DEVELOPMENT OF MORSE THEORY

MATTHEW STEED

Abstract. In this paper, we develop Morse theory, which allows us to determine topological information about manifolds using certain real-valued functions defined on the manifolds. We first prove the Morse lemma, which says that, near critical points, such functions can be written in a useful way that gives us topological information. We then show how the homotopy type of the manifold is related to the information obtained from the function. Next we show that functions satisfying the required conditions exist for every manifold. Finally, as an application of the theory, we find the homotopy type of complex projective space.

Contents

1. The Morse Lemma 2
2. Critical Points and Homotopy Type 4
3. The Homotopy Type of a Manifold 8
4. The Existence of Appropriate Functions 9
5. The Homotopy Type of the Complex Projective Plane 13
Acknowledgments 14
References 14

Morse theory is concerned with the topological information that we can learn about a manifold from the study of functions on the manifold. In particular, by studying functions that satisfy certain conditions, we are able to determine the homotopy type of a manifold. Critical points of functions play a major role in the homotopy type of a manifold. They are defined as follows:

Definition 0.1. Given a manifold M, and a smooth function $f : M \to \mathbb{R}$, a point $p \in M$ is a critical point of f if all of the partial derivatives of f are 0 at p in a local coordinate system around p.

The Hessian matrix of a function f at a point p is the matrix $H(f)(p)$ whose entries are given by $(\frac{\partial^2 f}{\partial u_i \partial u_j}(p))$ in a local coordinate system $(u_1, \dots, u_n)$ in a neighborhood of p.

Definition 0.2. A critical point p of f is called nondegenerate if the Hessian matrix $H(f)(p)$ is nonsingular. Otherwise it is called degenerate.

Another important concept for determining the homotopy type of a manifold is the index of the Hessian at a point. It is defined as follows:

Date: August 29, 2016.

Definition 0.3. The index of the Hessian $H(f)(p)$ at a critical point p is the maximal dimension of a subspace of the tangent space on which $H(f)(p)$ is negative definite. For notational convenience, the index of $H(f)(p)$ will also be called the index of f at p.

In this paper M will be an n-dimensional manifold unless otherwise stated. We will sometimes denote coordinates such as $(x_1, \ldots, x_n)$ or $(u_1, \ldots, u_n)$ by the single letters x and u. If $f : M \to \mathbb{R}$ is a smooth function on M, we will define the set M^a as

$$M^a = \{q \in M | f(q) \leq a\}.$$

It is the portion of the manifold on which the value of f is at most a.

The proofs and discussion in this paper are based off of those in John Milnor's treatment of the subject, [1].

1. The Morse Lemma

In this section, we state and prove the Morse Lemma, which tells us that an appropriate smooth, real-valued function on a manifold can be written in a nice form. The new form gives us the index of the function at the critical point, which, as we will see, is related to the homotopy type of the manifold. We first prove the following lemma:

Lemma 1.1. *Let f be a real-valued C^∞ function in a convex neighborhood V of $0 \in \mathbb{R}^n$ such that $f(0) = 0$. Then, there exist C^∞ functions $g_1, \ldots, g_n$ such that*

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

Proof. By the fundamental theorem of calculus and the chain rule,

$$\begin{aligned} f(x_1, \ldots, x_n) &= \int_0^1 \frac{d}{dt}(f(tx_1, \ldots, tx_n))dt \\ &= \int_0^1 \left[\sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n) \right] dt \\ &= \sum_{i=1}^{n} \left[x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n)dt \right]. \end{aligned}$$

With $g_i(x_1, \ldots, x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \ldots, tx_n)dt$, we have the desired result. $\square$

We now state and prove the Morse Lemma:

Lemma 1.2. *Let $f : M \to \mathbb{R}$ be smooth, and let p be a nondegenerate critical point for f. Then there is a local coordinate system $(y_1, \ldots, y_n)$ in a neighborhood U of p such that $y_i(p) = 0$ for all i and such that $f(q) = f(p) - (y_1(q))^2 - \cdots - (y_\lambda(q))^2 + (y_{\lambda+1}(q))^2 + \cdots + (y_n(q))^2$ for all $q \in U$. Furthermore, λ is the index of f at p.*

Proof. Let U be a neighborhood of p with local coordinates. We may assume that $p = 0$ in these coordinates. With a redefinition $\tilde{f}(q) = f(q) - f(p)$, we may assume that $f(0) = f(p) = 0$. By the previous lemma, there exist C^∞ functions $g_1, \ldots, g_n$ such that $f(x) = \sum_{i=1}^{n} x_i g_i(x)$. Note that

$$\frac{\partial f}{\partial x_i} = \sum_j \left[\frac{\partial x_j}{\partial x_i} g_j + x_j \frac{\partial g_j}{\partial x_i} \right] = g_i + \sum_j x_j \frac{\partial g_j}{\partial x_i}.$$

Since 0 is a critical point of f, we have

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

Then the g_i's also satisfy the conditions of Lemma 1.1. There exist C^∞ functions h_{ij} such that $g_j(x) = \sum_i x_i h_{ij}(x)$. Then

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

We may assume that $h_{ij} = h_{ji}$ be replacing both of them with $\frac{1}{2}(h_{ij} + h_{ji})$. Also, since $\frac{\partial^2 f}{\partial x_i \partial x_j} = 2h_{ij}$, the matrix $(h_{ij}(0))$ is $\frac{1}{2}Hf(0)$, which is nonsingular.

