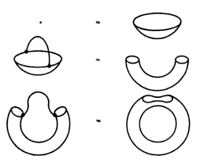
# From Morse theory to Bott periodicity

Aaron Mazel-Gee

In the original proof of complex Bott periodicity, Bott applied Morse theory to show that $\Omega^2 U \simeq U$ (where $U = \mathrm{colim}_n U(n)$ is the infinite unitary group). We survey the machinery and techniques on which Bott's proof relies. This will break into four sections: classical Morse theory, Riemannian geometry, the geometry of path spaces, and finally the proof itself. In the first two sections we will be very sketchy and will assume some familiarity with what is being covered, but in the second two sections we will be more in-depth and careful.

## 1 Morse theory

The idea of Morse theory is that for any smooth manifold $M$, studying the local behavior of a *Morse function* $f : M \to \mathbb{R}$ allows us to recover the homotopy type of $M$ in the form of a homotopy equivalent CW-complex. Each critical point of $f$ has an *index*, and each critical point of index $\lambda$ corresponds to a $\lambda$-cell.



The classic example is the height function on a torus standing on one end.[1] The index is easy to describe: at a critical point $p$, the index of $f$ is the dimension of the subspace of $T_pM$ of directions in which $f$ decreases. We move up the level sets of the function, attaching cells as we see critical points. So we begin with a 0-cell, because at its minimum $f$ increases in every direction. Then there is a saddle point, which has 1 downward dimension, so we attach a 1-cell. The next saddle point gives us another 1-cell, and the maximum gives us a 2-cell. Of course the attaching maps work out the way they should, and in the end this is just the usual CW-decomposition $e^0 \cup e^1 \cup e^1 \cup e^2$ of the torus.

### 1.1 Definitions

Let $M$ be a smooth manifold, and suppose $f : M \to \mathbb{R}$ is a smooth function. Then $p \in M$ is called a *critical point* of $f$ if $f_* : T_pM \to T_{f(p)}\mathbb{R}$ is not surjective. We call $f(p)$ a *critical value*.

A critical point $p$ is *non-degenerate* if the Hessian matrix of second partial derivatives $(\frac{\partial^2 f}{\partial x_i \partial x_j}(p))$ is nonsingular. This is independent of the coordinate system, and we give an alternative definition that will prove this and that will be important for our generalization of Morse theory to path spaces. We define a symmetric bilinear form (which we also call the Hessian) $f_{**} : T_pM \otimes T_pM \to \mathbb{R}$ as follows. Given $v, w \in T_pM$, we extend to vector fields $\tilde{v}, \tilde{w}$, and then we set $f_{**}(v, w) = \tilde{v}_p(\tilde{w}(f)) = v(\tilde{w}(f))$. This is symmetric because

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

since $p$ is a critical point of $f$, and then this is independent of the extensions because $\tilde{v}_p(\tilde{w}(f))$ is independent of the extension $\tilde{v}$ while $\tilde{w}_p(\tilde{v}(f))$ is independent of the extension $\tilde{w}$.

We now define the *index* $\lambda_f(p)$ of $f$ at $p$ to be the maximal dimension of subspace of $T_pM$ on which $f_{**}$ is negative-definite. The *nullity* is the dimension of the nullspace (i.e. $\{v \in T_pM : f_{**}(v, -) = 0\}$). So $p$ is non-degenerate if and only if $f_{**}$ has nullity 0. We call $f$ a *Morse function* if all its critical points are non-degenerate.

We now come to the fundamental result of Morse theory.

**Lemma** (Morse)**.** *If $p$ is a non-degenerate critical point of $f : M^n \to \mathbb{R}$ of index $\lambda$, then there exist coordinates around $p$ with $p = 0$ such that locally*

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

Thinking back to the torus, this means that (up to constants) the height function is locally given by $f = x_1^2 + x_2^2$ at the minimum, by $f = -x_1^2 + x_2^2$ at the saddle points, and by $f = -x_1^2 - x_2^2$ at the maximum. Note that this characterization implies immediately that non-degenerate critical points are isolated (from all other critical points).

[1] Image lifted shamelessly from Milnor's *Morse Theory*.

## 1.2 Theorems

We now classify $M$ based on the critical points of $f$. For convenience, we write $M^a = f^{-1}((-\infty, a])$.

**Theorem.** *Suppose $f^{-1}([a,b])$ is compact and contains no critical points. Then $M^a$ is diffeomorphic to $M^b$, and $M^a \subseteq M^b$ is a deformation retract. So the inclusion $M^a \hookrightarrow M^b$ is a homotopy equivalence.*

This is essentially because we can "push down" the points in $M^b \backslash M^a$ along trajectories that are orthogonal to the hypersurfaces $\{p \in M : f(p) = c\}$ for $c \in [a,b]$.[2]

**Theorem.** *Suppose $p$ is a non-degenerate critical point with index $\lambda$. Write $c = f(p)$. Suppose $f^{-1}([c-\varepsilon, c+\varepsilon])$ is compact and the only critical point it contains is $p$. Then for $\varepsilon$ sufficiently small, $M^{c+\varepsilon}$ deformation retracts onto $M^{c-\varepsilon} \cup e^\lambda$.*[3]

**Corollary.** *If $f : M \to \mathbb{R}$ is differentiable with no degenerate critical points and if each $M^a$ is compact, then $M$ has the homotopy type of a CW-complex with one $\lambda$-cell for each critical point of index $\lambda$.*[4]

We have the following convenient result on the existence of Morse functions.

**Theorem.** *If we embed $M \hookrightarrow \mathbb{R}^N$, then (the restriction to $M$ of) the square-distance function $L_p(x) = |x-p|^2$ is Morse for almost all $p \in \mathbb{R}^N$.*[5][6]

In fact, there is a precise sense in which Morse functions are generic among all smooth functions from $M$ to $\mathbb{R}$.

# 2 Riemannian geometry

Smooth manifolds are not very rigid, as they have no intrinsic notion of the lengths of curves or the angles between them. But if we embed our manifold into Euclidean space, we can recover such notions. On the other hand, it is powerful and enlightening to remove this assumption of embeddedness, and a Riemannian metric is exactly the intrinsically-defined object that preserves these notions. Here we give a rapid review of the Riemannian geometry that we will need.

## 2.1 Metrics

A *Riemannian metric* on a smooth manifold is a smooth choice of inner product $\langle\ ,\ \rangle : T_pM \otimes T_pM \to \mathbb{R}$ for each $p \in M$; smoothness means that if $X, Y \in \Gamma(TM)$, then $\langle X, Y\rangle \in C^\infty(M)$.[7] A manifold equipped with a Riemannian metric is called a *Riemannian manifold*; these are our primary objects of study.

Once we can talk about the magnitude of a tangent vector ($|X_p| = \langle X_p, X_p\rangle^{1/2}$), we can talk about the length of a curve. Let $c : [a,b] \to M$, and let $\frac{d}{dt}$ be the standard vector field on $\mathbb{R}$. We then define the *velocity vector field* of $c$ to be $\frac{dc}{dt} = dc(\frac{d}{dt})$. Note that $\frac{dc}{dt} \in \Gamma(c^*TM)$; this is what we call a *vector field along the curve $c$*. We now define the *length* of $c$ from $a$ to $b$ to be $\int_a^b |\frac{dc}{dt}| dt$.

---

[2] To say "orthogonal" we need to choose a Riemannian metric. Once we have this, we just flow at constant speed along $-\nabla f$. Note that we need compactness, otherwise for example we could be looking at a cylinder missing a single point.

