AN INTRODUCTORY TREATMENT OF MORSE THEORY ON MANIFOLDS

AMY HUA

Abstract. Morse Theory is greatly utilized to find and decompose structures on manifolds. The concept of a manifold, as a generalization of objects into n dimensions, is crucial in many areas of geometry. Manifolds range from the very simple (e.g. the plane) to the very complicated and unimaginable (e.g. the Klein bottle). However, despite their various forms and complexities, Morse Theory allows us to capture them in terms of Euclidean space, hence allowing us to conduct calculus on manifolds. Much of what Morse Theory can do will be beyond the scope of this paper. Our ultimate goal will be to familiarize ourselves with manifolds and be able to understand and prove the Morse Lemma. The Morse Lemma is a gateway theorem of Morse Theory which allows us to directly analyze the neighborhood of a nondegenerate critical point on a manifold in a useful, intuitive manner, akin to the slopes around a hole in 3-D space. This paper will lend an introductory discussion of manifolds and their preoccupations with Euclidean geometry. We will start by assuming the basic tenets of linear algebra and multivariable calculus.

Contents

1. Preliminaries: What is a Manifold?

Before we begin to define a manifold, we need a few preliminary definitions.

Definition 1.1. Let U and V be open sets. A function $f : U \to V$ is C^p **at a point** $u \in U$ (read "f is class p at a point u") if all partial derivatives of order $k \leq p$ exist and are continuous at u. We say that the function f is **of class** C^p if f is C^p at every point in U.

Definition 1.2. f is **smooth at a point** $u \in U$ if all partial derivatives of any finite order exist and are continuous at u. Such functions are also called C^∞ *at* u. We say that f is a **smooth function**, or f *is of class* C^∞, if it is smooth at every point in U.

Date: September 7, 2010.

Definition 1.3. Let X and Y be topological spaces. The map $h : X \to Y$ is a **homeomorphism** if h is bijective and continuous, and it has a continuous inverse. We consider X to be **homeomorphic** to Y if there exists a homeomorphism between X and Y.

Remark 1.4. Notice that the inverse of a homeomorphism and a composition of homeomorphisms are also homeomorphisms.

Definition 1.5. A **diffeomorphism** is a smooth homeomorphism with a smooth inverse.

We are now ready to define manifolds.

Definition 1.6. A space M is an **n-dimensional manifold** (without boundary), or an *n-manifold*, if there exists an open cover $\{U_\alpha\}_{\alpha \in I}$ on M (where I is some interval on $\mathbb{N}$) and open sets $\{V_\alpha\}_{\alpha \in I} \subset \mathbb{R}^n$ such that for every $\alpha \in I$, a homeomorphism $\varphi_\alpha : U_\alpha \to V_\alpha$ exists.

Surfaces are merely defined as two-dimensional manifolds. Examples of surfaces include the sphere, plane, cylinder, prism, and torus.

One example of a manifold is the **(n-1)-dimensional unit sphere**

$$S^n = \left\{(x_1), \ldots, x_n) | x_1^2 + \ldots + x_n^2 = 1\right\} \tag{1.7}$$

S^n is a generalization of the unit circle (S^1) and the unit sphere (S^2) to higher dimensions. We may think of the Earth as the surface S^2 whereby the terrain of a small area around every point can be captured in a "chart", and the entire topology of the Earth may be expressed in terms of a collection of charts, or a two-dimensional "atlas".

Definition 1.8. Such a function φ_α (defined above) with its domain U_α is called a **chart** or a **local coordinate system**, denoted as $(U_\alpha, \varphi_\alpha)$. If U_α is a neighborhood around a point p in M, then we call $(U_\alpha, \varphi_\alpha)$ a *chart (or local coordinate system) centered at p.*

A chart is endowed with a set of **local coordinates**, say, $(x_1, \ldots, x_n)$, where $x_i \in \mathbb{R}$ for every $i \in [1, n]$, and for which U_α can be "charted" upon. That is, if $\varphi_\alpha : U_\alpha \to V_\alpha$ is a chart defined by $\varphi(x) = (x_1, \ldots, x_n)$ for $x \in U_\alpha$, then $(x_1, \ldots, x_n)$ are local coordinates.

Definition 1.9. An **atlas** of M, denoted as $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$, is defined by open sets U_α on M and a set of charts $\varphi_\alpha : U_\alpha \to V_\alpha$ such that

$$M = \bigcup_\alpha \varphi_\alpha(U_\alpha)$$

In other words, the charts form an open cover of M.

