fields. In particular, the index λ is always finite.*

Let γ_τ denote the restriction of γ 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:

Lemma 3.5.6. *$\lambda(\tau)$ is a monotone function of τ.*

Proof. For if $\tau<\tau'$ then there exists a $\lambda(\tau)$ dimensional space $\mathscr{V}$ of vector fields along γ_τ which vanishes at $\gamma(0)$ and $\gamma(t)$ such that the Hessian $(E_0^\tau)_{**}$ is negative definite on this vector space. Each vector field in $\mathscr{V}$ extends to a vector field along $\gamma_{\tau'}$, which vanishes identically between $\gamma(\tau)$ and $\gamma(\tau')$. Hence $\lambda(\tau)\leq\lambda(\tau')$. □

Lemma 3.5.7. *$\lambda(\tau)=0$ for small values of τ.*

Proof. For if τ is sufficiently small then γ_τ is a minimal geodesic, hence $\lambda(\tau)=0$ by lemma 3.3.5. □

Now let us examine the discontinuities of the function $\lambda(\tau)$. First note that $\lambda(\tau)$ is continuous from the left:

Lemma 3.5.8. *For all sufficiently small $\varepsilon > 0$ we have $\lambda(\tau - \varepsilon) = \lambda(\tau)$.*

Proof. According to lemma 3.5.4 the number of $\lambda(1)$ can be interpreted as the index of a quadratic form on a finite dimensional vector space $T\Omega_\gamma(t_0, t_1, \ldots, t_k)$. 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_τ on a corresponding vector space of broken Jacobi fields along γ_τ. This vector space is to be constructed using the subdivision $0 < t_1 < t_2 < \cdots < t_i < \tau$ of $[0, \tau]$. Since a broken Jacobi field is uniquely determined by its values at the break points $\gamma(t_i)$, this vector space is isomorphic to the direct sum

$$\sum = TM_{\gamma(t_1)} \oplus TM_{\gamma(t_2)} \oplus \cdots \oplus TM_{\gamma(t_i)}$$

Note that this vector space $\sum$ is independent of τ. Evidently the quadratic form H_τ on $\sum$ varies continously with τ.

Now H_τ is negative definite on a subspace $\mathscr{V} \subset \sum$ of dimension $\lambda(\tau)$. For all τ' sufficiently close to τ it follows that H_{τ}, is negative definite on $\mathscr{V}$. Therefore $\lambda(\tau') \geq \lambda(\tau)$. But if $\tau' = \tau - \varepsilon < \tau$, then we also have $\lambda(\tau - \varepsilon) \leq \lambda(\tau)$ by lemma 3.5.6. Hence $\lambda(\tau - \varepsilon) = \lambda(\tau)$. $\square$

Lemma 3.5.9. *Let ν 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 $\lambda(\tau)$ jumps by ν when the variable t passes a conjugate point of multiplicity ν; and is continuous otherwise. Clearly this lemma will complete the proof of the index theorem.

Proof. Proof that $\lambda(\tau+\varepsilon) \leq \lambda(\tau)+\nu$. Let H_τ and $\sum$ be as in the proof above. Since $\dim \sum = ni$ we see that H_τ is positive definite on some subspace $\mathscr{V}' \subset \sum$ of dimension $ni - \lambda(\tau) - \nu$. For all τ' sufficiently close to τ, it follows that H_{τ}, is positive definite on $\mathscr{V}'$. Hence

$$\lambda(\tau') \leq \dim \sum - \dim \mathscr{V}' = \lambda(\tau) + \nu$$

Proof that $\lambda(\tau + \varepsilon) \geq \lambda(\tau) + \nu$. Let $W_1, \ldots, W_{\lambda(\tau)}$ be $\lambda(\tau)$ vector fields along γ_τ, vanishing at the endpoints, such that the matrix

$$((E_0^\tau)_{**}(W_i, W_j))$$

is negative definite. Let $J_1, \ldots, J_\nu$ be ν linear independent Jacobi fields along γ_τ, also vanishes at the endpoints. Note that the ν vectors

$$\frac{\mathrm{D}J_h}{\mathrm{d}t}(\tau) \in TM_{\gamma(\tau)}$$

are linear independent. Hence it is possible to choose ν vector fields $X_1, \dots, X_\nu$ along $\gamma_{\tau+\varepsilon}$, vanishing at the endpoints of $\gamma_{\tau+\varepsilon}$, so that

$$(\langle \frac{\mathrm{D}J_h}{\mathrm{d}t}(\tau), X_k(\tau)\rangle)$$

is equal to the $\nu \times \nu$ identity matrix. Extend the vector fields W_i and J_h over $\gamma_{\tau+\varepsilon}$ by setting these fields equal to zero for $\tau \leq t \leq \tau + \varepsilon$.

Using the second variation formula we see easily that

$$(E_0^{\tau+\varepsilon})_{**}(J_h, W_i) = 0$$

$$(E_0^{\tau+\varepsilon})_{**}(J_h, X_k) = 2\delta_{hk}$$

Now let c be a small number, and consider the $\lambda(\tau) + \nu$ vector fields

$$W_1, \dots, W_{\lambda(\tau)}, c^{-1}J_1 - cX_1, \dots, c^{-1}J_\nu - cX_\nu$$

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

$$\begin{pmatrix} ((E_0^\tau)_{**}(W_i, W_j)) & cA \\ cA^t & -4I + c^2 B \end{pmatrix}$$

where A and B are fixed matrices. If c is sufficiently small, this compound matrix is certainly negative definite, this completes the proof. □

The index theorem clearly follows from above lemmas. □

3.6. **A Finite Dimensional Approximation to Ω^c.** Let M be a connected Riemannian manifold and let p and q be two points of M. The set $\Omega\Omega(M,p,q)$ of piecewise C^∞ paths from p to q can be topologized as follows. Let ρ denote the topological metric on M coming from its Riemannian metric.

Definition 3.6.1. *Given two $\omega, \omega' \in \Omega$ with arc-lengths $s(t), s'(t)$ respectively, define the **distancee** $d(\omega,\omega')$ to be*

$$\max_{0\le t\le 1} \rho(\omega(t),\omega'(t)) + [\int_0^1 (\frac{\mathrm{d}s}{\mathrm{d}t} - \frac{\mathrm{d}s'}{\mathrm{d}t})^2 \mathrm{d}t]^{\frac{1}{2}}$$

Remark 3.6.2. *The last term is added on so that the energy function will be a continuous function from Ω to $\mathbb{R}$. This metric induces the required topology on Ω.*

Definition 3.6.3. *Given $c > 0$ let Ω^c denote the closed subset $E^{-1}([0,c]) \subset \Omega$ and let $\operatorname{Int}\Omega^c$ denote the open subset $E^{-1}([0,c))$.*

Now we will study the topology of Ω^c by constructing a finite dimensional approximation to it.

Choose some subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$ of the unit interval. Let $\Omega(t_0, t_1, \ldots, t_k)$ be the subspace of Ω consisting of paths $\omega : [0,1] \to M$ such that

1. $\omega(0) = p, \omega(1) = q$,
2. $\omega|_{[t_{i-1},t_i]}$ is a geodesic for each $i = 1, \ldots, k$.

Finally we define the subspaces

$$\Omega(t_0, t_1, \ldots, t_k)^c = \Omega^c \cap \Omega(t_0, t_1, \ldots, t_k)$$
$$\operatorname{Int}\Omega(t_0, t_1, \ldots, t_k)^c = (\operatorname{Int}\Omega)^c \cap \Omega(t_0, t_1, \ldots, t_k)$$

Lemma 3.6.4. *Let M be a complete Riemannian manifold; and let c be a fixed positive number such that $\Omega^c \neq \varnothing$. Then for all sufficiently fine subdivisions $(t_0, t_1, \ldots, t_k)$ of $[0,1]$ the set $\operatorname{Int}\Omega(t_0, t_1, \ldots, t_k)^c$ can be given the structure of a smooth finite dimensional manifold in a natural way.*

Proof. Let S denote the ball

$$\{x \in M : \rho(x,p) \le \sqrt{c}\}$$

Note that every path $\omega \in \Omega^c$ lies within this subset $S \subset M$. This follows from inequality $L^2 \le E \le c$.

Since M is complete, S is a compact set. Hence by 2.3.17 there exists $\varepsilon > 0$ so that whenever $x, y \in S$ and $\rho(x,y) < \varepsilon$ there is a unique geodesic from x to y of length $< \varepsilon$; and so that this geodesic denpends differentiably on x and y.

Choose the subdivision $(t_0, t_1, \ldots, t_k)$ of $[0,1]$ so that each difference $t_i - t_{i-1}$ is less than ε^2/c. Then for each broken geodesic

$$\omega \in \Omega(t_0, t_1, \ldots, t_k)^c$$

we have

$$(L_{t_{i-1}}^{t_i}\omega)^2 = (t_i - t_{i-1})(E_{t_{i-1}}^{t_i}\omega) \leq (t_i - t_{i-1})(E\omega) \leq (t_i - t_{i-1})c < \varepsilon^2$$

Thus the geodesic $\omega|_{[t_{i-1},t_i]}$ is uniquely and differentiably determined by the two end points.

The broken geodesic ω is uniquely and differentiably by the $(k-1)$-tuple

$$\omega(t_1), \omega(t_2), \dots, \omega(t_{k-1}) \in M \times M \times \cdots \times M$$

Evidently this correspondence

$$\omega \to (\omega(t_1), \dots, \omega(t_{k-1}))$$

defines a homeomorphism between $\operatorname{Int}\Omega(t_0, t_1, \dots, t_k)^c$ and a certain open subset of the $(k-1)$-fold product $M \times \cdots \times M$. Taking over the differentiable structure from the product, this completes the proof. □

To shorten the notation, let us denote this manifold $\operatorname{Int}\Omega(t_0, t_1, \dots, t_k)^c$ of broken geodesics by B. Let

$$E' : B \to \mathbb{R}$$

denote the restriction to B of the energy function.

Theorem 3.6.5. *This function $E' : B \to \mathbb{R}$ is smooth. Furthermore, for each $a < c$ the set $B^a = E^{-1}[0, a]$ is compact, and is a deformation retract of the corresponding set Ω^a. The critical point of E' are precisely the same as the critical points of E in $\operatorname{Int}\Omega^c$: namely the unbroken geodesic from p to q of length less than $\sqrt{c}$. The index [or the nullity] of the Hessian E'_{**} at each such critical point γ is equal to the index [or the nullity] of E_{**} at γ.*

Thus the finite dimensional manifold B provides a faithful model for the infinite dimensional path space $\operatorname{Int}\Omega^c$. As an immediate consequence we have the following result.

Theorem 3.6.6. *Let M be a complete Riemannian manifold and let $p, q \in M$ be two points which are not conjugate along any geodesic of length $\leq \sqrt{a}$. Then Ω^a has the homotopy type of a finite CW-complex, with one cell of dimension λ for each geodesic in Ω^a at which E_{**} has index λ.*

Proof. Since the broken geodesic $\omega \in B$ depends smoothly on the $(k-1)$-tuple

$$\omega(t_1), \dots, \omega(t_{k-1})$$

it is clear that the energy $E'(\omega)$ also depends smoothly on this $(k-1)$-tuple. In fact we have the explict formula

$$E'(\omega) = \sum_{i=1}^{k} \rho(\omega(t_{i-1}), \omega(t_i))^2/(t_i - t_{i-1})$$

For $a < c$ the set B^a is homeomorphic to the set of all $(k-1)$-tuples $(p_1, \dots, p_{k-1}) \in S \times \cdots \times S$ such that

$$\sum_{i=1}^{k} \rho(\omega(p_{i-1}), \omega(p_i))^2/(p_i - p_{i-1}) \le a$$

As a closed subset of a compact set, this is certainly compact.

A retraction $r : \operatorname{Int}\Omega^c \to B$ is defined as follows. Let $r(\omega)$ denote the unique broken geodesic in B such that each $r(\omega)|_{[t_{i-1},t_i]}$ is a geodesic of length less than ε from $\omega(t_{i-1})$ to $\omega(t_i)$. The inequality

$$\rho(p, \omega(t))^2 \le (L\omega)^2 \le E\omega < c$$

implies that $\omega[0,1] \subset S$. Hence the inequality

$$\rho(\omega(t_{i-1}), \omega(t_i))^2 \le (t_i - t_{i-1})(E_{t_{i-1}}^{t_i}\omega) \le \frac{\varepsilon^2}{c} \cdot c = \varepsilon^2$$

implies that $r(\omega)$ can be so defined.

