INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
RICE UNIVERSITY

Generalized Billiard Paths and Morse Theory for Manifolds with Corners

by

David Gerard Christian Handron

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Robin Forman, Chair
Associate Professor of Mathematics

Robert Hardt, W. L. Moody Professor of Mathematics

Anthony Chan, Assistant Professor of Space Physics and Astronomy

Houston, Texas

May, 1999
UMI Number: 9928542

UMI Microform 9928542
Copyright 1999, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103
Abstract

Generalized Billiard Paths and Morse Theory for Manifolds with Corners

by

David Ccrard Christian Handron

A billiard path on a manifold $M$ embedded in Euclidean space is a series of line segments connecting reflection points on $M$. In a generalized billiard path we also allow the path to pass through $M$. The two segments at a 'reflection' point either form a straight angle, or an angle whose bisector is normal to $M$. Our goal is to estimate the number of generalized billiard paths connecting fixed points with a given number of reflections.

We begin by broadening our point of view and allowing line segments that connect any sequence of points on $M$. Since a path is determined by its 'reflection' points, the length of a path with $k$ reflections may be thought of as a function on $M^k$. Generalized billiard paths correspond to critical points of this length function. The length function is not smooth on $M^k$, having singularities along some of its diagonals. Following the procedure of Fulton and MacPherson we may blow up $M^k$ to obtain a compact manifold with corners to which the length function extends smoothly.

We then develop a version of Morse theory for manifolds with corners and use it to study the length function. There are already versions of Morse theory that may
be used in this case, but ours is a generalization of the work of Braess, retaining both a global 'gradient' flow and the intrinsic stratification of a manifold with corners.

We find that the number of generalized billiard paths with \( k \) reflections connecting two points in \( \mathbb{R}^N \) can be estimated in terms of the homology of the Fulton-MacPherson blow up. Proceeding further, we find lower bounds for the number of such paths directly in terms of the homology of \( M \). In part, we show the number of these paths is at least

\[
\sum_{j=0}^{n-1} \sum_{i_1 + \ldots + i_k = j} b_{i_1}(M) \ldots b_{i_k}(M).
\]
Acknowledgments

I would like to thank my Mother and Father for always encouraging me to follow my academic interests. I also thank my wife, Kerry, for her constant support and encouragement.

I would also like to thank my teachers, at Rice and elsewhere, from whom I have learned so much. I am grateful as well to my Math 211 students. I hope they learned as much from me as I did from them.

Maxine Turner, Marie Magee and Sharon McDonough have simplified many tasks for me during my time at Rice. I also thank the friends I have made here, for giving me fond memories to take away with me. I especially appreciate the efforts of those good friends who have been looking out for me these last few years.

I appreciate the thoughtful suggestions that Robert Hardt and Anthony Chan have given me. Sandy Ganzell also deserves recognition for his proofreading efforts. I owe the greatest debt of gratitude to my advisor, Robin Forman, for all of his efforts on my behalf.

This thesis is dedicated to the memory of Ann Dower Handron, 1901-1999.
Contents

Abstract ii
Acknowledgments iv
List of Illustrations vii

1 Introduction 1
1.1 Statement of the Problem ............................... 1
1.2 A Brief History of Morse Theory .......................... 6
1.3 Manifolds with Boundary ................................. 11

2 Morse Theory on Manifolds with Corners 15
2.1 Manifolds with Corners .................................. 15
2.2 Stratified Spaces ......................................... 19
2.3 Morse Functions on Manifolds with Corners .......... 22
2.4 Modifying the Gradient Vector Field ................... 23
2.5 The Modified Gradient Vector Field Induces a Continuous Flow ... 27
2.6 Essential Critical Points ................................. 33
2.7 The Morse Theorems ..................................... 36
2.8 The Morse Inequalities .................................. 43

3 Generalized Billiard Paths 46
3.1 Statement of the Problem ................................ 46
3.2 The Length of a Path .................................... 47
3.3 The Blow Up Space ...................................... 50
3.4 When is $-L_k$ a Morse function? ....................... 55
3.5 Application of the Morse Inequalities 71

4 Morse Theory for Some Other Stratified Spaces 72
  4.1 $M \times \cdots \times M$ as a Stratified Space 72
  4.2 A Continuous Flow on $M^k$ 75
  4.3 The Morse Theorems 76

5 Billiard Paths Revisited 82
  5.1 A Comparison of Essential Critical Points in $X_k$ and $M^k$ 82
  5.2 Application of the Morse Inequalities 83

Bibliography 85
Illustrations

1.1 Light reflecting off a half-silvered surface. ........................................ 2
1.2 A path with five reflections connecting $p$ to $q$. ............................. 3
1.3 The four critical points of the height function on a torus. .................. 7
1.4 The descending cells of the height function on a torus. ....................... 8
1.5 Changes in topology as a level set crosses a critical point. ................. 9
1.6 A Morse-Smale function on a torus. ............................................... 11
1.7 The height function on a sphere with a disk removed. ......................... 12
1.8 A non-essential critical point on a sphere with a disk removed. ........... 13
1.9 An essential critical point on a sphere with a disk removed. .......... 13

2.1 An example of a manifold with corners. ......................................... 16
2.2 The spaces $\mathbb{H}_0^2$, $\mathbb{H}_1^2$, and $\mathbb{H}_2^2$. ......................... 18
2.3 Some inward pointing vectors and outward pointing vectors. ............... 19
2.4 $M = [0, 1] \times [0, 1]$ as a union of strata. ................................... 21
2.5 A function that is not a Morse function, and one that is. ................... 23
2.6 A vector field that does not produce a flow $M \times [0, \infty) \to M$. .... 24
2.7 Another vector field that does not produce a continuous flow. .......... 25
2.8 A vector field will produce a continuous flow. ................................. 25
2.9 Flow lines for one function that does not and one function that does
    satisfy Property (3). ............................................................... 28
2.10 Here $r$ and $\tau_0$ have been chosen so that $\mu \tau_0 = R - r$. ........... 29
2.11 The modified gradient flow near an uptick. .................................... 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.12</td>
<td>An essential critical point and a non-essential critical point.</td>
<td>34</td>
</tr>
<tr>
<td>2.13</td>
<td>A homotopy from $M_b$ to $M_a$.</td>
<td>37</td>
</tr>
<tr>
<td>2.14</td>
<td>The local model near a critical point.</td>
<td>39</td>
</tr>
<tr>
<td>2.15</td>
<td>The region $H$.</td>
<td>40</td>
</tr>
<tr>
<td>3.1</td>
<td>A path with five reflections connecting $p$ to $q$.</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>The effect of moving $\beta$ away from $\alpha$.</td>
<td>50</td>
</tr>
<tr>
<td>3.3</td>
<td>The point $\delta \in \Delta$ is blown up to the two points $\delta'$ and $\delta''$.</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>The infinitesimal tangent space diagram for two consecutive reflections colliding in $M$.</td>
<td>51</td>
</tr>
<tr>
<td>3.5</td>
<td>Possible infinitesimal tangent space diagrams for three consecutive reflections colliding in $M$.</td>
<td>53</td>
</tr>
<tr>
<td>3.6</td>
<td>The angle $\psi_{z_k}$.</td>
<td>59</td>
</tr>
<tr>
<td>3.7</td>
<td>The three angles $\sigma, \psi_{z_k}$, and $\theta_{z_k}$.</td>
<td>62</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Statement of the Problem

Imagine that we have chosen a 2-manifold, and had a glassblower produce a model of it. Suppose that, having obtained such a model, we were to have it half-silvered. Then any time a beam of light struck the surface, half of the light would reflect off the surface and half would pass through.

If such a model were placed on a shelf next to a lamp, it would produce a pattern of light and dark patches on the wall behind it. The intensity of the light at a particular point on the wall is a result of two factors: first, the number of paths the light may travel from the light source to the point in question, and second, the attenuation of the light along this path. (The intensity is reduced by 50% each time it hits the surface.)

The goal of this chapter is to study these issues. Given two points in the vicinity of such a model, how many paths may a beam of light travel connecting one point to the other, with a given number of reflections.

Figure 1.1 shows such a surface, and a few possible paths connecting two points. The beam of light may travel straight from \( p \) to \( q \), passing through the surface on its way. It may also reflect off the point \( \alpha \), or make two reflections, at \( \beta \) and \( \gamma \).

All of these paths have the property that wherever a reflection occurs, the angle of incidence is equal to the angle of reflection. This can be stated equivalently by saying that the bisector of the angle is perpendicular to the surface. How can we go about looking for these paths?
Before we try to answer that question, let's look at the situation in a little more generality. Let $M \hookrightarrow \mathbb{R}^N$ be a smooth $n$-manifold embedded in Euclidean space of dimension $N$. Between consecutive reflection points, a beam of light will travel in a straight line. Therefore, to describe an entire path, it is sufficient to identify the reflection points and the order in which they occur. This being so, it makes sense to define a path connecting $p \in \mathbb{R}^N$ to $q \in \mathbb{R}^N$ to be a sequence of points $\alpha_1, \ldots, \alpha_k \in M$. Note that this definition allows the line segment $\overline{\alpha_i \alpha_{i+1}}$ to intersect the manifold.

The path in Figure 1.2 is clearly not a path that a beam of light is likely to follow. We need some way of characterizing these paths. This can be taken care of by the following

**Definition 3.1** A path $P = \{\alpha_1, \ldots, \alpha_k\}$ connecting $p = \alpha_0$ to $q = \alpha_{k+1}$ is a generalized billiard path with $k$-reflections if for each $i$ one of the following is true:

1. The bisector of $\angle \alpha_{i-1} \alpha_i \alpha_{i+1}$ is normal to $T_{\alpha_i}M$. 
Figure 1.2 A path with five reflections connecting $p$ to $q$. Here the embedded manifold is $S^1 \hookrightarrow \mathbb{R}^2$.

2. $\angle \alpha_{i-1}\alpha_i\alpha_{i+1}$ is a straight angle.

If $M$ happens to be a convex hypersurface, then this definition reduces to the usual notion of a billiard path. The task at hand now may be thought of as counting generalized billiard paths.

Our grand strategy for solving this problem is as follows: (1) Choose a function on the space of paths whose critical points are precisely the generalized billiard paths. (2) Use Morse theory to get a lower bound for the number of critical points. We will run into a few pitfalls, however, as we proceed.

The space of paths can be thought of as the product $M^k = M \times \cdots \times M$. We can define the length of a path $P = \{\alpha_1, \ldots, \alpha_k\}$ to be the sum of the Euclidean distances between consecutive reflections:

$$L_k(P) = \sum_{i=0}^{k} d_{\text{Eucl}}(\alpha_i, \alpha_{i+1}).$$

This function will be central to our arguments, because of
Lemma 3.1. A path \( P = \{ \alpha_1, \ldots, \alpha_k \} \) with \( \alpha_i \neq \alpha_{i+1} \) for \( 0 \leq i \leq k \) satisfies

\[ \nabla L_k(P) = 0 \]

if and only if it is a generalized billiard path.

Unfortunately, the function \( L_k \) has a serious drawback. Wherever consecutive reflections coincide, \( L_k \) has a singularity that looks like \( |x - y| \).

In Section 3.3 we describe how to 'blow up' \( M^k \). The blow up we use was developed by Fulton and MacPherson ([FM]), and allows us to remove from \( M^k \) the diagonals \( \{ \alpha_i - \alpha_{i+1} \} \) that are causing difficulty and replace them with something that is easier for us to deal with. The result is a manifold with corners, \( X_k \) (Definition 2.2)(see also [AS] and [BT]).

Chapter 2 is devoted to developing a version of Morse theory for use with manifolds with corners. The treatment of Morse theory in [Mi2] relies heavily on the gradient flow of a function, but on a manifold with corners the gradient vector field may not induce a flow. Section 2.4 and Section 2.5 show that a modification to the gradient vector field allows a continuous flow to be defined. An essential critical point is then defined to be a stationary point of that flow. In Section 2.7, we prove the Morse theorems in this setting:

Theorem 2.1. Let \( f : M \to \mathbb{R} \) be a Morse function on a manifold with corners \( M \). If \( a < b \) and \( f^{-1}([a,b]) \) contains no essential critical points, then \( M_a \) is a deformation retract of \( M_b \), so the inclusion map \( M_a \hookrightarrow M_b \) is a homotopy equivalence.

Theorem 2.2. Let \( f : M \to \mathbb{R} \) be a Morse function on a manifold with corners \( M \). Let \( p \) be an essential critical point with index \( \lambda \). Set \( f(p) = c \). Suppose that, for some \( \epsilon > 0 \), \( f^{-1}([c-\epsilon, c+\epsilon]) \) contains no essential critical points other than \( p \). Then \( M_{c+\epsilon} \) is homotopy equivalent to \( M_{c-\epsilon} \) with a \( \lambda \)-cell attached.

These theorems imply the Morse Inequalities, which we will use to deduce the existence of generalized billiard paths.
In Section 3.4, we show that with some restrictions on the embedding $M \hookrightarrow \mathbb{R}^N$, most choices of endpoints $p$ and $q$ result in a length function that satisfies the definition of a Morse function (Definitions 2.5 and 2.7):

**Lemma 3.11** For an acceptable embedding $M \hookrightarrow \mathbb{R}^N$ the set of pairs $(p, q) \in \mathbb{R}^N \times \mathbb{R}^N$ such that $-L_k^{(p, q)}$ is a Morse function is open and dense in $\mathbb{R}^{2N}$.

Section 3.5 applies the results of Chapter 2 to the function $-L_k$ on $X_k$. The result is given by

**Theorem 3.1** The number of generalized billiard paths with $k$ reflections is at least

$$\sum_{i=0}^{kn} b_i(X_k),$$

where $b_i(X_k)$ denotes the $i^{th}$ Betti number of $X_k$.

Chapters 4 and 5 show how we can get a lower bound for the number of paths in terms of the topology of $M$ itself. We still use the length function, but rather than blowing up the bad set, we use it to define a 'stratification' of $M^k$. This stratification divides up $M^k$ into a union of submanifolds or 'strata' on which $L_k$ can be analyzed. Section 4.2 defines a continuous flow on $M^k$ that respects the behavior of $L_k$ on the various strata.

Again, an essential critical point is defined to be a stationary point of this flow. The Morse Theorems, and hence the Morse Inequalities, hold in this setting as well. Applying the Morse Inequalities, we get

**Theorem 5.1** The number of generalized billiard paths with $k$ reflections connecting two points in the vicinity of a manifold $M$ is at least

$$\sum_{j=0}^{n-1} \sum_{i_1 + \ldots + i_k = j} b_{i_1}(M) \ldots b_{i_k}(M),$$

where $b_j(M)$ is the $j^{th}$ Betti number of $M$. 
1.2 A Brief History of Morse Theory

Most of the information in this section can be found in Bott’s excellent survey article, *Morse Theory Indomitable* ([Bo4]).

The foundations of Morse theory were laid in the 1920’s by Marston Morse ([Mo1]). His original work relates information about the critical points of a smooth function on a smooth manifold to information about the topology of the manifold. This relationship was presented at that time as a collection of inequalities, known as the Morse Inequalities.