Definition 1.10. Let $\varphi_\alpha : U_\alpha \to V_\alpha$ and $\varphi_\beta : U_\beta \to V_\beta$ be charts on the n-manifold M. Suppose that $U_\alpha \cap U_\beta \neq \emptyset$. Then we say that the function

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$$

is a **transition map** or a **gluing map**, which maps from a set in $\mathbb{R}^n$ to the manifold and then back to another (or the same, if $\varphi_\alpha = \varphi_\beta$) set in $\mathbb{R}^n$.

We may think of an atlas as a union of charts glued together with gluing maps.

Remarks 1.11. Since φ_α^{-1} and φ_β are homeomorphisms, then the transition map is also a homeomorphism. So we can change from the local coordinates of one local coordinate system to the local coordinates of another local coordinate system continuously. A smooth function $f : M \to \mathbb{R}$ can be expressed in a certain way with respect to a local coordinate system by implicitly applying a transition map to it.

This paper will only discuss the case of general manifolds, but there are also other types of manifolds which are important in Morse Theory. Some of them are defined below.

Definition 1.12. A manifold M with an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is a **smooth manifold** if every transitions map on $\mathcal{A}$ is a diffeomorphism. On a smooth manifold, the atlas $\mathcal{A}$ is called a **smooth atlas** of M, since all charts in $\mathcal{A}$ are smooth.

Definition 1.13. A manifold M with an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha)\}$ is an n**-dimensional manifold with boundary** if every open set U_α is homeomorphic to an open set on the *upper half-space* $\mathbb{R}^{n^+}$, where

$$\mathbb{R}^{n^+} = \{(x_1, \ldots, x_n) \subseteq \mathbb{R}^n : x_n \geq 0\}$$

It is conventional to say that an n-manifold is a manifold without boundary unless specified.

Definition 1.14. A manifold M as a topological space is **compact** if every infinite open cover of M has a finite subcover. In other words, for every infinite collection of open sets $\{U_\alpha\}_{\alpha \in I}$ where $|I| = \infty$ and

$$M = \bigcup_{\alpha \in I} U_\alpha$$

there is a finite collection of open sets $\{U_\alpha\}_{\alpha \in J}$ where $|J| < \infty$ and

$$M = \bigcup_{\alpha \in J} U_\alpha$$

2. Basics of Morse Theory

For this section, let M be an n-manifold and $f : M \to \mathbb{R}$ be a smooth function.

Definition 2.1. If we choose local coordinates $(x_1, \ldots, x_n)$ centered at a point p_0 on M, then p_0 is a **critical point** of f if all of the first-order derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are zero when evaluated at p_0. That is,

$$\frac{\partial f}{\partial x_1}(p_0) = \ldots = \frac{\partial f}{\partial x_n}(p_0) = 0$$

In this case, we say that the number $f(p_0) \in \mathbb{R}$ is a **critical value** of f.

Definition 2.2. Let $(x_1, \ldots, x_n)$ be local coordinates centered at a point p. Then the **Hessian** of the function f at p, denoted as $H_f(p)$, is the $n \times n$ matrix of second-order derivatives

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

evaluated at the point p.

The Hessian is a useful tool for defining characteristics on critical points, such as degeneracy.

Definition 2.3. Let $p_0 \in M$ be a critical point on a function $f : M \to \mathbb{R}$ which is C^2 at p_0. p_0 is a **nondegenerate critical point** if the determinant of its Hessian $det(H_f(p_0)) \neq 0$. On the other hand, p_0 is a **degenerate critical point** if $det(H_f(p_0)) = 0$.

Remark 2.4. We previously defined the Hessian in such a way that the Hessian of f at a point p depends on the local coordinates chosen at p, and so it may seem that the degeneracy (that is, the characteristic of being degenerate or non-degenerate) of a function at a point would also depend upon the choice of local coordinates. However, the opposite is actually true. We will verify this statement in the discussion below.

Definition 2.5. Let $(x_1, \dots, x_n)$ and $(X_1, \dots, X_n)$ be local coordinate systems around a point $p \in M$. Then the **Jacobian matrix** at p, denoted as $J(p)$, is the $n \times n$ matrix of the coordinate transformation from one local coordinate system to the other

$$J = \begin{pmatrix} \frac{\partial x_1}{\partial X_1} & \frac{\partial x_1}{\partial X_2} & \cdots & \frac{\partial x_1}{\partial X_n} \\ \frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} & \cdots & \frac{\partial x_2}{\partial X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial X_1} & \frac{\partial x_n}{\partial X_2} & \cdots & \frac{\partial x_n}{\partial X_n} \end{pmatrix}$$

evaluated at the point p.