[3] Smale worked hard to strengthen this to a "handle decomposition", where he attaches handles $D^\lambda \times D^{n-\lambda}$ along $S^{\lambda-1} \times D^{n-\lambda}$. This was crucial to his theory of differentiable manifolds, which allowed him to prove the Poincaré conjecture for dimensions 5 and above.

[4] It is immediate that if $M$ is compact and $f : M \to \mathbb{R}$ is smooth and has only two critical points that are both non-degenerate, then $M$ is homeomorphic to a sphere. (This is still true if the critical points are allowed to be degenerate, but it's harder to prove.) On the other hand, $M$ need not be diffeomorphic to the standard sphere. Indeed, in Milnor's classic paper he shows that his exotic 7-spheres are spheres using Morse theory.

[5] This proves that the vector field definition of the Euler characteristic $\chi(M)$ agrees with the definition as the signed sum of the number of cells. We know the sum of the indices of a transverse vector field is a topological invariant, and for $-\nabla f$ this agrees with the CW-decomposition coming from Morse theory.

[6] Note that if $M$ is closed (as a topological space) then $M^a$ will always be compact, by the Heine-Borel theorem. If $M$ is not, then the we can still achieve this desirable property of $f$ by demanding that the "ends" of $M$ are unboundedly embedded.

[7] This is a local notion, but since we have partitions of unity at our disposal we won't distinguish between local vector fields and global vector fields.

## 2.2 Connections

If we want to take the derivative of a vector field according to the usual definition of a derivative, we need to compare tangent vectors in different tangent spaces. With what we have so far there is no canonical way to do this, but a connection fixes this: it will give us a way of connecting the tangent spaces.

A *connection* on a smooth manifold $M$ is a function $\nabla : \Gamma(TM) \otimes \Gamma(TM) \to \Gamma(TM)$, written $\nabla_X Y = \nabla(X \otimes Y)$, with the properties that

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$,
2. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$,
3. $\nabla_X(fY) = X(f)Y + f\nabla_X Y$.

So the value of $\nabla_X Y$ at $p$ depends on $X$ only at $p$, and for $X_p \in T_pM$ we call $\nabla_{X_p} Y$ the *covariant derivative* of $Y$ at $p$ in the direction $X_p$.

Once we have a connection on $M$, for any vector field $V \in \Gamma(c^*TM)$ along a curve $c$ we can define its *covariant derivative* $\frac{DV}{dt}$, which is characterized by the properties

1. $\frac{D(V+W)}{dt} = \frac{DV}{dt} + \frac{DW}{dt}$,
2. $\frac{D(fV)}{dt} = \frac{df}{dt}V + f\frac{DV}{dt}$,
3. $\frac{D(c^*Y)}{dt} = \nabla_{\frac{dc}{dt}} Y$ for any $Y \in \Gamma(TM)$.

For example, $\frac{D}{dt}\frac{dc}{dt}$ is the *acceleration vector field* of the curve $c$.

The vector field $V \in \Gamma(c^*TM)$ is called *parallel* if $\frac{DV}{dt} = 0$, and $V_b \in T_{c(b)}M$ is called the *parallel transport* of $V_a \in T_{c(a)}M$ along $c$. Given a vector $V_a \in T_{c(a)}M$, there is a unique parallel vector field $V$ along $c$ beginning with $V_a$, as it is just the solution to a particular ODE. So the curve $c$ gives us an invertible linear map $T_{c(a)}M \to T_{c(b)}M$.[8]

A useful method for making calculations involving vector fields along a curve is to begin with an orthonormal basis $E_1, \ldots, E_n$ of $T_{c(a)}M$ and extend the vectors to parallel vector fields $E_1(t), \ldots, E_n(t)$. We can write any $V \in \Gamma(c^*TM)$ as $V = \sum_{i=1}^n f_i(t)E_i(t)$, and then we recover the simple formula $\frac{DV}{dt} = \sum_{i=1}^n \frac{df_i}{dt}E_i(t)$.

A connection on $M$ is called *compatible* with a given Riemannian metric if parallel transport preserves inner products.[9] A connection $\nabla$ is called *symmetric* if $\nabla_X Y - \nabla_Y X = [X, Y]$ for all $X, Y \in \Gamma(TM)$.[10] The Levi-Civita theorem tells us that every metric admits a unique compatible symmetric connection, called the *Levi-Civita connection*. From now on, we will automatically assume that any Riemannian manifold comes equipped with its Levi-Civita connection, which is characterized by the *Koszul formula*

$$2\langle \nabla_Y X, Z\rangle = X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle - \langle [X, Z], Y\rangle - \langle [Y, Z], X\rangle - \langle [X, Y], Z\rangle .$$

Since we will want to compare derivatives in different directions, it will be useful to introduce the notion of a *parametrized surface*, which is a function $s : U \to M$ for some open subset $U \subseteq \mathbb{R}^2$. Just as for curves, a *vector field along the surface* $s$ is a section $V \in \Gamma(s^*TM)$. For example, if $u$ and $v$ are the coordinates on $\mathbb{R}^2$ and $\frac{\partial}{\partial u}$ and $\frac{\partial}{\partial v}$ are the standard coordinate vector fields, then we have $\frac{\partial s}{\partial u} = ds(\frac{\partial}{\partial u})$ and $\frac{\partial s}{\partial v} = ds(\frac{\partial}{\partial v})$, the *velocity vector fields* of $s$. For any $V \in \Gamma(s^*TM)$ we define the *covariant derivative* $\frac{DV}{\partial u} \in \Gamma(s^*TM)$ at each point $(u_0, v_0) \in V$ by restricting $V$ to the curve $s(u, v_0)$ and taking the covariant derivative as before. We define $\frac{DV}{\partial v} \in \Gamma(s^*TM)$ similarly. For example, $\frac{D}{\partial u}\frac{\partial s}{\partial u}$ and $\frac{D}{\partial v}\frac{\partial s}{\partial v}$ are the accelerations of the appropriate coordinate curves.

[8] If $c(a) = c(b)$, this automorphism (which is generally nontrivial) is called the *holonomy* of $c$.

[9] Compatibility is convenient, for example, because $\nabla$ is compatible with the metric if and only if for any $V, W \in \Gamma(c^*TM)$,

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

which is also true if and only if for any $X, Y, Z \in \Gamma(TM)$,

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

So compatibility is sort of just a souped-up version of the Leibniz rule.

[10] This doesn't look so symmetric, but the name comes from the fact that in this case the *Christoffel symbols* $\Gamma^k_{ij}$, which bookkeeping devices we're totally sweeping under the rug, are symmetric in $i$ and $j$. Moreover, in a coordinate system $(x_1, \ldots, x_n)$, this implies that $\nabla_{\partial/\partial x_i}\partial/\partial x_j - \nabla_{\partial/\partial x_j}\partial/\partial x_i = [\partial/\partial x_i, \partial/\partial x_j] = 0$; that is, partial derivatives commute.

## 2.3 Curvature

Curvature measures the extent to which covariant derivative operators don't commute with each other. Explicitly, the *Riemann curvature tensor* $R : \Gamma(TM) \times \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ is given by

$$R(X,Y)Z = -\nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z + \nabla_{[X,Y]} Z.^{11}$$

Saying that this is a tensor means that $R$ is $C^\infty(M)$-trilinear, so that the value of $R(X,Y)Z$ at a point $p$ only depends on $X_p, Y_p, Z_p \in T_pM$.

This satisfies the properties

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

## 2.4 Geodesics

Geodesics are the "straight lines" in our general setting. Their most fundamental property is that they are locally length-minimizing (in a sense to be made precise in a moment), just like straight lines in Euclidean space (which are indeed its geodesics). But after that, the geometry of a manifold can force its geodesics to exhibit remarkably different behavior.