The standard example for describing Morse theory is a torus in $\mathbb{R}^3$ standing on end in the $xy$-plane as shown in figure 1.3. The function we investigate is $f(p) = z(p)$. The critical points of this function are the places where the tangent plane is horizontal — one minimum, $p_1$; one maximum, $p_4$; and two saddle points, $p_2$ and $p_3$. These may also be seen in figure 1.3. Each of these critical points is assigned a number $\lambda_{p_i}$, called its index. The index measures the number of directions in which the principal curvature is downward, thus $\lambda_{p_1} = 0$, $\lambda_{p_2} = \lambda_{p_3} = 1$, and $\lambda_{p_4} = 2$. We write $m_i(f)$ for the number of critical points having index $i$. Here $m_0(f) = m_2(f) = 1$ and $m_1(f) = 2$.

The Morse Inequalities compare the numbers $m_i(f)$ to the Betti numbers of the torus. The zeroth Betti number measures the number of connected components of $T$, so $b_0(T) = 1$. The first Betti number $b_1(T) = 2$ indicates that there are two types of non-trivial loops in the torus — around the ‘small’ way, or around the ‘big’ way. The second Betti number of a manifold counts how many types of ‘non-trivial closed 2-cycles’ may be found in the manifold. In our case $b_2(T) = 1$. 
The Morse inequalities may then be written as:

\[ m_0(f) \geq b_0(T) \]
\[ m_1(f) - m_0(f) \geq b_1(T) - b_0(T) \]
\[ m_2(f) - m_1(f) + m_0(f) = b_2(T) - b_1(T) + b_0(T). \]

The equality on the last line reflects the fact that the torus is a 2-dimensional manifold. Note also that in our example, equality holds in all cases. These are sometimes called the Strong Morse Inequalities. Adding to each line the line preceding it yields the Weak Morse Inequalities.

\[ m_0(f) \geq b_0(T) \]
\[ m_1(f) \geq b_1(T) \]
\[ m_2(f) \geq b_2(T). \]

With the help of these inequalities, Morse was able to show that for any Riemann structure on an n-sphere, there are an infinite number of geodesics joining any two
points ([Mo2]). This result generated further interest in critical point theory, and in the decades to follow many people, including Morse himself, revised and extended his initial results. My intention here is to describe a few of the major steps along the way.

By the late 1940's, the gradient flow of the function was coming into the picture more forcefully. Once a Riemannian metric has been chosen, each point in the manifold lies in exactly one gradient flow line, and each such flow line begins and ends at a critical point. Thom noticed that by bundling together all the flow lines having the same initial point, the manifold can be decomposed into a collection of 'descending cells' — one for each critical point ([Th]). The dimension of the cell associated to a critical point is equal to the index of that critical point.

![Figure 1.4](image)

**Figure 1.4** The descending cells of the function $f(p) = z(p)$. The cell corresponding to $p_1$ is just $p_1$ itself. The cell corresponding to $p_2$ is shown by the dark ellipse connecting $p_2$ to $p_1$. The cell corresponding to $p_3$ is the dark inner circle connecting $p_3$ to $p_2$. The descending cell for $p_4$ is the remainder of the torus.
From this perspective, it is easier to investigate visually how the critical points relate to the topology of the manifold. Choose \( a \) and \( b \) so that \( a < f(p_2) < b \) and \( p_2 \) is the only critical point in \( f^{-1}[a, b] \). We obtain \( M_b = f^{-1}(-\infty, b] \) from \( M_a \) by adding the set \( f^{-1}(a, b] \).

![Diagram](image)

**Figure 1.5** As we pass from \( M_a \) to \( M_b \), the topology changes. \( M_b \) may be deformed into the union of \( M_a \) with the descending cell from \( p_2 \).

We can see in figure 1.5 that \( M_b \) may be deformed into

\[
M_a \cup \{\text{descending cell for } p_2\}.
\]

This is referred to as 'adding a handle to \( M_a \)'. Because the index of \( p_2 \) is one, we add a one dimensional handle.

The trouble with this point of view is that this cell decomposition need not be 'nice' in the sense that the boundary of a cell may not be a union of lower dimensional cells. The boundary of the cell associated to \( p_3 \), for instance, is the point \( p_2 \), in the middle of another 1-dimensional cell.

It was Smale who, in 1959, showed how one might ensure that the descending cells provide a 'nice' cell decomposition ([Sm1],[Sm2]). The requirement is that the
'ascending cell' of each critical point intersect transversely with each descending cell it meets. A function that meets this criterion is referred to as a Morse-Smale function. In our example this can be achieved by tipping the torus forward ever so slightly. This is shown in figure 1.6. In fact, the transversality condition may always be met by such a slight perturbation.

Under this transversality condition, the descending cells form a CW-complex. By studying the gradient paths carefully, it is possible to write down the boundary maps of the corresponding chain complex. Computing the homology of this chain complex then yields the homology of the underlying manifold.

This view provides a complete picture of Morse theory for smooth functions on smooth manifolds with isolated critical points. It is the context in which most people (at least most mathematicians) are introduced to Morse theory. This formulation of Morse theory was used by Smale to prove the Poincaré conjecture in dimensions five and higher ([Sm3]). This achievement remains one of the most exciting applications of Morse Theory. (Smale actually used a combination of Morse theory and handlebody theory in his proof. Milnor presented the proof entirely in the language of Morse theory in [Mi1].)

Morse theory has been generalized in a number of different directions. Bott developed a method for dealing with higher dimensional critical sets ([Bo1]). Witten, a physicist, showed how to think about Morse theory in a quantum mechanical way ([Wi]). Floer, Conley and others extended the results of Morse, Thom and Smale to include infinite dimensional spaces ([Co1],[F11],[F12]).

The direction that has the most direct relevance to this work, though, is treating functions on spaces other than smooth manifolds. Some results were available early on concerning manifolds with boundary. Braess, in 1974, published a 'complete' version of Morse theory for manifolds with boundary ([Br]). The most remarkable
achievement in this area, though, is Goresky and MacPherson’s stratified Morse theory. This version of Morse theory applies to a class called Whitney stratified spaces, which include manifolds with boundary and manifolds with corners ([GM]).

In the stratified Morse theory of Goresky and MacPherson, however, the gradient flow does not appear as prominently as it does in other versions. Indeed, the functions on these spaces need not even allow a gradient flow to be defined globally. My intention in this work is to produce a Morse theory for manifolds with corners, a type of stratified space, that retains the point of view developed by Thom and Smale.

1.3 Manifolds with Boundary

The easiest example of a manifold with corners is a manifold with boundary. Let’s see how Morse theory works in this case. Take as an example the round sphere with an open disk removed from the side shown in figure 1.7. The function we will consider
is again the height function. The first task is to decide what it means for a point to be critical in this setting.

![Diagram showing critical points on a sphere with a disk removed](image)

**Figure 1.7** The (potential) critical points of the height function on a sphere with a disk removed. The descending cells for $p_2$ and $p_3$ are also shown.

The points $p_1$ and $p_4$ are critical points of the height function on the sphere. They remain critical points when the disk is removed. We must also consider the points $p_2$ and $p_3$. They are critical points of the height function restricted to the boundary of the manifold. We need to examine how these two points affect the topology.

There is not, strictly speaking, a gradient flow for the height function on this manifold. Such a flow would 'run off the edge' of the manifold. We can imagine, however, that when the flow hits the boundary of the manifold it flows in the boundary until it reaches a point where the negative gradient points back into the interior. It is then possible again to talk about the descending cell of a critical point.
Choose $a$ and $b$ such that $a < f(p_2) < b$ and $p_2$ is the only 'critical' point in the set $f^{-1}[a, b]$. We see in figure 1.8 that $M_b$ can be deformed into $M_a$, so $p_2$ does not contribute anything to the topology.

![Diagram for figure 1.8](image)

**Figure 1.8** $M_b$ can be deformed into $M_a$.

On the other hand, there is a different story when we look at $p_3$. Again, choose $a$ and $b$ such that $a < f(p_3) < b$ and $p_3$ is the only 'critical' point in the set $f^{-1}[a, b]$. We find that in this case $M_b$ cannot be deformed into $M_a$. It can however be deformed into the union of $M_a$ with the descending cell from $p_3$.

![Diagram for figure 1.9](image)

**Figure 1.9** $M_b$ can be deformed into the union of $M_a$ with the descending cell from $p_3$. 
What is the difference between $p_2$ and $p_3$? At $p_2$ the gradient vector $-\nabla f(p_2)$ points into the interior of the manifold. At $p_3$ the gradient vector $-\nabla f(p_3)$ points outward. We will see that this is what makes all the difference. Critical points of the function restricted to the boundary only contribute to the topology if $-\nabla f$ points outward at that point.
Chapter 2

Morse Theory on Manifolds with Corners

In this chapter, we will develop a version of Morse theory on a class of spaces called manifolds with corners. The first three sections consist mainly of definitions and background. The rest of the chapter is devoted to showing how the techniques of classical Morse theory may be extended to these more general spaces.

2.1 Manifolds with Corners

A good example to keep in mind throughout this section is the square, \([0, 1] \times [0, 1]\) pictured in figure 2.1. Here we see that points in the interior have a neighborhood that looks just like a piece of \(\mathbb{R}^2\). A point lying in one of the edges has a neighborhood that looks like the upper half plane. \(H = \mathbb{R} \times [0, \infty)\). A corner of the square has a neighborhood that looks like a quadrant in \(\mathbb{R}^2\). The term 'manifold with corners' draws its inspiration from this last type of point.

Now we will proceed to make this idea explicit. A more thorough (and slightly more general) treatment of manifolds with corners may be found in [MO].

Let \(\mathcal{V} \subset \mathbb{R}^n\) denote an orthonormal collection of vectors in \(\mathbb{R}^n\), \(\mathcal{V} = \{v_1, \ldots, v_j\}\). Define \(\mathbb{H}_\mathcal{V}^n\) to be the set

\[\mathbb{H}_\mathcal{V}^n = \{w \in \mathbb{R}^n : w \cdot v_i \geq 0 \text{ for all } 1 \leq i \leq j\},\]

where \(\cdot\) denotes the standard inner product on \(\mathbb{R}^n\).

Suppose \(U \subset \mathbb{H}_\mathcal{V}^{n} \subset \mathbb{R}^n\) is open in \(\mathbb{H}_\mathcal{V}^{n}\), and \(g : U \to \mathbb{R}^m\). Say that \(g\) is differentiable if there is a neighborhood \(\tilde{U}\) of \(U\) in \(\mathbb{R}^n\) and a differentiable map \(\tilde{g} : \tilde{U} \to \mathbb{R}^m\), such that \(\tilde{g}|_U = g\). In this case, we write \(g \in C^1(U, \mathbb{R}^m)\). More generally, we can make the
Definition 2.1 Suppose $U \subset \mathbb{H}_p \subset \mathbb{R}^n$ is an open set (in $\mathbb{H}_p$), and $g : U \to \mathbb{R}^m$. Say that $g$ is smooth if there is a neighborhood $\bar{U}$ of $U$ in $\mathbb{R}^n$ and a map $\bar{g} : \bar{U} \to \mathbb{R}^m$, such that $\bar{g}$ possesses continuous partial derivatives of all orders and $\bar{g}|_U = g$. In this case, we write $g \in C^\infty(U, \mathbb{R}^m)$.

Let $M$ be a topological space. A coordinate chart on $M$ is a map

$$x : U \to \mathbb{H}_V,$$

for some $V \subset \mathbb{R}^n$, satisfying

1. $U$ is an open subset of $M$.

2. The image $x(U)$ is open in $\mathbb{H}_V$.

3. The map $x$ is a homeomorphism onto its image.
Two coordinate charts, \( x : U_x \to \mathbb{R}^n_x \) and \( y : U_y \to \mathbb{R}^n_y \), are said to be compatible if the transition functions
\[
x \circ y^{-1} : y(U_z \cap U_y) \to x(U_z \cap U_y)
\]
and
\[
y \circ x^{-1} : x(U_z \cap U_y) \to y(U_z \cap U_y)
\]
are both smooth.

A collection of such coordinate charts \( \mathcal{A} \) is called an atlas if the coordinate charts in \( \mathcal{A} \) are pairwise compatible and every \( p \in M \) is in the domain of some coordinate chart \( x \in \mathcal{A} \). Now we are ready to make the

**Definition 2.2**  An \( n \)-dimensional manifold with corners, \( M \), is a topological space together with an atlas, \( \mathcal{A} \), of charts \( x_a : U_a \to \mathbb{R}^n_a \) such that
\[
\bigcup_{a \in \mathcal{A}} U_a = M.
\]

A slightly more general definition of 'manifold with corners' may be found in [MO], but Definition 2.2 is sufficient for our purposes. In practice we will always choose \( \mathcal{V} = \{e_{n-j+1}, \ldots, e_n\} \) for some \( j \). In this case we will write \( \mathbb{R}^n_y = \mathbb{R}^n_f \), and
\[
\mathbb{R}^n_f = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_{n-j+1} \geq 0, \ldots, x_n \geq 0\}.
\]

Call a coordinate chart whose domain is \( \mathbb{R}^n_f \) a standard coordinate chart. We will always assume that a manifold with corners is equipped with a maximal standard atlas \( \mathcal{A} \), i.e. if a standard coordinate chart \( x \) is smoothly compatible with every coordinate chart in \( \mathcal{A} \), then \( x \in \mathcal{A} \). Moreover, \( \mathcal{A} \) contains only standard coordinate charts.

If \( p \in M \), we will say a coordinate chart at \( p \) is a chart \( x_p \in \mathcal{A} \) such that \( x_p(p) = 0 \in \mathbb{R}^n_f \). In this case, the number \( j \) is uniquely determined by the point \( p \). Thus we can write \( j = j(p) \).
Figure 2.2  The spaces $\mathbb{H}^2_0$ (left), $\mathbb{H}^2_1$ (center), and $\mathbb{H}^2_2$ (right).

The tangent space of a manifold with corners can be defined ([MO]) just as is for a smooth manifold ([Sp2]). Consider the set of pairs

$$C_p(M) = \{(x, v) : x \text{ is a coordinate chart at } p \in M \text{ and } v \in \mathbb{R}^n\}.$$ 

The relation $\sim$ is defined by

$$(x, v) \sim (y, w) \iff D(x \circ y^{-1})(y(p))(w) = v.$$ 

This relation satisfies the axioms of an equivalence relation since

1. Since $D(x \circ x^{-1})$ is the identity map, $(x, v) \sim (x, v)$.

2. The Inverse Function Theorem implies $D(x \circ y^{-1})^{-1} = D(y \circ x^{-1})$. Consequently $D(x \circ y^{-1})(y(p))(w) = v$ implies that $D(y \circ x^{-1})(x(p))(v) = w$, so $(x, v) \sim (y, w) \implies (y, w) \sim (x, v)$

3. Transversality follows from the chain rule.

We write $T_pM = C_p(M)/\sim$ and call this the tangent space to $M$ at $p$. Just as in the more familiar case of a manifold with boundary, if $p \in \partial M$ then some of the vectors in $T_pM$ point away from the manifold with corners.
Definition 2.3 A tangent vector in $T_p M$ points outward if some representative $(x, v)$ has $v \notin \mathbb{R}^n_j(p)$. A tangent vector is $T_p M$ points inward (or into $M$) if some representative $(x, v)$ has $v \in \mathbb{R}^n_j(p)$.

Note that the definition of an inward pointing vector includes those vectors which are tangent to the boundary of $M$. These terms are well defined, since for any two coordinate charts at $p$, the transition functions preserve $\mathbb{R}^n_j(p)$.