We now proceed inductively. Assume that there are coordinates $(u_1, \ldots, u_n)$ in a neighborhood U of 0 such that $f(u) = \pm u_1^2 \pm \cdots \pm u_{r-1}^2 + \sum_{i,j \geq r} u_i u_j h_{ij}(u)$ for some r and functions h_{ij} such that $\det(h_{ij}(0)) \neq 0$. Since this determinant is nonzero, there is a linear change in the last $n-r+1$ variables such that $h_{rr}(0) \neq 0$. We assume this to be the case. Then, there is an open neighborhood $U' \subset U$ of 0 such that $h_{rr} \neq 0$ in U'. We define new coordinates $(v_1, \ldots, v_n)$ as follows:

$$v_i(u) = \begin{cases} u_i, & i \neq r \\ \sqrt{|h_{rr}(u)|}\left[u_r + \sum_{i>r} \frac{u_i h_{ir}(u)}{h_{rr}(u)}\right], & i = r \end{cases}.$$

First, if $i \neq r$, then $\frac{\partial v_i}{\partial u_j} = \delta_{ij}$. Thus the determinant of the Jacobian at 0 is equal to $\frac{\partial v_r}{\partial u_r}(0)$. This is given by

$$\frac{\partial v_r}{\partial u_r}(0) = \sqrt{|h_{rr}(0)|},$$

since every other term is multiplied by some u_i. This is nonzero, so, by the inverse function theorem, $(v_1, \ldots, v_n)$ are coordinate functions for the manifold.

We now assume that $h_{rr} > 0$. The negative case is similar. Note that

$$v_r^2 = h_{rr}\left(u_r + \sum_{i>r} u_i \frac{h_{ir}}{h_{rr}}\right)^2 = h_{rr}u_r^2 + 2\sum_{i>r} u_r u_i h_{ir} + \frac{\left(\sum_{i>r} u_i h_{ir}\right)^2}{h_{rr}}$$

Then

$$\begin{aligned} f &= \pm u_1^2 \pm \cdots \pm u_{r-1}^2 + \sum_{i,j \geq r} u_i u_j h_{ij}(u) \\ &= \pm u_1^2 \pm \cdots \pm u_{r-1}^2 + u_r^2 h_{rr} + 2\sum_{i>r} u_r u_i h_{ir} + \sum_{i,j>r} u_i u_j h_{ij}(u) \\ &= \pm v_1^2 \pm \cdots \pm v_{r-1}^2 + v_r^2 + \sum_{i,j>r} u_i u_j \tilde{h}_{ij}(u), \end{aligned}$$

for new functions $\tilde{h}_{ij}$. These new functions also satisfy the inductive hypothesis. The sign of v_r^2 would have been negative if h_{rr} were negative. This establishes the inductive step, so f takes the desired form by induction.

Finally, assuming f is in the form $f = f(p) - (y_1)^2 - \cdots - (y_\lambda)^2 + (y_{\lambda+1})^2 + \cdots + (y_n)^2$, the Hessian is given by

$$H(f) = \begin{pmatrix} -2 & & & & & \\ & \ddots & & 0 & & \\ & & -2 & & & \\ & & & 2 & & \\ & 0 & & & \ddots & \\ & & & & & 2 \end{pmatrix},$$

where there are λ -2's and $(n - \lambda)$ 2's. Clearly this is negative definite on a subspace of dimension λ and positive definite on a subspace of dimension $n - \lambda$. Any subspace of dimension larger than λ must intersect the subspace of dimension $n - \lambda$, so the maximal dimension of a subspace on which it is negative definite is λ. Hence λ is the index of f at p. This complete the proof of Morse's Lemma. □

Corollary 1.3. *Nondegenerate critical points of smooth, real-valued functions on manifolds are isolated.*

Proof. There is a neighborhood of a critical point p with local coordinates such that $p = 0$ and f takes the form given above. Clearly 0 is the only point where all of the partial derivatives of f vanish. □

The next corollary will be used later.

Corollary 1.4. *Suppose that $f : M \to \mathbb{R}$ is smooth and that each M^a is compact. Then the set of critical values has no accumulation point.*

Proof. Let c be a real number. Suppose that M^{c+1} has an infinite number of critical points. Then there is a convergent subsequence since M^{c+1} is compact. Since f is smooth, the limit of this sequence is a critical point. This cannot happen because each critical point is isolated. Then M^{c+1} contains only a finite number of critical points, so there are only a finite number of critical values below $c + 1$. Then the critical values cannot accumulate around c. □

2. Critical Points and Homotopy Type

In this section, we prove the theorems that tell us how the homotopy type of a manifold changes as we pass through critical points of appropriate functions on the manifold. For the first theorem, we need some facts about 1-parameter groups of diffeomorphisms, which we define as follows:

Definition 2.1. A 1-parameter group of diffeomorphisms is a map $\varphi : M \times \mathbb{R} \to \mathbb{R}$ such that

(a) For all $t \in \mathbb{R}$, $\varphi_t : M \to \mathbb{R}$ defined by $\varphi_t(q) = \varphi(t, q)$ is a diffeomorphism.
(b) For all $s, t \in \mathbb{R}$, $\varphi_{t+s} = \varphi_t \circ \varphi_s$.

Given a 1-parameter group of diffeomorphisms φ, we can define a vector field on the manifold as

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

where f is any smooth, real-valued function defined on the manifold. We say that X generates the group of diffeomorphisms. We claim the following fact without proof. A proof can be found in [1], section 2.

Lemma 2.2. *A smooth vector field on M that vanishes outside of a compact set $K \subset M$ generates a unique 1-parameter group of diffeomorphisms φ of M. The group φ generated is such that the vector field satisfies the equation given in the paragraph above.*

Given a smooth curve $t \mapsto c(t) \in M$, we define the velocity vector in the tangent space by the identity

$$\frac{dc}{dt}(f) = \lim_{h \to 0} \frac{f(c(t+h)) - f(c(t))}{h},$$

where f is again any real-valued function on the manifold. Let X be a vector field satisfying the conditions of the lemma, and let φ be generated by X. Then,

$$\begin{aligned}\frac{d\varphi_t(q)}{dt}(f) &= \lim_{h \to 0} \frac{f(\varphi_{t+h}(q)) - f(\varphi_t(q))}{h} \\ &= \lim_{h \to 0} \frac{f(\varphi_h(\varphi_t(q))) - f(\varphi_t(q))}{h} \\ &= X_{\varphi_t(q)}(f).\end{aligned}$$

Thus, $\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)}$. Intuitively, φ gives a path that takes each point in the direction that X is pointing.