A path $c : [a,b] \to M$ is called a *geodesic* if its acceleration $\frac{D}{dt}\frac{dc}{dt}$ is identically zero. This is equivalent to saying that its velocity $\frac{dc}{dt}$ is parallel. In this case, $\frac{d}{dt}\left\langle \frac{dc}{dt}, \frac{dc}{dt}\right\rangle = 2\left\langle \frac{D}{dt}\frac{dc}{dt}, \frac{dc}{dt}\right\rangle = 0$, so the speed $|\frac{dc}{dt}|$ must be constant.

For all $p \in M$, every tangent vector $X_p \in T_pM$ defines a unique short-time geodesic, as this is just the solution to a particular ODE. This is written $\exp_p(tX_p)$, and for some open set $V \subseteq T_pM$ containing the origin we have a smooth map $\exp_p : V \to M$ which is a diffeomorphism onto its image. But geodesics may not be infinitely extendible; the easiest example is that by uniqueness, any geodesic on $M$ through $p$ will not be infinitely extendible on the manifold $M\backslash\{p\}$. If all geodesics are infinitely extendible we say $M$ is *geodesically complete*. The Hopf-Rinow theorem tells us that $M$ is geodesically complete if and only if it is metrically complete.[12]

We may now state precisely the sense in which geodesics are length-minimizing. For any $\tau \in (a,b)$, there is some $\varepsilon > 0$ such that for all $t \in (\tau - \varepsilon, \tau + \varepsilon)$, the portion of $c$ between $c(\tau)$ and $c(t)$ is a shortest path between those two points. In fact, we can say more. Every point $p \in M$ has a *normal neighborhood* $U \subseteq M$ such that any two points in $U$ have a unique shortest geodesic between them, and we can even demand that $U$ be *uniformly normal*, meaning that each such geodesic lies entirely in $U$.

However, geodesics need not be globally length-minimizing. For example, the geodesics on $S^n$ (with the standard Riemannian metric) are exactly the great circles; between two antipodal points there is an $S^{n-1}$ worth of geodesics, and when these are extended in either direction they cease to be length-minimizing.

## 2.5 Jacobi fields

Suppose $c : [0,a] \to M$ is a geodesic. We say that $J \in \Gamma(c^*TM)$ is a *Jacobi field* if it satisfies the equation $\frac{D^2 J}{dt^2} + R(\frac{dc}{dt}, J)\frac{dc}{dt} = 0$. As these are just the solutions to a particular ODE, they are uniquely determined by their initial conditions $J_0$ and $\frac{DJ}{dt}(0)$, and hence they form a vector space of dimension $2n$.

We say that $c(0)$ and $c(a)$ are *conjugate* along $c$ if there exists a nonzero Jacobi field vanishing at these endpoints. The *multiplicity* of the conjugacy is the dimension of the vector space of all such Jacobi fields.[13] The conjugate points to $p$ are exactly the critical values of $\exp_p : T_pM \to M$ (where a conjugacy at $\exp_p(v)$ is along the geodesic $\exp_p(tv)$).

The geometric significance of Jacobi fields is as follows. Suppose that $s : (-\varepsilon, \varepsilon) \times [0,1] \to M$ is a parametrized surface such that $s(u,t)$ is a geodesic for each $u \in (-\varepsilon, \varepsilon)$. Then the variation vector field $V_t = \frac{\partial s}{\partial u}(0,t)$ is a Jacobi

[11] This last term is necessary to make $R$ into a tensor, but if $X_i$ and $X_j$ are coordinate vector fields then $[X_i, X_j] = 0$ so it vanishes anyways.

[12] For $p, q \in M$, the distance is the infimum among the lengths of all paths connecting $p$ and $q$.

[13] Observe that $\frac{dc}{dt}$ and $t\frac{dc}{dt}$ are both nonzero Jacobi fields along $c$. The latter vanishes at $t = 0$ but not for any $t > 0$, so the multiplicity of a conjugacy can be at most $n - 1$.

field along $s(0,t)$. Conversely, every Jacobi field along a geodesic $c:[0,1]\to M$ may be obtained by such a variation of $c$ through geodesics.[14]

We will use the fact that in a uniformly normal neighborhood, a Jacobi field is uniquely determined by its values at the endpoints of the geodesic. (In fact, this holds as long as $c(0)$ and $c(a)$ are not conjugate along $c$.)

# 3 Path spaces

In order to prove that $\Omega^2 U\simeq U$, we'll first need to develop an understanding of the path space of a Riemannian manifold $M$. This will mimic the yoga of classical Morse theory, relying heavily on the Riemannian structure of $M$. We note here that the space of paths $\Omega^*(M;p,q)$ in $M$ from $p$ to $q$ is homotopy equivalent to the space of loops $\Omega^*(M;p,p)$ (which is homotopy equivalent to any $\Omega^*(M;r,r)$ by conjugation by any path between $r$ and $p$) since these are both fibers of the *path fibration* $f:\mathcal{P}M\twoheadrightarrow M$, where $\mathcal{P}M=\{c:[0,1]\to M \text{ s.t. } c(0)=p\}$ and $f(c)=c(1)$ is the endpoint-evaluation map. We will care about the actual loop space for our homotopy theoretic final goal, but the geometry will be much nicer if we choose $p$ and $q$ strategically and try to understand $\Omega^*(M;p,q)$ instead.

## 3.1 Definition of the path space

When we write $\Omega(M;p,q)$ (or often just $\Omega$), we will mean the space of piecewise smooth paths $\omega:[0,1]\to M$ from $p$ to $q$. (This choice will be justified later.) This is roughly an infinite-dimensional manifold, and more importantly it is a suitable setting for applying Morse theory. We define the *tangent space* $T_\omega\Omega$ at a path $\omega\in\Omega(M;p,q)$ to be the vector space of piecewise smooth vector fields along $\omega$ which vanish at the endpoints. (The points where the vector field is not smooth need not be points where $\omega$ itself is not smooth.)

## 3.2 The energy functional

The analogue of a Morse function that we will use is the *energy functional* $E:\Omega\to\mathbb{R}$. For $\omega\in\Omega$ we define the *energy* from $a$ to $b$ of $\omega$ by $E_a^b(\omega)=\int_a^b|\frac{d\omega}{dt}|^2dt$. The analysis works better with the energy functional than with the length functional, but it is not hard to see that on a complete manifold $M$ with $p,q\in M$ distance $d$ apart, $E:\Omega\to\mathbb{R}$ takes its minimum value $d^2$ precisely on the set of minimal geodesics. (In fact, we will see in a moment that something even better is true.)

We define the induced map $E_*:T_\omega\Omega\to T_{F(\omega)}\mathbb{R}$ as follows. Given $W\in T_\omega\Omega$, we choose a *variation* $\overline{\alpha}:(-\varepsilon,\varepsilon)\to\Omega$ of $\omega$ integrating $W$ (i.e. a parametrized surface $\alpha:(-\varepsilon,\varepsilon)\times[0,1]\to M$, thought of a family of paths $\alpha_u(t)=\alpha(u,t)$, for which $\alpha(u,0)=\omega(0)$, $\alpha(u,1)=\omega(1)$, $\alpha(0,t)=\omega(t)$, and $\frac{\partial\alpha}{\partial u}(0,t)=W(t)$, such that $\alpha$ is differentiable on each strip $(-\varepsilon,\varepsilon)\times[t_i,t_{i+1}]$ for some partition $0=t_0\leq\ldots\leq t_k=1$). Then, we set $E_*(W)(t)=\frac{dE(\overline{\alpha}(u))}{du}|_{u=0}$. This is not well-defined for an arbitrary function $F:\Omega\to\mathbb{R}$, but the following lemma will imply that it is well-defined here. A path $\omega\in\Omega$ is called a *critical path* if $E_*(W)=0$ for every $W\in T_\omega\Omega$.