2.2 Stratified Spaces

There are a number of different (related) notions of what constitutes a stratified space ([We],[GM]). We will not be using any results pertaining any particular theory of stratified spaces, but we will find the language to be convenient. Consequently, we will present a fairly general definition of 'stratified space'.

Definition 2.4 A stratified space consists of a topological space $X$, a partially ordered set $S$ and a collection $\{H_i\}_{i \in S}$ of subspaces of $X$ satisfying

1. Each $H_i$ is a manifold.

2. $X = \cup_{i \in S} H_i$.

![Figure 2.3](image)

Some inward pointing vectors (left) and outward pointing vectors (right).
3. \( H_i \cap \overline{H_j} \neq \emptyset \iff H_i \subseteq \overline{H_j} \iff i \preceq j \). In this case we also write \( H_i \preceq H_j \).

Each of the manifolds \( H_i \) is a \textit{stratum} of \( X \).

It is not difficult to see that a CW-complex is a stratified space. A \( \lambda \)-cell is a manifold with dimension \( \lambda \). The second condition is satisfied, since a CW-complex is the union of its cells. The third condition asserts that the attaching map of a \( \lambda \)-cell \( e^\lambda \) carries \( \partial e^\lambda \) to a union of cells with dimension at most \( \lambda - 1 \).

Another easy example of a stratified space is a manifold with boundary. In this case there are two strata. The interior of the manifold is one stratum and the boundary is the other.

For us, the most important example of a stratified space is a manifold with corners. For a manifold with corners \( M \), let \( \mathcal{E}_j(M) = \{ p \in M : j(p) = j \} \). It is not difficult to see that \( \mathcal{E}_j(M) \) is a manifold of dimension \( n - j \): If \( p \in \mathcal{E}_j(M) \), take a coordinate chart at \( p \), \( x_p : U \to \mathbb{R}^j \) and let \( \pi : \mathbb{R}^n \to \mathbb{R}^{n-j} \) be the projection onto the first \( n - j \) coordinates. Then \( \pi \circ x_p : U \cap \mathcal{E}_j(M) \to \mathbb{R}^{n-j} \) is a coordinate chart for \( \mathcal{E}_j(M) \) at \( p \).

We may think of each connected component of \( \mathcal{E}_j(M) \) as a stratum with dimension \( n - j \). Then each stratum is a manifold. \( M \) is equal to the union of these strata, since every \( p \in M \) is in \( \mathcal{E}_j(M) \) for some \( j \). It takes a little more work to show that condition three is satisfied, but not too much.

Consider a coordinate chart \( x : U \to \mathbb{H}^j \) near a point \( p \) in a component \( H \subset \mathcal{E}_j(M) \). Then the image of \( H \) is
\[
\{(x_1, \ldots, x_j, 0, \ldots, 0)\} \cap x(U).
\]
The image of any component of \( K \subset \mathcal{E}_\ell(M) \) having \( p \) in its closure is of the form
\[
\{(x_1, \ldots, x_j, 0, \ldots, 0, x_{i_1}, 0, \ldots, 0, x_{i_k}, 0, \ldots, 0)\} \cap x(U),
\]
where \( x_{i_1}, \ldots, x_{i_k} > 0 \). If \( q = (x_1, \ldots, x_j, 0, \ldots, 0) \) is another point in the image of \( H \) and \( \epsilon > 0 \) is given, we can choose \( x_{i_1}, \ldots, x_{i_k} \) arbitrarily small to produce a point of \( K \) inside \( B_\epsilon(q) \). So \( H \cap \overline{K} \) is open.

Now suppose that \( p \) is not in the closure of some component \( K' \subset \mathcal{E}_\ell(M) \). The image of any component having non-trivial intersection with \( U \) is of the form

\[
\{ (x_1, \ldots, x_j, 0, \ldots, 0, x_{i_1}, 0, \ldots, 0, x_{i_k}, 0, \ldots, 0) \} \cap \pi(U).
\]

Since each of these has \( p \) in its closure, \( K' \cap U = \emptyset \). It follows, then, that \( H \cap \overline{K} \) is closed. Since \( H \) is connected, either \( H \cap \overline{K} = H \) or \( H \cap \overline{K} = \emptyset \), as required by condition three.

![Diagram](image.png)

**Figure 2.4** \( M = [0, 1] \times [0, 1] \) as a union of strata.

For a concrete example, consider \( M = [0, 1] \times [0, 1] \subset \mathbb{R}^2 \). Let \((x_1, x_2)\) be the usual coordinates on \( \mathbb{R}^2 \). \( M \) can be broken down into nine strata. There is one stratum with dimension two, \( F = (0, 1) \times (0, 1) \). There are four strata with dimension one, \( E_1 = (0, 1) \times \{1\}, E_2 = \{0\} \times (0, 1), E_3 = (0, 1) \times \{0\} \) and \( E_4 = \{1\} \times (0, 1) \). Finally, there are four strata with dimension zero, \( V_1 = \{1\} \times \{0\}, V_2 = \{1\} \times \{1\}, V_3 = \{0\} \times \{1\}, V_4 = \{0\} \times \{0\} \). Figure 2.4 shows the breakdown of \( M \) into its strata.
2.3 Morse Functions on Manifolds with Corners

If $M$ is a manifold with corners, we say that a function $f : M \to \mathbb{R}$ is a smooth function if for each coordinate chart $x : U \to \mathbb{R}^n_j$ the function $f \circ x^{-1} : x(U) \to \mathbb{R}$ can be extended to a smooth function on a neighborhood of $x(U)$ in $\mathbb{R}^n$. This is well defined since the transition functions $x \circ y^{-1}$ are all smooth. If $H$ is a stratum of $M$, and $p \in H$, we say $p$ is a critical point of $f$ whenever $p$ is a critical point of $f|_H$.

If $p \in H$ is in the closure of another stratum, $K$, we can define the generalized tangent space

$$T_pK = \{w \in T_pM : w = \lim_{i \to \infty} v_i \in T_{q_i}K \text{ for some sequence } \{q_i\} \to p\}.$$ 

We may also write this as $T_pK = \lim_{q \to p} T_qK$.

**Definition 2.5** We call the function $f : M \to \mathbb{R}$ a Morse function if it has the following properties:

1. If $p \in H$ is a critical point of $f|_H : H \to \mathbb{R}$, then either
   
   (a) $p$ is a non-degenerate critical point of $f|_H : H \to \mathbb{R}$, i.e. the Hessian has non-zero determinant, or
   
   (b) the vector $-\nabla f(p)$ points into $M$.

2. If $p \in H$ is a critical point, then for any stratum $K \neq H$ with $p$ in the closure of $K$, $df_p$ is not identically zero on $T_pK$.

As an example, consider $f : \mathbb{R}^2 \to \mathbb{R}$ by $F(x_1, x_2) = x_1 + 2x_2$. We may restrict this function to the manifold $M$ from §1.1. The function $f|_M$ has four critical points, at each of the four zero dimensional strata. (A zero dimensional stratum is always a critical point.) Since there are no critical points in the higher dimensional strata,
Figure 2.5 On the left, the height function $h(x_1, x_2) = x_2$. This is not a Morse function. On the right is the Morse function $h(x_1, x_2) = x_1 + 2x_2$.

condition (1) is satisfied. At each critical point, $df(e_1) = 1$ and $df(e_2) = 2$. As a result, condition (2) is satisfied.

On the other hand, the function $f(x_1, x_2) = x_2$ is not a Morse function. The function is constant on the set $(0,1) \times \{0\}$, and so each of these points is a degenerate critical point of $f$ restricted to that stratum. The gradient vector points outward along this entire stratum, though, so the function does not satisfy condition (1a) or (1b).

2.4 Modifying the Gradient Vector Field

In classical Morse theory, a Morse function $f : M \to \mathbb{R}$ is studied by choosing a Riemannian metric on $M$ and examining the flow induced by the vector field $-\nabla f$. When we allow the manifold $M$ to have corners (or even just a boundary) a difficulty arises. If, as seen in figure 2.6, $-\nabla f$ points outward from any point in $\partial M$, the vector field can not produce a flow that carries $M$ to $M$. As a result, we must modify the gradient vector field to produce a new vector field that does induce such a flow. As we do this, we must keep in mind the two properties this flow must have. First, it
must be continuous, and second, the value of the function \( f \) must decrease along the flow lines.

**Figure 2.6** A vector field that does not produce a flow \( M \times [0, \infty) \rightarrow M \).

The point of view we will take is that we want to follow the gradient vector field as closely as possible. We could, at a point \( p \) where \( -\nabla f(p) \) points outward, define \( G(p) = -\nabla (f|_S)(p) \), where \( S \) is the stratum containing \( p \). The resulting vector field, however, does not always induce a continuous flow. For an example see figure 2.7, where this procedure is applied to the vector field in figure 2.6. What we must do is project the vector \( -\nabla f(p) \) onto the maximal stratum such that the resulting vector does not point outward from \( M \). The result is shown in figure 2.8.

At first sight, it makes no sense to talk about \( -\nabla (f|_H)(p) \), when \( p \notin H \). When \( p \in \overline{H} \), however, there are two ways in which this could reasonably be defined. Happily these two ideas give the same result. First, we could define

\[
-\nabla (f|_H)(p) = \lim_{q \in H \rightarrow p} -\nabla (f|_H)(q),
\]

extending the vector field continuously. The second approach requires us to remember that \( f|_{U_p} \) can be thought of as \( f \circ x_p^{-1} : x_p(U_p) \rightarrow \mathbb{R} \) and extended to a function
Figure 2.7  This vector field does not produce a continuous flow. The vertex at the top is an equilibrium point, but nearby points flow to the other vertices in finite time.

Figure 2.8  This vector field will produce a continuous flow.
$\tilde{f}: \mathbb{R}^n \to \mathbb{R}$. Thus we can extend the stratum $H$ to a manifold $\tilde{H} \supset H$ that contains $p$. Then we can define

$$-\nabla(f|_H)(p) = -\nabla(f|_{\tilde{H}})(p).$$

Since $-\nabla\tilde{f}$ is continuous on $\tilde{H}$, these two definitions must agree.

How do we know there must be a maximal stratum $K$ such that $-\nabla(f|_K)(p)$ does not point outward from $M$? Suppose we have two strata, $H_1$ and $H_2$, such that $H_1 \neq H_2$ and $-\nabla(f|_{H_1})(p)$ points inward toward $H_i$. Choose a standard coordinate chart $x: U_p \to \mathbb{R}^n$ at $p$. Then the coordinate $x_i$ will be positive whenever $i > n - j$. Let $A_i = \{ \ell \in \{1, \ldots, n\} : e_\ell \text{ points into } H_i \}$. Then $x(H_i \cap U_p)$ is an open subset of span$\{e_\ell\}_{\ell \in A_i}$. Since

$$-\nabla(f|_{H_i})(p) = \sum_{\ell \in A_i} \frac{\partial f}{\partial x_\ell}(p)e_\ell$$

and $-\nabla(f|_{H_i})(p)$ points into $M$, we must have $\frac{\partial f}{\partial x_\ell}(p) \geq 0$ when $i \geq n - j(p)$ and $\ell \in A_i$. (Recall that according to Definition 2.3 vectors tangent to a stratum in $\partial M$ are considered inward pointing.)

Let $B = A_1 \cup A_2$. Then

$$\sum_{\ell \in B} \frac{\partial f}{\partial x_\ell}(p)e_\ell = -\nabla(f|_K)(p)$$

for some stratum $K$, and $-\nabla(f|_K)(p)$ points into $M$.

Then either $\dim(K) > \dim(H_i)$ for $i = 1, 2$, or one of the two strata is contained in the boundary of the other. Consequently for each point $p \in M$, there is a unique maximal stratum $K_p$ such that $-\nabla(f|_{K_p})(p)$ points into $M$. This allows us to make the

**Definition 2.6** At each point $p \in M$, let $K_p$ be the unique maximal stratum such that $p \in K_p$ and $-\nabla(f|_{K_p})(p)$ does not point outward from $M$. Set $G(p) = -\nabla(f|_{K_p})(p)$. 
Then $G$ is a well defined vector field on $M$. From the above construction we see that the directional derivative $G(p)[f] < 0$, so the value of $f$ will decrease along the flow lines of any flow induced by $G$. What we must show is that such a flow exists and is continuous.

2.5 The Modified Gradient Vector Field Induces a Continuous Flow

First we will show that even though the modified gradient vector field $G$ is not continuous, it does induce a flow. An analogy can be made to the path of a raindrop. The rain falls until it hits a roof. It then follows the slope of the roof until it reaches the edge, whereupon it falls off the edge and back into free fall. The $G$-flow will follow the $-\nabla f$-flow until it hits a stratum $H$ in the boundary. It then follows the $-\nabla f|_H$-flow until it either hits a lower dimensional stratum, or flows back into the interior. To make this idea precise, we must impose another condition on our Morse functions.

**Definition 2.7** We say that a Morse function $f : M \to \mathbb{R}$ satisfies Property (3) if for any standard coordinate chart $x$, whenever $-\nabla f(p)$ is tangent to a stratum $H \subseteq \partial M$ and $\frac{\partial f}{\partial x_i}(p) = 0$, the directional derivative of $\frac{\partial f}{\partial x_i}$ in the direction $-\nabla f(p)$ is not zero, i.e.

$$(-\nabla f(p)) \left[ \frac{\partial f}{\partial x_i} \right] = \sum_{j=1}^{n} -\frac{\partial f}{\partial x_j} \frac{\partial^2 f}{\partial x_j \partial x_i} \neq 0.$$

Definition 2.7 is illustrated by figure 2.9.

**Lemma 2.1** If $f : M \to \mathbb{R}$ is a function satisfying Property (3), then the modified gradient vector field $G$ induces a flow $\varphi : M \times [0, \infty) \to M$
Figure 2.9  On the left are shown the flow lines for the function \( f(x, y) = -x - y^2 \). This function does not satisfy Property (3), since the vector is tangent all along the bottom edge. On the right are the flow lines for the function \( f(x, y) = x^2 + y^2 + y \). This function does satisfy Property (3).

satisfying

\[
\varphi(\cdot, 0) = \text{identity} \\
\frac{\partial}{\partial t} \varphi(p, t)|_{t=t_0} = G(\varphi(p, t_0)).
\]

Proof  Recall that in Definition 2.6 we selected at each point \( p \) in a stratum \( S \subset M \) a stratum \( K_p \) and defined \( G(p) = -\nabla(f|_{K_p})(p) \). Since \( f|_S \) is a smooth function, the vector field \( -\nabla f|_S \) induces a continuous flow \( \varphi_S \) on \( S \). We can use this flow to define a stratum \( H_p = \lim_{t \to 0} K_{\varphi_S(t_0, p)} \). In general we will find that \( H_p = K_p \), but if \( \frac{\partial}{\partial x_i}(p) = 0 \) for some \( i \), this may not be the case.

We can solve the initial value problem \( \sigma'_p(t) = -\nabla (f|_{H_p}) (\sigma_p(t)) \), \( \sigma_p(0) = p \). The solution \( \sigma_p \) lies in \( H_p \). Let \( t_1 \) be given by

\[
t_1 = \sup\{ t \in \mathbb{R}_+ : \sigma'_p(t) = G(\sigma_p(t)) \}.
\]

Then for \( t \leq t_1 \), we set \( \varphi(t, p) = \sigma_p(t) \). For \( t > t_1 \), we must repeat this procedure starting from \( \sigma_p(t_1) \), and flowing for time \( t - t_1 \). □

This shows that the vector field \( G \) induces some flow on \( M \). Our goal now is to show that this flow is a continuous.
**Lemma 2.2** If \( f : M \to \mathbb{R} \) is a function satisfying Property (3), then the flow induced by the modified gradient \( G \) is continuous.