Clearly $E(r(\omega)) < E(\omega) < c$. This retraction r fits into a 1-parameter family of maps

$$r_u : \operatorname{Int}\Omega^c \to \operatorname{Int}\Omega^c$$

as follows. For $t_{i=1} \le u \le t_i$ let

$$\begin{cases} r_u(\omega)|_{[0,t_{i-1}]} = r(\omega)|_{[0,t_{i-1}]}, \\ r_u(\omega)|_{[t_{i-1},u]} = \text{minimal geodesic from } \omega(t_{i-1}) \text{ to } \omega(u), \\ r_u(\omega)|_{[u,1]} = \omega|_{[u,1]} \end{cases}$$

Then r_0 is the identity map of $\operatorname{Int}\Omega^c$, and $r_1 = r$. It is easily to verified that $r_u(\omega)$ is continuous as a function of both variables. This prove that B is a deformation retract of $\operatorname{Int}\Omega^c$.

Since $E(r_u(\omega)) \le E(\omega)$ it is clear that each B^a is also a deformation retract of Ω^a.

Every geodesic is also a broken geodesic, so it is clear that every "critical point" of E in $\operatorname{Int}\Omega^c$ automatically lies in the submanifold B. Using the first variation formula it is clear that the critical points of E' are precisely the unbroken geodesics.

Consider the tangent space TB_γ to the manifold B at a geodesic γ. This will be identified with the tangent space $T\Omega_\gamma(t_0, t_1, \dots, t_k)$ of broken Jacobi fields along γ. $\square$

3.7. **The Topology of the Full Path Space.** Let M be a Riemannian manifold with Riemannian metric g, and let ρ be the induced topological metric. Let p and q be two not necessarily distinct points of M.

In homotopy theory one studies the space Ω^* of all continuous paths

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

from p to q, in the compact open topology. This topology can also be described as that induced by metric

$$d^*(\omega, \omega') = \max_t \rho(\omega(t), \omega'(t))$$

On the other hand we have been studying the space Ω of piecewise C^∞ paths from p to q with the metric

$$d(\omega, \omega') = d^*(\omega, \omega') + [\int_0^1 (\frac{\mathrm{d}s}{\mathrm{d}t} - \frac{\mathrm{d}s'}{\mathrm{d}t})^2 \mathrm{d}t]^{\frac{1}{2}}$$

Since $d \geq d^*$ the natural map

$$i : \Omega \hookrightarrow \Omega^*$$

is continous.

Theorem 3.7.1. *This natural map i is a homotopy equivalence between Ω and Ω^*.*

Proof. □

It is known that the space Ω^* has the homotopy type of a CW-complex. Therefore

Corollary 3.7.2. *Ω has the homotopy type of a CW-complex.*

This statement can be sharpened as follows.

Theorem 3.7.3. (*Fundamental theorem of Morse Theory*) *Let M be a complete Riemannian manifold, and let $p, q \in M$ be two points which are not conjugate along any geodesic. Then $\Omega(M; p, q)$ has the homotopy type of a countable CW-complex which contains one cell of dimension λ for each geodesic from p to q of index λ.*

Proof. Choose a sequence $a_0 < a_1 < a_2 < \dots$ of real numbers which are not critical values of the energy function E, so that each interval (a_{i-1}, a_i) contains precisely one critical value. Consider the sequence

$$\Omega^{a_0} \subset \Omega^{a_1} \subset \Omega^{a_2} \subset \dots$$

where we may assume Ω^{a_0} is vacuous. It follows from theorem 3.6.5 together with Remark 1.2.9 and lemma 1.2.12 that each Ω^{a_i} has the homotopy type of $\Omega^{a_{i-1}}$ with a finite number of cells attached: one λ-cell for each geodesic of index λ in $E^{-1}(a_{i-1}, a_i)$. Now just as in the proof of theorem 1.2.10, one constructs a sequence $K_0 \subset K_1 \subset K_2 \subset \dots$ of CW-complexes with cells of the required descriptions, and a sequence

$$\begin{array}{ccccccc}
\Omega^{a_1} & \hookrightarrow & \Omega^{a_2} & \hookrightarrow & \Omega^{a_3} & \hookrightarrow & \dots \\
\downarrow & & \downarrow & & \downarrow & & \\
K^{a_1} & \hookrightarrow & K^{a_2} & \hookrightarrow & K^{a_3} & \hookrightarrow & \dots
\end{array}$$

of homotopy equivalences. Letting $f : \Omega \to K$ be the direct limit mapping, it is clear that f induces isomorphism of homotopy group in all dimensions. Since Ω is known to have the homotopy type of a CW-complex it follows from Whitehead's theorem that f is a homotopy equivalence. This completes the proof. □

Example 3.7.4. ***The path space of the sphere*** S^n. *Suppose that* p *and* q *are two non-conjugate points of* S^n. *That is, suppose that* $q \neq p, p'$ *where* p' *denotes the antipode of* p. *Then there are denumerably many geodesics* $\gamma_0, \gamma_1, \dots$ *from* p *to* q, *as follows. Let* γ_0 *denote the short great circle from* p *to* q; *let* γ_1 *denote the long great circle are* $pq'p'q$; *let* γ_2 *denote the arc* $pqp'q'pq$; *and so on. The subscript* k *denotes the number of times* p *or* p' *occurs in the interior of* γ_k.

The index $\lambda(\gamma_k) = \mu_1 + \cdots + \mu_k$ *is equal to* $k(n-1)$, *since each of the points* p *or* p' *in the interior is conjugate to* p *with multiplicity* $n-1$. *Therefore we have:*

Corollary 3.7.5. *The loop space* $\Omega(S^n)$ *has the homotopy type of a CW-complex with one cell each in the dimension* $0, n-1, 2(n-1), 3(n-1), \dots$.

For $n > 2$ *the homology of* $\Omega(S^n)$ *can be computed immediately from this information. Since* $\Omega(S^n)$ *has non-trivial homology in infinitely many dimensions, we can conclude:*

Corollary 3.7.6. *Let* M *have the homology type of* S^n, *for* $n > 2$. *Then any two non-conjugate points of* M *are joined by infinitely many geodesics.*

Proof. This follows since the homotopy type of $\Omega^*(M)$ and hence of $\Omega(M)$ depends only on the homotopy type of M. There must be at least one geodesic in $\Omega(M)$ with index 0, at least one with index $n-1, 2(n-1), 3(n-1)$, and so on. □

Remark 3.7.7. *More generally if* M *is any complete manifold which is not contractible then any two non-conjugate points of* M *are joined by infinitely many geodesics.*

3.8. **Existence of Non-Conjugate Points.**

Theorem 3.8.1. *The point* $\exp v$ *is conjugate to* p *along the geodesic* γ_v *from* p *to* $\exp v$ *if and only if the mapping* $\exp$ *is critical at* v.

Proof. Suppose that exp is critical at $v \in TM_p$. Then $\exp_*(X) = 0$ for some non-zero $X \in T(TM_p)_v$, the tangent space at v to TM_p, considered as a manifold. Let $u \to v(u)$ be a path in TM_p such that $v(0) = v$ and $\frac{\mathrm{d}v}{\mathrm{d}t}(0) = X$. Then the map α defined by $\alpha(u,t) = \exp tv(u)$ is a variation through geodesics of the geodesic γ_v given by $t \to \exp tv$. Therefore the vector field W given by $t \to \frac{\partial}{\partial u}(\exp tv(u))|_{u=0}$ is a Jacobi field along γ_v. Obviously $W(0) = 0$. We also have

$$W(1) = \frac{\partial}{\partial u}(\exp v(u))|_{u=0} = \exp_* \frac{\mathrm{d}v(u)}{\mathrm{d}u}(0) = \exp_* X = 0$$

But this field is not identically zero since

$$\frac{\mathrm{D}W}{\mathrm{d}t}(0) = \frac{\mathrm{D}}{\partial u}\frac{\partial}{\partial t}(\exp tv(u))|_{(0,0)} = \frac{\mathrm{D}}{\partial u}v(u)|_{u=0} \neq 0$$

So there is a non-trivial Jacobi field along γ_v from p to $\exp v$, vanishing at these points; hence p and $\exp v$ are conjugate along γ_v.

Now suppose that $\exp_*$ is non-singular at v. Choose n independent vectors $X_1, \dots, X_n$ in $T(TM_p)_v$. Then $\exp_*(X_1), \dots, \exp_*(X_n)$ are linearly independent. In TM_p choose paths $u \to v_1(u), \dots, u \to v_n(u)$ with $v_i(0) = v$ and $\frac{\mathrm{d}v_i(u)}{\mathrm{d}u}(0) = X_i$.

Then $\alpha_1, \dots, \alpha_n$ constructed as above, provids n Jacobi fields $W_1, \dots, W_n$ along γ_v, vanishing at p. Since the $W_i(1) = \exp_*(X_i)$ are independent, no non-trivial linear combiation of the W_i can vanish at $\exp v$. Since n is the dimension of the space of Jacobi fields along γ_v, which vanishes at p, clearly no non-trivial Jacobi field along γ_v vanishes at both p and $\exp v$. This completes the proof. □

Corollary 3.8.2. *Let* $p \in M$. *Then for almost all* $q \in M$, p *is not conjugate to* q *along any geodesic.*

Proof. This follows immediately from theorem 3.8.1 together with Sard's theorem. □

3.9. **Some Relations Between Topology and Curvature.** This section will describe the behavior of geodesics in a manifold with negative curvatures or with positive curvatures.

Lemma 3.9.1. *Suppose that $\langle R(A,B)A,B\rangle \leq 0$ for every pair of vectors A, B in the tangent space TM_p and for every $p \in M$. Then no two points of M are conjugate along any geodesic.*

Proof. Let γ be a geodesic with velocity vector field V; and let J be a Jacobi field along γ. Then

$$\frac{\mathrm{D}^2 J}{\mathrm{d}t^2} + R(V,J)V = 0$$

so that

$$\langle \frac{\mathrm{D}^2 J}{\mathrm{d}t^2}, J\rangle = -\langle R(V,J)V, J\rangle \geq 0$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \frac{\mathrm{D}J}{\mathrm{d}t}, J\rangle = \langle \frac{\mathrm{D}^2 J}{\mathrm{d}t^2}, J\rangle + \|\frac{\mathrm{D}J}{\mathrm{d}t}\|^2 \geq 0$$

Thus the function $\langle \frac{\mathrm{D}}{\mathrm{d}t}, J\rangle$ is monotonically increasing, and strictly so if $\frac{\mathrm{D}J}{\mathrm{d}t} \neq 0$

If J vanishes both at 0 and at $t_0 > 0$, then the function $\langle \frac{\mathrm{D}J}{\mathrm{d}t}, J\rangle$ also vanishes at 0 and t_0, and hence must vanish identically throughout the interval $[0, t_0]$. This implies that

$$J(0) = \frac{\mathrm{D}J}{\mathrm{d}t}(0) = 0$$

so that J is identically zero. This completes the proof. □

Remark 3.9.2. *If A and B are orthogonal unit vectors at p then the quantity $\langle R(A,B)A,B\rangle$ is called* ***sectional curvature*** *determined by A and B. It is equal to the Gaussian curvature of the surface*

$$(u_1, u_2) \to \exp_p(u_1 A + u_2 B)$$

spanned by the geodesics through p with velocity vector in the subspace spanned by A and B.

Intuitively the curvature of a manifold can be described in terms of "optics" within the manifold as follows. Suppose that we think of the geodesics as being the paths of light rays. Consider an obverser at p looking in the direction of the unit vector U towards a point $q = \exp(rU)$. A small line segment at q with length L, pointed in a direction corresponding to the unit vector $W \in TM_p$, would appear to the observer as a line of segment of length

$$L(1 + \frac{r^2}{6}\langle R(U,W)U,W\rangle + (\text{terms involving higher powers of } r))$$

Thus if sectional curvatures are negative then any object appears shorter than it really is. A small sphere of radius ε at q would appear to be an ellipsoid with principal radii $\varepsilon(1+\frac{r^2}{6}K_1+\dots),\dots,\varepsilon(1+\frac{r^2}{6}K_n+\dots)$ where $K_1,\dots,K_n$ denote the eigenvalues of the linear transformation $W\to R(U,W)U$. Any small object of volume v would appear to have volume $v(1+\frac{r^2}{6}(K_1+\dots+K_n)+\text{(higher terms)}$ where $K_1+\dots+K_n$ is equal to the "Ricci" curvature $K(U,U)$, as defined in this section.