We are now ready to prove the first theorem, which tells us that Homotopy type does not change when passing through an area with no critical points.

Theorem 2.3. *Let $f : M \to \mathbb{R}$ be smooth. Let $a < b$, and suppose that $f^{-1}[a, b]$ is compact and contains no critical points of f. Then, M^a is diffeomorphic to M^b. Furthermore, M^a is a deformation retract of M^b, so M^a and M^b have the same homotopy type.*

Proof. The idea of the proof is that without any critical points in $f^{-1}[a, b]$, the flow lines generated by ∇f are nice, and we can follow them from M^a into M^b. The 1-parameter group of diffeomorphisms generated by ∇f will be used to define the diffeomorphism and the deformation retract.

First, pick a Riemannian metric on M with inner product $\langle \cdot, \cdot \rangle$. The gradient ∇f is the vector field characterized by the equality $\langle X, \nabla f \rangle = X(f)$ for any vector field X. That is, at a point $q \in M$, $\langle X, \nabla f \rangle_q$ gives the directional derivative of f in the direction of the vector field X for any vector field X. Also, ∇f vanished at precisely the critical points of f, so it does not vanish at any point in $f^{-1}[a, b]$.

Note that if $c : \mathbb{R} \to M$ is a curve with velocity vector $\frac{dc}{dt}$, then

$$\left\langle \frac{dc}{dt}, \nabla f \right\rangle = \frac{dc}{dt}(f) = \lim_{h \to 0} \frac{f(c(t+h)) - f(c(t))}{h} = \frac{d(f \circ c)}{dt}.$$

Let $\rho : M \to \mathbb{R}$ be a function such that the following holds:

- for $q \in f^{-1}[a, b]$, $\rho(q) = \frac{1}{\langle \nabla f, \nabla f \rangle_q}$.
- $\rho = 0$ outside of a compact neighborhood of $f^{-1}[a, b]$.
- ρ is smooth and nonnegative.

Let X be the vector field defined by $X_q = \rho(q)(\nabla f)_q$. Then X is the normalized gradient in $f^{-1}[a, b]$, and it is a smooth vector field that vanishes outside of a compact set. Thus, it generates a 1-parameter group of diffeomorphisms $\varphi : \mathbb{R} \times M \to \mathbb{R}$ by Lemma 2.2.

For a fixed $q \in M$, we have a function $t \mapsto f(\varphi_t(q))$. This function gives the value of f after following the point q along its flow line for a time t. If $\varphi_t(q) \in f^{-1}[a,b]$, then

$$\frac{df(\varphi_t(q))}{dt} = \left\langle \frac{d\varphi_t(q)}{dt}, (\nabla f)_q \right\rangle = \langle X_q, (\nabla f)_q \rangle = 1,$$

from the definition of X. Thus, $t \mapsto f(\varphi_t(q))$ is linear with derivative 1. In $f^{-1}[a,b]$, $f(\varphi_t(q)) = f(\varphi_0(q)) + t = f(q) + t$.

We claim that the map $\varphi_{b-a} : M \to M$ restricted to M^a is a diffeomorphism onto M^b. It is a diffeomorphism on M, so we show that it maps M^a bijectively onto M^b. It is injective because the map on M is injective. Suppose that $q \in M^a$. Then, $f(q) \leq a$. If $\varphi_t(q) \notin f^{-1}[a,b]$ for all $t \in [0, b-a]$, then $\varphi_{b-a}(q) \in M^a$. Otherwise, by continuity, there is some $r \in [0, b-a]$ such that $f(\varphi_r(q)) = a$. Then $f(\varphi_{b-a}(q)) = f(\varphi_{b-a-r}(\varphi_r(q))) = f(\varphi_r(q)) + b - a - r = b - r \leq b$. Then $\varphi_{b-a}(q) \in M^b$ for any $q \in M^a$.

For surjectivity, suppose that $q \in f^{-1}(a,b]$. Then the value of f increases linearly along the path $\varphi_t(q)$ until it reaches b. At this point, the value of f never decreases because ρ is nonnegative, so, for any point in $q \in f^{-1}(a,b]$, $\varphi_{b-a}(q)$ is not in M^b. In the previous paragraph, we showed that points in M^a map into M^b. For $q \notin M^b$, we have $f(q) > b$. Since the value of f starts out above b, it will not fall below b along the line $\varphi_t(q)$ because, as above, the value of f along the line does not decrease when it nears b. Thus the only points that map to M^b are those that lie in M^a.

Since φ_{b-a} is surjective on the whole manifold, $\varphi_{b-a}|_{M^a}$ maps surjectively onto M^b. Then, φ_{b-a} is a diffeomorphism from M^a onto M^b.

We now give a deformation retract $r_t : M_b \to M_a$ defined by

$$r_t(q) = \begin{cases} q, & f(q) \leq a \\ \varphi_{t(a-f(q))}(q), & a \leq f(q) \leq b \end{cases}.$$

This map fixes M^a for all t. At $t = 0$, the map is the identity, since φ_0 is the identity. Points in $f^{-1}[a,b]$ are brought down along the flow lines as t increases, so they are brought to the boundary $f^{-1}(a)$ of M^a. Indeed, at $t = 1$, f maps the points to $f(\varphi_{a-f(q)}(q)) = f(q) + a - f(q) = a$. This completes the proof of the theorem. $\square$

The next theorem tells us how the homotopy type of the manifold changes as we cross a nondegenerate critical point. Crossing such a point results in the attaching of a λ-cell, which is a topological space homeomorphic to the open λ-dimensional unit disk in Euclidean space.

