PREFACE

This book gives a present-day account of Marston Morse's theory of the calculus of variations in the large. However, there have been important developments during the past few years which are not mentioned.

Let me describe three of these.

R. Palais and S. Smale have studied Morse theory for a real-valued function on an infinite dimensional manifold and have given direct proofs of the main theorems, without making any use of finite dimensional approximations. The manifolds in question must be locally diffeomorphic to Hilbert space, and the function must satisfy a weak compactness condition. As an example, to study paths on a finite dimensional manifold $M$ one considers the Hilbert manifold consisting of all absolutely continuous paths $\gamma: [0,1] \rightarrow M$ with square integrable first derivatives. Accounts of this work are contained in R. Palais, Morse Theory on Hilbert Manifolds, Topology, Vol. 2 (1963), pp. 299-340; and in S. Smale, Morse Theory and a Non-linear Generalization of the Dirichlet Problem, Annals of Mathematics, Vol. 80 (1964), pp. 382-396.

The Bott periodicity theorems were originally inspired by Morse theory (see part IV). However, more elementary proofs, which do not involve Morse theory at all, have recently been given. See M. Atiyah and R. Bott, On the Periodicity Theorem for Complex Vector Bundles, Acta Mathematica, Vol. 112 (1964), pp. 229-247, as well as R. Wood, Banach Algebras and Bott Periodicity, Topology, 4 (1965-66), pp. 371-389.

Morse theory has provided the inspiration for exciting developments in differential topology by S. Smale, A. Wallace, and others, including a proof of the generalized Poincaré hypothesis in high dimensions. I have tried to describe some of this work in Lectures on the h-cobordism theorem, notes by L. Siebenmann and J. Sondow, Princeton University Press, 1965.

Let me take this opportunity to clarify one term which may cause confusion. In §12 I use the word "energy" for the integral.
along a path \( \omega(t) \). V. Arnol'd points out to me that mathematicians for the past 200 years have called \( E \) the "action" integral. This discrepancy in terminology is caused by the fact that the integral can be interpreted, in terms of a physical model, in more than one way.

Think of a particle \( P \) which moves along a surface \( M \) during the time interval \( 0 \leq t \leq 1 \). The action of the particle during this time interval is defined to be a certain constant times the integral \( E \). If no forces act on \( P \) (except for the constraining forces which hold it within \( M \)), then the "principle of least action" asserts that \( E \) will be minimized within the class of all paths joining \( \omega(0) \) to \( \omega(1) \), or at least that the first variation of \( E \) will be zero. Hence \( P \) must traverse a geodesic.

But a quite different physical model is possible. Think of a rubber band which is stretched between two points of a slippery curved surface. If the band is described parametrically by the equation \( x = \omega(t), 0 \leq t \leq 1 \), then the potential energy arising from tension will be proportional to our integral \( E \) (at least to a first order of approximation). For an equilibrium position this energy must be minimized, and hence the rubber band will describe a geodesic.

The text which follows is identical with that of the first printing except for a few corrections. I am grateful to V. Arnol'd, D. Epstein and W. B. Houston, Jr. for pointing out corrections.

J.W.M.

Los Angeles, June 1968.

CONTENTS

PREFACE

PART I. NON-DEGENERATE SMOOTH FUNCTIONS ON A MANIFOLD

§1. Introduction ..................................... 1
§2. Definitions and Lemmas ............................ 4
§3. Homotopy Type in Terms of Critical Values ........ 12
§4. Examples ........................................ 25
§5. The Morse Inequalities .......................... 28
§7. The Lefschetz Theorem on Hyperplane Sections .... 39

PART II. A RAPID COURSE IN RIEMANNIAN GEOMETRY

§8. Covariant Differentiation ........................ 43
§9. The Curvature Tensor ............................ 51
§10. Geodesics and Completeness ...................... 55