To fix notation, we will henceforth write $\overline{\alpha}:(-\varepsilon,\varepsilon)\to\Omega$ for a variation of $\omega$ integrating $W$, $V_t=\frac{d\omega}{dt}$ (the velocity vector field of $\omega$), $A_t=\frac{D}{dt}\frac{d\omega}{dt}$ (the acceleration vector field of $\omega$), and $\Delta_tV=V_{t+}-V_{t-}$ (the discontinuity of $V$ at $t$).

**Lemma.** *The derivative of $E$ is given by the* first variation formula

$$E_*(W)=\frac{dE(\overline{\alpha}(u))}{du}(0)=-2\left(\sum_{t\in[0,1]}\langle W_t,\Delta_tV\rangle+\int_0^1\langle W_t,A_t\rangle\,dt\right).$$

*Proof.* We calculate

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

where the last equality comes from the fact that our connection is symmetric and so partial derivatives commute. On each interval $[t_{i-1},t_i]$ such that $\alpha|_{(-\varepsilon,\epsilon)\times[t_{i-1},t_i]}$ is differentiable, we integrate by parts using the identity

[14] Note, however, that just because $c(0)$ and $c(1)$ are conjugate along $c$ does not mean that we can vary $c$ through geodesics that fix the endpoints. We may only conclude that in such an integral variation of geodesics, the endpoints do not vary to first order. Of course, the function $f(t)=t^2$ illustrates that just because something vanishes to first order doesn't mean it stays fixed in any interval.

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

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

At $u=0$ this becomes $\langle W_t, V_t\rangle\,|_{t=t_{i-1}+}^{t=t_i-} - \int_{t_{i-1}}^{t_i}\langle W_t, A_t\rangle\, dt$, and summing over all intervals gives the result. ☐

Intuitively, the idea here is that the first term shows that varying a path inwards along a "kink" tends to decrease its energy, as does varying in the direction of $A_t$. (Indeed, the terms $\Delta_t V$ represent a discrete version of acceleration.)

**Corollary.** *A path $\omega \in \Omega$ is critical for $E$ if and only if $\omega$ is a geodesic, i.e. $A_t = 0$ and $\Delta_t V = 0$ for all $t$.*

We define the *Hessian* $E_{**} : T_\omega\Omega \otimes T_\omega\Omega \to \mathbb{R}$ at a critical path $\omega \in \Omega$ (i.e. a geodesic) by analogy with the secondary definition from before. Given $W, W' \in T_\omega\Omega$ we choose a 2-parameter variation $\overline{\alpha} : U \to \Omega$ for some neighborhood $U \subseteq \mathbb{R}^2$ of the origin which integrates $W$ and $W'$ (i.e. a 2-parameter variation $\alpha : U \times [0,1] \to M$ with $\alpha(u_1,u_2,0) = \omega(0)$, $\alpha(u_1,u_2,1) = \omega(1)$, $\alpha(0,0,t) = \omega(t)$, $\frac{\partial\alpha}{\partial u_1}(0,0,t) = W_t$, and $\frac{\partial\alpha}{\partial u_2}(0,0,t) = W'_t$) and then set

$$E_{**}(W,W') = \frac{\partial^2 E(\overline{\alpha}(u_1,u_2))}{\partial u_1 \partial u_2}\bigg|_{(u_1,u_2)=(0,0)}.$$

The following lemma implies that this is well-defined.

**Lemma.** *At a critical path of $E$, we have the* second variation formula

$$\frac{\partial^2 E}{\partial u_1\partial u_2}(0,0) = -2\left(\sum_{t\in[0,1]}\left\langle W'_t, \Delta_t \frac{DW}{dt}\right\rangle + \int_0^1 \left\langle W'_t, \frac{D^2W}{dt^2} + R(V_t,W_t)V_t\right\rangle dt\right).$$

*Proof.* From the first variation formula,

$$-\frac{1}{2}\frac{\partial E}{\partial u_2} = \sum_{t\in[0,1]}\left\langle \frac{\partial\alpha}{\partial u_2}, \Delta_t\frac{\partial\alpha}{\partial t}\right\rangle + \int_0^1 \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle dt.$$

Therefore,

$$-\frac{1}{2}\frac{\partial^2 E}{\partial u_1\partial u_2} = \sum_{t\in[0,1]}\left(\left\langle \frac{D}{\partial u_1}\frac{\partial\alpha}{\partial u_2}, \Delta_t\frac{\partial\alpha}{\partial t}\right\rangle + \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{\partial u_1}\Delta_t\frac{\partial\alpha}{\partial t}\right\rangle\right) + \int_0^1 \left(\left\langle \frac{D}{\partial u_1}\frac{\partial\alpha}{\partial u_2}, \frac{D}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle + \left\langle \frac{\partial\alpha}{\partial u_2}, \frac{D}{\partial u_1}\frac{D}{\partial t}\frac{\partial\alpha}{\partial t}\right\rangle\right) dt.$$

At $(u_1,u_2) = (0,0)$, both $\Delta_t\frac{\partial\alpha}{\partial t}$ and $\frac{D}{\partial t}\frac{\partial\alpha}{\partial t}$ vanish since $\omega = \overline{\alpha}(0,0)$ is an unbroken geodesic, so the first term of the sum and the first term of the integral both vanish. Thus the above equation simplifies to

$$-\frac{1}{2}\frac{\partial^2 E}{\partial u_1\partial u_2}(0,0) = \sum_{t\in[0,1]}\left\langle W', \Delta_t\frac{D}{dt}W\right\rangle + \int_0^1 \left\langle W', \frac{D}{\partial u_1}\frac{D}{\partial t}V\right\rangle dt.$$

We can interchange the operators $\frac{D}{\partial u_1}$ and $\frac{D}{\partial t}$ by using the curvature formula $\frac{D}{\partial u_1}\frac{D}{\partial t}V - \frac{D}{\partial t}\frac{D}{\partial u_1}V = R(\frac{\partial\alpha}{\partial t}, \frac{\partial\alpha}{\partial u_1})V = R(V,W)V$. Combined with the fact that $\frac{D}{\partial u_1}V = \frac{D}{\partial t}\frac{\partial\alpha}{\partial u_1} = \frac{D}{\partial t}W$, this gives us that $\frac{D}{\partial u_1}\frac{D}{\partial t}V = \frac{D^2W}{dt^2} + R(V,W)V$, which proves the result. ☐

It follows immediately that the Hessian is bilinear, and it is also symmetric since $\frac{\partial^2 E}{\partial u_1\partial u_2} = \frac{\partial^2 E}{\partial u_2\partial u_1}$.

**Corollary.** *For a geodesic $\omega$, a vector field $W \in T_\omega\Omega$ is in the null space of $E_{**}$ if and only if $W$ is a Jacobi field. So $E_{**}$ is degenerate at $\omega$ if and only if its endpoints are conjugate along it. The nullity is equal to the multiplicity of the conjugacy.*

Since the space of Jacobi fields along a geodesic is finite-dimensional, it follows that the nullity is always finite.

To calculate the index of $E_{**}$, which as before is defined to be the maximum dimensional subspace of $T_\omega\Omega$ on which $E_{**}$ is negative-definite, we have the Morse index theorem.