Theorem 3.9.3. (*Cartan*) *Suppose that M is a simply connected, complete Riemannian manifold, and that the sectional curvature $\langle R(A,B)A,B\rangle$ is everywhere ≤ 0. Then any two points of M are joined by a unique geodesic. Furthermore, M is diffeomorphic to the Euclidean space $\mathbb{R}^n$.*

Proof. Since there are no conjugate points, it follows from the index theorem that every geodesic from p to q has index $\lambda=0$. Thus theorem 3.7.3 asserts that the path space $\Omega(M;p,q)$ has the homotopy type of a 0-dimensional CW-complex, with one vertex for each geodesic.

The hypothesis that M is simply connected implies that $\Omega(M;p,q)$ is connection. Since a connected 0-dimensional CW-complex must consist of a single point, it follows that there is precisely one geodesic from p to q.

Therefore, the exponential map $\exp_p: TM_p\to M$ is one-one and onto. But $\exp_p$ is non-critical everywhere, since there exist no conjugate points. So that $\exp_p$ is locally a diffeomorphism. Combining these two facts, we see that $\exp_p$ is a global diffeomorphism. This completes the proof. □

More generally, suppose that M is not simply connected, but it is complete and has sectional curvature ≤ 0. Then theorem 3.9.3 applies to the universal covering space $\widetilde{M}$ of M. For it is clear that $\widetilde{M}$ inherits a Riemannian metric from M which is geodesically complete, and has sectional curvature ≤ 0.

Given two points p,q of M, it follows taht each homotopy class from p to q contains precisely one geodesic.

The fact that $\widetilde{M}$ is contractible puts strong restriction on the topology of M. For example:

Corollary 3.9.4. *If M is complete with $\langle R(A,B)A,B\rangle\leq 0$ then the homotopy group $\pi_i(M)$ are zero for $i>1$, and $\pi_1(M)$ contains no element of finite order other than identity.*

Proof. Clearly $\pi_i(M)=\pi_i(\widetilde{M})=0$ for $i>1$. Since $\widetilde{M}$ is contractible the cohomology group $H^k(M)$ can be identified with the cohomology group $H^k(\pi_i(M))$ of the group $\pi_1(M)$. Now suppose that $\pi_1(M)$ contains a non-trivial finite cyclic subgroup G. Then for a suitable space $\widehat{M}$ of M we have $\pi_1(\widehat{M})=G$; hence

$$H^k(G)=H^k(\widehat{M})=0,\quad \text{for } k>n.$$

But the cohomology group of a finite cyclic group are non-trivial in arbitrary high dimensions. This gives a contradiction and completes the proof. □

Now we will consider manifold with positive curvature. Instead of considering the sectional curvature, one can obtain sharper results in this case by considering the Ricci tensor (sometimes called the mean curvature tensor).

Definition 3.9.5. *The Ricci tensor at a point p of a Riemannian manifold M is a bilinear pairing*

$$K : TM_p \times TM_p \to \mathbb{R}$$

defined as follows. Let $K(U_1, U_2)$ be the trace of the linear transformation

$$W \to R(U_1, W)U_2$$

from TM_p to TM_p. It follows easily from lemma 2.2.4 that K is symmetric.

The Ricci curvature is related to sectional curvature as follows. Let $U_1, \ldots, U_n$ be an orthogonal basis for the tangent space TM_p.

Lemma 3.9.6. *$K(U_n, U_n)$ is equal to the sum of the sectional curvatures $\langle R(U_n, U_i)U_n, U_i\rangle$ for $i = 1, 2, \ldots, n-1$.*

Proof. By definition $K(U_n, U_n)$ is equal to the trace of the matrix $(\langle R(U_n, U_i)U_n, U_j\rangle$. Since the n-th diagonal term of this matrix is zero, we obtain a sum of $n-1$ sectional curvatures, as asserted. ☐

Theorem 3.9.7. (*Myers*) *Suppose that the Ricci curvature K satisfies*

$$K(U, U) \geq (n-1)/r^2$$

for every unit vector U at every point of M; where r is a positive constant. Then every geodesic on M of length $> \pi r$ contains conjugate points; and hence is not minimal.

Proof. Let $\gamma : [0, 1] \to M$ be a geodesic of length L. Choose parallel vector fields $P_1, \ldots, P_n$ along γ which are orthogonal at one point, and hence are orthogonal everywhere along γ. We may assume that P_n points along γ, so that

$$V = \frac{\mathrm{d}\gamma}{\mathrm{d}t} = LP_n, \quad \frac{\mathrm{D}P_i}{\mathrm{d}t} = 0$$

Let $W_i(t) = (\sin \pi t)P_i(t)$. Then

$$\begin{aligned}\frac{1}{2}E_{**}(W_i, W_i) &= -\int_0^1 \langle W_i, \frac{\mathrm{D}^2 W_i}{\mathrm{d}t^2} + R(V, W_i)V\rangle \mathrm{d}t \\ &= \int_0^1 (\sin \pi t)^2(\pi^2 - L^2\langle R(P_n, P_i)P_n, P_i\rangle)\mathrm{d}t\end{aligned}$$

Summing for $i = 1, 2, \ldots, n-1$ we obtain

$$\frac{1}{2}\sum_{i=1}^{n-1} E_{**}(W_i, W_i) = \int_0^1 (\sin \pi t)^2((n-1)\pi^2 - L^2 K(P_n, P_n))\mathrm{d}t$$

Now if $K(P_n, P_n) \geq (n-1)/r^2$ and $L > \pi r$ then this expression is < 0. Hence $E_{**}(W_i, W_i) < 0$ for some i. This implies that the index of γ is positive, and hence, by the index theorem, that γ contains conjugate points.

It follows also that γ is not a minimal geodesic. In fact if $\bar{\alpha} : (-\varepsilon, \varepsilon) \to \Omega$ is a variation with variation vector field W_i then

$$\frac{\mathrm{d}E(\bar{\alpha}(u))}{\mathrm{d}u} = 0, \quad \frac{\mathrm{d}^2 E(\bar{\alpha}(u))}{\mathrm{d}u^2} < 0$$

for $u = 0$. Hence $E(\bar{\alpha}(u)) < E(\gamma)$ for small values of $u \neq 0$. This completes the proof. □

Example 3.9.8. *If M is a sphere of radius r then every sectional curvature is equal to $\frac{1}{r^2}$. Hence $K(U, U)$ takes the constant value $(n-1)/r^2$. It follows from above that every geodesic of length $> \pi r$ contains conjugate points: a best possible result.*

Corollary 3.9.9. *If M is complete, and $K(U, U) \geq (n-1)/r^2 > 0$ for all unit vector U, then M is compact, with diameter $\leq \pi r$.*

Proof. If $p, q \in M$ let γ be a minimal geodesic from p to q. Then the length of γ must be $\leq \pi r$. Therefore, all points have distances $\leq \pi r$. Since closed bounded sets in a complete manifold are compact, it follows that M itself is compact. □

This corollary applies also to the universal space $\widetilde{M}$ of M. Since $\widetilde{M}$ is compact, it follows that the fundamental group $\pi_1(M)$ is finite. This assertion can be sharpened as follows.

Theorem 3.9.10. *If M is a compact manifold, and if the Ricci tensor K of M is everywhere positive definite, then the path space $\Omega(M; p, q)$ has the homotopy type of CW-complex having only finitely many cells in each dimension.*

Proof. Since the space consisting of all unit vectors U on M is compact, it follows that the continuous function $K(U, U) > 0$ takes on a minimum, which we can denote by $(n-1)/r^2 > 0$. Then every geodesic $\gamma \in \Omega(M; p, q)$ of length $> \pi r$ has index ≥ 1.

More generally consider a geodesic γ of length $> k\pi r$. Then a similar argument shows that γ has index $\geq k$. In fact for each $i = 1, 2, \ldots, k$ one can construct a vector field X_i along γ which vanishes outside of the interval $(\frac{i-1}{k}, \frac{i}{k})$, and such that $E_{**}(X_i, X_i) < 0$. Clearly $E_{**}(X_i, X_j) - 0$ for $i \neq j$; so that $X_1, \ldots, X_k$ span a k-dimensional subspace of $T\Omega_\gamma$ on which E_{**} is negative definite.

Now suppose that the points p and q are not conjugate along any geodesic. Then according to 3.6.6 there are only finitely many geodesics from p to q of length $\leq k\pi r$. Hence there only finitely many geodesics with index $< k$. Together with fundamental theory of morse theory, this completes the proof. □

4. APPLICATIONS TO LIE GROUP AND SYMMETRIC SPACES

4.1. Symmetric Spaces.

Definition 4.1.1. *A **symmetric space** is a connected Riemannian manifold M such that, for each $p \in M$ there is an isometry $I_p : M \to M$ which leaves p fixed and reverses geodesics through p, i.e., if γ is a geodesic and $\gamma(0) = p$ then $I_p(\gamma(t)) = \gamma(-t)$*

Lemma 4.1.2. *Let γ be a geodesic in M, and let $p = \gamma(0)$ and $q = \gamma(c)$. Then $I_q I_p(\gamma(t)) = \gamma(t + 2c)$. Moreover, $I_q I_p$ preserves parallel vector fields along γ.*

Proof. Let $\gamma'(t) = \gamma(t + c)$. Then γ' is a geodesic and $\gamma'(0) = q$. Therefore

$$I_q I_p(\gamma(t)) = I_q(\gamma(-t)) = I_q(\gamma'(-t - c)) = \gamma'(t + c) = \gamma(t + 2c)$$

If the vector field V is parallel along γ then $I_{p_*}(V)$ is parallel, since I_p is an isometry. And $I_{p_*}V(0) = -V(0)$; therefore $I_{p_*}V(t) = -V(-t)$. Therefore $I_{q_*}I_{p_*}(V(t)) = V(t + 2c)$. □

Corollary 4.1.3. *M is complete.*

Corollary 4.1.4. *I_p is unique.*

Corollary 4.1.5. *If U, V and W are parallel fields along γ, then $R(U, V)W$ is also a parallel field along γ.*

Proof. If X denotes a fourth parallel vector field along γ, note that the quantity $\langle R(U, V)W, X\rangle$ is constant along γ. In fact, give $p = \gamma(0), q = \gamma(c)$, consider this isometry $T = I_{\gamma(c/2)}I_p$ which carries p to q. Then

$$\langle R(U_q, V_q)W_q, X_q\rangle = \langle R(T_*U_p, T_*V_p)T_*W_p, T_*X_p\rangle$$

Since T is isometry, thus $\langle R(U, V)W, X\rangle$ is constant for every parallel vector field X. It clearly follows that $R(U, V)W$ is parallel. □

Definition 4.1.6. *Manifold with the property of Corollary* 4.1.5 *are called **locally symmetric**.*

In any locally symmetric manifold the Jacobi differential equations have simple explicit solutions. Let $\gamma : \mathbb{R} \to M$ be a geodesic in a locally symmetric manifold. Let $V = \dfrac{\mathrm{d}\gamma}{\mathrm{d}t}(0)$ be the velocity vector at $p = \gamma(0)$, define a linear transformation

$$K_V : TM_p \to TM_p$$

by $K_V(W) = R(V, W)V$. Let $e_1, \dots, e_n$ denote the eigenvalues of K_V.

Theorem 4.1.7. *The conjugate points to p along γ are the points $\gamma(\pi k/\sqrt{e_i})$ where k is any non-zero integer, and e_i is any positive eigenvalue of K_V. The multiplicity of $\gamma(t)$ as a conjugate point is equal to the number of e_i such that t is a multiple of $\pi/\sqrt{e_i}$.*

Proof. First observe that K_V is self-adjoint:

$$\langle K_V(W), W'\rangle = \langle W, K_V(W')\rangle$$

This follows immediately from the symmetry relation

$$\langle R(V,W)V', W'\rangle = \langle R(V',W')V, W\rangle$$

Therefore we may choose an orthogonal basis $U_1, \dots, U_n$ for TM_p so that

$$K_V(U_i) = e_i U_i$$

where $e_1, \dots, e_n$ are the orthogonal eigenvalues. Extend the U_i to vector fields along γ by parallel translation. Then since M is locally symmetric, the condition

$$R(V, U_i)V = e_i U_i$$

remains true everywhere along γ. Any vector field W along γ may be expressed uniquely as

$$W(t) = w_1(t)U_1(t) + \cdots + w_n(t)U_n(t)$$

Then the Jacobi equation $\dfrac{\mathrm{D}^2 W}{\mathrm{d}t} + K_V(W) = 0$ takes the form

$$\sum \frac{\mathrm{d}^2 w_i}{\mathrm{d}t^2} U_i + \sum e_i w_i U_i = 0$$

Since the U_i are everywhere linearly independent this is equivalent to the system of n equations

$$\frac{\mathrm{d}^2 w_i}{\mathrm{d}t^2} + e_i w_i = 0$$

We are interested in solutions that vanish at $t = 0$. If $e_i > 0$ then

$$w_i(t) = e_i \sin(\sqrt{e_i} t), \quad \text{for some constant } c_i$$

Then the zeros of $w_i(t)$ are at the multiples of $t = \pi/\sqrt{e_i}$.