Theorem 2.4. *Let $f : M \to \mathbb{R}$ be smooth. Let p be a nondegenerate critical point with index λ. Let $f(p) = c$. Suppose that for some $\epsilon > 0$ small enough, $f^{-1}[c-\epsilon, c+\epsilon]$ is compact and contains no critical points of f other than p. Then for all sufficiently small $\delta > 0$, $M^{c+\delta}$ has the homotopy type of $M^{c-\delta}$ with a λ-cell attached.*

Proof. For this proof, we use the Morse Lemma, Lemma 1.2. Choose a coordinate system $(u_1, \ldots, u_n)$ in a neighborhood U of p such that $u_i(p) = 0$ for all i and $f = c - u_1^2 - \cdots - u_\lambda^2 + u_{\lambda+1}^2 + \cdots + u_n^2$. For convenience, we define $\xi, \eta : U \to [0, \infty)$

as

$$\xi = u_1^2 + \cdots + u_\lambda^2$$
$$\eta = u_{\lambda+1}^2 + \cdots + u_n^2,$$

so $f = c - \xi + \eta$.

Note that since there is $\epsilon > 0$ so that $f^{-1}[c-\epsilon, c+\epsilon]$ is compact, for any $0 < \epsilon' < \epsilon$, $f^{-1}[c - \epsilon', c + \epsilon']$ is a closed subset of a compact set, so it is also compact. Since U is open, it contains a ball around 0. Then we can choose $\epsilon > 0$ small enough that both of the following hold:

(a) $f^{-1}[c - \epsilon, c + \epsilon]$ is compact and contains no critical points of f other then p.
(b) The ball $\{(u_1, \dots, u_n) \mid \sum u_i^2 \leq 2\epsilon\}$ is contained in the image of U in $\mathbb{R}^n$.

Let D_λ be the λ-cell defined by

$$D_\lambda = \left\{ (u_1, \dots, u_n) \;\middle|\; \sum_{i=1}^{\lambda} u_i^2 \leq \epsilon,\ u_{\lambda+1} = \cdots = u_n = 0 \right\}.$$

Note that $D_\lambda \cap M^{c-\epsilon}$ is precisely the boundary

$$\dot{D}_\lambda = \left\{ (u_1, \dots, u_n) \;\middle|\; \sum_{i=1}^{\lambda} u_i^2 = \epsilon,\ u_{\lambda+1} = \cdots = u_n = 0 \right\},$$

since $f = c - \epsilon$ on this set and the value of f increases as the value of ξ decreases in the interior of the disk.

Let $\mu : \mathbb{R} \to \mathbb{R}$ be a nonnegative smooth function satisfying

(a) $\mu(0) > \epsilon$
(b) $\mu(r) = 0$ for $r \geq 2\epsilon$
(c) $-1 < \mu'(r) \leq 0$ for all r

Define $F : M \to \mathbb{R}$ by

$$F = \begin{cases} f, & \text{outside } U \\ f - \mu(\xi + 2\eta), & \text{in } U \end{cases}.$$

Then $F = c - \xi + \eta - \mu(\xi + 2\eta)$ in U.

We claim that $F^{-1}(-\infty, c+\epsilon] = f^{-1}(-\infty, c+\epsilon] = M^{c+\epsilon}$. In the region, $\xi + 2\eta > 2\epsilon$, $\mu(\xi + 2\eta) = 0$, so $F = f$. In the region $\xi + 2\eta \leq 2\epsilon$,

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

so this region is contained in the preimages of $(-\infty, c + \epsilon]$ for both regions.

We also claim that the critical points of F are the same as those of f. In U, the derivatives of F are

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

In particular, both are nonzero. By the chain rule,

$$dF = \frac{\partial F}{\partial \xi} d\xi + \frac{\partial F}{\partial \eta} d\eta = \frac{\partial F}{\partial \xi} \begin{pmatrix} 2u_1 \cdots 2u_\lambda & 0 \cdots 0 \end{pmatrix} + \frac{\partial F}{\partial \eta} \begin{pmatrix} 0 \cdots 0 & 2u_{\lambda+1} \cdots 2u_n \end{pmatrix}.$$

This is 0 only at $p = 0$ in U. Then p is the only critical point of F in U. Outside of U, $F = f$, so they share the same critical points.

Since $F \leq f$, we have $f^{-1}(-\infty, a) \subset F^{-1}(-\infty, a)$ for all a. Since $F^{-1}(-\infty, c + \epsilon] = f^{-1}(-\infty, c+\epsilon]$, we have $F^{-1}[c-\epsilon, c+\epsilon] \subset f^{-1}[c-\epsilon, c+\epsilon]$. Then it is compact and contains no critical points of F except possibly p. However, $F(p) = F(0) = c - \mu(0) < c - \epsilon$, since $\mu(0) > \epsilon$. Then $F^{-1}[c - \epsilon, c + \epsilon]$ does not contain p, so this region contains no critical points of F. By Theorem 2.3, $F^{-1}(-\infty, c - \epsilon]$ is a deformation retract of $F^{-1}(-\infty, c + \epsilon]$.

Let $H = \overline{F^{-1}(-\infty, c - \epsilon] - M^{c-\epsilon}}$. Then $F^{-1}(-\infty, c - \epsilon] = M^{c-\epsilon} \cup H$.

We claim that $D_\lambda \subset H$. Suppose that $q \in D_\lambda$. Since $\frac{\partial F}{\partial \xi} < 0$, $F(q) \leq F(0) < c - \epsilon$. However, $f(q) \geq c - \epsilon$ since $f = c - \xi + \eta$.

We now claim that $M^{c-\epsilon} \cup D_\lambda$ is a deformation retract of $M^{c-\epsilon} \cup H$. We define the deformation retract $r_t : M^{c-\epsilon} \cup H \to M^{c-\epsilon} \cup H$ in four areas. Outside of U, it is defined as the identity. If $\xi \leq \epsilon$, it is defined as

$$(u_1, \ldots, u_n) \mapsto (u_1, \ldots, u_\lambda, (1-t)u_{\lambda+1}, \ldots, (1-t)u_n).$$

If $\epsilon \leq \xi \leq \eta + \epsilon$, it is defined as

$$(u_1, \ldots, u_n) \mapsto (u_1, \ldots, u_\lambda, s_t u_{\lambda+1}, \ldots, s_t u_n),$$

where the s_t are defined as

$$s_t = (1-t) + t\sqrt{\frac{\xi - \epsilon}{\eta}}.$$

Finally, if $\eta + \epsilon \leq \xi$, r_t is defined as the identity. This gives a deformation retraction.