PART III. THE CALCULUS OF VARIATIONS APPLIED TO GEODESICS

§11. The Path Space of a Smooth Manifold ............. 67
§12. The Energy of a Path ............................ 70
§13. The Hessian of the Energy Function at a Critical Path .......................... 74
§14. Jacobi Fields: The Null-space of \( E_{\omega}^{\prime} \) ........ 77
§15. The Index Theorem ................................ 83
§16. A Finite Dimensional Approximation to \( \omega^{C} \) ........ 88
§17. The Topology of the Full Path Space ................ 93
§18. Existence of Non-conjugate Points ................ 98
§19. Some Relations Between Topology and Curvature .... 100
In this section we will illustrate by a specific example the situation that we will investigate later for arbitrary manifolds. Let us consider a torus \( M \), tangent to the plane \( V \), as indicated in Diagram 1.

Let \( f: M \to \mathbb{R} \) (\( \mathbb{R} \) always denotes the real numbers) be the height above the \( V \) plane, and let \( M^a \) be the set of all points \( x \in M \) such that \( f(x) \leq a \). Then the following things are true:

1. If \( a < f(p) \), then \( M^a \) is vacuous.
2. If \( f(p) < a < f(q) \), then \( M^a \) is homeomorphic to a 2-cell.
3. If \( f(q) < a < f(r) \), then \( M^a \) is homeomorphic to a cylinder:
4. If \( f(r) < a < f(s) \), then \( M^a \) is homeomorphic to a compact manifold of genus one having a circle as boundary:
I. NON-DEGENERATE FUNCTIONS

§1. INTRODUCTION

1. NON-DEGENERATE FUNCTIONS

(5) If \( f(s) < a \), then \( M^a \) is the full torus.

In order to describe the change in \( M^a \) as \( a \) passes through one of the points \( f(p), f(q), f(r), f(s) \), it is convenient to consider homotopy type rather than homeomorphism type. In terms of homotopy types:

(1) \(-\) (2) is the operation of attaching a 0-cell. For as far as homotopy type is concerned, the space \( M^a, f(p) < a < f(q) \), cannot be distinguished from a 0-cell:

Here "\(-\)" means "is of the same homotopy type as."

(2) \(-\) (3) is the operation of attaching a 1-cell:

(3) \(-\) (4) is again the operation of attaching a 1-cell:

(4) \(-\) (5) is the operation of attaching a 2-cell.

The precise definition of "attaching a k-cell" can be given as follows. Let \( Y \) be any topological space, and let

\[
\mathcal{S}^k = \{ x \in \mathbb{R}^k : ||x|| = 1 \}
\]

be the k-cell consisting of all vectors in Euclidean k-space with length \( \leq 1 \).

The boundary

\[
\mathcal{S}^{k-1} = \{ x \in \mathbb{R}^k : ||x|| = 1 \}
\]

will be denoted by \( \mathcal{S}^{k-1} \). If \( g: \mathcal{S}^{k-1} \rightarrow Y \) is a continuous map then

\[
Y \cup_{g} \mathcal{S}^k
\]

(Y with a k-cell attached by \( g \)) is obtained by first taking the topological sum (= disjoint union) of \( Y \) and \( \mathcal{S}^k \), and then identifying each \( x \in \mathcal{S}^{k-1} \) with \( g(x) \in Y \). To take care of the case \( k = 0 \) let \( e^0 \) be a point and let \( e^0 \cup Y \) be vacuous, so that \( Y \) with a 0-cell attached is just the union of \( Y \) and a disjoint point.

As one might expect, the points \( p, q, r \) and \( s \) at which the homotopy type of \( M^a \) changes, have a simple characterization in terms of \( f \). They are the critical points of the function. If we choose any coordinate system \( (x, y) \) near these points, then the derivatives \( \frac{\partial f}{\partial x} \) and \( \frac{\partial f}{\partial y} \) are both zero. At \( p \) we can choose \( (x, y) \) so that \( f = x^2 + y^2 \), at \( s \) so that \( f = \text{constant} - x^2 - y^2 \), and at \( q \) and \( r \) so that \( f = \text{constant} + x^2 - y^2 \). Note that the number of minus signs in the expression for \( f \) at each point is the dimension of the cell we must attach to go from \( M^a \) to \( M^b \), where \( a < f(\text{point}) < b \). Our first theorems will generalize these facts for any differentiable function on a manifold.