If $e_i = 0$ then $w_i(t) = c_i t$ and if $e_i < 0$ then $w_i(t) = c_i \sinh(\sqrt{|e_i|} t)$ for some constant c_i. Thus if $e_i \leq 0$, $w_i(t)$ vanishes only at $t = 0$. This completes the proof. $\square$

4.2. **Lie Groups as Symmetric Space.** In this section we consider a Lie group G with a Riemannian metric which is invariant both under left translations

$$L_\tau : G \to G, \quad L_\tau(\sigma) = \tau\sigma$$

and right translations. If G is commutative such a metric certainly exists. If G is compact then such a metric can be constructed as follows: Let $\langle , \rangle$ ne any Riemannian metric on G, and let μ denote the Haar measure on G. Then μ is right and left invariant. Define a new inner product $\langle\langle , \rangle\rangle$ on G by

$$\langle\langle V, W\rangle\rangle = \int_{G\times G} \langle L_{\sigma_*} R_{\tau_*}(V), L_{\sigma_*} R_{\tau_*}(W)\rangle \mathrm{d}\mu(\sigma)\mathrm{d}\mu(\tau)$$

Lemma 4.2.1. *If G is a Lie group with a left and right invariant metirc, then G is a symmetric space. The reflection I_τ in any point $\tau \in G$ is given by the formula $I_\tau(\sigma) = \tau\sigma^{-1}\tau$.*

Proof. By hypothesis L_τ and R_τ are isometries. Define a map $I_e : G \to G$ by

$$I_e(\sigma) = \sigma^{-1}$$

Then $I_{e_*} : TG_e \to TG_e$ reverses the tangent space of e; so is certainly an isometry on this tangent space. Now the identity

$$I_e = R_{\sigma^{-1}} I_e L_{\sigma^{-1}}$$

shows that $I_{e_*} : TG_\sigma \to TG_{\sigma^{-1}}$ is an isometry for any $\sigma \in G$. Since I_e reverses the tangent space at e, it reverses geodesics through e.

Finally, defining $I_\tau(\sigma) = \tau\sigma^{-1}\tau$, the identity $I_\tau = R_\tau I_e R_\tau^{-1}$ shows that each I_τ is an isometry which reverses geodesics through γ. □

Definition 4.2.2. *A **1-parameter subgroup** of G is a C^∞ homomorphism of $\mathbb{R}$ into G.*

Remark 4.2.3. *It is well known that a 1-parameter subgroup of G is determined by its tangent vector at e.*

Lemma 4.2.4. *The geodesics γ in G with $\gamma(0) = e$ are precisely the 1-parameter subgroups of G*

Proof. Let $\gamma : \mathbb{R} \to G$ be a geodesic with $\gamma(0) = e$. By lemma 4.1.2 the map $I_{\gamma(t)} I_e$ takes $\gamma(u)$ into $\gamma(u + 2t)$. Now $I_{\gamma(t)} I_e(\sigma) = \gamma(t)\sigma\gamma(t)$ so $\gamma(t)\gamma(u)\gamma(t) = \gamma(u + 2t)$. By induction it follows that $\gamma(nt) = \gamma(t)^n$ for any integer n. If t'/t'' is rational so that $t' = n't$ and $t'' = n''t$ for some t and some integers n' and n'' then $\gamma(t' + t'') = \gamma(t)^{n'+n''} = \gamma(t)\gamma(t'')$. By continuity γ is a homomorphism.

Now let $\gamma : \mathbb{R} \to G$ be a 1-parameter subgroup. Let γ' be the geodesic through e such that the tangent vector of γ' at e is the tangent vector of γ at e. We have just seen that γ' is a 1-parameter subgroup. Hence $\gamma' = \gamma$ by remark 4.2.3. This completes the proof. □

Definition 4.2.5. *A vector field X on a Lie group G is called* ***left invariant*** *if and only if $(L_a)_*(X_b) = X_{ab}$ for every a and b in G.*

Remark 4.2.6. *If X and Y are left invariant then $[X, Y]$ is also. The Lie algebra $\mathfrak{g}$ of G is the vector space of all left invariant vector fields, made into an algebra by the bracket [ ].*

$\mathfrak{g}$ is actually a Lie algebra beacause the Jacobi identity

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$$

holds for all vector fields X, Y and Z.

Theorem 4.2.7. *Let G be a Lie group with a left and right invariant Riemannian metric. If X, Y, Z and W are left invariant vector fields on G, then:*

1. $\langle [X, Y], Z\rangle = \langle X, [Y, Z]\rangle$
2. $R(X, Y)Z = \frac{1}{4}[[X, Y], Z]$
3. $\langle R(X, Y)Z, W\rangle = \frac{1}{4}\langle [X, Y], [Z, W]\rangle$

Proof. For any left invariant X the identity

$$X \vdash X = 0$$

is satisfied, since the integral curves of X are left translates of 1-parameter subgroups, and therefore are geodesics.

Therefore

$$(X + Y) \vdash (X + Y) = (X \vdash X) + (X \vdash Y) + (Y \vdash X) + (Y \vdash Y)$$

is zero; Hence

$$X \vdash Y + Y \vdash X = 0$$

On the other hand

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

Adding these two conditions we obtain:

$$2(X \vdash Y) = [X, Y]$$

Now recall the identity

$$4. \quad Y\langle X, Z\rangle = \langle Y \vdash X, Z\rangle + \langle X, Y \vdash Z\rangle$$

The left side of this equation is zero, since $\langle X, Z\rangle$ is constant. Substituting formula 4 in this equation we obtain

$$0 = \langle [X, Y], Z\rangle + \langle X, [Y, Z]\rangle$$

Finally, using the skew commutativity of $[Y, X]$, we obtain the required formula

$$1. \quad \langle [X, Y], Z\rangle = \langle X, [Y, Z]\rangle$$

By definition, $R(X, Y)Z$ is equal to

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

Substituting formula 4, this becomes

$$-\frac{1}{4}[X,[Y,Z]]+\frac{1}{4}[Y,[X,Z]]+\frac{1}{2}[[X,Y],Z]$$

Using Jacobi identity, this yields the required formula

$$2. \quad R(X,Y)Z=\frac{1}{4}[[X,Y],Z]$$

And the formula 3 follows from 1 and 2. $\square$

Corollary 4.2.8. *The sectional curvature* $\langle R(X,Y)X,Y\rangle = \langle [X,Y],[X,Y]\rangle$ *is always* ≥ 0. *Equality holds if and only if* $[X,Y]=0$

Definition 4.2.9. *The* ***center*** *of a Lie algebra* $\mathfrak{g}$ *is defined to be the set of* $X\in\mathfrak{g}$ *such that* $[X,Y]=0$ *for all* $Y\in\mathfrak{g}$.

Corollary 4.2.10. *If* G *has a left and right invariant metric, and if the Lie algebra* $\mathfrak{g}$ *has trivial center, then* G *is compact, with finite fundamental group.*

Proof. $\square$

This result can be sharpened slightly as follows.

Corollary 4.2.11. *A simply connected Lie group* G *with left and right invariant metric splits as a Cartesian product* $G'\times\mathbb{R}^k$ *where* G' *is compact and* $\mathbb{R}^k$ *denotes the additive Lie group of some Euclidean space. Furthermore, the Lie algebra of* G' *has trivial center.*

Conversely it is clear that any such product possesses a left and right invariant metric.

Proof. $\square$

Theorem 4.2.12. (*Bott*) *Let* G *be a compact, simply connected Lie group. Then the loop space* $\Omega(G)$ *has the homotopy type of a CW-complex with no odd dimensional cells, and with only finitely many* λ*-cell for each even value of* λ. *Thus the homology group of* $\Omega(G)$ *is zero for* λ *odd, and is free abelian of finite rank for* λ *even.*

Remark 4.2.13. *This CW-complex will always be infinite dimensional. As an example, if* G *is the group* S^3 *of unit quaternions, then we have seen that the homology group* $H_i(\omega(S^3))$ *is infinite cyclic for all even values of* i.

Remark 4.2.14. *This theorem remains true even for a non-compact group. In fact any connected Lie group contains a compact subgroup as deformation retract.*

Proof. $\square$

4.3. **Whole Manifolds of Minimal Geodesics.** So far we have used a path space $\Omega(M; p, q)$ based on two points $p, q \in M$ which are in "general position". However, Bott has pointed out that every useful results can be obtained by considering pairs p, q in some special positions. As an example let M be the unit sphere S^{n+1}, and let p, q be antipodal points. Then there are infinitely many minimal geodesics from p to q. In fact the space Ω^{π^2} of minimal geodesics forms a smooth manifold of dimension n which can be identified with the equator $S^n \subset S^{n+1}$. We will see that this space of minimal geodesics provides a fairly good approximation to the enter loop space $\Omega(S^n)$.

Let M be a complete Riemannian manifold, and let $p, q \in M$ be two points with distance $\rho(p, q) = \sqrt{d}$.

Theorem 4.3.1. *If the space Ω^d of minimal geodesics from p to q is a topological manifold, and if every non-minimal geodesic from p to q has index $\geq \lambda_0$, then the relative homotopy group $\pi_i(\Omega, \Omega^d)$ is zero for $0 \leq i < \lambda_0$.*

Thus we obtain:

Corollary 4.3.2. *With the same hypothesis, $\pi_i(\Omega^d)$ is isomorphic to $\pi_{i+1}(M)$ for $0 \leq i \leq \lambda_0 - 2$.*

Let us apply this corollary to the case of two antipodal points on the $(n+1)$-sphere. Evidently the hypothesis are satisfied with $\lambda_0 = 2n$. For any non-minimal geodesic must wind one and a half times around S^{n+1}; and contains two conjugate poitns, each of multiplicity n, in its interior. This proves the following

Corollary 4.3.3. (*The fundamental suspension theorem*) *The homotopy group $\pi_i(S^n)$ is isomorphic to $\pi_{i+1}(S^{n+1})$ for $i \leq 2n - 2$.*

The theorem 4.3.1 also implies the homology groups of the loop space Ω are isomorphic to those of Ω^d in dimensions $\leq \lambda_0 - 2$. This fact follows from theorem 4.3.1 together with the relative Hurewicz theorem.

4.4. **The Bott Periodicity Theorem for the Unitary Group.**

Definition 4.4.1. *The* ***unitary group*** $U(n)$ *is defined to be the group of all linear transformation* $S : \mathbb{C}^n \to \mathbb{C}^n$ *which preserve the Hermitian inner product.*

Definition 4.4.2. *For any* $n \times n$ *complex matrix* A *the* ***exponential*** *of* A *is defined by the convergent power series expansion*

$$\exp A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$$

The following properties are easily verified

1. $\exp(A^*) = (\exp A)^*; \exp(TAT^{-1}) = T(\exp A)T^{-1}$
2. If A and B commute then

$$\exp(A + B) = (\exp A)(\exp B)$$

In particular:

3. $(\exp A)(\exp -A) = I$
4. The function exp maps a neighborhood of 0 in the space of $n \times n$ matrices diffeomorphically onto a neighborhood of I.

If A is skew-Hermitian, that is $A + A^* = 0$, then it follows from 1 and 3 that $\exp A$ is unitary. Conversely if $\exp A$ is unitary, and A belongs to a sufficiently small neighborhood of 0, then it follows $1, 3$ and 4 that $A + A^* = 0$. From these facts one easily proves that:

5. $U(n)$ is a smooth submanifold of the space of $n \times n$ matrices;
6. the tangent space $TU(n)_I$ can be identified with the space of $n \times n$ skew-Hermitian matrices.

Therefore the Lie algebra $\mathfrak{g}$ of $U(n)$ can also be identified with the space of skew-Hermitian matrices. Computaion shows that the bracket product of left invariant fields correspond to the product $[A, B] = AB - BA$ of matrices.

Since $U(n)$ is compact, it possesses a left and right invariant Riemannian metric. Note that the function

$$\exp : TU(n)_I \to U(n)$$

defined by exponentiation of matrices coincides with the function exp defined by following geodesics on the resulting Riemannian manifold. In fact for each skew-Hermitian matrix A the correspondence

$$t \to \exp(tA)$$

defines a 1-parameter subgroup of $U(n)$; and hence defines a geodesic.