**Proof** Near a point \( p \in M \), there is a coordinate system \( x : U_p \to V_p \subseteq \mathbb{R}^n \) such that \( x(q) = (x_1(q), \ldots, x_n(q)) \). As usual, the coordinates are chosen so that \( x_i \in (-\infty, \infty) \) for \( i \leq n - j \), \( x_i \in [0, \infty) \) for \( i > n - j \), and \( x(p) = (0, \ldots, 0) \).

Choose an \( R > 0 \) such that \( B_R(p) \subseteq U_p \). Let \( \mu = \sup_{q \in M} \|G(q)\| \). Then we can choose \( r \) and \( \tau_0 \) such that \( \mu \tau_0 < R - r \). Then \( \varphi(t, B_r(p)) \subseteq B_R(p) \) for every \( t < \tau_0 \). Consequently, it is sufficient to view the situation in terms of the coordinate system \( x \). This situation is pictured in figure 2.10.

![Figure 2.10](image)

**Figure 2.10** Here \( r \) and \( \tau_0 \) have been chosen so that \( \mu \tau_0 = R - r \).

Define a projection \( \pi : \mathbb{R}^n \to \mathbb{R}^j \) by \( \pi(v) = (\pi_1(v_1), \ldots, \pi_n(v_n)) \), where

\[
\pi_i(v_i) = \begin{cases} 
v_i, & \text{if } 1 \leq i \leq n - j \text{ or } v_i \geq 0 \\
0, & \text{else.}
\end{cases}
\]

Note that \( \pi \) is a continuous map and \( \text{dist}(\pi(x), y) \leq \text{dist}(x, y) \).

We have a vector field \( \widetilde{G} = x_*(G) \) on \( V_p \). There is also another vector field \( F = x_*(-\nabla f) \). Extend \( F \) to all of \( \pi^{-1}(V_p) \) by setting \( \widetilde{F} \) to be "constant" (i.e. parallel in the Euclidean metric) along each preimage, \( \pi^{-1}(q) \) for \( q \in V_p \). The extended vector field \( \widetilde{F} \) is Lipschitz, and so induces a continuous flow, denoted by \( \psi \).
Next, we want to define a map $T_i : \pi_{-1}(V) \to \{0, 1\}$ for $i = n-j+1, \ldots, n$. The idea is that $T_i$ will be zero where the flow $\varphi$ stays within a stratum where $x_i = 0$. $T_i$ changes to 1 when the flow enters a higher dimensional stratum where $x_i > 0$.

$$T_i(q) = \begin{cases} 0, & \text{if } y_i(q) \leq 0 \text{ and } \langle -\tilde{G}(q), \frac{\partial}{\partial y_i} \rangle \leq 0 \\ 1, & \text{else.} \end{cases}$$

**Definition 2.8** Say that $\psi(\cdot, q)$ has an uptick at time $t$ if for some $i \in \{n-j+1, \ldots, n\}$,

$$\lim_{s \to t^-} T_i(\psi(s, q)) < \lim_{s \to t^+} T_i(\psi(s, q)).$$

**Lemma 2.3**

1. If $q \in V_p$ and $\psi(\cdot, q)$ has no upticks in $(0, \tau)$, then

$$\varphi(\tau, q) = \pi \circ \psi(\tau, q).$$

2. If $\psi(\cdot, q)$ has an uptick in $(0, \tau)$, then

$$\text{dist}(\varphi(\tau, q), \pi \circ \psi(\tau, q)) < 2\mu\tau.$$

**Proof** Since $\mu$ is the maximal speed for both flows, the farthest apart any two points can get in time $\tau$ is $2\mu\tau$. That proves the second part of the Lemma.

Suppose, then, that $\psi(\cdot, q)$ has no upticks in $(0, \tau)$. Let $\varphi(t, q) = (x_1(t), \ldots, x_n(t))$ and $\psi(t, q) = (y_1(t), \ldots, y_n(t))$. It is sufficient to consider the case where the flow remains in a single stratum, say $H \subset E_j(M)$.

The $x_i$'s satisfy the system of differential equations

$$\frac{dx_i}{dt} = g_i(x_1, \ldots, x_n)$$

$$= g_i(x_1, \ldots, x_j, 0, \ldots, 0).$$
We are able to set $x_{j+1} = \cdots = x_n = 0$, since this flow remains in $H$.

The $y_i$'s, on the other hand, are determined by the system

$$\frac{dy_i}{dt} = f_i(y_1, \ldots, y_n) = f_i(y_1, \ldots, y_j, 0, \ldots, 0).$$

Here, we replace $y_{j+1}, \ldots, y_n$ with 0, because the $f_i$ are constant on

$$\pi^{-1}(y_1, \ldots, y_j, 0, \ldots, 0).$$

Moreover, for $1 \leq i \leq j$,

$$f_i(y_1, \ldots, y_j, 0, \ldots, 0) = g_i(y_1, \ldots, y_j, 0, \ldots, 0).$$

Consequently for $1 \leq i \leq j$, $y_i(t) = x_i(t)$.

For $i > j$, $\pi_i(y_i) = 0$. Since $x_{j+1} = \cdots = x_n = 0$, it follows that

$$\pi(y_1(t), \ldots, y_n(t)) = (x_1(t), \ldots, x_n(t)).$$

It follows then, that $\pi \circ \psi(\tau, q) = \varphi(\tau, q)$.

\[\square\]

**Lemma 2.4** For $q \in U$ and a suitably chosen $\tau$ there is a finite upper limit, $N$, to the number of upticks along $\psi(\cdot, q) : [0, \tau] \to M$.

**Proof** The set $f^{-1}((-\infty, f(q))]$ is compact and contains the image of the curve $\psi(\cdot, q) : [0, \tau] \to M$. Suppose that the set $\{p_i\}$ of points where $\psi(\cdot, q) : [0, \tau] \to M$ has an uptick is infinite. Then some subsequence of $\{p_i\}$ has a limit point $p_0$.

Property (3), however, ensures that there is a neighborhood of $p_0$ that contains no other upticks, deriving a contradiction. This shows that each such curve has a
finite number $N_q$ of upticks. We need to show that there is a finite upper bound for \$\{N_q : q \in M\} \$.

Suppose there is no such upper bound. Then choose a sequence \$\{q_i\} \$ so that \$N_{q_i} > i \$.
Finally, choose a pair \$\{a_i = \psi(\tau_{a_i}, q_i), b_i = \psi(\tau_{b_i}, q_i)\} \$ so that \$|\tau_{a_i} - \tau_{b_i}| \$ is minimized along the curve \$\psi(\cdot, q) : [0, \tau] \to M \$. Then \$|\tau_{a_i} - \tau_{b_i}| \to 0 \$ as \$i \to \infty \$.

Using compactness again, we can find a subsequence of pairs \$\{a_i, b_i\} \$ so that \$a_i \to p_0 \$ and \$b_i \to p_0 \$ as \$i \to \infty \$. It follows that in a standard coordinate system, for some \$n - j(p_0) < i \leq n, \$
\[ \frac{\partial f}{\partial x_i}(p_0) = 0 \]

since \$p_0 \$ is a limit of upticks, and \$\frac{\partial f}{\partial x_i} \$ is continuous. Moreover, because \$p_0 \$ is the limit of two consecutive upticks, the directional derivative of \$\frac{\partial f}{\partial x_i} \$ in the \$-\nabla f(p_0) \$ direction satisfies
\[ (-\nabla f(p_0)) \left[ \frac{\partial f}{\partial x_i} \right] = 0. \]

But this contradicts the fact that \$f \$ satisfies Property (3).

We can define maps \$\psi_k : [0, \tau_0) \times M \to M \$ by \$\psi_k(\tau, q) = [\pi \circ \psi(\frac{\tau}{k}, \cdot)]^k(q) \$. Figure 2.11 shows an example where \$k = 3 \$. If \$\psi(\tau, q) \$ has upticks at times \$t_1, \ldots, t_n \$,

![Figure 2.11](image)

**Figure 2.11** The solid line shows the path from \$q \$ to \$\varphi(\tau, q) \$. The dotted line shows the path from \$q \$ to \$\psi_3(q) = [\pi \circ \psi(\frac{\tau}{3}, \cdot)]^3(q) \$. 
we may write this as
\[
\psi_k(\tau, q) = [\pi \circ \psi (\frac{t}{k}, \cdot)]^{N_n} \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)]^{N_{n-1}} \circ \ldots \\
\ldots \circ [\pi \circ \psi (\frac{t}{k}, \cdot)]^{N_1} \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)]^{N_0} (q) \\
= [\pi \circ \psi (N_n \frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\pi \circ \psi (N_{n-1} \frac{t}{k}, \cdot)] \circ \ldots \\
\ldots \circ [\pi \circ \psi (N_1 \frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\pi \circ \psi (N_0 \frac{t}{k}, \cdot)] (q),
\]
where \( t_i \in \left( \frac{\tau k_{i-1}}{k}, \frac{\tau k_{i+1}}{k} \right) \). Using the first part of the lemma, we can write
\[
\psi_k(\tau, q) = [\varphi (N_n \frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\varphi (N_{n-1} \frac{t}{k}, \cdot)] \circ \ldots \\
\ldots \circ [\varphi (N_1 \frac{t}{k}, \cdot)] \circ [\pi \circ \psi (\frac{t}{k}, \cdot)] \circ [\varphi (N_0 \frac{t}{k}, \cdot)] (q).
\]
Now, \([0, \tau] \times M\) is compact, so there is a constant \( K > 0 \) such that
\[
\text{dist} (\psi(t, p), \psi(s, q)) \leq K (\text{dist}(p, q) + ||t - s||).
\]
Combining this estimate with the fact that \( \tau \) does not increase distances and the lemma, we get the estimate
\[
\text{dist} (\psi_k(\tau, q), \varphi(\tau, q)) \leq 2\mu \frac{\tau}{k} \sum_{i=1}^{n} K^i \leq \frac{1}{k} \left( 2\mu \tau_0 \sum_{i=1}^{N} K^i \right).
\]
This bound is independent of \( \tau \) and \( q \), so as \( k \to \infty \), \( \psi_k \) converges uniformly to \( \varphi \). Since the maps \( \psi_k \) are all continuous it follows that \( \varphi : [0, \infty) \times M \to M \) is a continuous flow.

\[ \square \]

### 2.6 Essential Critical Points

In classical Morse theory, critical points of a Morse function \( f \) appear as stationary points of the \(-\nabla f\)-flow. Analyzing the behavior of the flow near these points allows one to prove the Morse theorems. In §1.2 we defined a critical point to be any point \( p \) such that \(-\nabla (f|_S)(p) = 0\), where \( S \) is the stratum containing \( p \). Which of these critical points are stationary points of the modified gradient flow?
A point $p$ in the stratum $S \subseteq M$ will be a stationary point if $G(p) = 0$. This means that the projection of $-\nabla f(p)$ onto any stratum other than $S$ will point outward from $M$. This is equivalent to saying that if $x : U_p \to \mathbb{R}^n$ is a standard coordinate chart near $p$, then $e_i[f](p) > 0$ for $i > n - j$.

![Diagram](image)

**Figure 2.12** An essential critical point and a non-essential critical point. The function here is the height function, so $-\nabla h$ points straight down.

**Definition 2.9** An *essential critical point* is a point $p \in M$ satisfying $G(p) = 0$.

In classical Morse theory, a critical point of $f$ is labeled with a number called its index. The index $\lambda$ of a critical point $p$ is the number of negative eigenvalues of the Hessian matrix $H(p)$ of second partial derivatives of $f$ at $p$. The Lemma of Morse tells us that near such a critical point, there is a system of local coordinates $x_p$ such that $f = f(p) - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_n^2$. Our situation requires a slight modification of this lemma.

**Lemma 2.5** Let $p$ be an essential critical point of a Morse function $f$, and suppose that $p$ is contained in a stratum $S$ having dimension $n - j$. 
Then there is a local coordinate system \( x_p : U_p \to \mathbb{R}^n \) such that the identity

\[
f = f(p) - x_1^2 + \cdots + x_{j-1}^2 + x_{j+1}^2 + \cdots + x_{n-j+1}^2 + \cdots + x_n^2
\]

holds throughout \( U_p \).

**Proof** We may assume without loss of generality that \( f(p) = 0 \). Since \( p \) is a non-degenerate critical point of \( f|_S \), we can choose coordinates \((u_1, \ldots, u_{n-j})\) at \( p \) such that

\[
f|_S = -u_1^2 - \cdots - u_{j-1}^2 + u_{j+1}^2 + \cdots + u_{n-j}^2,
\]

and extend this to a standard coordinate chart \( \mathbf{u} \) on \( M \).

We can express \( f \) as a Taylor series in these coordinates:

\[
f = -u_1^2 - \cdots - u_{j-1}^2 + u_{j+1}^2 + \cdots + u_{n-j}^2 \]
\[
+ \sum_{i=n-j+1}^n u_i \left[ f_{u_i} + \frac{1}{2} \sum_{j=1}^n f_{x_i x_j} x_j + \ldots \right],
\]

then set \( x_i = u_i \) for \( 1 \leq i \leq n-j \) and for \( i > n-j \) define

\[
x_i = \left[ f_{u_i} + \frac{1}{2} \sum_{j=1}^n f_{x_i x_j} x_j + \ldots \right].
\]

Then for each \( i, x_i(u_1, \ldots, u_n) \) is a smooth function. Let \( h \) be the map that carries \((u_1, \ldots, u_n)\) to \( x(u_1, \ldots, u_n) \). Since \( h \) is smooth and

\[
\det(Dh(0)) = \det
\begin{bmatrix}
1 & & & \\
& \ddots & & \\
& & 1 & \\
& & & f_{u_{n-j+1}} \\
& & & \ddots \\
& & & f_{u_n}
\end{bmatrix},
\]
Since \( p \) is an essential critical point, \( f_{u_i} > 0 \) for all \( i > n - j \). Thus \( \det(Dh(0)) \neq 0 \). It follows from the Implicit Function Theorem ([Sp1]) that on some neighborhood of \( p \), \( x \) is a coordinate system, compatible with \( u \).

Moreover, it is clear from the definition of \( x_i \) that in the domain of the coordinate chart \( x \), \( x_i = 0 \) if and only if \( u_i = 0 \). Further, \( x_i > 0 \) if and only if \( u_i > 0 \). Consequently, \( x \) is a standard coordinate system. In this coordinate system, \( f \) is given by

\[
-x_1^2 - \cdots - x_{\lambda}^2 + x_{\lambda+1}^2 + \cdots + x_{n-j}^2 + x_{n-j+1}^2 + \cdots + x_n^2
\]

as required. \( \square \)

We will call the number \( \lambda \) the index of \( p \), and we will take this lemma to be the definition of the index of an essential critical point. The coordinate system \( x \) in the lemma induces a coordinate system \( \tilde{x} = (x_1, \ldots, x_{n-j}) \) on \( S \). From this it is clear that \( \lambda \) is the index of \( p \) as a critical point of \( f|_S \).

\section{2.7 The Morse Theorems}

In this section we see how the number and type of essential critical points a function on a manifold with corners may have is governed by the topology of the domain. We will use the following notation: \( M_a = f^{-1}((-\infty, a]) \). We will assume that \( M_a \) is compact for each \( a \in \mathbb{R} \).