REFERENCES

For further information on Morse Theory, the following sources are extremely useful.


R. Bott, Morse Theory and its application to homotopy theory, Lecture notes by A. van de Ven (mimeographed), University of Bonn, 1960.
I. NON-DEGENERATE FUNCTIONS

§2. Definitions and Lemmas.

The words "smooth" and "differentiable" will be used interchangeably to mean differentiable of class $C^\infty$. The tangent space of a smooth manifold $M$ at a point $p$ will be denoted by $T_pM$. If $g: M \to N$ is a smooth map with $g(p) = q$, then the induced linear map of tangent spaces will be denoted by $D_qg: T_pM \to T_qN$.

Now let $f$ be a smooth real valued function on a manifold $M$. A point $p \in M$ is called a critical point of $f$ if the induced map $f_*: T_pM \to T_{f(p)}N$ is zero. If we choose a local coordinate system $(x^1, ..., x^n)$ in a neighborhood $U$ of $p$ this means that

$$\frac{\partial f}{\partial x^i}(p) = 0.$$

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

We denote by $\mathcal{M}^a$ the set of all points $x \in M$ such that $f(x) < a$. If $a$ is not a critical value of $f$ then it follows from the implicit function theorem that $\mathcal{M}^a$ is a smooth manifold-with-boundary. The boundary $f^{-1}(a)$ is a smooth submanifold of $M$.

A critical point $p$ is called non-degenerate if and only if the matrix

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

is non-singular. It can be checked directly that non-degeneracy does not depend on the coordinate system. This will follow also from the following intrinsic definition.

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

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

where $\tilde{v}_p$ is, of course, just $v$. We must show that this is symmetric and well-defined. It is symmetric because

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

where $[\tilde{v}, \tilde{w}]$ is the Poisson bracket of $\tilde{v}$ and $\tilde{w}$, and where $[\tilde{v}, \tilde{w}]_p(f) = 0$.

Therefore we can let $g_1(x_1, ..., x_n) = \int_0^1 \frac{\partial f}{\partial x_1}(tx_1, ..., tx_n) \cdot x_1 \, dt$.

§2. DEFINITIONS AND LEMMAS
LEMMA 2.2 (Lemma of Morse). Let \( p \) be a non-degenerate critical point for \( f \). Then there is a local coordinate system \( (y^1, \ldots, y^N) \) in a neighborhood \( U \) of \( p \) with \( y^i(p) = 0 \) for all \( i \) and such that the identity
\[
f = f(p) - \left( y^1 \right)^2 - \ldots - \left( y^k \right)^2 + \left( y^{k+1} \right)^2 + \ldots + \left( y^N \right)^2
\]
holds throughout \( U \), where \( k \) is the index of \( f \) at \( p \).

PROOF: We first show that if there is any such expression for \( f \), then \( k \) must be the index of \( f \) at \( p \). For any coordinate system \( (z^1, \ldots, z^N) \), if
\[
f(q) = f(p) - \left( z^1(q) \right)^2 - \ldots - \left( z^k(q) \right)^2 + \left( z^{k+1}(q) \right)^2 + \ldots + \left( z^N(q) \right)^2
\]then we have
\[
\frac{2}{\partial z^i \partial z^j}(p) = \begin{cases} 
-2 & \text{if } i = j \leq k, \\
2 & \text{if } i = j > k, \\
0 & \text{otherwise,}
\end{cases}
\]
which shows that the matrix representing \( f_{**} \) with respect to the basis \( \frac{\partial}{\partial z^i} \big| p, \ldots, \frac{\partial}{\partial z^N} \big| p \) is
\[
\begin{pmatrix}
-2 & 2 & \ldots & 2 \\
2 & -2 & \ldots & 2 \\
\vdots & \vdots & \ddots & \vdots \\
2 & 2 & \ldots & -2
\end{pmatrix}
\]
Therefore there is a subspace of \( T_p \mathbb{R}^N \) of dimension \( k \) where \( f_{**} \) is negative definite, and a subspace \( V \) of dimension \( n-k \) where \( f_{**} \) is positive definite. If there were a subspace of \( T_p \mathbb{R}^N \) of dimension greater than \( k \) on which \( f_{**} \) were negative definite then this subspace would intersect \( V \), which is clearly impossible. Therefore \( k \) is the index of \( f_{**} \).