We conclude that $M^{c+\epsilon}$ deformation retracts onto $F^{-1}(-\infty, c - \epsilon] = M^{c-\epsilon} \cup H$, which deformation retracts onto $M^{c-\epsilon} \cup D_\lambda$, which completes the proof. $\square$

It is also true that if $f^{-1}(c)$ has nondegenerate critical points $p_1, \ldots, p_k$ with indices $\lambda_1, \ldots, \lambda_k$, then $M^{c+\epsilon}$ has the homotopy type of $M^{c-\epsilon} \cup D_{\lambda_1} \cup \cdots \cup D_{\lambda_k}$ for $\epsilon > 0$ small.

3. The Homotopy Type of a Manifold

In this section, we show that manifolds with appropriate functions on them have the homotopy types of CW complexes. We use the following two results from algebraic topology, which we cite without proof:

Lemma 3.1 (Whitehead)**.** *Let φ_0, φ_1 be homotopic maps from $\dot{D}_\lambda$ to a topological space X. Then the identity map on X extends to a homotopy equivalence*

$$K : (X \cup_{\varphi_0} D_\lambda) \to (X \cup_{\varphi_1} D_\lambda).$$

Lemma 3.2. *Let $\varphi : \dot{D}_\lambda \to X$ be an attaching map. Any homotopy equivalence $f : X \to Y$ extends to a homotopy equivalence*

$$F : (X \cup_\varphi D_\lambda) \to (Y \cup_{f\varphi} D_\lambda).$$

Proofs of these lemmas can be found in [1], section 3.

With these, we can prove the following theorem:

Theorem 3.3. *If f is a smooth, real-valued function on a manifold M with no degenerate critical points, and if M is compact, then M has the homotopy type of a CW complex*

Proof. Each M^a is a closed subset of M, so each M^a is also compact. Let $c_1 < c_2 < \cdots$ be the critical values of $f : M \to \mathbb{R}$. Since each M^a is compact, the critical values do not accumulate, by Corollary 1.4. For $a < c_1$, $M^a = \varnothing$. We proceed inductively. Suppose that a is not a critical value and that M^a has the homotopy type of a CW complex. Let c be the smallest critical value greater than a, which exists because they do not accumulate. Then, for ϵ small enough,

$$M^{c+\epsilon} \cong M^{c-\epsilon} \cup_{\varphi_1} D_{\lambda_1} \cup_{\varphi_2} \cdots \cup_{\varphi_{j(c)}} D_{\lambda_{j(c)}}$$

for some attaching maps φ_i by Theorem 2.4. Also, $M^{c-\epsilon} \cong M^a$ with some homotopy equivalence $h : M^{c-\epsilon} \to M^a$ by Theorem 2.3. By assumption, we also have a homotopy equivalence $h' : M^a \to K$, where K is a CW complex. Each $h' \circ h \circ \varphi_i : \dot{D}_{\lambda_i} \to K$ is homotopic to a cellular map

$$\Psi_i : \dot{D}_{\lambda_i} \to (\lambda_i - 1) - \text{skeleton of } K.$$

Then

$$K \cup_{\Psi_1} D_{\lambda_1} \cup_{\Psi_2} \cdots \cup_{\Psi_{j(c)}} D_{\lambda_{j(c)}}$$

is a CW complex. We now have

$$M^{c-\epsilon} \cong M^a \cong K$$

with homotopy equivalence $h' \circ h$. Then, by lemma 3.2 for the first equivalence and lemma 3.1 for the second,

$$M^{c-\epsilon} \cup_{\varphi_1} D_{\lambda_1} \cong K \cup_{h' \circ h \circ \varphi_1} D_{\lambda_1} \cong K \cup_{\Psi_1} D_{\lambda_1}.$$

Repeating this $j(c) - 1$ more times, we have

$$M^{c+\epsilon} \cong \left(K \cup_{\Psi_1} D_{\lambda_1} \cup_{\Psi_2} \cdots \cup_{\Psi_{j(c)}} D_{\lambda_{j(c)}}\right),$$

where the right hand side is a CW complex. Then, each M^a has the homotopy type of a CW complex. Since M is compact, it is equal to M^a for some a, so it has the homotopy type of a CW complex. □

Remark 3.4. If we remove the condition that M is compact and replace it with the condition that M^a is compact for each a, the result that M has the homotopy type of a CW complex still holds. The proof proceeds in the same manner until the end. At that point, we have a sequence of increasing portions of the manifold and corresponding CW complexes. An argument involving limits shows that M still has the homotopy type of a CW complex.

4. The Existence of Appropriate Functions

In this section we show that the theory developed above is not vacuous. We show that for any manifold embedded in Euclidean space, there are many smooth functions on the manifold with no degenerate critical points such that each M^a is compact. First, we need the concept of focal points of a manifold.

Let M be a k-dimensional manifold differentially embedded in $\mathbb{R}^n$, with $k < n$. Let $N \subset M \times \mathbb{R}^n$ be the normal vector bundle defined as

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

Then N is an n-dimensional manifold differentiably embedded in $\mathbb{R}^{2n}$. Let $E : N \to \mathbb{R}^n$ be the endpoint map defined by $E(q, v) = q + v$. We now define focal points as follows:

Definition 4.1. A point $e \in \mathbb{R}^n$ 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 $\mu > 0$. We will also call e a focal point of M if e is a focal point of (M, q) for some q.