**Lemma 2.6** If \( a < b \), and \( f^{-1}([a, b]) \) contains no essential critical points, then there is a time \( \tau > 0 \) such that \( \varphi(\tau, M_b) \subset M_a \).

**Proof** Suppose that there is a point \( q \in f^{-1}([a, b]) \) such that \( \varphi(t, q) \notin f^{-1}((-\infty, a]) \) for all \( t > 0 \). Let \( \{q_i\}_{i>0} \) be the sequence \( \varphi(i, q) \). Then \( \{q_i\} \subset f^{-1}([a, b]) \), a compact set. Consequently, there is a subsequence of \( \{q_i\} \) that converges to a limit \( q_0 \in f^{-1}([a, b]) \). We must have \( f(\varphi(t, q)) > f(q_0) \) for all \( t \) and \( \lim_{t \to -\infty} f(\varphi(t, q)) = f(q_0) \).
Since $q_0$ is not an essential critical point, $G(q_0)$ is non-zero. We can choose some time $t_0$ such that $f(\varphi(t_0, q_0)) < f(q_0)$. Let $U$ be a neighborhood of $\varphi(t_0, q_0)$ such that $f(U) < f(q_0)$. Since $\varphi(t_0, \cdot)$ is continuous, $\varphi(t_0, \cdot)^{-1}(U)$ is an open set containing $q_0$. It follows that there is some $i$ such that $f(\varphi(i + t_0, q)) < f(q_0)$ which contradicts the assumption that $\{q_i\} \subset f^{-1}([a, b])$. \hfill \Box

We are now in a position to prove three of the central theorems of Morse theory.

**Theorem 2.1** Let $f : M \to \mathbb{R}$ be a Morse function on a manifold with corners $M$. If $a < b$ and $f^{-1}([a, b])$ contains no essential critical points, then $M_a$ is a deformation retract of $M_b$, so the inclusion map $M_a \hookrightarrow M_b$ is a homotopy equivalence.

![Diagram](image)

**Figure 2.13** The arrows illustrate the homotopy from $M_b$ to $M_a$.

**Proof** Since there are no essential critical points in $f^{-1}([a, b])$ and the value of $f$ decreases along the flow lines of $\varphi$, for each point $p \in M_b$ there is a time $t$ such that $\varphi(t, p) \in M_a$. Let $t_p = \inf\{t \in \mathbb{R}_+: \varphi(t, p) \in M_a\}$. 
Now we can define a homotopy $H : M_b \times [0,1] \to M_a$ by

$$H(p,s) = \begin{cases} 
\varphi(p, \frac{s}{1-t}), & \text{if } \frac{s}{1-t} \leq t_p \\
\varphi(p, t_p), & \text{if } \frac{s}{1-t} \geq t_p 
\end{cases}$$

\[ \square \]

**Theorem 2.2** Let $f : M \to \mathbb{R}$ be a Morse function on a manifold with corners $M$. Let $p$ be an essential critical point with index $\lambda$. Set $f(p) = c$. Suppose that, for some $\epsilon > 0$, $f^{-1}([c - \epsilon, c + \epsilon])$ contains no essential critical points other than $p$. Then $M_{c+\epsilon}$ is homotopy equivalent to $M_{c-\epsilon}$ with a $\lambda$-cell attached.

**Proof** Choose a coordinate system $x : U_p \to \mathbb{R}^{n-j} \times [0, \infty)^j$ in which we can write

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

Then choose $\epsilon > 0$ sufficiently small so that $f^{-1}[c - \epsilon, c + \epsilon]$ contains no essential critical points other than $p$, and the image $x(U_p)$ contains the closed 'ball'

$$\{(x_1, \ldots, x_n) : \sum_{i=1}^{n-j} x_i^2 + \sum_{i=n-j+1}^n x_i^1 \leq 2\epsilon\}.$$ 

The proof from here will consist of the following three steps:

1. Defining a region $H$, as shown in figure 2.15.

2. Show $M_{c-\epsilon} \cup H \simeq M_{c+\epsilon}$.

3. Show $M_{c-\epsilon} \cup e^\lambda \simeq M_{c-\epsilon} \cup H$. 
We begin by tweaking the function $f$ a bit. Choose a $C^\infty$ function $\mu : \mathbb{R} \to \mathbb{R}$ that satisfies

\[
\mu(0) > \epsilon \\
\mu(r) = 0, \quad \text{for } r > 2\epsilon \\
-1 < \mu' \leq 0.
\]

If we write

\[
\xi = x_1^2 + \cdots + x_\lambda^2 \\
\eta = x_{\lambda+1}^2 + \cdots + x_{n-j}^2 \\
\zeta = x_{n-j+1}^2 + \cdots + x_n^2,
\]

then we can write $f = c - \xi + \eta + \zeta$.

Define a new function $F$ by

\[
F = f - \mu(\xi + 2\eta + 2\zeta) \\
= c - \xi + \eta + \zeta - \mu(\xi + 2\eta + 2\zeta).
\]

We will use this function (and its level sets) to define the region $H$.

**Claim 2.1** The essential critical points of $F$ and $f$ are identical.
Outside our 'ball' of 'radius' $2\epsilon$, $F = f$ and so any critical points there must coincide. Inside, the function $f$ has a single essential critical point at $p$. To find the essential critical points of $F$ we must compute $dF$.

$$dF = (-1 - \mu')d\xi + (1 - 2\mu')d\eta + (1 - 2\mu')d\zeta.$$  

The coefficients $(-1 - \mu')$ and $(1 - 2\mu')$ are nowhere zero and $d\xi$ and $d\eta$ are simultaneously zero only at $p$. Thus $p$ is an essential critical point provided that $e_i[F](p) > 0$ for $i > n - j$. A computation shows that

$$e_i[F](p) = dF(e_i)(p) = (1 - 2\mu')(e_i)(p) = (1 - 2\mu')(1) > 0,$$

so $p$ is indeed an essential critical point of $F$.

**Claim 2.2** $F^{-1}(-\infty, c + \epsilon) = f^{-1}(-\infty, c + \epsilon).$

**Figure 2.15** The heavy outline shows the set \{\(\xi + 2\eta + 2\zeta = \epsilon\)\}, and the medium shaded region is H.
Outside the set \( \{ \xi + 2\eta + 2\zeta \leq 2\epsilon \} \) we know that \( \mu = 0 \), so \( F = f \). Inside this set, we see that
\[
F \leq f = c - \xi + \eta + \zeta.
\]
Equality holds on the boundary of the 'ball'. Also,
\[
c - \xi + \eta + \zeta \leq c + \left( \frac{1}{2} \xi + \eta + \zeta \right).
\]
Here equality holds when \( \xi = 0 \). Finally, we note that
\[
c + \left( \frac{1}{2} \xi + \eta + \zeta \right) \leq c + \epsilon.
\]
Here, again, equality holds on the boundary of the 'ball'. So we see that within this set, \( F \leq c + \epsilon \) and \( f \leq c + \epsilon \) unless \( \xi = 0 \) and \( \eta + \zeta = \epsilon \), in which case \( F = f = c + \epsilon \).

Claim 2.3 \( F^{-1}(-\infty, c - \epsilon] \) is a deformation retract of \( M_{c+\epsilon} \).

Consider the region \( F^{-1}[c - \epsilon, c + \epsilon] \). It is compact, but does it contain any critical points? The only possibility is \( p \), but
\[
F(p) = c - \mu(0) < c - \epsilon,
\]
so \( p \notin F^{-1}[c - \epsilon, c + \epsilon] \) and Theorem 2.1 applies: \( F^{-1}(-\infty, c - \epsilon] \) is a deformation retract of
\[
F^{-1}(-\infty, c + \epsilon] = f^{-1}(-\infty, c + \epsilon] = M_{c+\epsilon}.
\]

Define the region \( H \) by
\[
H = F^{-1}(-\infty, c - \epsilon] - M_{c-\epsilon}.
\]
Recall that we have defined the \( \lambda \)-cell \( e^\lambda \) such that \( e^\lambda = \{ q : \xi(q) < \epsilon, \eta(q) = \zeta(q) = 0 \} \). Note that \( e^\lambda \subseteq H \), since \( \frac{\partial F}{\partial \xi} = -1 - \mu' < 0 \) implies for \( q \in e^\lambda \),
\[
F(q) < F(p) < c - \epsilon.
\]
Note also that \( e^\lambda \cap M_{c-\epsilon} = \partial e^\lambda \).
Claim 2.4 \( M_{c-\epsilon} \cup e^\lambda \) is a deformation retract of \( M_{c-\epsilon} \cup H \).

For each \( t \in [0,1] \) we define a map \( r_t : M_{c-\epsilon} \cup H \to M_{c-\epsilon} \cup H \) as follows:

Case 1. If \( q \in M_{c-\epsilon} \), set \( r_t(q) = q \) for all \( t \).

Case 2. If \( q \in H \) and \( \xi(q) < \epsilon \), then set

\[
r_t(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, (1-t)x_{\lambda+1}, \ldots, (1-t)x_n).
\]

Then \( r_0 \) is the identity map, and \( r_0 : M_{c-\epsilon} \cup H \to M_{c-\epsilon} \cup e^\lambda \). Moreover, \( r_t(q) \in e^\lambda \) for each \( t \), because \( \frac{\partial F}{\partial \eta} > 0 \) and \( \frac{\partial F}{\partial \zeta} > 0 \). (Moving toward \( e^\lambda \) decreases \( F \).)

Case 3. If \( c \leq \xi(q) \leq \eta(q) + \zeta(q) + \epsilon \), then define \( r_t \) by

\[
r_t(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, s_t x_{\lambda+1}, \ldots, s_t x_n).
\]

where

\[
s_t = (1 - t) + t \left[ \frac{\xi - \epsilon}{\eta + \zeta} \right]^{\frac{1}{2}}.
\]

Here again, \( r_0 \) is the identity, and now \( r_1(q) \in f^{-1}(c-\epsilon) \). We must show that the functions \( s_t x_i \) are continuous as \( \xi \to \epsilon \), \( \eta \to 0 \), \( \zeta \to 0 \). Since \( x_i \to 0 \) as \( \eta + \zeta \to 0 \),

\[
\lim_{\eta + \zeta \to 0} \left[ \frac{\xi - \epsilon}{\eta + \zeta} \right]^{\frac{1}{2}} x_i = \left[ \lim_{\eta + \zeta \to 0} \frac{\xi - \epsilon}{\eta + \zeta} \right]^{\frac{1}{2}} (0).
\]

Since

\[
0 = \frac{(\epsilon - \epsilon)}{\eta + \zeta} \leq \frac{\xi - \epsilon}{\eta + \zeta} \leq \frac{(\eta + \zeta + \epsilon) - \epsilon}{\eta + \zeta} = 1,
\]

the limit is zero, and it follows that each \( s_t x_i \) is continuous.

Note that this definition agrees with Case 1 when \( \xi = \epsilon \) and with Case 2 when \( \xi - \eta - \zeta = \epsilon \). Thus \( r \) provides a deformation retraction of \( M_{c-\epsilon} \cup H \) to \( M_{c-\epsilon} \cup e^\lambda \).

This concludes the proof of Theorem 2.2. \( \Box \)

With these two theorems in hand, we can now state the main theorem of this chapter.
Theorem 2.3 (Main Theorem)  If $M$ is a manifold with corners, $f : m \to \mathbb{R}$ a Morse function on $M$ and $f^{-1}(-\infty, c]$ is compact for each $c$, then $M$ has the homotopy type of a CW complex with one cell of dimension $\lambda$ for each essential critical point with index $\lambda$.

The proof reads identically to the analogous proof in [Mi2]. The interested reader is referred there.

2.8  The Morse Inequalities

Let $f : M^n \to \mathbb{R}$ be a Morse function on an $n$-dimensional manifold with corners. Applying Theorem 2.3, $M$ is homotopy equivalent to a CW complex $X$, having a $\lambda$-cell for each essential critical point with index $\lambda$.

A cellular $p$-chain is a formal linear combination (with coefficients in $\mathbb{R}$)

$$r_1 \sigma_1^{(p)} + \cdots + r_m \sigma_m^{(p)}$$

of oriented $p$-cells in $X$. Let $C_p = C_p(M; \mathbb{R})$ denote the space of cellular $p$-chains. There exist boundary maps $\partial_i : C_i \to C_{i-1}$ such that the corresponding chain complex

$$C : \quad 0 \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

satisfies $H_*(C) \cong H_*(M; \mathbb{R})$ ([Mu]).

Here we are using the fact that homology is a homotopy invariant, i.e. that $H_*(X; \mathbb{R}) \cong H_*(M; \mathbb{R})$. Let $b_i(M)$ denote the $i^{\text{th}}$ Betti number of $M$. Then

$$b_i(M) = \dim(H_i(M; \mathbb{R})) = \dim(\ker \partial_i / \text{im} \partial_{i+1}) = \dim(\ker \partial_i) - \dim(\text{im} \partial_{i+1}).$$

Denote by $m_i(f)$ the number of essential critical points of $f$ having index $i$. 
**Theorem 2.4 (The Strong Morse Inequalities)** If $M$ is an $n$-dimensional manifold with corners and $f : M \to \mathbb{R}$ a Morse function on $M$, then for each $k$,

$$\sum_{i=0}^{k} (-1)^{k+i} b_i(M) \leq \sum_{i=0}^{k} (-1)^{k+i} m_i(f).$$

Equality holds when $k \geq n$.

**Proof** Compute

$$\sum_{i=0}^{k} (-1)^{k+i} b_i(M) = \sum_{i=0}^{k} (-1)^{k+i} [\dim(\ker \partial_i) - \dim(\operatorname{im} \partial_{i+1})]$$

$$= \sum_{i=0}^{k} (-1)^{k+i} \dim(\ker \partial_i) + (-1)^{k+i+1} \dim(\operatorname{im} \partial_{i+1})$$

$$= (-1)^k \dim(\ker \partial_0) + \sum_{i=1}^{k} (-1)^{k+i} [\dim(\operatorname{im} \partial_i) + \dim(\ker \partial_i)]$$

$$+ (-1)^{2k+1} \dim(\operatorname{im} \partial_1).$$

Applying the Rank and Nullity Theorem from linear operator theory([Sc]), this can be reduced to

$$\sum_{i=0}^{k} (-1)^{k+i} b_i(M) = (-1)^k \dim(C_0) + \sum_{i=1}^{k} (-1)^{k+i} \dim(C_i)$$

$$- \dim(\operatorname{im} \partial_{k+1})$$

$$= \sum_{i=0}^{k} (-1)^{k+i} m_i(f) - \dim(\operatorname{im} \partial_{k+1})$$

$$\leq \sum_{i=0}^{k} (-1)^{k+i} m_i(f).$$

We see that equality holds exactly when $\dim(\operatorname{im} \partial_{k+1}) = 0$. In particular, this is so when $k \geq n$. \qed

**Theorem 2.5 (The Weak Morse Inequalities)** If $M$ is a manifold with corners and $f : m \to \mathbb{R}$ a Morse function on $M$, then $m_k(f) \geq b_k(M)$ for each $k \geq 0$. 

Proof  This follows directly from the existence of the chain complex $C$.

\[ b_k(M) = \dim(\ker(\partial_k)) - \dim(\text{im}(\partial_{k+1})) \leq \dim(\ker(\partial_k)) \leq \dim(C_k) = m_k(f). \]

\[ \square \]
Chapter 3

Generalized Billiard Paths

In this chapter, we introduce the notion of a generalized billiard path, and use the results of Chapter 2 to count these paths.