A specific Riemannian metric on $U(n)$ can be defined as follows. Given matrices $A, B \in \mathfrak{g}$, let $\langle A, B\rangle$ denote the real part of the complex number

$$\operatorname{trace}(AB^*) = \sum_{i,j} A_{ij} B_{ij}$$

Clearly this inner product is positive defined on $\mathfrak{g}$.

This inner product on $\mathfrak{g}$ determines a unique left invariant Riemannian metric on $U(n)$. To verify that the resulting metirc is also right invariant, we must check that it is invariant under the adjoint action of $U(n)$ on $\mathfrak{g}$.

Definition 4.4.3. *Each $S \in U(n)$ determines an inner automorphism*

$$X \to SXS^{-1} = (L_S R_S^{-1})X$$

of the group $U(n)$. The induced linear mapping

$$(L_S R_S^{-1})_* : TU(n)_I \to TU(n)_I$$

*is called **adjoint action**, denoted by* $\mathrm{Ad}(S)$.

Thus $\mathrm{Ad}(S)$ is an automorphism of the Lie algebra of $U(n)$. In fact, we have the explict formula

$$\mathrm{Ad}(S)A = SAS^{-1}, \quad \forall A \in \mathfrak{g}, S \in U(n)$$

The inner product $\langle A, B\rangle$ is invariant under each such automorphism $\mathrm{Ad}(S)$. In fact if $A_1 = \mathrm{Ad}(S)A, B_1 = \mathrm{Ad}(S)B$, then the identity

$$A_1 B_1^* = SAS^{-1}(SBS^{-1})^* = SABS^{-1}$$

implies that

$$\mathrm{trace}(A_1 B_1^*) = \mathrm{trace}(SAB^*S^{-1}) = \mathrm{trace}(AB^*)$$

and hence that

$$\langle A_1, B_1\rangle = \langle A, B\rangle$$

It follow that the corresponding left invariant metric on $U(n)$ is also right invariant.

Given $A \in \mathfrak{g}$ we know by ordinary matrix theory that there exists $T \in U(n)$ such that TAT^{-1} is in diagonal form

$$TAT^{-1} = \mathrm{diag}(ia_1, \ldots, ia_n)$$

where the a_i' s are real. Also, given any $S \in U(n)$, there is a $T \in U(n)$ such that

$$TST^{-1} = \mathrm{diag}(e^{ia_1}, \ldots, e^{ia_n})$$

where again the a_i's are real. Thus we see directly that $\exp : \mathfrak{g} \to U(n)$ is onto.

One may treat the special unitary group $SU(n)$ in the same way. If exp is regarded as the ordinary exponential map of matrices, it is easy to show, using the diagonal form, that

$$\det(\exp A) = e^{\mathrm{trace}\, A}$$

Using this equation, one may show that $\mathfrak{g}'$, the Lie algebra of $SU(n)$ is the set of all matrices A such that $A + A^* = 0$ and $\mathrm{trace}\, A = 0$.

In order to apply Morse theory to the topology of $U(n)$ and $SU(n)$, we begin by considering the set of all geodesics in $U(n)$ from I to $-I$. In other words, we look for all $A \in TU(n)_I = \mathfrak{g}$ such that $\exp A = -I$. Suppose A

is such a matrix; if it is not already in diagonal form, let $T \in U(n)$ be such that TAT^{-1} is in diagonal form. Then

$$\exp TAT^{-1} = T(\exp A)T^{-1} = T(-I)T^{-1} = -I$$

so that we may as well assume that A is already in diagonal form

$$A = \operatorname{diag}(ia_1, \dots, ia_n)$$

In this case

$$\exp A = \operatorname{diag}(e^{ia_1}, \dots, e^{ia_n})$$

so that $\exp A = -I$ if and only if A has the form

$$\exp A = \operatorname{diag}(k_1 i\pi, \dots, k_n i\pi)$$

for some odd integers $k_1, \dots, k_n$.

Since the length of the geodesic $t \to \exp tA$ from $t = 0$ to $t = 1$ is $|A| = \sqrt{\operatorname{trace} AA^*}$, the length of the geodesic determined by A is $\pi\sqrt{k_1^2 + \cdots + k_n^2}$. Thus A determines a minimal geodesic if and only if each k_i equals ± 1, and in that case, the length is $\pi\sqrt{n}$. Now, regarding such A as a linear map of $\mathbb{C}^n$ to $\mathbb{C}^n$ observe that A is competely determined by specifying eigen$(i\pi)$, the vector space consisting of all $v \in \mathbb{C}^n$ such that $Av = i\pi v$ and eigen$(-i\pi)$. Since $\mathbb{C}^n$ splits as the orthogonal sum eigen$(i\pi) \oplus$ eigen$(-i\pi)$, the matrix A is then complete determined by eigen$(i\pi)$, which is an arbitrary subspace of $\mathbb{C}^n$. Thus the space of all minimal geodesics in $U(n)$ from I to $-I$ may be identified with the space of all sub-vector-spaces of $\mathbb{C}^n$.

Unfortunately, this space is rather inconvenient to use since it has components of varying dimensions. This difficulty may be removed by replacing $U(n)$ by $SU(n)$ and setting $n = 2m$. In this case, all the above considerations remain valid. But the additional condition that $a_1 + \cdots + a_{2m} = 0$ with $a_i = \pm\pi$ restricts eigen$(i\pi)$ to being an arbitrary m dimension sub-vector-space of $\mathbb{C}^{2m}$. This proves the following:

Lemma 4.4.4. *The space of minimal geodesics from I to $-I$ in the special unitary group $SU(2m)$ is homeomorphic to the complex Grassmann manifold $G_m(\mathbb{C}^{2m})$, consisting of all dimensional vector subspaces of $\mathbb{C}^{2m}$*

We will prove the following result at the end of this section

Lemma 4.4.5. *Every non-minimal geodesic from I to $-I$ in $SU(2m)$ has index $\geq 2m + 2$.*

Combining these two lemmas with section 4.3 we obtain:

Theorem 4.4.6. (*Bott*) *The inclusion map $G_m(\mathbb{C}^{2m}) \to \omega(SU(2m); I, -I)$ induces isomorphism of homotopy groups in dimension $\leq 2m$. Hence*

$$\pi_i G_m(\mathbb{C}^{2m}) \cong \pi_{i+1} SU(2m)$$

for $i \leq 2m$

On the other hand using standard method of homotopy theory one obtains somewhat different isomorphisms.

Lemma 4.4.7. *The group $\pi_i G_m(\mathbb{C}^{2m})$ is isomorphic to $\pi_{i-1}U(m)$ for $i \leq 2m$. Furthermore, $\pi_{i-1}U(m) \cong \pi_{i-1}(m+1) \cong \pi_{i-1}U(m+2) \cong \dots$ for $i \leq 2m$; and $\pi_j U(m) \cong \pi_j SU(m)$ for $j \neq 1$.*

Proof. First note that for each m there exists a fibration

$$U(m) \to U(m+1) \to S^{2m+1}$$

From the homotopy exact sequence

$$\dots \to \pi_i S^{2m+1} \to \pi_{i-1}U(m) \to \pi_{i-1}U(m+1) \to \pi_{i-1}S^{2m+1} \to \dots$$

of this fibration we see that

$$\pi_{i-1}U(m) \cong \pi_{i-1}U(m+1)$$

for $i \leq 2m$. It follows that the inclusion homomorphisms

$$\pi_{i-1}U(m) \to \pi_{i-1}U(m+1) \to \pi_{i-1}U(m+2) \to \dots$$

are all isomorphisms for $i \leq 2m$. These mutually isomorphic group are called the $(i-1)$-th **stable homotopy group** of the unitary group. They will be denoted briefly by $\pi_{i-1}U$

The same exact sequence shows that, for $i = 2m+1$, the homomorphism $\pi_{2m}U(m) \to \pi_{2m}U(m+1) \cong \pi_{2m}U$ is onto. □

Combining lemma 4.4.7 and theorem 4.4.6 we see that

$$\pi_{i-1}U = \pi_{i-1}U(m) \cong \pi_i G_m(\mathbb{C}^{2m}) \cong \pi_{i+1}SU(2m) \cong \pi_{i+1}U$$

for $1 \leq i \leq 2m$. Thus we obtain

Theorem 4.4.8. (*Periodicity Theorem*)

$$\pi_{i-1}U \cong \pi_{i+1}U, \quad \text{for } i \geq 1$$

To evaluate these group it is now sufficient to observe that $U(1)$ is a circle; so that

$$\pi_0 U = \pi_0 U(1) = 0$$
$$\pi_1 U = \pi_1 U(1) \cong \mathbb{Z}$$

As a check, since $SU(2)$ is a 3-sphere, we have:

$$\pi_2 U = \pi_2 SU(2) = 0$$
$$\pi_3 U = \pi_3 SU(2) \cong \mathbb{Z}$$

Thus we have proved the following result.

Theorem 4.4.9. (*Bott*) *The stable homotopy group $\pi_i U$ of the unitary group are periodic with periodic with period 2. In fact the groups*

$$\pi_0 U \cong \pi_2 U \cong \pi_4 U \cong \dots$$

are zero, and the groups

$$\pi_1 U \cong \pi_3 U \cong \pi_5 U \cong \dots$$

are infinitely cyclic.

4.5. **The Periodicity Theorem for the Orthogonal Group.** This section will carry out an analogous study of the iterated loop space of the orthogonal group. However the treatment is rather sketchy, and many details are left out.

Definition 4.5.1. *Consider the vector space $\mathbb{R}^n$ with the usual inner product. The orthogonal group $O(n)$ consists of all linear maps*

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

which preserves this inner product.

Remark 4.5.2. *Alternatively $O(n)$ consists of all real $n \times n$ matrices T such that $TT^* = I$. This group can be considered as a smooth subgroup of the unitary group $U(n)$; and therefore inherits a right and left invariant Riemannian metric.*

Now suppose that n is even.

Definition 4.5.3. *A* ***complex structure*** *J on $\mathbb{R}^n$ is a linear transformation $J : \mathbb{R}^n \to \mathbb{R}^n$, belonging to the orthogonal group, which satisfies the identity $J^2 = -I$. The space consisting of all such complex structures on $\mathbb{R}^n$ will be denoted by $\Omega_1(n)$.*

Remark 4.5.4. *Given some fixed $J_1 \in \Omega_1(n)$ let $U(n/2)$ be the subgroup of $O(n)$ consisting of all orthogonal transformation which commutes with J_1. Then $\Omega_1(n)$ can be identifiedwith the quotient space $\Omega(n)/U(n/2)$.*

Lemma 4.5.5. *The space of minimal geodesics from I to $-I$ on $O(n)$ is homeomorphic to the space $\Omega_1(n)$ of complex structures on $\mathbb{R}^n$.*

Proof. $\square$

Part 2. A newer approach: gradient flow lines

In this part, we will introduce an auxiliary Riemannian metric g on X, and consider the negative gradient vector field of f with respect to g, and denoted by V. One then looks at flow lines of the vector field V which start at one critical point and end at another. If the metric is generic, then there are finitely many gradient flow lines from a critical point of index i to a critical point of index $i-1$. Then one can define a chain complex $C_*^{\mathrm{Morse}}(f,g)$ over $\mathbb{Z}$, called the Morse complex, whose chain group C_i is generated by the critical points of index i, and whose differential counts gradient flow lines between critical points of index difference one. A fundamental result is that the homology of this chain complex is canonically isomorphic to the singular homology of X. Roughly speaking, the isomorphism from Morse homology to singular homology sends a critical point to its descending manifold. One can easily deduces the Morse inequalities from this: there have to be enough critical points to generate the homology.

The significance of the language of gradient flow lines is that, as realized by Floer, it extends to important infinite dimensional cases where the classical approach is useless. These are cases where the critical points have infinite index, so that passing through a critical point does not change the topology of X_a. However sometimes the index difference between two critical points is still finite, in that one can make sense of gradient flow lines between two critical points, and these form a finite dimensional moduli space.

5. The definition of Morse homology

5.1. The gradient flow.

Let g be a metric on finite-dimensional compact manifold X, and let $-\operatorname{grad} f$ denote the negative gradient of f with respect to g, that is, $\operatorname{grad} f$ is defined by

$$\langle \operatorname{grad} f, \cdot \rangle = \mathrm{d}f$$

When X is embedded in $\mathbb{R}^N$, this vector field is given explictly by

$$\operatorname{grad} f(x) = (\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_N}(x))$$

For any vector field V, we can define the following concepts for it

Definition 5.1.1. *An* ***flow line*** *of the vector field V is a differentiable function of the form*

$$\gamma : U \to X$$

for $U \subset \mathbb{R}$ an open interval with the property that its tangent vector at any $t \in U$ equals the value of the vector field V at the point $\gamma(t)$.