Definition 2.6. Let A be an $l \times k$ matrix, $A = (\alpha_{ij})_{k \times l}$. Then the **transpose** of A is the matrix $A^T = (\alpha_{ji})_{l \times k}$. Moreover, if $l = k$ and $A = A^T$, then A is a **symmetric** matrix.

Definition 2.7. A function $f : M \to \mathbb{R}$ on an n-manifold M is a **Morse function** if every critical point of f is nondegenerate.

Observe that on a smooth, real-valued function $f : M \to \mathbb{R}$ on an n-manifold M, the Hessian $H_f(p_0)$ is symmetric, because $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ for every $i, j \in [1, n]$.

We now have the necessary knowledge to establish the following lemma.

Lemma 2.8. *Let f be a real-valued, smooth function defined on the n-manifold M. Let $(x_1, \dots, x_n)$ and $(X_1, \dots, X_n)$ be two coordinate systems at a critical point p_0 of f with Hessians $H_f(p_0)$ and $\mathcal{H}_f(p_0)$, respectively. Then,*

$$H_f(p_0) = J^T(p_0)\mathcal{H}_f(p_0)J(p_0). \tag{2.9}$$

The proof is a rather complicated calculation of the left and right-hand sides of the statement after applying the chain rule. To avoid unnecessary tediousness, only the general outline of the proof is stated below.

Proof. In order to abbreviate notation, we will only consider the case when $n = 2$. The steps of the proof are the same for any $n \geq 2$. Let p_0 be a critical point of the smooth function $f : M \to \mathbb{R}$, where M is a two-dimensional manifold. In the case of $n = 2$, equation 2.1 translates to

$$\begin{pmatrix} \frac{\partial f}{\partial x_1^2}(p_0) & \frac{\partial f}{\partial x_1 \partial x_2}(p_0) \\ \frac{\partial f}{\partial x_2 \partial x_1}(p_0) & \frac{\partial f}{\partial x_2^2}(p_0) \end{pmatrix}$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial X_1}(p_0) & \frac{\partial x_2}{\partial X_1}(p_0) \\ \frac{\partial x_1}{\partial X_2}(p_0) & \frac{\partial x_2}{\partial X_2}(p_0) \end{bmatrix} \times \begin{bmatrix} \frac{\partial f}{\partial X_1^2}(p_0) & \frac{\partial f}{\partial X_1 \partial X_2}(p_0) \\ \frac{\partial f}{\partial X_2 \partial X_1}(p_0) & \frac{\partial f}{\partial X_2^2}(p_0) \end{bmatrix} \times \begin{bmatrix} \frac{\partial x_1}{\partial X_1}(p_0) & \frac{\partial x_1}{\partial X_2}(p_0) \\ \frac{\partial x_2}{\partial X_1}(p_0) & \frac{\partial x_2}{\partial X_2}(p_0) \end{bmatrix}$$

At this point, we may think about using the chain rule we have all learned from high school calculus. We apply it here to coordinate systems (x_1, x_2) and (X_1, X_2) centered at p_0.

$$\frac{\partial f}{\partial X_1} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial X_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial X_1}$$

$$\frac{\partial f}{\partial X_2} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial X_2} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial X_2} \tag{2.10}$$

We may use the chain rule to express $\frac{\partial^2 f}{\partial X_1^2}$, $\frac{\partial^2 f}{\partial X_1 \partial X_2}$, and $\frac{\partial^2 f}{\partial X_2^2}$ in terms of $\frac{\partial^2 f}{\partial x_1^2}$, $\frac{\partial^2 f}{\partial x_1 \partial x_2}$, and $\frac{\partial^2 f}{\partial x_2^2}$. Since p_0 is a critical point, we observe that $\frac{\partial f}{\partial x_1}(p_0) = \frac{\partial f}{\partial x_2}(p_0) = 0$. Then, we use this fact to establish the equivalence in equation 2.1, and we are done. □

Corollary 2.11. *The degeneracy of a critical point p_0 is independent of the choice of local coordinates.*

Proof. This follows directly from the lemma established above, because

$$H_f(p_0) = J^T(p_0)\mathcal{H}_f(p_0)J(p_0)$$
$$\implies det(H_f(p_0)) = det(J^T(p_0))det(\mathcal{H}_f(p_0))det(J(p_0))$$

Observe that

$$J(p_0)J^{-1}(p_0) = Id$$

where Id is the identity map. Therefore,

$$det[J(p_0)J^{-1}(p_0)] = det(Id) = 1.$$

The determinant is multiplicative, and so

$$det(J(p_0))det(J^{-1}(p_0)) = det(Id) = 1$$

Therefore, $det(J(p_0)) \neq 0$. Since the Jacobian determinant is always nonzero, then it follows that $det(H_f(p_0))$ is zero if and only if $det(\mathcal{H}_f(p_0))$ is zero, and vise versa. □

The Hessian is now clearly well-defined.