**Theorem** (Morse)**.** *The index $\lambda$ of $E_{**}$ at a geodesic $\omega : [0,1] \to M$ is the number of points $\omega(t)$ with $t \in (0,1)$ such that $\omega(t)$ is conjugate to $\omega(0)$ along $\omega$, counted with multiplicity. This index is always finite.*[15]

*Proof sketch.* The proof uses broken Jacobi fields whose smooth components are contained in normal neighborhoods (where they are uniquely determined by their endpoints). We choose a subdivision $0 = t_0 < \ldots < t_k = 1$ of $[0,1]$ such that $\omega([t_{i-1}, t_i])$ lies in a normal neighborhood for all $i$, and we let $T_\omega\Omega(\vec{t}) \subseteq T_\omega\Omega$ be the subspace of all $W$ such that $W|_{[t_{i-1},t_i]}$ is a Jacobi field for all $i$. Then, we let $T' \subseteq T_\omega\Omega$ be the subspace of all $W$ such that $W_{t_i} = 0$ for all $i$.

One checks that $T_\omega\Omega \cong T_\omega\Omega(\vec{t}) \oplus T'$, that this is an orthogonal decomposition with respect to $E_{**}$, and that $E_{**}$ is positive definite on $T'$. This implies that the index and nullity of $E_{**}$ can both be computed by restricting to $T_\omega\Omega(\vec{t})$, and in particular that the index is finite.

The result then follows from four assertions.

1. The index $\lambda(\tau)$ of $\omega|_{[0,\tau]}$ is increasing in $\tau$.[16]
2. $\lambda(\tau) = 0$ for small values of $\tau$.[17]
3. For any $\tau$, there is some $\varepsilon > 0$ such that $\lambda(\tau - \varepsilon) = \lambda(\tau)$.[18]
4. If $\nu$ is the nullity of the Hessian $(E_0^\tau)_{**}$ (i.e. $\omega(0)$ and $\omega(\tau)$ are conjugate along $\omega$ with multiplicity $\nu$), then for all sufficiently small $\varepsilon > 0$, $\lambda(\tau + \varepsilon) = \lambda(\tau) + \nu$.[19]

□

### 3.3 Recovering the homotopy type of $\Omega$

We now endow $\Omega(M; p, q)$ with a metric. Let $\rho$ be the metric on $M$ induced by its Riemannian metric. Given $\omega, \omega' \in \Omega$, we define the distance to be

$$d(\omega, \omega') = \max_{t\in[0,1]} \rho(\omega(t), \omega'(t)) + \left( \int_0^1 \left( \left| \frac{d\omega}{dt} \right| - \left| \frac{d\omega'}{dt} \right| \right)^2 dt \right)^2 .$$

(The second term is there so that the energy functional $E : \Omega \to \mathbb{R}$ is continuous.)

For $c > 0$, we will study $\Omega^c = E^{-1}([0, c]) \subseteq \Omega$. We can approximate the interior of this space by a subset that carries a natural manifold structure. For a sufficiently fine subdivision $0 = t_0 < \ldots < t_k = 1$ of $[0,1]$, we let $B$ be the subset of $\mathrm{Int}(\Omega^c)$ consisting of paths $\omega$ such that $\omega|_{[t_i,t_{i+1}]}$ is a geodesic for all $i$. We require our subdivision to be fine enough that each geodesic segment $\omega|_{[t_i,t_{i+1}]}$ is uniquely determined by the points $\omega(t_i)$ and $\omega(t_{i+1})$. These points can vary within small open sets of $M$, and so $B$ is naturally an $n(k-1)$-dimensional manifold. We write $E' = E|_B$.

**Theorem.** *The function $E' : B \to \mathbb{R}$ is smooth, and for each $a < c$ the set $B^a = E'^{-1}([0, a])$ is compact and is a deformation retract of $\Omega^a$. The critical points of $E'$ are exactly those of $E$ in $\mathrm{Int}(\Omega^c)$, namely (unbroken) geodesics from $p$ to $q$ of length less than $\sqrt{c}$. The index/nullity of $E'_{**}$ at a critical point $\omega$ is exactly the index/nullity of $E_{**}$.*

This is proved by deforming all paths to broken geodesics in a continuous way that fixes the images of the $t_i$, giving a deformation retraction $\mathrm{Int}(\Omega^a) \to B^a$. (Remember that minimal geodesics minimize energy.)

**Corollary.** *Let $M$ be a complete Riemannian manifold and let $p, q \in M$ be two points not conjugate along any geodesic of length less than or equal to $\sqrt{a}$.*[20] *Then $\Omega^a$ has the homotopy type of a finite CW-complex with one $\lambda$-cell for each geodesic $\omega \in \Omega^a$ at which $E_{**}$ has index $\lambda$.*[21]

---

[15] Thus a geodesic segment $\omega([0,1])$ can only contain finitely many points which are conjugate along $\omega$ to $\omega(0)$.

[16] This comes from the fact that if we have a subspace of $T_{\omega|_{[0,\tau]}}\Omega$ on which the restriction $(E_0^\tau)_{**}$ is negative definite, we can extend its vector fields by zero along the rest of $\omega$. Then for $\tau' > \tau$, $(E_0^{\tau'})_{**}$ will be negative definite on this extension.

[17] This comes from the fact that for sufficiently small $\tau$, $\omega|_{[0,\tau]}$ is a minimal geodesic.

[18] This comes from rewriting the space of broken Jacobi fields on $\omega|_{[0,\tau]}$ that vanish at the endpoints as $\Sigma = T_{\omega(t_1)}M \oplus \ldots \oplus T_{\omega(t_i)}M$, where are assuming without loss of generality that $\tau \in (t_i, t_{i+1})$. The associated quadratic form varies continuously with $\tau \in (t_i, t_{i+1})$, so for $\tau'$ sufficiently near $\tau$ we have that $\lambda(\tau') \geq \lambda(\tau)$. But for $\tau' < \tau$, we also have that $\lambda(\tau') \leq \lambda(\tau)$.

[19] This involves similar techniques to what we have used before, but it is what makes the proof too involved to give full details.

[20] For any $p \in M$ there exists such a $q \in M$ even without a length bound on the geodesics, basically by an application of Sard's theorem to $\exp : T_pM \to M$.

[21] In particular, $\Omega^a$ contains only finitely many geodesics.

### 3.4 The full path space

So far we have been talking about $\Omega(M;p,q)$, the space of piecewise smooth paths in $M$ from $p$ to $q$, but what we actually care about is $\Omega^*(M;p,q)$, the space of all paths in $M$ from $p$ to $q$. This is endowed with the compact open topology, which is induced by the metric $d^*(\omega,\omega') = \max_{t\in[0,1]} \rho(\omega(t),\omega'(t))$.) Luckily, we have the following result.

**Theorem.** *The inclusion $\Omega \hookrightarrow \Omega^*$ is a homotopy equivalence.*

*Proof sketch.* First, note that this map is continuous since $d \geq d^*$. We choose an atlas $\{N_\alpha\}$ for $M$ of uniformly normal neighborhoods, and we define filtrations

$$\begin{array}{ccccccccc}
\Omega_1^* & \hookrightarrow & \Omega_2^* & \hookrightarrow & \Omega_3^* & \hookrightarrow & \cdots & \xrightarrow{\text{colim}} & \Omega^* \\
\uparrow & & \uparrow & & \uparrow & & & & \uparrow \\
\Omega_1 & \hookrightarrow & \Omega_2 & \hookrightarrow & \Omega_3 & \hookrightarrow & \cdots & \xrightarrow{\text{colim}} & \Omega
\end{array}$$

by setting $\Omega_k^* = \left\{\omega \in \Omega^* : \text{ for all } j,\ \omega\left(\left[\frac{j-1}{2^k},\frac{j}{2^k}\right]\right) \subset N_\alpha \text{ for some } \alpha\right\}$ and $\Omega_k = \Omega \cap \Omega_k^*$. If for each $\omega \in \Omega_k^*$ we hold the points $\omega(\frac{j}{2^k})$ fixed and straighten to a piecewise geodesic between these points, this gives a deformation retraction of $\Omega_k^*$ onto $\Omega_k$.