A focal point is essentially a place where normal vectors of nearby points intersect. Since the Jacobian of E has a nontrivial null space, moving to nearby points of N in certain dimensions does not change where the endpoints of the normal vectors lie. The locations of focal points are related to the curvatures and radii of curvatures at a place in the manifold. Intuitively, at a place with a certain curvature in the direction of a normal vector, we expect a focal point to lie along the vector at a distance equal to the radii of curvature. For example, the center of a circle is a focal point of every point of the circle. To make this rigorous, we introduce the first and second fundamental forms on a manifold. We first need local coordinates on the manifold and coordinates that arise from embedding the manifold in $\mathbb{R}^n$.

Let $(u_1, \ldots, u_k)$ be local coordinates for a region of M. The inclusion of M into $\mathbb{R}^n$ gives n smooth functions $x_1(u_1, \ldots, u_k), \ldots, x_n(u_1, \ldots, u_k)$. We will denote them as $\vec{x}(u_1, \ldots, u_k) = (x_1, \ldots, x_n)$. We will also denote other points in $\mathbb{R}^n$ with the vector hats. We define the derivative of $\vec{x}$ as

$$\frac{\partial \vec{x}}{\partial u_i} = \left(\frac{\partial x_1}{\partial u_i}, \ldots, \frac{\partial x_n}{\partial u_i} \right),$$

with the second derivative defined analogously.

The first fundamental form is defined as the matrix

$$(g_{ij}) = \left(\frac{\partial \vec{x}}{\partial u_i} \cdot \frac{\partial \vec{x}}{\partial u_j} \right),$$

with respect to the dot product in Euclidean space. For the second fundamental form, the vector $\frac{\partial^2 \vec{x}}{\partial u_i \partial u_j}$ has components tangent to M and normal to M. Let $\vec{l_{ij}}$ be the component normal to M. The second fundamental form is the matrix of vector-valued functions given by $(\vec{l_{ij}})$. Let $\vec{v}$ be a vector that is normal to M at $\vec{q}$. Then the second fundamental form in the direction of $\vec{v}$ is the matrix $(\vec{v} \cdot \vec{l_{ij}})$, which is also equal to $(\vec{v} \cdot \frac{\partial^2 \vec{x}}{\partial u_i \partial u_j})$ since the tangent part of the derivative disappears when it is dotted with a normal vector.

We can assume that the coordinates u have been chosen so the first fundamental form (g_{ij}) is the identity matrix at some point $\vec{q} \in M$, by, for example, a Gram-Schmidt process. Let $\vec{v}$ be a vector normal to M at $\vec{q}$. The eigenvalues $K_1, \ldots, K_k$ of the second fundamental form $(\vec{v} \cdot \vec{l_{ij}})$ are called the principal curvatures of M at $\vec{q}$ in the direction of $\vec{v}$. For the nonzero principal curvatures, the inverses, K_i^{-1}, are called the principal radii of curvature.

We fix a point $\vec{q} \in M$ and a normal vector $\vec{v}$. We choose coordinates so that the first fundamental form equals the identity at $\vec{q}$. The following lemma tells us the locations of any focal points of M on the line $l = \vec{q} + t\vec{v}$ for $t \in \mathbb{R}$.

Lemma 4.2. *The focal points of $(M, \vec{q})$ along the line l are precisely the points $\vec{q} + K_i^{-1}\vec{v}$, where $1 \leq i \leq k$ and $K_i \neq 0$.*

Proof. First, choose $n - k$ vector fields $\vec{w}_1(u_1, \ldots, u_k), \ldots, \vec{w}_{n-k}(u_1, \ldots, u_k)$ that are orthonormal vectors normal to M. At any point they form a basis for the normal space. We can define local coordinates $(u_1, \dots, u_k, t_1, \dots, t_{n-k})$ on N by

$$(u_1, \dots, u_k, t_1, \dots, t_k) \mapsto \left(\vec{x}(u_1, \dots, u_k), \sum_{\alpha=1}^{n-k} t_\alpha \vec{w}_\alpha(u_1, \dots, u_k) \right) \in N.$$

The endpoint map $E : N \to \mathbb{R}$ is then given by

$$(u_1, \dots, u_k, t_1, \dots, t_k) \mapsto \vec{x}(u_1, \dots, u_k) + \sum_{\alpha=1}^{n-k} t_\alpha \vec{w}_\alpha(u_1, \dots, u_k)$$

Let $\vec{e}(u_1, \dots, u_k, t_1, \dots, t_k) = \vec{x}(u_1, \dots, u_k) + \sum_{\alpha=1}^{n-k} t_\alpha \vec{w}_\alpha(u_1, \dots, u_k)$ be the resulting vector. We are interested in the rank of its Jacobian at the points $(\vec{q}, t\vec{v})$. Its partial derivatives are given by

$$\frac{\partial \vec{e}}{\partial u_i} = \frac{\partial \vec{x}}{\partial u_i} + \sum_\alpha t_\alpha \frac{\partial \vec{w}_\alpha}{\partial u_i}$$
$$\frac{\partial \vec{e}}{\partial t_\alpha} = \vec{w}_\alpha.$$

The vectors $\frac{\partial \vec{x}}{\partial u_i}$ are linearly independent tangent vectors that form a basis for the tangent space. Then $\frac{\partial \vec{x}}{\partial u_1}, \dots, \frac{\partial \vec{x}}{\partial u_k}, \vec{w}_1, \dots, \vec{w}_{n-k}$ is a collection of n linearly independent vectors in $\mathbb{R}^n$. Then the matrix with these vectors as columns has full rank, so the product of the Jacobian with this matrix has the rank of the Jacobian. This product is given, in block form, by

$$\begin{pmatrix} \frac{\partial \vec{e}}{\partial u_i} \\ \vdots \\ \frac{\partial \vec{e}}{\partial t_{n-k}} \end{pmatrix} \begin{pmatrix} \frac{\partial \vec{x}}{\partial u_1} & \cdots & \vec{w}_{n-k} \end{pmatrix} = \begin{pmatrix} \left[\frac{\partial \vec{x}}{\partial u_i} \cdot \frac{\partial \vec{x}}{\partial u_j} + \sum_\alpha t_\alpha \frac{\partial \vec{w}_\alpha}{\partial u_i} \cdot \frac{\partial \vec{x}}{\partial u_j} \right] & \left[\sum_\alpha t_\alpha \frac{\partial \vec{w}_\alpha}{\partial u_i} \cdot \vec{w}_\beta \right] \\ 0 & Id \end{pmatrix}.$$