Definition 5.1.2. *A (**global**[‡]) **flow** of V is a function of form*

$$\Psi : X \times \mathbb{R} \to X$$

such that for any $x \in X$ the function $\Psi(x, -) : \mathbb{R} \to X$ is a flow line of V.

The flow of the vector field $-\operatorname{grad} f$ defines a one-parameter group of diffeomorphisms $\Psi_s : X \to X$ for $s \in \mathbb{R}$ with $\Psi_0 = \mathrm{id}$ and $\mathrm{d}\Psi_s/\mathrm{d}t(x) = -\operatorname{grad} f$.

If p is a critical point, then we define the following two concepts:

Definition 5.1.3. *The **descending manifold** is defined as follows*

$$\mathscr{D}(p) := \{x \in X \mid \lim_{s \to -\infty} \Psi_s(x) = p\}$$

Definition 5.1.4. *The **ascending manifold** is defined as follows*

$$\mathscr{A}(p) := \{x \in X \mid \lim_{s \to +\infty} \Psi_s(x) = p\}$$

Remark 5.1.5. *Returning to the case where $X \subset \mathbb{R}^N$ and f is the height function, it is clear why $\mathscr{D}(p)$ and $\mathscr{A}(p)$ are called descending manifold and ascending manifold. We can think of pouring syrup on X, and the points of $\mathscr{D}(p)$ are the points that move towards p and $\mathscr{A}(p)$ are the points that move away from p.*

Remark 5.1.6. *These are sometimes also called the "unstable manifold" and "stable manifold", respectively, of the flow V.*

Proposition 5.1.7. *The ascending manifold $\mathscr{A}(p)$ is diffeomorphic to an open disc of dimension $n - \operatorname{ind}(p)$ and the descending manifold $\mathscr{D}(p)$ is diffeomorphic to an open disc of dimension $\operatorname{ind}(p)$.*

Proof. (sketch) The idea is to look at the linearized flow of $-\operatorname{grad} f$ near p. The index, which is the number of negative eigenvalues of $H_p(f)$, reprensents the number of linearly independent directions that the gradient flow is pointing towards p. Since there are no eigenvalue of 0, the number of "in" directions plus the number of "out" directions must be n. This determines the dimension of $\mathscr{D}(p)$, and similarly the complementary dimension of $\mathscr{A}(p)$. ☐

We assume for the rest of this section that the pair (f, g) is **Morse Smale**: namely, f is Morse and for every pair of critical points p and q, the descending manifold $\mathscr{D}(p)$ is transverse to the ascending manifold $\mathscr{A}(q)$. We will see later that this condition holds generically.

Definition 5.1.8. *If p and q are critical points, a **flow line** from p to q is a path $\gamma : \mathbb{R} \to X$ with $\gamma'(s) = -\operatorname{grad} f(\gamma(s))$ and $\lim_{s \to -\infty} \gamma(s) = p$ and $\lim_{s \to +\infty} \gamma(s) = q$.*

[‡]In general, not every vector field admits a global flow, such vector field is called complete vector field. However, when the manifold X is compact, then any vector field on X is complete.

Remark 5.1.9. *Note that any point $x \in \mathscr{A}(q) \cap \mathscr{D}(p)$ determines a path γ from p to q, given by integrating the vector field $-\operatorname{grad} f$ with initial x. Therefore we can identity the space of paths from p to q with $\mathscr{A}(q) \cap \mathscr{D}(p)$. Note that starting at any $x' \in \gamma(t_0), t_0 \neq 0$ and integrating will give the same path γ that x give. Therefore we equip $\mathscr{A}(p) \cap \mathscr{D}(p)$ with the relation $x \sim x' \iff x' = \gamma(t_0)$ for some $t_0 \in \mathbb{R}$.*

Definition 5.1.10. *Let $\mathscr{M}(p,q)$ denote the moduli space of flow lines from p to q, modulo translation, that is,*

$$\mathscr{M}(p,q) = \mathscr{A}(q) \cap \mathscr{D}(p)/\sim$$

Proposition 5.1.11. *When (f,g) is Morse-Smale, the moduli space $\mathscr{M}(p,q)$ for distinct critical points p,q is a manifold with dimension*

$$\dim \mathscr{M}(p,q) = \operatorname{ind}(p) - \operatorname{ind}(q) - 1$$

Proof. The Morse-Smale condition implies that the intersection $\mathscr{A}(q) \cap \mathscr{D}(p)$ is transverse, and therefore a manifold of dimension $n - \operatorname{ind}(q) + \operatorname{ind}(p) - n = \operatorname{ind}(p) - \operatorname{ind}(q)$. The $\mathbb{R}$ action is free and smooth, therefore the quotient is a smooth manifold of dimension

$$\operatorname{ind}(p) - \operatorname{ind}(q) - \dim \mathbb{R} = \operatorname{ind}(p) - \operatorname{ind}(q) - 1$$

□

It will be useful to have an orientation on $\mathscr{M}(p,q)$. Note that we have a natural isomorphism of tangent spaces

$$\begin{aligned} T\mathscr{D}(p) &\cong T_x(\mathscr{D}(p) \cap \mathscr{A}(q)) \oplus T_x(X/\mathscr{A}(q)) \\ &\cong T_\gamma \mathscr{M}(p,q) \oplus T_x\gamma \oplus T_x\mathscr{D}(q) \end{aligned}$$

We orient $\mathscr{M}(p,q)$ so that the isomorphism is orientation-preserving.

5.2. **Compactification by broken flow lines.** To motivate why we would like $\mathscr{M}(p,q)$ to be compact, consider the case where $\operatorname{ind}(p) - \operatorname{ind}(q) = 1$. By above, the moduli space will be dimension 0. If $\mathscr{M}(p,q)$ is compact, this would mean that it is just a finite collection of points. This will be the basis for defining the Morse homology chain complex. But in general, $\mathscr{M}(p,q)$ is not compact, so we must look at its compactification.

Theorem 5.2.1. *If X is closed and (f,g) is Morse-Smale, then for any two critical points p,q, the moduli space $\mathscr{M}(p,q)$ has a natural compactification to a smooth manifold with corners. The k-dimensional corners of it are*

$$\overline{\mathscr{M}(p,q)}_k = \bigcup_{\substack{r_1,\ldots,r_k \in \operatorname{Crit}(f)\\ p\neq r_i \neq q}} \mathscr{M}(p,r_1)\times\mathscr{M}(r_1,r_2)\times\cdots\times\mathscr{M}(r_{k-1},r_k)\times\mathscr{M}(r_k,q_1)$$

This is saying that we can compactify the $\mathscr{M}(p,q)$ by adding the space of flow lines between p and q that break at points r_i of intermediate index.

Corollary 5.2.2. *If $p,q \in \operatorname{Crit}(f)$ are such that $\operatorname{ind}(p) - \operatorname{ind}(q) = 1$, then $\mathscr{M}(p,q)$ is compact.*

Proof. Using the fact there are no points r_i of intermediate index when $\operatorname{ind}(p) - \operatorname{ind}(q) = 1$. $\square$

Corollary 5.2.3. *If p,q are two distinct critical points such that $\operatorname{ind}(p) - \operatorname{ind}(q) = 2$, then $\overline{\mathscr{M}(p,q)}$ is a compact 1-manifold with boundary and*

$$\partial\overline{\mathscr{M}(p,q)} = \bigcup_{\operatorname{ind}(r)=\operatorname{ind}(q)+1} \mathscr{M}(p,r)\times\mathscr{M}(r,q)$$

5.3. **The chain complex.**

Definition 5.3.1. *Morse complex $(C_*^{Morse}(f,g), \partial^{Morse})$ is defined as follows:*

Let $\mathrm{Crit}_i(f)$ *denote the set of index i critical points of f. The chain module C_i is the free $\mathbb{Z}$-module generated by this finite set:*

$$C_i^{Morse}(f,g) := \mathbb{Z}\,\mathrm{Crit}_i(f)$$

The differential counts gradient flow lines. That is, if $p \in \mathrm{Crit}_i(f)$, then

$$\partial^{Morse}(p) := \sum_{q \in \mathrm{Crit}_{i-1}(f)} \#\mathscr{M}(p,q) \cdot q.$$

Here $\#\mathscr{M}(p,q) \in \mathbb{Z}$ denotes the number of points in $\mathscr{M}(p,q)$, counted with the signs given by the orientation on $\mathscr{M}(p,q)$.

Lemma 5.3.2. $(\partial^{Morse})^2 = 0$

Proof.

$$\begin{aligned}
\partial_{i-1}^M \circ \partial_i^M(p) &= \sum_{r \in \mathrm{Crit}_{i-1}(f)} \sum_{q \in \mathrm{Crit}_{i-2}(f)} \#\mathscr{M}(p,r)\#\mathscr{M}(r,q) \cdot q \\
&= \sum_{q \in \mathrm{Crit}_{i-2}(f)} \#(\bigcup_{r \in \mathrm{Crit}_{i-1}(f)} \mathscr{M}(p,r) \times \mathscr{M}(r,q)) \cdot q \\
&= \sum_{q \in \mathrm{Crit}_{i-1}(f)} \#\partial\overline{\mathscr{M}(p,q)} \cdot q \\
&= 0
\end{aligned}$$

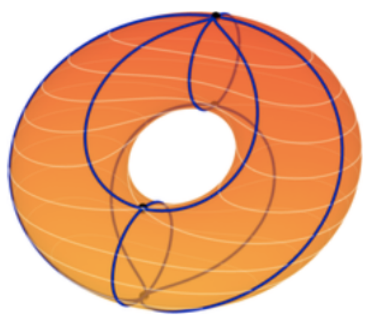
Here we used the Corollary5.2.3 and the fact that the signed count of the boundary if a 1-manifold is zero. □

Definition 5.3.3. *The **Morse homology** is defined to be the homology of Morse chain complex.*

5.4. **Examples and Applications.** Morse homology gives a handy way of understanding singular homology, according to the isomorphism which we will establish later. Insofar as Morse functions, like the height function, are more more intuitive than simplices, Morse homology can be used to restate results in singular homology in a more intuitive way.

We start by computing the Morse homology of a few shapes using height function to show it is isomorphic to singular homology indeed.

Example 5.4.1. *Let T be the torus embedded in $\mathbb{R}^3$ as usually, but tilted at a small angle so that it is Morse-Smale as following figure*



Let p be the index 2 critical point, q be the index 0 critical point, and r_1, r_2 be the index 1 critical points, Then

$$C_0 = \langle q \rangle, \quad C_1 \langle r_1, r_2 \rangle, \quad C_2 = \langle p \rangle$$

From the figure, we see that there are two flow lines with oppsite sign from p to r_1. Similarly, there are two flow lines from p to r_2, also with oppsite sign. This means

$$\#\mathscr{M}(p, r_1) = \#\mathscr{M}(p, r_2) = 0$$

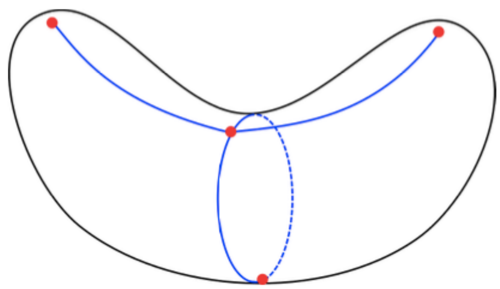
So we have $\partial_2^M = 0$. By the same reasoning, we have $\partial_1^M = 0$ as well.

Therefore our chain complex is

$$0 \to \mathbb{Z} \xrightarrow{0} \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{0} \mathbb{Z} \to 0$$

And the homology group are $\mathbb{Z}, \mathbb{Z} \oplus \mathbb{Z}$ and $\mathbb{Z}$ as desired according to singular homology.