3.1 Statement of the Problem

We now return to the problem posed in Section 1.1. We have a compact \( n \)-manifold embedded in some Euclidean space, \( M \hookrightarrow \mathbb{R}^N \). We imagine that the manifold is half-silvered, so that light may either reflect off the surface or pass through with no change in direction. We are interested in the paths that a beam of light might follow from one point \( p \in \mathbb{R}^n \) to another point \( q \in \mathbb{R}^n \).

Between consecutive reflection points, a beam of light will travel in a straight line. Therefore, to describe an entire path, it is sufficient to identify the reflection points and the order in which they occur (as well as the endpoints \( p \) and \( q \)). This being so, it makes sense to define a path connecting \( p \in \mathbb{R}^N \) to \( q \in \mathbb{R}^N \) to be a sequence of points \( \alpha_1, \ldots, \alpha_k \in M \).

This definition allows the line segment \( \overline{\alpha_i \alpha_{i+1}} \) to intersect the manifold. It also allows 'reflections' that do not follow the "angle of incidence equals the angle of reflection" rule. The path in Figure 3.1 illustrates both of these possibilities. The following definition characterizes the paths in which we are interested.

**Definition 3.1** A path \( P = \{\alpha_1, \ldots, \alpha_k\} \) connecting \( p = \alpha_0 \) to \( q = \alpha_{k+1} \) is a generalized billiard path with \( k \)-reflections if for each \( i \) one of the following is true:
Figure 3.1 A path with five reflections connecting $p$ to $q$. Here the embedded manifold is $S^1 \rightarrow \mathbb{R}^2$.

1. The bisector of $\angle \alpha_{i-1} \alpha_i \alpha_{i+1}$ is normal to $T_{\alpha_i} M$.

2. $\angle \alpha_{i-1} \alpha_i \alpha_{i+1}$ is a straight angle.

If $M$ happens to be a convex hypersurface, then this definition reduces to the usual notion of a billiard path. The task at hand now may be thought of as counting generalized billiard paths.

3.2 The Length of a Path

We can define the length of a path $P = \{\alpha_1, \ldots, \alpha_k\}$ connecting $p$ and $q$ to be

$$L_k^{(p,q)}(P) = \sum_{i=0}^{k} d_{\text{Euc}}(\alpha_i, \alpha_{i+1}),$$

and think of $L_k^{(p,q)}$ as a function

$$L_k^{(p,q)} : M \times \cdots \times M \rightarrow \mathbb{R},$$

$k$ copies
When there is no confusion regarding the endpoints, we will write \( L_k \) for \( L_k^{(n,q)} \).

This length function has one bad property. Wherever consecutive points of a path coincide, \( L_k \) has a singularity that looks like \(|x - y|\). It has another property, though, that makes us willing to put up with this. Away from this bad set, we can compute \( \nabla L_k \). Paths for which \( \nabla L_k = 0 \) will be of special interest, as shown by the following

**Lemma 3.1** A path \( P = \{\alpha_1, \ldots, \alpha_k\} \) with \( \alpha_i \neq \alpha_{i+1} \) for \( 0 \leq i \leq k \) satisfies
\[
\nabla L_k(\alpha_1, \ldots, \alpha_k) = 0
\]
if and only if it is a generalized billiard path.

**Proof** Near each point \( \alpha_i \) we can choose a coordinate chart
\[
x^i = (x^i_1, \ldots, x^i_n) : U_{\alpha_i} \to \mathbb{R}^n.
\]
Then we see that \( \nabla L_k \) is given by
\[
\nabla L_k = \left( \frac{\partial L_k}{\partial x_1^i}, \ldots, \frac{\partial L_k}{\partial x_n^i}, \frac{\partial L_k}{\partial x_1}, \ldots, \frac{\partial L_k}{\partial x_n} \right).
\]
Now in order for \( \nabla L_k \) to be zero, each of the terms \( \frac{\partial L_k}{\partial x_j} \) must be zero. Since this term depends only on \( d_{\text{Euc}}(\alpha_{i-1}, \alpha_i) \) and \( d_{\text{Euc}}(\alpha_i, \alpha_{i+1}) \), it is sufficient to look at a path with only one 'reflection'. Hence we need only look at the function
\[
f(\alpha) = d_{\text{Euc}}(p, \alpha) + d_{\text{Euc}}(\alpha, q) : M \to \mathbb{R}.
\]
In order to compute \( \frac{\partial f}{\partial x_i} (\alpha) \) we can choose a curve in \( M \) (which we will also call \( \alpha \)) satisfying
\[
\alpha(0) = \alpha \quad \text{and} \quad \alpha'(0) = e_i.
\]
In this way, we think of \( \alpha \) as varying along the curve, rather than as a fixed point.

Then
\[
\frac{\partial f}{\partial x_i}(\alpha) = \frac{\partial}{\partial t} f \circ \alpha(t)|_{t=0}.
\]
Then we can compute

\[
\frac{\partial}{\partial t} (p - \alpha(t) \cdot (p - \alpha(t)))^{1/2} \frac{\partial}{\partial t} [(\alpha(t) - q) \cdot (\alpha(t) - q)]^{1/2} = \frac{-\alpha'(t)(p - \alpha(t))}{[(p - \alpha(t))(p - \alpha(t))]^{1/2}} + \frac{\alpha'(t)(\alpha(t) - q)}{[(\alpha(t) - q)(\alpha(t) - q)]^{1/2}}
\]

\[
= \alpha'(t) \cdot (\frac{p - \alpha(t)}{||p - \alpha(t)||} + \frac{\alpha(t) - q}{||\alpha(t) - q||}).
\]

Evaluating at \( t = 0 \) we find

\[
\frac{\partial f}{\partial x^i}(\alpha) = e_i^\alpha \cdot \left( \frac{p - \alpha}{||p - \alpha||} + \frac{\alpha - q}{||\alpha - q||} \right).
\]

This will be zero for all \( i \) provided that the vector

\[
\left( \frac{p - \alpha}{||p - \alpha||} + \frac{\alpha - q}{||\alpha - q||} \right)
\]

is either normal to \( T_\alpha M \) or zero. When this vector is non-zero it is a bisector of the angle \( \angle p o q \). When it is zero, \( \angle p o q \) is a straight angle. Thus the gradient is zero exactly when the path is a generalized gradient path.

Having established that \( L_k \) is a function worth considering, let's look more closely at its behavior near the diagonals \( \{ \alpha_i = \alpha_{i+1} \} \). Consider a path \( \{ \alpha, \beta, \gamma \} \) where \( \alpha = \beta \). If \( \beta \) moves slightly to \( \beta' \), as in figure 3.2, the triangle inequality tells us that we have increased the length of the path.

We know that the vector field \( -\nabla L_k \) points in the direction of decreasing length. Consequently, under any modified gradient flow, nearby consecutive reflections would tend to flow toward each other. It is possible to define such a flow (as in Section 4.2), but at the moment we are not interested in what happens in the bad set. Instead what we will do is look at the function \( -L_k \). It has the same critical points, but the \( \nabla L_k \)-flow tends away from the bad set. Then we will 'blow up' the product \( M^k \) along the appropriate diagonals.
3.3 The Blow Up Space

The notion of blowing up was introduced by Fulton and MacPherson in [FM]. To understand what is meant by blow up, let’s think about a simple example. Take $M = S^1$ and $k = 2$. Assume that $p, q \notin M$. Then $-L_2$ is a function on the torus. The bad set is the diagonal $\Delta \subset S^1 \times S^1$.

Notice that as $(\alpha, \beta) \to (\alpha, \alpha)$ the limit of $\nabla L_2$ depends on the direction of approach. The gradient vector $\nabla L_2(\alpha, \beta)$ consists of a vector in $T_{\alpha}M$ pointing away from $\beta$ and a vector in $T_{\beta}M$ pointing away from $\alpha$. If $\beta$ is allowed to approach $\alpha$ from the opposite side, the gradient vector is reversed.

We need to produce a closure of $S^1 \times S^1 - \Delta$ on which we can extend $\nabla L_2$ continuously. Consequently, as $(\alpha, \beta)$ approaches $\Delta$, we keep track not only of the limiting point, but also of the relative positions of $\alpha$ and $\beta$.

Now let’s consider a path with $k$ reflections on an $n$-manifold $M$. Figure 3.4 shows the situation when two consecutive points coincide. Such a collision is described by the limiting point and an infinitesimal tangent space diagram. This diagram is
Figure 3.3 The point $\delta \in \Delta$ is blown up to the two points $\delta'$ and $\delta''$.

represented by points $v_\alpha$ and $v_\beta$ in the tangent space of the limiting point. Two such diagrams are equivalent if they differ by translation and multiplication by a positive constant. We can translate the diagram so that $v_\alpha$ is at the origin, and then scale it so $v_\beta$ is on the unit circle. This shows that each such point will be blown up into a copy of $S^{n-1}$.

Figure 3.5 shows what may happen when $\alpha$, $\beta$, and $\gamma$ coincide at a point $\theta \in M$. The situation is a bit more complicated now. Again, we can translate the diagram so that $v_\alpha$ is at the origin, and then scale it so $v_\gamma$ is on the unit circle. The point $v_\beta$ now may lie anywhere in $T_\theta M \cup \infty$. It would seem that each such point $\theta$ is

Figure 3.4 The infinitesimal tangent space diagram for two consecutive reflections colliding in $M$. 
blown up to a copy of $S^{n-1} \times S^n$. In fact this is not the case. Whenever $v_\beta = v_\alpha$ or $v_\beta = v_\gamma$, the resulting double point must also be blown up. On the other hand, if scaling the diagram so that $v_\gamma$ is on the unit circle pushes $v_\beta$ off to infinity, we would do better to scale the diagram so that $v_\beta$ is on the unit circle. (We are free to choose, since all these diagrams are equivalent.) When we rescale in this fashion, we will find that $v_\gamma = v_\alpha$. This point need not be blown up further, because $\alpha$ and $\gamma$ are not consecutive reflections. All of these special situations correspond to a situation where two of the points approach each other much more quickly than they approach the third.

The more points that collide, the more of these cascading diagrams there may be. In addition, two collections of points may collide independently at different points in the manifold. In this case we have two separate collections of infinitesimal diagrams corresponding to the two collections of points. We will denote the space that results from blowing up $M^k$ in this way by $X_k = X_k(M)$. The spaces that result are somewhat difficult to describe. There is one thing we can say about these spaces which will be of particular importance to us.

**Lemma 3.2** For any smooth manifold $M$, the space $X_k = X_k(M)$ is a manifold with corners.

**Proof** It is shown in [FM] (see also [AS] and [BT]) that the result of blowing up all the diagonals is a manifold with corners. In our case, we are only concerned with the diagonals corresponding to the collision of consecutive reflections. Here we show that blowing up only these diagonals also leads to a manifold with corners.

First, we define some convenient notation for referring to a stratum of the blow up $X_k$. When we write

$$(\alpha_1, \ldots, \alpha_{i-1}, \{\alpha_i, \ldots, \alpha_{i+j}\}, \ldots, \alpha_k),$$
Figure 3.5 Possible infinitesimal tangent space diagrams for three consecutive reflections colliding in $M$. In (a), $\alpha$, $\beta$ and $\gamma$ all approach each other at approximately the same rate. In (b), $\beta$ and $\gamma$ approach each other much faster than they approach $\alpha$. In (c), $\alpha$ and $\beta$ approach each other much faster than they approach $\gamma$. The situation where $\alpha$ and $\gamma$ approach each other faster than they approach $\beta$ need not be considered separately, since these reflections are not consecutive.
we mean that \( \alpha_i = \cdots = \alpha_{i+j} \), and all these points come together at commensurable rates. This stratum will be described by an infinitesimal diagram in \( T_\theta M \) in which \( v_i \neq \cdots \neq v_{i+j} \). Furthermore, when we write

\[
(\ldots, \{\alpha_i, \ldots, \{\alpha_{i+\ell}, \ldots, \alpha_{i+\ell+m}\}, \ldots, \alpha_{i+j}\}, \ldots)
\]

we mean that \( v_{i+\ell} = \cdots = v_{i+\ell+m} \) in the first infinitesimal diagram, requiring a second diagram.

We can group things together however we desire. Each grouping designates a stratum with as many infinitesimal diagrams as there are pairs of braces. Moreover, for each pair of braces we add, the codimension of the stratum is increased by one. To see this lets look an example. Take

\[
(\alpha_1, \ldots, \alpha_{i-1}, \{\alpha_i, \ldots, \alpha_{i+j}\}, \ldots, \alpha_k).
\]

This describes a \((k-j+1)n\)-dimensional set in \( M^k \), in which the points \( \alpha_i, \ldots, \alpha_{i+j} \) are all equal to a point \( \theta \in M \).

Now we can calculate the dimension of the stratum in the blow up \( X_k \) that corresponds to this stratum in the product \( M^k \). In the infinitesimal diagram, \( v_\alpha \), can be scaled to the origin, and \( v_{\alpha_{i+1}} \) can be scaled to the unit circle, providing \( n-1 \) dimensions. Each of the other \( j-2 \) points can be anywhere in \( T_\theta M \), another \((j-2)n\) dimensions. Adding this up, we get

\[
(k-j+1)n + n - 1 + (j-2)n = kn - 1.
\]

To see how a coordinate chart \( u \) may be defined near a point on this stratum, first choose a coordinate chart at each of the distinct points in \( M \). The chart at \( \theta \) induces coordinates on \( T_{\theta'} M \) for \( \theta' \) near \( \theta \). Then we may choose coordinates on the unit sphere in \( T_{\theta'} M \) that vary smoothly with \( \theta' \).
When $\alpha_i, \ldots, \alpha_{i+j}$ are all sufficiently close together we can write uniquely

$$(\alpha_i, \ldots, \alpha_{i+j}) = (\exp_{\theta}(tv_i), \ldots, \exp_{\theta}(tv_{i+j})),$$

by requiring $v_i = 0$ (so that $\theta' = \alpha_i$), $|v_{i+1}| = 1$ and $t \geq 0$. Then the limit as $t \to 0$ is the infinitesimal diagram defined by $\{v_i, \ldots, v_{i+1}\}$. Set, for $v_i = 0$ and $|v_{i+1}| = 1$ and $\theta'$ in a small neighborhood of $\theta$,

$$x(\alpha_1, \ldots, \alpha_{i-1}, \theta', \alpha_{i+j+1}, \ldots, \alpha_k, v_i, \ldots, v_{i+j}, t) =$$

$$= (\alpha_1, \ldots, \alpha_{i-1}, \exp_{\theta'}(tv_i), \ldots, \exp_{\theta'}(tv_{i+j}), \alpha_{i+j+1}, \ldots, \alpha_k).$$

Then on a neighborhood of $(\alpha_1, \ldots, \alpha_k)$ this map defines a coordinate chart.

The same procedure can be used for any grouping of the $\alpha_i$'s, using one parameter $0 \leq t_i \in \mathbb{R}$ for each pair of braces. \qed

There is a map $g : X_k \to M^k$ that assigns to each point in $X_k$ the corresponding limiting point in $M^k$. We can define (abusing notation in the process)

$$-L_k : X_k \to \mathbb{R}$$

by

$$-L_k(q) = -L_k \circ g(q).$$

Now, we wish to study this function on $X_k$. There is just one more order of business to attend to first.