We now show that a suitable coordinate system \( (y^1, \ldots, y^N) \) exists. Obviously we can assume that \( p \) is the origin of \( \mathbb{R}^N \) and that \( f(p) = 0 \). By 2.1 we can write
\[
f(x_1, \ldots, x_N) = \sum_{j=1}^{N} x_j g_j(x_1, \ldots, x_N)
\]
for \( (x_1, \ldots, x_N) \) in some neighborhood of \( 0 \). Since \( 0 \) is assumed to be a critical point:
\[
g_j(0) = \frac{\partial}{\partial x^j}(0) = 0.
\]
Therefore, applying 2.1 to the \( g_j \) we have
\[
g_j(x_1, \ldots, x_N) = \sum_{i=1}^{N} x_i h_{ij}(x_1, \ldots, x_N)
\]
for certain smooth functions \( h_{ij} \). It follows that
\[
f(x_1, \ldots, x_N) = \sum_{i, j=1}^{N} x_i x_j h_{ij}(x_1, \ldots, x_N).
\]
We can assume that \( h_{jj} = h_{jj}^* \) since we can write \( f_{jj} = \frac{1}{2}(h_{jj}^* + h_{jj}) \), and then have \( f_{jj} = f_{jj}^* \) and \( f = \sum x_i x_j h_{ij}^* \). Moreover the matrix \( (h_{ij}(0)) \) is equal to \( (\frac{1}{2} \frac{\partial^2 f}{\partial x^i \partial x^j}(0)) \), and hence is non-singular.

There is a non-singular transformation of the coordinate functions which gives us the desired expression for \( f \), in a perhaps smaller neighborhood of \( 0 \). To see this we just imitate the usual diagonalization proof for quadratic forms. (See for example, Birckhoff and MacLane, "A survey of modern algebra," p. 271.) The key step can be described as follows.

Suppose by induction that there exist coordinates \( u_1, \ldots, u_n \) in a neighborhood \( U_1 \) of \( 0 \) so that
\[
f = \sum (u_i)^2 + \sum_{1 \leq j < r} u_i u_j h_{ij}(u_1, \ldots, u_n)
\]
throughout \( U_1 \) where the matrices \( (h_{ij}(u_1, \ldots, u_n)) \) are symmetric. After a linear change in the last \( n-r \) coordinates we may assume that \( H_{rr}(0) \neq 0 \).

Let \( g(u_1, \ldots, u_n) \) denote the square root of \( |H_{rr}(u_1, \ldots, u_n)| \). This will be a smooth, non-zero function of \( u_1, \ldots, u_n \) throughout some smaller neighborhood \( U_2 \subset U_1 \) of \( 0 \). Now introduce new coordinates \( v_1, \ldots, v_n \) by
\[
v_i = u_i \quad \text{for } i \neq r
\]
and
\[
v_r(u_1, \ldots, u_n) = g(u_1, \ldots, u_n) \left[ u_r + \sum_{1 \leq j < r} u_i H_{ij}(u_1, \ldots, u_n) / H_{rr}(u_1, \ldots, u_n) \right].
\]
It follows from the inverse function theorem that \( v_1, \ldots, v_n \) will serve as coordinate functions within some sufficiently small neighborhood \( U_3 \) of \( 0 \). It is easily verified that \( f \) can be expressed as
\[
f = \sum_{1 \leq i \leq r} (v_i)^2 + \sum_{1 \leq i, j > r} v_i v_j h_{ij}^*(v_1, \ldots, v_n).
\]
I. NON-DEGENERATE FUNCTIONS

throughout $U_3$. This completes the induction; and proves Lemma 2.2.