Example 5.4.2. *Now compute $H_*^M(S^2)$. To do this, we can embed S^2 into $\mathbb{R}^2$ in the standard way, but it's a quite boring way, so instead we embed S^2 with two local maximum, as shown in the following figure*



Let p_1, p_2 be two local maximum, r be the saddle, and q be the local minimum. These have index $2, 1$ and 0, respectively. Therefore

$$C_2 = \langle p_1, p_2 \rangle, \quad C_1 = \langle r \rangle, \quad C_0 = \langle q \rangle$$

There is only one flow line from p_1 to r, and likewise only one from p_2 to r. Since these have the same sign, by symmetry, we have $\partial_2^M(p_1) = \partial_2^M(p_2) = r$. Thus the map $\partial_2^M : C_2 \to C_1$ is given by summing coefficients. And like with the case in torus, there are two flow lines from r to q, each with opposite sign. Therefore $\partial_1^M = 0$, then our chain complex is

$$0 \to \mathbb{Z} \oplus \mathbb{Z} \xrightarrow{+} \mathbb{Z} \xrightarrow{0} \mathbb{Z} \to 0$$

The Morse homology groups are $\mathbb{Z}, 0$ and $\langle a, -a \rangle \cong \mathbb{Z}$, as desired according to singular homology.

And now we will show applications in Poincaré duality and Morse inequalities under the assumption there exists isomorphism between Morse homology and singular homology, which will make the proof elegant.

Application 5.4.3. (*Poincaré Duality*) *We can define the Morse cohomology in the same way, except with the pair $(-f, g)$. We can define the chain group C_i in the same way as before. Furthermore, since negating f causes the index of each critical point to negate $\pmod n$, we can define maps $\delta_i^M : C^i \to C^{i+1}$ by*

$$\delta^M(p) := \sum_{\operatorname{ind}(p)=i+1} \#\mathscr{M}(p, q) \cdot q$$

Thus we get a cochain complex, from which we get the Morse cohomology. Since the critical points of $-f$ are the same as f, there is an isomorphism between C_i and C^{n-i}. This isomorphism then descends onto the level of homotopy groups since ∂ and δ are the same, then we arrive at Poincaré duality:

$$H_*^M(X) \cong (H^M)^{n-*}(X)$$

In a geometric point of view, Poincaré duality is the intuitive statement that looking at X upside down doesn't change its homotopy.

Application 5.4.4. (*Morse Inequalities*) *Isomorphism between Morse homology and singular homology allow us to relate the number of critical points of a Morse function on X to the Betti number of X.*

Let c_i be the number of critical of f with index i, and let b_i be its i-th Betti number. Since $H_i^M(X)$ is a quotient of a subgroup of C_i, its rank is at most c_i, then we have the inequalities $c_i \geq b_i$, which are known as Morse inequalities. Moreover, we have the following bound for the number of critical points of f

$$|\operatorname{Crit}(f)| \geq \sum_i b_i$$

6. Morse homology and Singular homology

We will now prove the following theorem, which is one of the most fundamental facts about finite-dimensional Morse theory.

Theorem 6.0.1. *If X is a closed smooth manifold and (f, g) is a Morse Smale pair on X, then there is a canonical isomorphism*

$$H_*^{Morse}(f, g) \cong H_*(X)$$

6.1. **Outline of the proof.**

Part 3. GIT and Symplectic Reduction for Bundles and Varieties

In this part we will give some Introductions on Geometric Invariant Theory and symplectic reduction, and apply them to moduli of bundles and varieties.

Before we enter the main content, there are many definitions we need to clearify.

7. What's GIT?

7.1. **Introduction.** It is the study of quotients in the context of algebraic geometry. Many objects we want to take a quotient always have some sort of geometric structures, and GIT allow us to construct quotients that preserve geometric structure.

We let a group G act on a geometric object X. The action of G gives a partition of X into G-orbits, so we can take quotient X/G. However, it is not always the case that the set of G-orbits has a geometric structure. We take the situation in smooth manifold for an example:

Suppose G is a Lie group and X is a smooth manifold, the quotient X/G will not always have the structure of a smooth manifold(For example, the presence of non-closed orbits, usually gives a non-Hausdorff quotient). However, if G acts properly and freely, then X/G has a smooth manifold structure, such that natural projection $\pi : X \to X/G$ is a smooth submanifold.

GIT consider the same thing under the context of algebraic geometry. As we can see in smooth manifold, only certain types of group(compared with Lie group) and group actions(compared with properly and freely) are allowed in the construction of GIT.

Now we consider a quite simple example, to catch the ideal of GIT.

Example 7.1.1. *Let M_n be the $n \times n$ matrices over $\mathbb{C}$, then we can give M_n a geometric structure by regarding it as an affine variety. Consider the conjugate action of GL_n on M_n. Can we regard $\mathrm{M}_n / \mathrm{GL}_n$ as a variety? The answer is yes and we will show it later.*

However, good thing does not happen always, consider

Example 7.1.2. *Let $\mathbb{C}^\times$ acts on $\mathbb{C}^2$ by $\lambda.(x, y) := (\lambda x, \lambda y)$. The $\mathbb{C}^\times$-orbits are $\{(\lambda x, \lambda y) : \lambda \in \mathbb{C}^\times, (x, y) \neq (0, 0)\}$ as well as the origin $\{(0, 0)\}$. However, this set of orbits can not have a structure of variety.*

So we need to be more careful when we constructing quotients in the category of varieties.

Definition 7.1.3. *A morphism $f : X \to Y$ is called G-invariant morphism, if it is constant on orbits.*

Definition 7.1.4. *In any category, we call a G-invariant morphism $\pi : X \to Y$ is categorical quotient of X by G, when for any G-invariant morphism*

$f : X \to Z$, we have that f factors uniquely through π. That is, there exists a unique $\bar{f}$ such that $\bar{f} \circ \pi = f$, for any G-invariant morphism f.

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr, "\pi"'] & & Z \\ & Y \arrow[ur, "\bar{f}"'] & \end{tikzcd}$$

Since categorical quotient is defined by its universal property, so it is unique when it exists. However, for a quotient in the category of varieties, simple being a categorical quotient may not have a good geometric properties, so we need to define good categorical quotient.

If G acts on a variety X, then we can get an action on the regular functions on X as follows. For $f \in \mathscr{O}(U), U \subset X$, we define

$$g.f(x) = f(g^{-1}.x)$$

For the types of group action we are interested in, we require

$$g.f \in \mathscr{O}(U), \quad \forall f \in \mathscr{O}(U)$$

Definition 7.1.5. *A surjective G-invariant map of varieties $p : X \to Y$ is called a good categorical quotient of X by G, if the following three properties holds*

1. *For all open $U \subset Y$, $p^* : \mathscr{O}(U) \to \mathscr{O}(p^{-1}(U))^G$ is an isomorphism.*
2. *If $W \subseteq X$ is closed and G-invariant, then $p(W) \subset Y$ is closed.*
3. *If $V_1, V_2 \subseteq X$ are closed, G-invariants, and $V_1 \cap V_2 = \varnothing$, then $p(V_1) \cap p(V_2) = \varnothing$.*

Remark 7.1.6. *Note that the first requirement implies a good categorical quotient must be a categorical one: If $f : X \to Z$ is a G-invariant morphism, then $f^* : \mathscr{O}(Z) \to \mathscr{O}(X)$ must embed in $\mathscr{O}(X)^G$. If p is a good categorical quotient, then p^* is an isomorphism to $\mathscr{O}(X)^G$, so*

$$\begin{tikzcd} \mathscr{O}(Z) \arrow[rr, "f^*"] \arrow[dr, dotted, "\bar{f}^*"'] & \arrow[d, phantom, "\circlearrowleft"] & \mathscr{O}(X)^G \arrow[r, hook] & \mathscr{O}(X) \\ & \mathscr{O}(Y) \arrow[ur, hook, "p^*"'] & & \end{tikzcd}$$

So f^* can factor through $\mathscr{O}(Y)$, and this factoring is unique since p^* is an isomorphism. By the anti-equivalence of category, the dual $f = \bar{f} \circ p$ is a unique factoring of f through p.

Remark 7.1.7. *As we can see in the above Remark, the first requirement already implies categorical quotient, the more restrictions intend to avoid bad situation in geometry, such as Example*7.1.2.

We denote by $X//G$ the good categorical quotient, or GIT quotient, of a variety X by a group G.

In the following, we will first construct GIT quotient in affine case, and this serves as a guide for projective case: we want to glue affine quotients to get projective one, since every projective variety admits an affine covering. Unfortunately, we can not cover the whole of a projective variety, which leads to the concept of semistability.

It's natural to define $X//G = \operatorname{Spec}\mathscr{O}(X)^G$ in affine cases, since $X = \operatorname{Spec}\mathscr{O}(X)$, so G-invariant regular functions may representate the quotient we desire, but for this we require that $\mathscr{O}(X)^G$ is finitely generated.

Historically, whether the ring of invariants is finitely generated or not is knowns as Hilbert's 14-th problem. For general linear group over $\mathbb{C}$, Hilbert showed that the invariant rings are always finitely generated. However, Nagata gave an counterexample that $\mathscr{O}(X)^G$ is not finitely generated, and proved that for any reductive group, $\mathscr{O}^G$ is finitely generated.

That's why we come to reductive groups now!

7.2. **Reductive groups.** Now we foucus on the reductive group which we can use to construct GIT quotient. We will define when a linear algebraic group is reductive and give some properties of it.

Definition 7.2.1. *A (linear) algebraic group is a subgroup of* $\mathrm{GL}(n,k)$ *which is an irreducible algebraic set.*

Example 7.2.2. *The set of unitary matrices with determinant* 1

$$\mathrm{SO}(2) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc - 1 = 0 \right\}$$

is an algebraic group§*.*

Example 7.2.3. $k^\times$ *is also an algebraic group, by the embedding* $\lambda \to \lambda I$

Example 7.2.4. *General linear group* $\mathrm{GL}(n,k)$ *is an algebraic group*¶*.*

Definition 7.2.5. *A linear algebraic group G over k is reductive if every representation* $\rho : G \to \mathrm{GL}(n,k)$ *has a decomposition as a direct sum of irreducible representations.*

In fact, many classical groups such as $\mathrm{GL}(n,\mathbb{C}), \mathrm{SL}(n,\mathbb{C})$ are reductive, now we give a proof of $\mathbb{C}^\times$ is a reductive group, the proof given here follows [5].

Proposition 7.2.6. *The multiplicative group* $\mathbb{C}^\times$ *is reductive.*

Proof. Let $\rho : \mathbb{C}^\times \to \mathrm{GL}(n,\mathbb{C})$ be a representation of $\mathbb{C}^\times$, we will show ρ has a decomposition as a direct sum of irreducible representations. Assume ρ is not irreducible.

Let $\langle , \rangle$ denote the standard inner product on $V = \mathbb{C}^n$, then define

$$\langle x; y \rangle := \int_0^{2\pi} \langle \rho(e^{i\theta})x, \rho(e^{i\theta})y \rangle d\theta$$

---

§In general, special linear group $\mathrm{SL}(n)$ is always an algebraic group by considering the irreducible polynomial $\det -1$.

¶We can check this by introducing a new variable T and consider irreducible polynomial $T \cdot \det -1$ with $n^2 + 1$ variables.

This form has the following property; $\langle\rho(g)x;\rho(g)y\rangle=\langle x;y\rangle$, where $x,y\in V, g=e^{i\psi}\in S=\{z\in\mathbb{C}^\times:|z|=1\}$. Indeed,

$$\begin{aligned}\langle\rho(e^{i\psi})x;\rho(e^{i\psi})y\rangle&=\int_0^{2\pi}\langle\rho(e^{i\theta}\rho(e^{i\psi}))x,\rho(e^{i\theta})\rho(e^{i\psi})y\rangle d\theta\\&=\int_0^{2\pi}\langle\rho(e^{i(\theta+\psi)})x,\rho(e^{i(\theta+\psi)})y\rangle d\theta\\&=\int_0^{2\pi}\langle\rho(e^{i\phi})x,\rho(e^{i\phi})y\rangle d\phi,\text{ where }\phi=\theta+\psi\\&=\langle x;y\rangle\end{aligned}$$

And also note that $\langle;\rangle$ is an inner product. If ρ is not irreducible, then there exists some $\mathbb{C}^\times$-invariant subspace U of V, let $W=U^\perp$ be the orthogonal complement of U with respect to $\langle;\rangle$. Then we can see W is S-invariant as follows

$$\begin{aligned}\langle u;\rho(g)w\rangle&=\left\langle\rho\left(g^{-1}\right)u;\rho\left(g^{-1}\right)\rho(g)w\right\rangle\\&=\left\langle\rho\left(g^{-1}\right)u;w\right\rangle\\&=0\end{aligned}$$

where $w\in W,u\in U,g\in S$. The last equality holds since U is S-invariant. What we need to do is to show W is $\mathbb{C}^\times$-invaraint.