### 3.4 When is $-L_k$ a Morse function?

We want to show that $-L_k$ satisfies the properties in Definition 2.5 and Definition 2.7. First of all, we must show that $\nabla L_k$ extends continuously to $X_k$. Recall the definition of $L_k$:

$$L_k(P) = \sum_{i=0}^{k} d_{E_{uc}}(\alpha_i, \alpha_{i+1}).$$
It is sufficient to show that $\nabla d_{\text{Euc}}(\alpha_i, \alpha_{i+1})$ extends continuously for each $i$. If $\alpha_i$ and $\alpha_{i+1}$ do not approach each other, then $d_{\text{Euc}}(\alpha_i, \alpha_{i+1})$ is smooth as it approaches the boundary of $X_k$, and $\nabla d_{\text{Euc}}(\alpha_i, \alpha_{i+1})$ can be extended continuously. If $\alpha_i$ and $\alpha_{i+1}$ do approach each other, we must show that $\nabla d_{\text{Euc}}(\alpha_i, \alpha_{i+1})$ approaches a limit.

As the points $\alpha_i$ and $\alpha_{i+1}$ approach each other, we can write

$$\alpha_j = \exp_{\theta'}(tv'_j),$$

where $\theta' \to \theta$ and $v'_j \to v_j \in T_{\theta} M$ as $t \to 0$. Then, since $D\left(\exp_{\theta}\right)_0$ is the identity on $T_{\theta'} M$, we can write,

$$\alpha_j = \theta' + tv'_j + O(t^2).$$

Here we are thinking of $T_{\theta'} M$ as a linear subspace of $\mathbb{R}^N$. Since $M$ is compact, $O(t^2)$ is a uniform bound for bounded $v'_j$. Then the distance from $\alpha_i$ to $\alpha_{i+1}$ is given by

$$d_{\text{Euc}}(\alpha_i, \alpha_{i+1}) = t|v'_i - v'_{i+1}| + O(t^2).$$

So $\nabla d_M(\alpha_i, \alpha_{i+1})$ is given by

$$\left(\frac{v'_i - v'_{i+1}}{|v'_i - v'_{i+1}|}\right) + O(t^2) \oplus \left(\frac{v'_{i+1} - v'_i}{|v'_{i+1} - v'_i|}\right) + O(t^2).$$

The first vector in the sum is in $T_{\alpha_i} M$. The second lies in $T_{\alpha_{i+1}} M$. The limit as $t \to 0$ exists and is equal to

$$\left(\frac{v_i - v_{i+1}}{|v_i - v_{i+1}|}\right) \oplus \left(\frac{v_{i+1} - v_i}{|v_{i+1} - v_i|}\right) \in T_{\theta} M \oplus T_{\theta'} M,$$

so $\nabla L_k$ extends continuously to all of $X_k$.

One thing that may prevent $-L_k$ from being Morse is a billiard path that has a tangential reflection, i.e. the angle of incidence and angle of reflection are both zero. We will find in Lemma 3.7 that this is a rare occurrence. As long as there are no tangential billiard paths connecting $p$ to $q$ with $k$ reflections, the vector field $G$ points
inward at every point in the boundary of $M$. In this case $-L_k$ satisfies Property (3) (Definition 2.7), and for every stratum $H \subset \partial M$ a critical point of $f|_H$ satisfies Properties (1b) and (2) in Definition 2.5. It remains only to show that a critical point in the interior of $M$ satisfies Property (1a).

This property requires that the critical points be non-degenerate, i.e. the determinant of the Hessian at a critical point must be non-zero. To begin with lets look at $-L_2 : X_2 \to \mathbb{R}$.

Here there are two reflection points, $\alpha$ and $\beta$. Choose an orthonormal coordinate system $x$ satisfying $x(\beta) = 0$, and a coordinate system $y$ with $y(\alpha) = 0$. The function then can be written as

$$-L_2 = -||p - \alpha|| - ||\alpha - \beta|| - ||\beta - q||.$$ 

Our first goal is to get an explicit representation for the Hessian.

**Lemma 3.3** The Hessian of $-L_2$ is given by

$$H(\alpha, \beta) = \begin{bmatrix} A & K \\ K & B \end{bmatrix}$$

where

$$A_{i,j} = \alpha_{y_j,y_i} \cdot (v_{p\alpha} - v_{\alpha\beta}) + [\cos(\psi_{y_j}) \cos(\psi_{y_i}) - \delta_{ij}] \left( \frac{1}{||p - \alpha||} + \frac{1}{||\alpha - \beta||} \right),$$

$$B_{i,j} = \beta_{x_i,x_j} \cdot (v_{\alpha\beta} - v_{\beta q}) + [\cos(\psi_{x_j}) \cos(\psi_{x_i}) - \delta_{ij}] \left( \frac{1}{||\alpha - \beta||} + \frac{1}{||\beta - q||} \right),$$

and

$$K_{i,j} = \frac{1}{||\alpha - \beta||} \left( \beta_{x_i} \cdot \alpha_{y_j} - \cos(\psi_{x_i}) \cos(\psi_{y_j}) \right).$$

In the above expressions $v_{\alpha\beta} = \frac{\alpha - \beta}{||\alpha - \beta||}$ and $\delta_{ij}$ is the Kronecker delta.
Proof  We need to compute the second partial derivatives. First compute
\[
\frac{\partial^2(-L_2)}{\partial x_i \partial y_j} = \frac{\partial}{\partial x_i} \left( \frac{\partial}{\partial y_j} \left( -\|\alpha - \beta\| - \|\beta - q\| \right) \right)
\]
\[
= \frac{-\beta_{y_j} \cdot (\alpha - \beta)}{\|\alpha - \beta\|} - \frac{\beta_{y_j} \cdot (\beta - q)}{\|\beta - q\|}
\]
\[
= \beta_{z_i} \cdot \left( \frac{\alpha - \beta}{\|\alpha - \beta\|} - \frac{\beta - q}{\|\beta - q\|} \right).
\]

In order to compute \( K \) we must differentiate in one \( x \) variable and one \( y \) variable. In this case, the result is
\[
\frac{\partial^2(-L_2)}{\partial y_j \partial x_i} = \beta_{z_i} \cdot \frac{\partial}{\partial y_j} \left( \frac{\alpha - \beta}{\|\alpha - \beta\|} \right)
\]
\[
= \beta_{z_i} \cdot \left( \frac{\|\alpha - \beta\| \alpha_{y_j} - (\alpha - \beta) \frac{\alpha_{y_j} \cdot (\alpha - \beta)}{\|\alpha - \beta\|^2}}{\|\alpha - \beta\|^2} \right).
\]

This expression can be simplified by setting \( v_{\alpha \beta} = (\alpha - \beta)/\|\alpha - \beta\| \) and letting \( \psi_{z_i} \) be the angle between \( v_{\alpha \beta} \) and \( \beta_{z_i} \). Similarly, \( \psi_{y_j} \) will be the angle between \( v_{\alpha \beta} \) and \( \alpha_{y_j} \). Then
\[
\frac{\partial^2(-L_2)}{\partial y_j \partial x_i} = \beta_{z_i} \cdot \left( \frac{\alpha_{y_j}}{\|\alpha - \beta\|} - \frac{\alpha - \beta}{\|\alpha - \beta\|^2} \cos(\psi_{y_j}) \right)
\]
\[
= \frac{1}{\|\alpha - \beta\|} \left( \beta_{z_i} \cdot (\alpha_{y_j} - \cos(\psi_{z_i}) \cos(\psi_{y_j})) \right).
\]

In order to compute \( B \), we must differentiate by two different \( x \)-variables. In that case, we find
\[
\frac{\partial^2(-L_2)}{\partial x_j \partial x_i} = \beta_{z_i} \cdot \left( \frac{\alpha - \beta}{\|\alpha - \beta\|} - \frac{\beta - q}{\|\beta - q\|} \right) + \beta_{z_i} \cdot \frac{\partial}{\partial x_j} \left( \frac{\alpha - \beta}{\|\alpha - \beta\|} - \frac{\beta - q}{\|\beta - q\|} \right)
\]
and
\[
\frac{\partial}{\partial x_j} \left( \frac{\alpha - \beta}{\|\alpha - \beta\|} - \frac{\beta - q}{\|\beta - q\|} \right) = \frac{-\|\alpha - \beta\| \beta_{z_j} - (\alpha - \beta) \frac{-\beta_{z_j} \cdot (\alpha - \beta)}{\|\alpha - \beta\|^2} - \|\beta - q\| \beta_{z_j} - (\beta - q) \frac{\beta_{z_j} \cdot (\beta - q)}{\|\beta - q\|^2}}{\|\alpha - \beta\|^2} \]
\[
= \frac{-\beta_{z_j} \cdot v_{\alpha \beta} \cos(\psi_{z_j}) + \beta_{z_j} \cdot \frac{\beta_{z_j} \cdot \cos(\psi_{z_j})}{\|\alpha - \beta\|} - \beta_{z_j} \cdot \frac{\beta_{z_j} \cdot \cos(\psi_{z_j})}{\|\beta - q\|}}{\|\alpha - \beta\|} + \frac{\beta_{z_j} \cdot \cos(\psi_{z_j})}{\|\beta - q\|}.
\]
Since we are at a critical point, it is easily shown that \( \beta_z \cdot v_{a\beta} = \beta_z \cdot v_{\beta q} \). From this it follows that the angle between \( \beta_z \) and \( v_{\beta q} \) is also \( \psi_{z_i} \). Using this and the orthonormality of the coordinate system \( \mathbf{x} \), we can write

\[
\frac{\partial^2(-L_2)}{\partial x_j \partial x_i} = \beta_{z_j z_i} \cdot (v_{a\beta} - v_{\beta q}) + \cos(\psi_{z_i}) \cos(\psi_{z_j}) \left( \frac{1}{||\alpha - \beta||} + \frac{1}{||\beta - q||} \right).
\]

The last derivative we must compute is \( \frac{\partial^2(-L_2)}{\partial z_i^2} \). Here it is:

\[
\frac{\partial^2(-L_2)}{\partial z_i \partial z_i} = \beta_{z_i z_i} \cdot \left( \frac{\alpha - \beta}{||\alpha - \beta||} - \frac{\beta - q}{||\beta - q||} \right) + \beta_{z_i} \cdot \frac{\partial}{\partial x_i} \left( \frac{\alpha - \beta}{||\alpha - \beta||} - \frac{\beta - q}{||\beta - q||} \right)
\]

\[
= \beta_{z_i z_i} \cdot (v_{a\beta} - v_{\beta q}) + \beta_{z_i} \cdot \left( \frac{-\beta_{z_i}}{||\alpha - \beta||} + \frac{v_{a\beta} \cos(\psi_{z_i})}{||\alpha - \beta||} - \frac{\beta_{z_i}}{||\beta - q||} + \frac{v_{\beta q} \cos(\psi_{z_i})}{||\beta - q||} \right)
\]

\[
= \beta_{z_i z_i} \cdot (v_{a\beta} - v_{\beta q}) + [\cos^2(\psi_{z_i}) - 1] \left( \frac{1}{||\alpha - \beta||} + \frac{1}{||\beta - q||} \right).
\]

The matrix \( A \) can also be obtained from these last two expressions by substituting \( y \)'s for \( x \)'s and changing \( \alpha \sim p \), \( \beta \sim \alpha \), and \( q \sim \beta \). \( \square \)
Now we may begin to investigate the Hessian matrix \( H(\alpha, \beta) \). We can write this as

\[
H(\alpha, \beta) = \begin{bmatrix}
\left( \frac{\partial^2(-L_1)}{\partial y_i \partial y_j} \right)_{i,j} & \left( \frac{\partial^2(-L_1)}{\partial y_i \partial x_j} \right)_{i,j} \\
\left( \frac{\partial^2(-L_1)}{\partial x_i \partial y_j} \right)_{i,j} & \left( \frac{\partial^2(-L_1)}{\partial x_i \partial x_j} \right)_{i,j}
\end{bmatrix} =
\begin{bmatrix}
A & K \\
K & B_1
\end{bmatrix} - \frac{1}{\|\beta - q\|} \begin{bmatrix}
0 & 0 \\
0 & C
\end{bmatrix},
\]

where

\[
B_1 = \left( \beta_{x_i x_i} \cdot (v_{\alpha \beta} - v_{\beta q}) + \frac{\cos(\psi_{x_i})\cos(\psi_{x_j}) - \delta_{i,j}}{\|\alpha - \beta\|} \right)_{i,j}
\]

and

\[
C = (\delta_{i,j} - \cos(\psi_{x_i})\cos(\psi_{x_j}))_{i,j}.
\]

At this time, we shall shift our point of view slightly. Rather than think of the end of the path as the fixed point \( q \), we will allow ourselves to choose a point \( q' \) along the line \( \beta + tv_{\beta q} \). We may decide how far along the path we would like to go, and stop there. Set \( \ell = \|\beta - q'\| \).

Recall that the purpose of this discussion was to determine when the Hessian in non-degenerate. The Hessian is non-degenerate if it has no zero eigenvalues. How can we ensure that there will be no zero eigenvalues? We will show in Lemma 3.9 that as \( \ell \) increases, the eigenvalues of \( H \) either increase or remain constant. Moreover, an eigenvalue can remain constant under these circumstances only if it has a corresponding eigenvector of the form

\[
\begin{pmatrix}
v \\
0
\end{pmatrix}
\]
A vector of this form can be an eigenvector only if \( K \mathbf{v} = 0 \). The matrix \( K \) can be written

\[
K = \frac{1}{\| \alpha - \beta \|} \left( \beta_{x_i} \cdot \alpha_{y_j} - \cos(\psi_{x_i}) \cos(\psi_{y_j}) \right)_{i,j}
\]

\[
= \frac{K_1 - K_2}{\| \alpha - \beta \|}.
\]

From this it follows that \((\mathbf{v},0)^T\) is an eigenvector if and only if \( K_1 \mathbf{v} = K_2 \mathbf{v} \).

**Lemma 3.4** If there is a vector \( \mathbf{v} \) satisfying \( K \mathbf{v} = 0 \), then the path is tangent to \( M \) at either \( \alpha \) or \( \beta \).

**Proof** Let us investigate \( K_2 \mathbf{v} \) first. The \( k^{th} \) component is given by

\[
(K_2 \mathbf{v})_k = \sum_j v_j \cos(\psi_{x_k}) \cos(\psi_{y_j})
\]

\[
= \cos(\psi_{x_k}) \sum_j v_j \cos(\psi_{y_j})
\]

\[
= (\beta_{x_k} \cdot \mathbf{v}_{\alpha \beta}) \sum_j v_j (\alpha_{y_j} \cdot \mathbf{v}_{\alpha \beta})
\]

\[
= (\beta_{x_k} \cdot \mathbf{v}_{\alpha \beta}) \left( \sum_j v_j \alpha_{y_j} \right) \cdot \mathbf{v}_{\alpha \beta}
\]

\[
= (\beta_{x_k} \cdot \mathbf{v}_{\alpha \beta})(\mathbf{v} \cdot \mathbf{v}_{\alpha \beta}).
\]

We also have

\[
(K_1 \mathbf{v})_k = \sum_j v_j \beta_{x_k} \cdot \alpha_{y_j}
\]

\[
= \beta_{x_k} \cdot \sum_j v_j \alpha_{y_j}
\]

\[
= \beta_{x_k} \cdot \mathbf{v}.
\]

When is \( \beta_{x_k} \cdot \mathbf{v} = (\beta_{x_k} \cdot \mathbf{v}_{\alpha \beta})(\mathbf{v} \cdot \mathbf{v}_{\alpha \beta}) \)? Let \( \theta_{x_k} \) be the angle between \( \beta_{x_k} \) and \( \mathbf{v} \), and let \( \sigma \) be the angle between \( \mathbf{v}_{\alpha \beta} \) and \( \mathbf{v} \). The statement then reduces to

\[
\cos(\theta_{x_k}) = \cos(\psi_{x_k}) \cos(\sigma)
\]
Figure 3.7 The three angles $\sigma$, $\psi_{x_k}$, and $\theta_{x_k}$.