The rank of this matrix, which is the rank of the Jacobian of E, is equal to the rank of the upper left hand block. Since $\frac{\partial \vec{x}}{\partial u_i}$ is a tangent vector and $\vec{w}_\alpha$ is a normal vector, we have $\frac{\partial \vec{x}}{\partial u_i} \cdot \vec{w}_\alpha = 0$. Then,

$$0 = \frac{\partial}{\partial u_j} \left(\vec{w}_\alpha \cdot \frac{\partial \vec{x}}{\partial u_i} \right) = \frac{\partial \vec{w}_\alpha}{\partial u_j} \cdot \frac{\partial \vec{x}}{\partial u_i} + \vec{w}_\alpha \cdot \frac{\partial^2 \vec{x}}{\partial u_i \partial u_j}.$$

Thus

$$\sum_\alpha t_\alpha \frac{\partial \vec{w}_\alpha}{\partial u_i} \cdot \frac{\partial \vec{x}}{\partial u_j} = -\sum_\alpha t_\alpha \vec{w}_\alpha \cdot \frac{\partial^2 \vec{x}}{\partial u_i \partial u_j} = -\sum_\alpha t_\alpha \vec{w}_\alpha \cdot \vec{l}_{ij}$$

So we can rewrite the upper left hand block of the matrix as

$$\left(g_{ij} - \sum_\alpha t_\alpha \vec{w}_\alpha \cdot \vec{l}_{ij} \right).$$

Thus, a point $\vec{q} + t\vec{v}$ is a focal point of multiplicity μ if and only if the matrix $(g_{ij} - t\vec{v} \cdot \vec{l}_{ij})$ is singular with nullity μ at $\vec{q}$. Since we have assumed that (g_{ij}) is the identity matrix, the above matrix is singular if and only if $\frac{1}{t}$ is an eigenvalue of $(\vec{v} \cdot \vec{l}_{ij})$. Then the multiplicity of μ is equal to the multiplicity of $\frac{1}{t}$ as an eigenvalue. The only focal points are those of the form $\vec{q} + K_i^{-1}\vec{v}$, as claimed. This completes the proof. □

Remark 4.3. The above proof shows that there at most k focal points of $(M, \vec{q})$ along l, each counted with multiplicity.

Fix a point $p \in \mathbb{R}^n$. We now consider the function $L_{\vec{p}} = f : M \to \mathbb{R}$ defined as

$$f(\vec{x}(u_1, \ldots, u_k)) = ||\vec{x} - \vec{p}||^2 = \vec{x} \cdot \vec{x} - 2\vec{x} \cdot \vec{p} + \vec{p} \cdot \vec{p}.$$

We will show that there are many functions of this form that have no degenerate critical points. Clearly, each M^a is compact for a function of this form. The following lemma characterizes the degenerate critical points of f.

Lemma 4.4. *A point $\vec{q} \in M$ is a degenerate critical point of $L_{\vec{p}}$ if and only if $\vec{p}$ is a focal point of $(M, \vec{q})$.*

Proof. The derivatives of f are given by

$$\frac{\partial f}{\partial u_i} = 2\vec{x} \cdot \frac{\partial \vec{x}}{\partial u_i} - 2\vec{p} \cdot \frac{\partial \vec{x}}{\partial u_i} = 2\frac{\partial \vec{x}}{\partial u_i} \cdot (\vec{x} - \vec{p}).$$

Since the $\frac{\partial \vec{x}}{\partial u_i}$ form a basis for the tangent space, this is 0 for each i if and only if $\vec{x} - \vec{p}$ is a normal vector. Thus, a point $\vec{q}$ is a critical point of f if and only if $\vec{q} - \vec{p}$ is normal to M at $\vec{q}$. The second derivatives of f are

$$\frac{\partial^2 f}{\partial u_i \partial u_j} = 2\left[\frac{\partial \vec{x}}{\partial u_i} \cdot \frac{\partial \vec{x}}{\partial u_j} + \frac{\partial^2 \vec{x}}{\partial u_i \partial u_j} \cdot (\vec{x} - \vec{p})\right].$$

Assuming we are evaluating this at a critical point $\vec{q}$, $\vec{p} = \vec{q} + t\vec{v}$ for a normal unit vector $\vec{v}$ since $\vec{q} - \vec{p}$ is normal to M. Then the second derivative can be rewritten as

$$\frac{\partial^2 f}{\partial u_i \partial u_j} = 2\left[g_{ij} - t\vec{v} \cdot \vec{l}_{ij}\right].$$

Then the Hessian matrix is singular if and only if $\vec{p}$ is a focal point of $(M, \vec{q})$, as we saw in the proof of Lemma 4.2.

□

To show that many of these functions have no degenerate critical points, we need Sard's theorem, which we state here without proof:

Theorem 4.5 (Sard)**.** *If M_1 and M_2 are differentiable manifolds of the same dimension, each with a countable basis, and if $f : M_1 \to M_2$ is C^1, then the image of the set of critical points of f has measure 0.*

This theorem shows that the set of focal points of a manifold has measure 0 in $\mathbb{R}^n$:

Corollary 4.6. *For almost all $x \in \mathbb{R}^n$, x is not a focal point of M.*

Proof. N is a n-dimensional manifold, and $x \in \mathbb{R}^n$ is a focal point of M if and only if x is in the image of the set of critical points of $E : N \to \mathbb{R}^n$. This set has measure 0 by Sard's Theorem. □

Since the set of focal points of a manifold has measure 0, we have the following theorem:

Theorem 4.7. *For almost all $\vec{p} \in \mathbb{R}^n$, the function $L_{\vec{p}} : M \to \mathbb{R}$ has no degenerate critical points.*