□

So we can use $\Omega$ for our homotopy theoretic purposes even when what we really mean is $\Omega^*$. Henceforth, we will ignore the distinction.

## 4 Bott periodicity

We now embark on the proof of Bott periodicity, picking up the last few facts we'll need along the way.

### 4.1 Symmetric spaces

A *symmetric space* is a a connected Riemannian manifold $M$ such that for each $p \in M$ we have an isometry $I_p : M \to M$ fixing $p$ with $dI_p : T_pM \to T_pM$ equal to $-I$. (So $I_p$ reverses geodesics through $p$.) Any such $M$ is necessarily complete since using the involutions we can infinitely extend geodesics, and the choice of $I_p$ is unique since by completeness any point of $M$ is joined to $p$ by a geodesic.

**Lemma.** *Let $c : [0,1] \to M$ be a geodesic in a symmetric space. If $X, Y, Z \in \Gamma(c^*TM)$ are parallel vector fields, then $R(X,Y)Z \in \Gamma(c^*TM)$ is also a parallel vector field.*

*Proof.* Suppose $W \in \Gamma(c^*TM)$ is also parallel. Write $p = c(0)$ and $q = c(1)$, and consider the isometry $T = I_{c(1/2)}I_{c(0)}$ which takes $c(0)$ to $c(1)$. It is easy to see that $T$ preserves parallel vector fields along $c$, and so $\langle R(X_q,Y_q)Z_q, W_q\rangle = \langle R(T_*(X_p),T_*(Y_p))T_*(Z_p),T_*(W_p)\rangle = \langle R(X_p,Y_p)Z_p,W_p\rangle$. Thus $\langle R(X,Y)Z,W\rangle$ is constant for every parallel vector field $W$, and hence $R(X,Y)Z$ is parallel. □

In a symmetric space the Jacobi equation has simple explicit solutions, and this gives us the following result.

**Theorem.** *Suppose $c : \mathbb{R} \to M$ is a geodesic in a symmetric space. Write $p = c(0)$ and $V = \frac{dc}{dt}(0)$, and define a linear transformation $K_V : T_pM \to T_pM$ by $K_V(W) = R(V,W)V$. Let $e_1,\ldots,e_n$ denote the eigenvalues of $K_V$. Then the conjugate points to $p$ along $c$ are the points $c(\frac{\pi k}{\sqrt{e_i}})$ for any $k \in \mathbb{Z}\backslash\{0\}$ and any positive eigenvalue $e_i$. The multiplicity of the conjugacy is equal to the number of eigenvalues $e_j$ such that $\frac{\pi k}{\sqrt{e_i}}$ is a multiple of $\frac{\pi}{\sqrt{e_j}}$.*

*Proof.* Note that $K_V$ is self-adjoint; this means that $\langle K_V(W),W'\rangle = \langle W, K_V(W')\rangle$, which is just one of the symmetries of the curvature tensor. Hence we can choose an orthonormal basis of eigenvectors $E_1,\ldots E_n$ of $T_pM$ with $K_V(E_i) = e_iE_i$. Extend $V$ and the $E_i$ to parallel vector fields along $c$. Since $M$ is symmetric, then $K_V(E_i) = R(V,E_i)V = e_iE_i$ holds all along $c$. Now any $W \in \Gamma(c^*TM)$ is a linear combination $W(t) = \sum_{i=1}^n w_i(t)E_i(t)$, and the Jacobi equation $\frac{D^2W}{dt^2} + K_V(W) = 0$ becomes

$$\sum_{i=1}^n \frac{d^2w_i}{dt^2}E_i + \sum_{i=1}^n e_iw_iE_i = 0.$$

But the $E_i$ are everywhere linearly independent, so this reduces to a system of $n$ differential equations $\frac{d^2 w_i}{dt^2} + e_i w_i = 0$. Since we want solutions vanishing at $t = 0$, we obtain a basis of solutions given by

$$w_i(t) = \begin{cases} \sin(\sqrt{e_i}t), & e_i > 0 \\ t, & e_i = 0 \\ \sinh(\sqrt{|e_i|}t), & e_i < 0. \end{cases}$$

So to get a conjugacy we can only have (scalar) linear combinations of solutions for $e_i > 0$ (since $\sinh(t) > 0$ for $t > 0$). The theorem follows easily. ☐

## 4.2 Compact Lie groups

Let $G$ be a compact Lie group. Given any arbitrary inner product on $\mathfrak{g}$, we can average over the group to get a bi-invariant Riemannian metric on $G$.[22] This makes $G$ into a symmetric space with involution associated to any $g_0 \in G$ given by $I_{g_0}(g) = g_0 g^{-1} g_0$. Note that $I_e$ reverses geodesics through the identity element $e \in G$.[23]

The following theorem identifies a crucial relationship between the curvature of $G$ and the Lie bracket on $\mathfrak{g}$.

**Theorem.** *If $G$ is a Lie group with a bi-invariant Riemannian metric and $X, Y, Z \in \mathfrak{g}$ (considered as the left-invariant vector fields on $G$), then $R(X,Y)Z = \frac{1}{4}[[X,Y],Z]$.*

*Proof sketch.* We will take as given the fact that $\langle [X,Y],Z\rangle = \langle X,[Y,Z]\rangle$.[24] We begin with the Koszul formula

$$2\langle \nabla_Y X, Z\rangle = X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle - \langle [X,Z],Y\rangle - \langle [Y,Z],X\rangle - \langle [X,Y],Z\rangle.$$

If we set $X = Y$, then since the inner product of any two left-invariant vector fields is constant we obtain $2\langle \nabla_X X, Z\rangle = -2\langle [X,Z],X\rangle$, which implies that $\langle \nabla_X X, Z\rangle = \langle [Z,X],X\rangle = \langle Z,[X,X]\rangle = 0$. Since this is true for any $Z$, then $\nabla_X X = 0$ for any $X \in \mathfrak{g}$. Thus $0 = \nabla_{X+Y}(X+Y) = \nabla_X Y + \nabla_Y X = \nabla_X Y + (\nabla_X Y - [X,Y])$ by the symmetry of the connection, and so $\nabla_X Y = \frac{1}{2}[X,Y]$. We now calculate that

$$\begin{aligned} R(X,Y)Z &= \nabla_Y\nabla_X Z - \nabla_X\nabla_Y Z + \nabla[X,Y]Z \\ &= \frac{1}{4}[Y,[X,Z]] - \frac{1}{4}[X,[Y,Z]] + \frac{1}{2}[[X,Y],Z] \\ &= \frac{1}{4}\left([[Z,X],Y] + [[Y,Z],X]\right) + \frac{1}{2}[[X,Y],Z] \\ &= \frac{1}{4}(-[[X,Y],Z]) + \frac{1}{2}[[X,Y],Z] \\ &= \frac{1}{4}[[X,Y],Z] \end{aligned}$$

by the Jacobi identity. ☐

## 4.3 Manifolds of minimal geodesics

Consider $\Omega(S^n; p, -p)$. These antipodal points are not in "general position" (i.e. they are conjugate), but Bott observed that this can actually be helpful. Indeed, the space of minimal geodesics in $\Omega(S^n; p, -p)$ can be identified with the equatorial $S^{n-1}$. This will give us another way of approximating $\Omega S^n$.