3. Proof of the Morse Lemma

To establish the Morse Lemma, we first need the following two theorems. One is a fundamental fact of multivariate calculus.

Theorem 3.1. *Let f be a smooth function in a neighborhood N_x of $x = (x_1, \dots, x_n)$ in $\mathbb{R}^n$. Suppose $f(0, \dots, 0) = 0$. Then, there exist n smooth functions $g_i, \dots, g_n$ defined on N_x such that $g_i(0, \dots, 0) = \frac{\partial f}{\partial x_i}(0, \dots, 0)$ for every i, and*

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

Proof. Fix a point $(x_1, \dots, x_n)$ in $\mathbb{R}^n$. Consider the function $f(tx_1, \dots, tx_n)$ with the parameter t. By knowing the chain rule, we observe the following.

$$f(x_1, \dots, x_n) = \int_0^1 \frac{df(tx_1, \dots, tx_n)}{dt} dt = \int_0^1 \sum_{i=1}^{n} x_i \frac{\partial f(tx_1, \dots, tx_n)}{\partial x_i} dt$$

Now, for every i, let

$$g_i(x_1, \dots, x_n) = \int_0^1 \sum_{i=1}^{n} \frac{\partial f(tx_1, \dots, tx_n)}{\partial x_i} dt$$

□

The other prerequisite is what is known as the Inverse Function Theorem. We will omit the proof here but state the theorem below.[1]

Theorem 3.2. *(Inverse Function Theorem)*[2] *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth function on an open set U containing $a \in \mathbb{R}^n$. Suppose that $det J_f(a) \neq 0$.*

Then there is an open set $V \subset \mathbb{R}^n$ containing a and an open set $W \subset \mathbb{R}^n$ containing $f(a)$ such that $f : V \to W$ is a diffeomorphism.

Theorem 3.3. *(Morse Lemma) Let p_0 be a non-degenerate critical point of a smooth function $f : M \to \mathbb{R}$, where M is an n-manifold. Then we can choose a local coordinate system $(x_1, \dots, x_n)$ centered at p_0 such that*

$$f = -x_1^2 - x_2^2 - \dots - x_\lambda^2 + x_\lambda^2 + \dots + x_n^2 + c \tag{3.4}$$

where $c = f(p_0)$ is some constant and λ is the index of f at p_0.

The definition of an index is more formally stated below.

Definition 3.5. Let $f : V \to \mathbb{R}$ be a bilinear map defined on the real vector space V. Then the **index** of f is the maximal dimension of a subspace of V on which H is negative definite.[2]

Finally, we may now go about proving the Morse Lemma. The proof of the Morse Lemma is mostly a proof by induction. The base case for $k = 1$ is given below.

[1]The proof may be found in Michael Spivak's *Calculus on Manifolds.*
[2]Principally adopted from Spivak's *Calculus on Manifolds.*
[2]This definition was adapted from John Milnor's text *Morse Theory*, 1963.

Proof. Let p_0 be a non-degenerate critical point of the function $f : M \to \mathbb{R}$, where M is an n-manifold. As stated before, the degeneracy of the point p_0 on f is determined independent of our choice of a local cooordinate system. Therefore, we may assume that when we pick a local coordinate system $(x_1, \ldots, x_n)$ defined in a neighborhood N_{p_0},

$$\frac{\partial^2 f}{\partial x_1^2}(p_0) \neq 0 \tag{3.6}$$

or that we may pick a suitable linear transformation of the local coordinate system such that equation 3.6 is true. We may further assume that p_0 corresponds to the origin $(0, \ldots, 0) \in \mathbb{R}^n$ on the local coordinate system and that $f(p_0) = 0$, replacing f with $f - f(p_0)$ if necessary.

By Theorem 3.1, there exist n smooth functions $g_i, \ldots, g_n$ defined on N_{p_0} such that

$$g_i(0, \ldots, 0) = \frac{\partial f}{\partial x_i}(0, \ldots, 0) \tag{3.7}$$

and

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i g_i(x_1, \ldots, x_n) \tag{3.8}$$

But since p_0 is a critical point, equation 3.7 turns out to be zero on both sides at p_0. So we can apply Theorem 3.1 again to get n smooth functions $h_{i1}, \ldots, h_{in}$ for every i that is defined on N_{p_0} such that

$$\sum_{j=1}^{n} x_j h_{ij}(x_1, \ldots, x_n) = g_i(x_1, \ldots, x_n). \tag{3.9}$$