Let N be the subset of $\mathbb{C}^\times$ which leaves W invariant, it contains S obviously. We will show that this set is closed in the Zariski topology. If we can do this, since all Zariski closed subset in $\mathbb{C}^\times$ are finite sets and whole space, so we can conclude $N=\mathbb{C}^\times$, as desired.

Let $W=\operatorname{span}\{e_1,\ldots,e_r\}$, and extends this basis to a basis $\{e_1,\ldots,e_n\}$ of V. Then we can regard W as solutions of equations

$$\langle v,e_i\rangle=0,\quad i=r+1,\ldots,n$$

these define polynomials which take the coordinate of v as variables, which we call it f_i, so we can see W as a zero set of $\{f_{r+1},\ldots,f_n\}$.

For each $i\in\{1,\ldots,r\},j\in\{r+1,\ldots,n\}$, consider the set $\{T\in\mathrm{GL}(V)\mid f_j(Te_i)=0\}$. If we fix i,j, this set is the zero set of a polynomial in the coordinates of T. So it's a closed set in $\mathrm{GL}(V)$, with respect to Zariski topology. Then we have $\{T\in\mathrm{GL}(V)\mid Te_i\in W\}=\bigcap_{j=r+1}^n\{T\in\mathrm{GL}(V)\mid f_j(Te_i)=0\}$ is closed, so

$$\{T\in\mathrm{GL}(V)\mid Te_i\in W,\forall i\in\{1,\ldots,r\}\}=\bigcap_{i=1}^r\{T\in\mathrm{GL}(V)\mid Te_i\in W\}$$

is closed, thus we have

$$\begin{aligned}\{T\in\mathrm{GL}(V)\mid Tw\in W,\forall w\in W\}&=\{T\in GL(V):T\left(\lambda_1e_1+\ldots+\lambda_re_r\right)\in U\text{ for all }\lambda_i\in\mathbb{C}\}\\&=\{T\in GL(V):\lambda_1\left(Te_1\right)+\ldots+\lambda_r\left(Te_r\right)\in U\text{ for all }\lambda_i\in\mathbb{C}\}\\&=\{T\in GL(V):Te_i\in W\text{ for each }i\in\{1,2,\ldots,r\}\}\end{aligned}$$

is closed with respect to Zariski topology, so $N = \rho^{-1}(\{T \in \mathrm{GL}(V) \mid Tw \in W, \forall w \in W\})$ is closed, as we desired. $\square$

Thanks to Maschke's theorem, we also have the following result.

Proposition 7.2.7. *Let G be a finite group, then G is reductive.*

Now we many examples of reductive alegbraic group, so we can define how to act on a variety X properly.

Definition 7.2.8. *For a reductive alegbraic group, we say that G acts rationally on a variety X if it acts by a morphism of varieties $G \times X \to X$*

But why we need reductive groups? and why this action? There are two key properties which might answer these questions.

Lemma 7.2.9. *Let G be a reductive group acting rationally on an affine variety X, then $\mathscr{O}(X)^G$ is finitely generated.*

Proof. See [6] for a proof. $\square$

The following lemma is used in the construction of GIT quotient. It allows us to find a G-invariant function which separates disjoint G-invariant sets.

Lemma 7.2.10. *Let G be a reductive group acting rationally on an affine variety $X \subset \mathbb{A}^n$. Let Z_1, Z_2 be two closed G-invariant subsets of X with $Z_1 \cap Z_2 = \varnothing$. Then there exists a G-invariant function $F \in \mathscr{O}(X)^G$ such that $F(Z_1) = 1, F(Z_2) = 0$.*

Proof. See [5] for a proof. $\square$

7.3. **The affine quotient.** We now have enough tools to construct the quotient of an affine variety by a reductive group.

For an affine variety X, the quotient of X by a reductive group G is just $\operatorname{Spec} \mathscr{O}(X)^G$. We will prove that this construction satisfies the required conditions being a good categorical quotient.

Theorem 7.3.1. *Let X be an affine variety and G be a reductive group acting rationally on X. Let $p^* : \mathscr{O}(X)^G \to \mathscr{O}(X)$ be defined by the inclusion $\mathscr{O}(X)^G \subseteq \mathscr{O}(X)$. Then the dual of this map, $p : X \to Y := \operatorname{Spec} \mathscr{O}(X)^G$ is a good categorical quotient.*

Proof. $\square$

Now we give a concrete example to show how powerful the GIT construction is, and gives the answer to the Example7.1.1 we mentioned at first.

Example 7.3.2. *Consider the set X of 2×2 matrices over $\mathbb{C}$, embedded in $\mathbb{C}^4$ by*

$$\begin{pmatrix} w & x \\ y & z \end{pmatrix} \mapsto (w, x, y, z)$$

It is an affine variety obviously, and consider the general linear group acts on it by conjugate action, then as the theorem above implies

$$X//G = \operatorname{Spec} k[w,z,y,z]^G$$

We know that there are two important invariants under conjugate action, that is, determinant and trace. In this case they are $\det = wz - xy$ *and* $\operatorname{tr} = w+z$, *so we have an obvious inclusion*

$$k[wz-xy, w+z] \subset k[w,x,y,z]^G$$

We will show that we in fact have equality.

Let $\lambda \in \mathbb{C}^\times$ *be arbitrary and consider the matrix* $A = \begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix}$. *For all matrices* $M = \begin{pmatrix} w & x \\ y & z \end{pmatrix}$, *we can calculate as follows*

$$\begin{aligned} A^{-1}MA &= \begin{pmatrix} 0 & -\frac{1}{\lambda} \\ -1 & 0 \end{pmatrix} \begin{pmatrix} w & x \\ y & z \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix} \\ &= \begin{pmatrix} z & \frac{y}{\lambda} \\ \lambda x & w \end{pmatrix} \end{aligned}$$

Let $f \in k[w,x,y,z]^G$, *i.e., we require* f *satisfy that* $f(w,x,y,z) = A.f(M) = f(A.M) = f(A^{-1}MA) = f(z, \frac{y}{\lambda}, \lambda x, w)$. *That is*

$$f(w,x,y,z) = f\left(z, \frac{y}{\lambda}, \lambda x, w\right)$$

From this equality, we can make the following observations

1. x *must appear in the form* xy *to cancel* λ *in* $A.f$.
2. z *and* w *must appear in an symmetric way, i.e., must in the forms of* $z+w$ *or* zw.

So we conclude $f \in k[xy, wz, z+w]$.

Similarly consider matrix $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. *And after the same calculation we can get*

$$f(w,x,y,z) = f(w-x, w+x-y-z, y, y+z)$$

As we already have $f \in k[xy, wz, w+z]$, *we can reformulate this requirement into*

$$f(xy, wz, w+z) = f\left(wy + xy - y^2 - z, wy + wz - y^2 - yz, w+z\right)$$

We can see that this formular holds only when extra terms in $B.f$ *must cancel with each other, which implies* $f \in k[wz-xy, w+z]$, *as desired. So we have the construction*

$$\begin{aligned} X//G &= \operatorname{Spec} k[w,x,y,z]^G \\ &= \operatorname{Spec} k[wz-xy, w+z] \\ &= \operatorname{Spec} k[u,v] \\ &= \mathbb{C}^2 \end{aligned}$$

Remark 7.3.3. *There is a high-dimensional analogous: if* $\mathrm{GL}(n, \mathbb{C})$ *acts on* $M(n, \mathbb{C})$ *by conjugate action, then*

$$M(n, \mathbb{C}) // \mathrm{GL}(n, \mathbb{C}) = \mathbb{C}^n$$

See [7] *for more details.*

7.4. **The projective quotient.** Now we construct projective quotient by gluing together affine quotients.

Let X be a projective variety, then X can be covered by some affine varieties X_{f_i}. In order to construct GIT quotient of X by G, it's natural for us to take quotient for every affine variety of G of the form $X_{f_i}//G = \mathrm{Spec}(\mathscr{O}(X_{f_i})^G)$, and cover the projective quotient by them. To do this, we need an action of G on the coordinates of X.

Our approach is to embed X in $\mathbb{P}^m$ for some m such that the action of G can be extended to a linear action on $\mathbb{A}^{m+1}$. This is called a linearisation of the action of G.

Definition 7.4.1. *Let the group G act rationally on a projective variety X. Let $\varphi : X \hookrightarrow \mathbb{P}^m$ be an embedding of X that extends the group action, i.e., we have a rationally group action on $\mathbb{P}^m$ such that $\varphi(g.x) = g.\varphi(x)$. Let $\pi : \mathbb{A}^{m+1} \to \mathbb{P}^m$ be the natural projection. A linearisation of the action of G with respect to φ is a linear action of G on $\mathbb{A}^{m+1}$ that is compatible with the action of G on X in the following sense*

1. *For any $y \in \mathbb{A}^{m+1}, g \in G$*

$$\pi(g.y) = g.(\pi(y))$$

2. *For all $g \in G$, the map*

$$\mathbb{A}^{m+1} \to \mathbb{A}^{m+1}, \quad y \mapsto g.y$$

   is linear.

We write φ_G for a linearisation of the action of G with respect to φ.

Remark 7.4.2. *Note that such action induces an action of G on $\mathscr{O}(X)$. we have $\mathscr{O}(X) \cong k[x_0, \ldots, x_m]/I$ for some homogeneous ideal I, since X is isomorphic to the image $\varphi(X) \subseteq \mathbb{P}^m$. Using the fact that G acts on $k[x_0, \ldots, k_m]$ by $g.f(x_0, \ldots, x_m) := f(g^{-1}.(x_0, \ldots, x_m))$, we can know that G also acts on $\mathscr{O}(X)$, and it's well-defined, since $g.f' \in I$ for $f' \in I$.*

Example 7.4.3. *Let $\mathbb{C}^\times$ act on $\mathbb{P}^1$ by $\lambda.(x_0, x_1) = (x_0 : \lambda x_1)$. A linearisation can be given by the obvious action on $\mathbb{A}^2$ with $\lambda.(x_0, x_1) = (x_0, \lambda x_1)$.*

The above example illustrates a quite important issue when we are constructing projective quotient: good categorical quotient may not exist. The only possible G-invariant morphism sends all orbits to a point, since $(1, 0), (0, 1)$ are both in the closure of $(1, t)$. But this fails to seperate closed orbits, so is not a good categorical quotient.

The solution to such problem is to take an open G-invariant subset which has a good categorical quotient. We desire this subset to be covered by

G-invariant open affine subsets so that we can cover the quotient by gluing together affine quotients. This leads us to the notion of semistability,

Definition 7.4.4. *Let G be a reductive group acting on a projective variety X which has an embedding $\varphi : X \to \mathbb{P}^m$. A point $x \in X$ is called semistable (with respect to the linearisation φ_G) if there exists some G-invariant homogeneous polynomial f of degree greater than 0 in $\mathscr{O}(X)$, such that $f(x) \neq 0$ and X_f is affine.*

Remark 7.4.5. *Write $X^{as}(\varphi_G)$ for the set of semistable points of X with respect to φ_G, or just X^{as} when it's not ambiguous.*

For Example7.4.3, the set of semistable points of X with respect to φ_G is $X^{\mathrm{as}} = X_{x_0} = \mathbb{P}^1 \backslash \{(0:1)\}$. On this subset, the map to a point $p : X^{\mathrm{as}} \to \mathbb{P}^0$ is indeed a good categorical quotient.

Theorem 7.4.6. *Let G be a reductive group acting rationally on a projective variety X embedded in $\mathbb{P}^m$ with a linearisation φ_G. Let R be the coordinate ring of X, then there is a good categorical quotient*

$$p : X^{as}(\varphi_G) \to X^{as(\varphi_G)}//G \cong \operatorname{Proj} R^G$$

References

[1] J.Milnor, Morse theory, Princeton University Press, 1969
[2] Michael Hutchings, Lecture notes on Morse homology (with an eye towards Floer theory and pseudoholomorphic curve), 2002
[3] George Torres, Morse homology, 2016
[4] R.P.Thomas, Notes on GIT and symplectic reduction for bundles and varieties, 2006
[5] Matthew Brassil, Geometric invariant theory, 2012
[6] Harm Derksen and Gregor Kemper. Computational invariant theory. Invariant Theory and Algebraic Transformation Groups, I. Springer-Verlag, Berlin, 2002. Encyclopaedia of Mathematical Sciences, 130.
[7] Michel Brion. Introduction to actions of algebraic groups. Institut Fourier, 2009
[8] Victoria Hoskins. Moduli problems and Geometric Invariant Theory.

School of Mathematics, Shandong University, Jinan, 250100, P.R. China,
Email address: `bowenl@mail.sdu.edu.cn`