**Theorem.** *Suppose $M$ is a complete Riemannian manifold, and let $p, q \in M$ with distance $\rho(p,q) = \sqrt{d}$. If the space $\Omega^d$ of minimal geodesics is a (topological) manifold and if every non-minimal geodesic in $\Omega$ has index greater than $k$, then $\Omega^d$ can be taken to be the $k$-skeleton of $\Omega$.*[25]

We can also think of this in the following way. Suppose $f : M \to \mathbb{R}$ is smooth, and suppose $V \subseteq M$ is a smooth, connected, compact submanifold. We say $V$ is a *non-degenerate critical submanifold* of $f$ if $V \subseteq \mathrm{Crit}(f)$ and for

[22] That every compact Lie group admits a bi-invariant metric is crucial for most of their theory. Non-compact Lie groups do not in general admit bi-invariant metrics.

[23] For any bi-invariant metric, the geodesics in $G$ through $e$, which are given by $\exp(tx)$ for $x \in \mathfrak{g}$, are precisely the one-parameter subgroups of $G$.

[24] This follows from an interpretation of the vector field bracket as a special case of the Lie bracket.

[25] This implies the *Freudenthal suspension theorem*, that $\pi_i(S^n) \cong \pi_{i+1}(S^{n+1})$ for $i \leq 2n-2$. Indeed, it is not hard to see that the indices of geodesics in $\Omega(S^{n+1}; p, -p)$ are 0, $2n$, $4n$, etc., so $S^n \cong \Omega^\pi S^{n+1} \to \Omega S^{n+1}$ induces isomorphisms on $\pi_{\leq 2n-2}$.

all $p \in V$, $\text{Null}(f_{**}(p) : T_pM \to T_pM) = T_pV$. The *index* $\lambda_f(V)$ of $f$ on $V$ is by definition the index of any $p \in V$. So for example, a torus lying on a plane has two non-degenerate critical submanifolds, a circle of index 0 at the bottom and a circle of index 1 at the top. We say $f$ is *Morse-Bott* if $\text{Crit}(f)$ is a union of non-degenerate critical submanifolds.

**Theorem.** *If $f$ is Morse-Bott and all the critical points in $M^b \backslash M^a$ together form a non-degenerate critical submanifold $V$, then $M^b \simeq M^a \cup \nu^-$, where $\nu^-$ is the negative normal bundle of $V$. If we perturb $f$ to be Morse, then $M^b \simeq M^a \cup e_1 \cup \ldots \cup e_s$, where $\dim(e_i) \geq \lambda_f(V)$.*

So manifolds of minimal geodesics in Morse theory on path spaces play exactly the role of these non-degenerate critical submanifolds in classical Morse theory.

## 4.4 The proof

Recall that the $n^{th}$ unitary group is $U(n) = \{A \in M_n(\mathbb{C}) : AA^* = I\}$, with Lie algebra $\mathfrak{u}(n) = T_IU(n) = \{x \in M_n(\mathbb{C}) : x + x^* = 0\}$. We endow $U(n)$ with a Riemannian metric by first defining the inner product $\langle x, y\rangle = \text{tr}(xy^*) = \sum_{i,j} x_{ij}\overline{y}_{ij}$ on $\mathfrak{u}(n)$. This extends to a unique left-invariant metric on $U(n)$. This is also right-invariant, since it is Ad-invariant (for $\text{Ad} : U(n) \to GL(\mathfrak{u}(n))$ the adjoint action). So this metric is bi-invariant.

To understand $\Omega(U(n))$, we will look for minimal geodesics in $U(n)$ from $I$ to $-I$, and we will see that all non-minimal geodesics have rather high index. So the space of minimal geodesics in $\Omega(U(n); I, -I)$ will be a good model for all of $\Omega(U(n); I, -I)$.

To begin our analysis, recall that the exponential map from Riemannian geometry coincides with the exponential map from Lie theory, so this means that we're just looking for $x \in \mathfrak{u}(n)$ with $\exp(x) = -I$. Recall too that all matrices in $U(n)$ and $\mathfrak{u}(n)$ can be conjugated by elements of $U(n)$ to diagonal matrices, and that conjugation commutes with the matrix exponential. Since diagonal matrices in $U(n)$ are exactly those with unimodular diagonal entries while diagonal matrices in $\mathfrak{u}(n)$ are exactly those with purely-imaginary diagonal entries, we see that $\exp : \mathfrak{u}(n) \to U(n)$ is surjective, and

$$\exp^{-1}(-I) = \text{Ad}(U(n))(\{diag(i\pi k_1, \ldots, i\pi k_n) : k_i \in 2\mathbb{Z} + 1\}).$$

For $x = diag(i\pi k_1, \ldots, i\pi k_n)$, the length of of $\exp(tx)$ (for $x \in [0, 1]$) is just $|x| = \sqrt{\text{tr}(xx^*)} = \pi\sqrt{k_1^2 + \ldots + k_n^2}$ since geodesics have constant speed. So $\exp(tx)$ is a minimal geodesic iff $k_i = \pm 1$ for all $i$, in which case its length is $\pi\sqrt{n}$. We therefore write $\Omega^{\pi\sqrt{n}}(U(n); I, -I)$ for the space of minimal geodesics in $\Omega(U(n); I, -I)$.

Observe that if $\exp(tx)$ is a minimal geodesic, then $x$ is completely determined once we specify its eigenspaces $\text{Eigen}(i\pi)$ and $\text{Eigen}(-i\pi)$. But $\mathbb{C}^n$ decomposes as an orthogonal direct sum of these eigenspaces (since in the diagonalization this is true, and conjugating by elements of $U(n)$ preserves this) so in fact it suffices to specify $\text{Eigen}(i\pi)$. Of course this can be any subspace, so we get a homeomorphism $\Omega^{\pi\sqrt{n}}(U(n); I, -I) \cong \coprod_{k=0}^n Gr_k(\mathbb{C}^n)$, where $Gr_k(\mathbb{C}^n)$ denotes the Grassmannian manifold of $k$-planes in $\mathbb{C}^n$.

This is a good start, but it is inconvenient to work with a manifold with components of different dimensions. However, we can use $SU(2m) = \{A \in U(2m) : \det(A) = 1\}$ in place of $U(n)$. Since $\mathfrak{su}(2m) = T_ISU(2m) = \{A \in \mathfrak{u}(2m) : \text{tr}(A) = 0\}$, if $\exp(tx)$ is a minimal geodesic from $I$ to $-I$ in $SU(2m)$ then we must additionally have that $\text{tr}(x) = \sum_{i=1}^{2m} k_i = 0$. Thus it must be that $\dim \text{Eigen}(i\pi) = \dim \text{Eigen}(-i\pi) = m$, so we get a homeomorphism $\Omega^{\pi\sqrt{n}}(SU(2m); I, -I) \cong Gr_m(\mathbb{C}^{2m})$, which is much more tractable.

The following technical lemma constitutes the heart of the proof.