By plugging equation 3.9 into equation 3.8, we get

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j h_{ij}(x_1, \ldots, x_n) \tag{3.10}$$

We may assume that $h_{ij} = h_{ji}$, rewriting h_{ij} as $H_{ij} = \frac{h_{ij}+h_{ji}}{2}$ if necessary. Furthermore,

$$(h_{ij}(0, \ldots, 0))_{n \times n} = (\frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(0, \ldots, 0))_{n \times n} \tag{3.11}$$

And since we assumed equation 3.6 to be true, then $h_{11}(0, \ldots, 0) \neq 0$. h_{11} is a smooth, hence continuous function, and so h_{11} is not zero in a neighborhood of the origin. Let us call this neighborhood $\bar{N}_0$.

Our ultimate goal is to express f in the standard quadratic form of equation 3.4. We do this by eliminating all terms which are not of the form $\pm x_i^2$ via induction over $k \leq n$ steps. While we are currently dealing with $k = 1$, in the general case of k, we wish to express f as a sum of terms such that k terms are of the form $\pm x_i^2$ and the rest of the terms depend on coordinates in the set $\{x_i | i \neq k\}$. To this end, let

$$G(x_1, \ldots, x_n) = \sqrt{|h_{11}(x_1, \ldots, x_n)|}. \tag{3.12}$$

G is a smooth, non-zero function of $x_1, \ldots, x_n$ on $\bar{N}_0$.

Now suppose by induction that there exists a local coordinate system $(y_1, \ldots, y_n)$ defined on $\bar{N}_0$ such that

$$y_i = x_i \ (i \neq 1) \tag{3.13}$$

$$y_1 = G * (x_1 + \sum_{i>1}^{n} \frac{x_i h_{1i}}{h_{11}}). \tag{3.14}$$

It follows from the Inverse Function Theorem that $y_1, \ldots, y_n$ is a local coordinate system defined on a smaller neighborhood $\widetilde{N}_0 \subset \bar{N}_0$, since the determinant of the Jacobian of the transformation from $(x_1, \ldots, x_n)$ to $(y_1, \ldots, y_n)$ may be verified to be nonzero.

When we square y_1, we get

$$y_1^2 = \pm h_{11} x_1^2 \pm 2 \sum_{i=2}^{n} x_1 x_i h_{1i} \pm \frac{(\sum_{i=2}^{n} x_i h_{1i})^2}{h_{11}} \tag{3.15}$$

where the signs are either positive if $h_{11} > 0$ or negative if $h_{11} < 0$. Using equation 3.10, we can verify that f can be expressed in the following way with respect to this coordinate system on the restricted domain $\widetilde{N}_0$.

$$f = \pm y_1^2 + \sum_{i=2}^{n} \sum_{j=2}^{n} x_i x_j h_{ij} - \frac{(\sum_{i=2}^{n} x_i h_{1i})^2}{h_{11}} \tag{3.16}$$

where the sign of the y_1^2 term is positive if $h_{11} > 0$ or negative if $h_{11} < 0$. Staying consistent with our goals, we notice that the first term is in the standard quadratic form seen in equation 3.4, whereas the rest of the terms depend on local coordinates x_i whereby $i \neq k$ $(k = 1)$. By induction from $k = 1$ to $k = n$, we prove the Morse Lemma. □

As important consequences,

Corollary 3.17. *nondegenerate critical points are isolated.*

Corollary 3.18. *A Morse function defined on a compact manifold only has finitely many critical points.*

By using a convenient coordinate system, we can see how the real-valued function f behaves on the manifold near a nondegenerate critical point, allowing us to classify an area around a nondegenerate critical points according to the index of the function at that point. We have now proven the Morse Lemma, and we may use our understanding of it to advance our study of Morse Theory as a whole.

Acknowledgments. I greatly acknowledge the assistance I had from my two mentors, Laurie Field and Ben Fehrman, in completing this paper and for introducing Morse Theory and the general field of geometric topology to me. Benson Farb's class on Geometry also great furthered my grasp of geometry and was overall very enlightening. I thank Laszlo Babai for his rigorous teaching style and invested interest in the education of undergraduate math students. I would also like to thank Peter May and other organizers of the VIGRE REU program for making my whole experience here possible and for helping me confirm my interest for mathematics as a whole. Thank you, all.

References

[1] John Milnor. Morse Theory Princeton University Press. 1963.
[2] Michael Spivak. Calculus on Manifolds Benjamin/Cummings Publishing Co. 1965.
[3] Yukio Matsumoto. An Introduction to Morse Theory Iwanani Shoten, Publishers. 1997.