It is a fact that any manifold can be differentiably embedded in Euclidean space, so on any manifold, there is a function with no degenerate critical points for which each M^a is compact. Thus, we have the following theorem:

Theorem 4.8. *A manifold has the homotopy type of a CW complex.*

Recall that the index of a critical point gives the dimension of the cell attached when crossing that critical point. The index of $L_{\vec{p}}$ is related to focal points, as shown in the following lemma:

Lemma 4.9. *The index of $L_{\vec{p}}$ at a nondegenerate critical point $\vec{q} \in M$ is equal to the number of focal points of $(M, \vec{q})$ that lie on the line segment from $\vec{q}$ to $\vec{p}$, with each focal point being counted with multiplicity.*

Proof. With $L_{\vec{p}} = f$ and notation as above, the Hessian matrix is given by

$$\left(\frac{\partial^2 f}{\partial u_i \partial u_j}\right) = 2\left(g_{ij} - t\vec{v} \cdot \vec{l}_{ij}\right),$$

where $t > 0$. This has index equal to the number of negative eigenvalues. Since $(g_{ij}) = Id$, the eigenvalue equation for an eigenvalue λ is $0 = (\lambda - \frac{1}{t})Id + (\vec{v} \cdot \vec{l}_{ij}) = (\frac{1}{t} - \lambda)Id - (\vec{v} \cdot \vec{l}_{ij})$. Then λ is an eigenvalue of the Hessian if and only if $\frac{1}{t} - \lambda$ is an eigenvalue of the second fundamental form. Then the number of negative eigenvalues of the Hessian is equal to the number of eigenvalues of the fundamental form that are greater than $\frac{1}{t}$. An eigenvalue greater than $\frac{1}{t}$ corresponds to a principal radius less than t, so the corresponding focal point lies on the line between $\vec{q}$ and $\vec{p}$. Thus the number of negative eigenvalues of the Hessian is equal to the number of focal points on the line between $\vec{q}$ and $\vec{p}$. This establishes the index theorem. □

5. The Homotopy Type of the Complex Projective Plane

In this section, we give an application of the theory that has been developed so far. We find the homotopy type of the complex projective plane $\mathbb{CP}^n$. We think of $\mathbb{CP}^n$ as the set of equivalence classes of $(n+1)$-tuples $(z_0, \ldots, z_n)$ such that $\sum |z_i|^2 = 1$. Two tuples are equivalent if they differ by multiplication by a scalar complex number of absolute value 1. We denote the equivalence class of $(z_0, \ldots, z_n)$ by $(z_0 : \cdots : z_n)$. Let $f : \mathbb{CP}^n \to \mathbb{R}$ be defined by

$$f(z_0 : \cdots : z_n) = \sum c_i |z_i|^2,$$

where the c_i are distinct real numbers. Let $U_0 = \{(z_0 : \cdots : z_n) : z_0 \neq 0\}$. For a given point $(z_0 : \cdots : z_n)$ and $j \neq 0$, let $x_j + iy_j = |z_0|\frac{z_j}{z_0}$. To see that this is well-defined, let $(uz_0 : \cdots : uz_n)$ be another representative of the equivalence class, where u has norm 1. Then $|uz_0|\frac{uz_j}{uz_0} = |z_0|\frac{z_j}{z_0}$. Thus the functions $x_1, y_1, \ldots, x_n, y_n : U_0 \to \mathbb{R}$ are well-defined and map U_0 diffeomorphically onto an open unit ball in $\mathbb{R}^{2n}$. Note that $|z_j|^2 = x_j^2 + y_j^2$, so that

$$|z_0|^2 = 1 - \sum (x_j^2 + y_j^2).$$

Then, in U_0, f is of the form

$$f = c_0 + \sum_{j=1}^{n} (c_j - c_0)(x_j^2 + y_j^2).$$

Note that $\frac{\partial f}{\partial x_j} = 2x_j(c_j - c_0)$, and similarly for y_j. Since the c_i are all distinct, $(1 : 0 : \cdots : 0)$ is the only critical point of f in U_i. The Hessian is of the form

$$H(f) = \begin{pmatrix} c_1 - c_0 & & & & & & & \\ & c_1 - c_0 & & & & & 0 & \\ & & c_2 - c_0 & & & & & \\ & & & c_2 - c_0 & & & & \\ & 0 & & & \ddots & & & \\ & & & & & c_n - c_0 & & \\ & & & & & & c_n - c_0 \end{pmatrix}$$

The index is the number of negative eigenvalues, which is equal to twice the number of j such that $c_j < c_0$.

We can also consider the other coordinate systems $U_i = \{(z_0 : \cdots : z_n) : z_i \neq 0\}$. A similar arguement for each shows that the points $p_1 = (0 : 1 : 0 : \cdots : 0), p_2, \ldots, p_n$, with $p_i \in U_i$, are the only critical points of f in each U_i. The union of each U_i is $\mathbb{CP}^n$, so the p_i are the only critical points of f. The Hessian of f at each p_i is similar to the one above at p_0. The index of f at p_i is twice the number of j's with $c_j < c_i$. Thus, every even number between 0 and $2n$ is the index of f at exactly one of the p_i's. For example, if $c_0 > c_1 > \cdots > c_n$, then the index at p_0 is $2n$, the index at p_1 is $2n-2$, ... , and the index at p_n is 0. More generally, for an ordering $c_{i_0} > c_{i_1} > \cdots > c_{i_n}$, the index of f at p_{i_0} is $2n$, the index at p_{i_1} is $2n-2$, and so on. Thus $\mathbb{CP}^n$ has the homotopy type of a CW complex of the form

$$D_0 \cup D_2 \cup \cdots \cup D_{2n}.$$

Acknowledgments. It is a pleasure to thank my mentor, Jonathan Wang, for helping me understand many of the proofs in this paper and new mathematical concepts that I had not encountered before.

References

[1] John Milnor. Morse Theory. Princeton University Press.