**Lemma.** *Every non-minimal geodesic from $I$ to $-I$ in $SU(2m)$ has index greater than $2m + 1$.*

*Proof.* Suppose $\exp(tx)$ is a non-minimal geodesic in $SU(2m)$ from $I$ to $-I$ for some $x \in \mathfrak{su}(2m)$. To find its index, we must find the conjugate points to $I$ along it. These are determined by the positive eigenvalues of $K_x : \mathfrak{su}(2m) \to \mathfrak{su}(2m)$, where $K_x(y) = R(x, y)x = \frac{1}{4}[[x, y], x]$. We may assume (up to the adjoint action of $U(n)$) that $x = diag(i\pi k_1, \ldots, i\pi k_{2m})$ for $k_i \in 2\mathbb{Z} + 1$ with $k_1 \geq \ldots \geq k_n$. Writing $y = (y_{pq}) \in \mathfrak{su}(2m)$, we get that

$$[x, y] = xy - yx = (i\pi k_p y_{pq}) - (i\pi k_q y_{pq}) = (i\pi(k_p - k_q)y_{pq}),$$

so

$$4K_x(y) = [[x, y], x] = [x, y]x - x[x, y] = ((i\pi k_q)(i\pi(k_p - k_q))y_{pq}) - ((i\pi k_p)(i\pi(k_p - k_q))y_{pq}) = (\pi^2(k_p - k_q)^2 y_{pq}).$$

Thus, we can describe a basis of $\mathfrak{su}(2m)$ of eigenvectors for $K_x$ as follows.

1. For $p < q$, the matrix with 1 in the $(p, q)$ position and $-1$ in the $(q, p)$ position is an eigenvector with eigenvalue $\frac{\pi^2}{4}(k_p - k_q)^2$.
2. For $p < q$, the matrix with $i$ in the $(p, q)$ and $(q, p)$ positions is an eigenvector with eigenvalue $\frac{\pi^2}{4}(k_p - k_q)^2$.
3. All diagonal matrices are eigenvectors with eigenvalue 0.

So the positive eigenvalues of $K_x$ are exactly the numbers $\frac{\pi^2}{4}(k_p - k_q)^2$ where $k_p > k_q$ (since this implies by our assumptions that $p < q$), and each of these has multiplicity 2.

Now, the condition that $\exp(tx)$ is conjugate to $I$ is exactly the condition that $t$ is a multiple of

$$\frac{\pi}{\sqrt{\frac{\pi^2}{4}(k_p - k_q)^2}} = \frac{2}{k_p - k_q}$$

for some $k_p > k_q$. Since we need $t \in (0, 1)$, each pair $(p, q)$ with $k_p > k_q$ gives us $\frac{k_p - k_q}{2} - 1$ conjugate points. But each positive eigenvalue has multiplicity 2, so this pair contributes $k_p - k_q - 2$ to the index of $\exp(tx)$. Thus the index is given by

$$\lambda = \sum_{k_p > k_q} (k_p - k_q - 2).$$

Since $\sum_{i=1}^{2m} k_i = 0$, if more than half of the $k_i$ are negative then $k_1 \geq 3$, so $\lambda \geq (m+1)(3 - (-1) - 2) = 2m + 2$. Similarly, if more than half of the $k_i$ are positive that $k_{2m} \leq -3$, so $\lambda \geq (m+1)(1 - (-3) - 2) = 2m + 2$. Lastly, if half of the $k_i$ are positive and half are negative, then since $\exp(tx)$ is not minimal we know that $k_1 \geq 3$ and $k_{2m} \leq -3$, and so $\lambda \geq (m-1)(3 - (-1) - 2) + (m-1)(1 - (-3) - 2) + (3 - (-3) - 2) = 4m \geq 2m + 2$. So in any case $\lambda \geq 2m + 2$, which proves the claim.

□

From here, the proof of Bott periodicity is a straightforward series of applications of the long exact sequence in homotopy for a fibration.

**Theorem** (Bott periodicity). *$U \xleftarrow{\sim} \Omega(\mathrm{colim}_m Gr_m(\mathbb{C}^{2m})) \xrightarrow{\sim} \Omega^2 U$.*

*Proof.* To show that the first map is a homotopy equivalence, we proceed as follows. First, the fibration $U(m) \hookrightarrow U(m+1) \twoheadrightarrow S^{2m+1}$ shows that the inclusion $U(m) \hookrightarrow U(m+1)$ induces isomorphisms on $\pi_{\leq 2m-1}$. So in fact all the inclusions $U(m) \hookrightarrow U(m+1) \hookrightarrow U(m+2) \hookrightarrow \ldots$ induce isomorphisms on $\pi_{\leq 2m-1}$. Hence, from the fibration $U(m) \hookrightarrow U(2m) \twoheadrightarrow U(2m)/U(m)$ we can conclude that $\pi_i(U(2m)/U(m)) = 0$ for $i \leq 2m - 1$. We can identify $Gr_m(\mathbb{C}^{2m}) = U(2m)/(U(m) \times U(m))$, so from the fibration $U(m) \hookrightarrow U(2m)/U(m) \twoheadrightarrow Gr_m(\mathbb{C}^{2m})$ we can conclude that $\Omega(Gr_m(\mathbb{C}^{2m})) \to U(m)$ induces isomorphisms on $\pi_{\leq 2m-2}$.[26] Taking colimits, this means that

$$\Omega(\mathrm{colim}_m(Gr_m(\mathbb{C}^{2m}))) = \mathrm{colim}_m \Omega(Gr_m(\mathbb{C}^{2m})) \to U$$

induces isomorphisms on $\pi_*$ and hence is a homotopy equivalence by Whitehead's theorem.[27]

To show that the second map is a homotopy equivalence, we proceed as follows. In the previous lemma we showed that $Gr_m(\mathbb{C}^{2m})$ can be taken to be the $(2m+1)$-skeleton of $\Omega(SU(2m))$. Thus the inclusion $Gr_m(\mathbb{C}^{2m}) \hookrightarrow \Omega(SU(2m))$ induces isomorphisms on $\pi_{\leq 2m}$, and hence the induced inclusion $\Omega(Gr_m(\mathbb{C}^{2m})) \hookrightarrow \Omega^2(SU(2m))$ induces isomorphisms on $\pi_{\leq 2m-1}$. Moreover, from the fibration $SU(2m) \hookrightarrow U(2m) \twoheadrightarrow S^1$ we can conclude that the inclusion $SU(2m) \hookrightarrow U(2m)$ induces isomorphisms on $\pi_{\neq 1}$, so the induced inclusion $\Omega(SU(2m)) \hookrightarrow \Omega(U(2m))$ induces isomorphisms on $\pi_{\neq 0}$, and the induced inclusion $\Omega^2(SU(2m)) \hookrightarrow \Omega^2(U(2m))$ induces isomorphisms on $\pi_*$ and hence is a homotopy equivalence by Whitehead's theorem. So the composite $\Omega(Gr_m(\mathbb{C}^{2m})) \hookrightarrow \Omega^2(SU(2m)) \hookrightarrow \Omega^2(U(2m))$ induces isomorphisms on $\pi_{\leq 2m-1}$. Taking colimits, this means that

$$\Omega(\mathrm{colim}_m(Gr_m(\mathbb{C}^{2m}))) = \mathrm{colim}_m \Omega(Gr_m(\mathbb{C}^{2m})) \to \mathrm{colim}_m \Omega^2(U(2m)) = \Omega^2 U$$

induces isomorphisms on $\pi_*$ and hence is a homotopy equivalence by Whitehead's theorem.

□

[26] Given a fibration $F \hookrightarrow E \twoheadrightarrow B$, if we convert the map $F \to E$ to a fibration then its fiber will end up being homotopy equivalent to $\Omega B$. The inclusion $\Omega B \to F$ will allow us to form the composite $\pi_i(B) \cong \pi_{i-1}(\Omega B) \to \pi_{i-1}(F)$, which is perhaps to cleanest way to write down the connecting homomorphism of the long exact sequence.

[27] Taking colimits commutes with taking loops because the circle is compact.