COROLLARY 2.3 Non-degenerate critical points are isolated.

Examples of degenerate critical points (for functions on $\mathbb{R}$ and $\mathbb{R}^2$) are given below, together with pictures of their graphs.

(a) $f(x) = x^3$. The origin is a degenerate critical point.

(b) $F(x) = e^{-1/x^2} \sin^2(1/x)$. The origin is a degenerate, and non-isolated, critical point.

(c) $f(x,y) = x^3 - 3xy^2$. Real part of $(x + iy)^3$. $(0,0)$ is a degenerate critical point (a "monkey saddle").

(d) $f(x,y) = x^2$. The set of critical points, all of which are degenerate, is the $x$ axis, which is a sub-manifold of $\mathbb{R}^2$.

(e) $f(x,y) = x^2y^2$. The set of critical points, all of which are degenerate, consists of the union of the $x$ and $y$ axes, which is not even a sub-manifold of $\mathbb{R}^2$.

We conclude this section with a discussion of 1-parameter groups of diffeomorphisms. The reader is referred to K. Nomizu, "Lie Groups and Differential Geometry," for more details.

A 1-parameter group of diffeomorphisms of a manifold $M$ is a $C^\infty$ map

$\phi: \mathbb{R} \times M \rightarrow M$
such that

1) for each \( t \in \mathbb{R} \) the map \( \varphi_t : M \to M \) defined by 
\[
\varphi_t(q) = \varphi(t,q)
\]
is a diffeomorphism of \( M \) onto itself,

2) for all \( t,s \in \mathbb{R} \) we have \( \varphi_{t+s} = \varphi_t \circ \varphi_s \)

Given a 1-parameter group \( \varphi \) of diffeomorphisms of \( M \) we define a vector field \( X \) on \( M \) as follows. For every smooth real valued function \( f \) let
\[
X_q(f) = \lim_{h \to 0} \frac{f(\varphi_h(q)) - f(q)}{h}.
\]
This vector field \( X \) is said to generate the group \( \varphi \).

**Lemma 2.4.** A smooth vector field on \( M \) which vanishes outside of a compact set \( K \subset M \) generates a unique 1-parameter group of diffeomorphisms of \( M \).

**Proof:** Given any smooth curve \( t \to c(t) \in M \)
it is convenient to define the velocity vector
\[
\frac{dc}{dt} \in T_{c(t)}M
\]
by the identity
\[
\frac{dc}{dt}(f) = \lim_{h \to 0} \frac{f(c(t+h)) - f(c(t))}{h}.
\]
(Compare §8.) Now let \( \varphi \) be a 1-parameter group of diffeomorphisms, generated by the vector field \( X \). Then for each fixed \( q \) the curve
\[
t \to \varphi_t(q)
\]
satisfies the differential equation
\[
\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)},
\]
with initial condition \( \varphi_0(q) = q \). This is true since
\[
\frac{d\varphi_t(q)}{dt}(f) = \lim_{h \to 0} \frac{f(\varphi_{t+h}(q)) - f(\varphi_t(q))}{h} = \lim_{h \to 0} \frac{f(\varphi_h(p)) - f(p)}{h} = X_p(f),
\]
where \( p = \varphi_0(q) \). But it is well known that such a differential equation, locally, has a unique solution which depends smoothly on the initial condition. (Compare Graves, "The Theory of Functions of Real Variables," p. 166.

Note that, in terms of local coordinates \( u^1, \ldots, u^n \), the differential equation takes on the more familiar form:
\[
\frac{du^i}{dt} = X^i(u^1, \ldots, u^n),
\]
with \( i = 1, \ldots, n \).