(since the vectors in question are all unit vectors).

Let $\pi_\beta$ denote the projection onto $T_\beta M$. Then the following identities hold.

$$
\pi_\beta(v_{\alpha\beta}) = \sum_k \cos(\psi_{x_k}) \beta_{x_k}
$$

$$
\pi_\beta(v) = \sum_k \cos(\theta_{x_k}) \beta_{x_k}
$$

From this we see

$$
\pi_\beta(v) = \sum_k \cos(\psi_{x_k}) \cos(\sigma) \beta_{x_k}
$$

$$
= \cos(\sigma) \sum_k \cos(\psi_{x_k}) \beta_{x_k}
$$

$$
= \cos(\sigma) \pi_\beta(v_{\alpha\beta})
$$

We can write $v = v_{\alpha\beta} \cos(\sigma) + (v_{\alpha\beta})_\perp \sin(\sigma)$ for some $(v_{\alpha\beta})_\perp$ orthogonal to $v$. It then follows that either $\pi_\beta(v_{\alpha\beta})_\perp = 0$ or $\sin(\sigma) = 0$. In the first case the conclusion is that $v_{\alpha\beta} \in T_\beta M$, and so the path is tangent to $M$ at the point $\beta$. The second condition implies that $v_{\alpha\beta} = \pm v$, and so $v_{\alpha\beta} \in T_\alpha M$. Here the path is tangent to $M$ at $\alpha$. \qed

So generalized billiard paths with this type of tangential reflection are a cause for concern. This is not without reason, as shown by the following
Lemma 3.5 If the path \{\alpha, \beta\} has a tangential reflection at \alpha, then the vector
\[
\begin{pmatrix}
\mathbf{v}_{\alpha \beta} \\
0
\end{pmatrix}
\]
is an eigenvector with eigenvalue zero.

Proof We can compute
\[
H(\alpha, \beta) \begin{pmatrix}
\mathbf{v}_{\alpha \beta} \\
0
\end{pmatrix} = \begin{pmatrix}
A & K \\
K & B - \frac{C}{r}
\end{pmatrix} \begin{pmatrix}
\mathbf{v}_{\alpha \beta} \\
0
\end{pmatrix} = \begin{pmatrix}
A \mathbf{v}_{\alpha \beta} \\
K \mathbf{v}_{\alpha \beta}
\end{pmatrix} = \begin{pmatrix}
A \mathbf{v}_{\alpha \beta} \\
0
\end{pmatrix}.
\]

So what is \(A \mathbf{v}_{\alpha \beta}\)? The matrix \(A\) is given by
\[
A = \left[ \beta_{y_i y_i} \cdot (\mathbf{v}_{p\alpha} - \mathbf{v}_{\alpha \beta}) + [\cos(\psi_{y_i}) \cos(\psi_{y_i}) - \delta_{ij}] \left( \frac{1}{||p - \alpha|| + \frac{1}{||\alpha - \beta||}} \right) \right]_{i,j}.
\]
Since the reflection at \(\alpha\) is tangential, \(\mathbf{v}_{p\alpha} = \mathbf{v}_{\alpha \beta}\) and the first term drops out. Then, using \((\mathbf{v}_{\alpha \beta})_j = \mathbf{v}_{\alpha \beta} \cdot \alpha_{y_j} = \cos(\psi_{y_j})\) the \(k^{th}\) component of \(A \mathbf{v}_{\alpha \beta}\) is given by
\[
(A \mathbf{v}_{\alpha \beta})_k = \sum_j [\cos(\psi_{y_k}) \cos(\psi_{y_j}) - \delta_{kj}] \left( \frac{1}{||p - \alpha|| + \frac{1}{||\alpha - \beta||}} \right) (\mathbf{v}_{\alpha \beta})_j
\]
\[
= \cos(\psi_{y_k}) \left( \sum_j \cos^2(\psi_{y_j}) \right) \left( \frac{1}{||p - \alpha|| + \frac{1}{||\alpha - \beta||}} \right) - \cos(\psi_{y_k}) \left( \frac{1}{||p - \alpha|| + \frac{1}{||\alpha - \beta||}} \right)
\]
\[
= \left[ \cos(\psi_{y_k})(1 - \cos(\psi_{y_k})) \right] \left( \frac{1}{||p - \alpha|| + \frac{1}{||\alpha - \beta||}} \right) = 0.
\]
This shows that whenever the first reflection of a two reflection path is a tangential reflection, the Hessian has a zero eigenvalue. Consequently a generalized billiard path with a tangential reflection is a degenerate critical point of $-L_k$. The same is true in general.

**Lemma 3.6** If there is a generalized billiard path with $k$ reflections connecting $p$ to $q$ and one or more of the reflections is tangential, then the length function $-\nabla L_k$ is not a Morse function.

**Proof** Let $\{\alpha_1, \ldots, \alpha_k\}$ be a generalized billiard path connecting $p$ to $q$, and suppose that there is a tangential reflection at $\alpha_i$. Let $v = \frac{\alpha_i - \alpha_{i+1}}{||\alpha_i - \alpha_{i+1}||}$. When the Hessian acts on the vector $(0; \ldots; 0; v; 0; \ldots; 0)^T$ (where $v$ is in the $i^{th}$ position) the result is

$$
\begin{bmatrix}
A_1 & K_1 \\
K_1 & A_2 & K_2 \\
& K_2 & \ddots \\
& & \cdots & K_{i-1} \\
& & K_{i-1} & A_i & K_i \\
& & & K_i & \ddots \\
& & & & \cdots & K_{k-1} \\
& & & & K_{k-1} & A_k
\end{bmatrix}
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
= 
\begin{pmatrix}
0 \\
A_i v \\
A_i v
\end{pmatrix}.
$$

The calculations above show that $A_i v = 0$ and $K_i v$. Since $\alpha_i$ is a tangential reflection, we know that

$$
v = \frac{\alpha_i - \alpha_{i+1}}{||\alpha_i - \alpha_{i+1}||} = \frac{\alpha_{i-1} - \alpha_i}{||\alpha_{i-1} - \alpha_i||}.
$$

As a result, the calculations also show that $K_{i-1} v = 0$. Consequently, zero is an eigenvalue of the Hessian. \qed
Lemma 3.6 shows that tangential reflections are bad news, but just how prevalent are they? Suppose $M$ is an $n$-manifold embedded in $\mathbb{R}^N$. We are free to choose the endpoints $p$ and $q$ as we see fit. Such a choice corresponds to choosing a point in $\mathbb{R}^{2N}$. Let us first consider the simplest case — a path with one reflection which is tangential — and count dimensions.

First, we must choose a point $\alpha$ on the manifold to be the reflection point. This is an $n$-dimensional space. Second, we can choose a unit vector $v_\alpha$ in $T_\alpha M$ to be the direction to the point $p$. This gives another $n-1$ dimensions of freedom. Third, we must choose a distance in this direction to be the length $\|p - \alpha\|$, contributing one more dimension. The direction to the point $q$ is already determined (it is the opposite of the direction to $p$). The distance from $\alpha$ to $q$ is the last choice we must make. It is another one-dimensional choice.

Adding this all up, the space of such paths is of dimension $n + (n-1)+1+1 = 2n+1$. If the codimension of $M$ is $u$, then $2n + 1 = 2N - (2u - 1)$. Since $u \geq 1$ this space has codimension at least one. It follows that the set of all $(p,q) \in \mathbb{R}^N \times \mathbb{R}^N$ such that the line segment connecting $p$ to $q$ is not tangent to $M$ is open and dense.

Now we must consider what happens when we allow multiple reflections. Again we need to choose a reflection point $\alpha$ and a unit vector $v_\alpha$, a $2n-1$ dimensional choice, but after that the situation changes a bit. When the choice of a length $\ell$ is made, we must check to see if the line segment $\{tv_\alpha : t \in (0,\ell)\}$ intersects $M$. If it does, there is a choice to be made, whether to pass through the manifold or reflect. If there is no reflection, then everything is as before. If there is we need to look a little more closely.

Recall how a path must behave at a reflection point. The bisector of the angle $\angle \alpha \beta \gamma$ must be normal to $T_\beta M$. Another way to say this is to require the outgoing
unit vector \( v \) to satisfy

\[
\pi_\beta (v) = \pi_\beta \left( \frac{\beta - \alpha}{\|\beta - \alpha\|} \right).
\]

The space of such vectors is isomorphic to the unit circle in the normal bundle \( N_\beta M \). If \( M \) has codimension \( u \) then this space has dimension \( u - 1 \). This looks like bad news. We want the dimension of the space of endpoints of tangential paths to be less than \( 2N \). If the space of possible choices increases for each reflection, it will eventually exceed this number.

The outlook brightens, though, when we remember that the vector from our reflection point must point toward another point in \( M \) (or the endpoint \( p \)). The cone

\[
V = \{ tv : t > 0 \text{ and } v \text{ is a possible reflection direction} \}
\]

has dimension \( u \). Consequently a transverse intersection \( V \cap M \) will have dimension zero. In this case there will be a countable number of possible reflections that will lead toward another point in \( M \). If we wish to forgo another reflection, there is still a \( u \)-dimensional space of choices. Let's add all this up:

\[
\begin{align*}
 n &= N - u & \text{choice of a tangential reflection point} \\
 n - 1 &= N - u - 1 & \text{choice of a tangential direction} \\
 0 &= \text{reflections en route to } p \\
 u &= \text{choice of endpoint } p \\
 0 &= \text{reflections en route to } q \\
 u &= \text{choice of endpoint } q
\end{align*}
\]

\[
2N - 1.
\]

This count depends, of course, on the intersection of \( M \) with the appropriate cones to be transverse. This prompts us to make the following
Definition 3.2 Say the embedding $M \hookrightarrow \mathbb{R}^N$ is \textit{cone-transverse} if for each $p \in M$ and unit vector $v \in T_pM$ and $c \in [0, 1]$ the cone

$$\{ w \in T_p \mathbb{R}^N : \frac{w}{||w||} \cdot v = c \}$$

intersects $M$ transversely.

Now we can state the following

Lemma 3.7 If $M \hookrightarrow \mathbb{R}^N$ is a cone-transverse embedding, then the set of $(p; q) \in \mathbb{R}^N \times \mathbb{R}^N$ such that all generalized billiard paths from $p$ to $q$ with up to $k$ reflections are non-tangential is open and dense.

Proof The set of endpoints

$$T_k = \{(p; q) \in \mathbb{R}^{2N} : \text{a tangential billiard path with } k \text{ reflections connects } p \text{ to } q\}$$

is a finite union of sets with codimension one. The complement of this set is then a finite intersection of open dense sets. Since a finite intersection of open dense sets is open and dense, $\mathbb{R}^{2N} - T_k$ is open and dense in $\mathbb{R}^{2N}$. \qed

Lemma 3.8 For a given non-degenerate (i.e. eigenvalues of the corresponding Hessian at that point are all non-zero) generalized billiard path $P$, as the endpoint $q$ is varied, the eigenvalues of the Hessian vary continuously.

Proof Let $x$ be a coordinate system near $\alpha_k$, the last reflection of $P$. For the other reflections, $\alpha_i$, let $y^i = (y^i_1, \ldots, y^i_n)$ be a coordinate system in a neighborhood. We may, without loss of generality, assume that the domain of $(y^1; \ldots; y^{n-1}; x)$ is contained in the interior of $X_k$. 
Recall the formulae for the entries of the Hessian, given in Lemma 3.3. If the endpoint $q$ is moved to $q'$, the vector

$$v_k = \frac{\alpha_k - q'}{||\alpha_k - q'||}$$

varies continuously with $q'$. The vectors $v_{k-1} = \frac{\alpha_{k-1} - \alpha_k}{||\alpha_{k-1} - \alpha_k||}$, $\alpha_k, x_j$, and $\alpha_i, y_j$ are all constant. It follows that $\frac{\partial (-L_{k_q}^{(p,q)})}{\partial y_j}$ is zero and $\frac{\partial (-L_{k_q}^{(p,q')})}{\partial x_i}$ varies continuously with $q'$. So the gradient $\nabla (-L_{k_q}^{(p,q')})$ varies continuously with $q'$. It follows that there is a generalized billiard path $P'$ whose reflections are close to the reflections of $P$. Moreover, $P'$ varies continuously with $q'$.

In addition, the quantity $||\alpha_k - q'||$ and the angle $\psi_{x_i}$ vary continuously with $q'$. It follows that the entries of the Hessian $H(P')$ vary continuously, and hence so do the eigenvalues.

Lemma 3.8 shows that for a given cone-transverse embedding, $M \hookrightarrow \mathbb{R}^N$, the set of pairs $(p, q) \in \mathbb{R}^N \times \mathbb{R}^N$ such that $-L^{(p,q)}$ is a Morse function is open in $\mathbb{R}^{2N}$.

**Lemma 3.9** For each non-tangential generalized billiard path $P$, as the length $\ell$ from the last reflection to $q$ is increased, the eigenvalues of the Hessian $H(P_{\ell})$ increase strictly monotonically.

**Proof** Recall from the proof of Lemma 3.3 that the Hessian can be written as

$$H(P_{\ell}) = N - \frac{1}{\ell} \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C \end{bmatrix},$$

where $N$ is a constant matrix and

$$C = (\delta_{ij} - \cos(\psi_{x_i}) \cos(\psi_{x_j}))_{i,j}.$$
A standard result of linear operator theory tells us that the $i^{th}$ eigenvalue (in increasing order) is given by

$$\lambda_i = \sup_{\{X_1, \ldots, X_{i-1}\} \text{ linearly independent}} \inf_{V \in \{X_1, \ldots, X_{i-1}\}^\perp} \left( \frac{\langle HV, V \rangle}{\langle V, V \rangle} \right),$$

where $V = (v_1; \ldots; v_k)$.

The value $\lambda_i$ can be realized by choosing $x_j$ to be an eigenvector corresponding to $\lambda_j$ and $v$ to be an eigenvector corresponding to $\lambda_i$. Because of this, we may restrict the inf to those vectors $V \in \{X_1, \ldots, X_{i-1}\}^\perp$ with $v_k \neq 0$. (All eigenvectors are of this form.) We may also restrict our attention to those vectors with $\|v_k\| = 1$.

If $v$ is a unit vector, then a calculation shows

$$\langle Cv, v \rangle = \langle v, v \rangle - v^T \cdot \cos(\psi_{x_i}) \cos(\psi_{x_j})_{i,j} \cdot v$$

$$= 1 - (v \cdot v_{\alpha_k})^2.$$

Thus $\langle Cv, v \rangle$ is positive unless $v = \pm v_{\alpha_k}$. This cannot be the case, though, since $P$ is a non-tangential reflection. It follows that when $\ell$ increases, the value of

$$\frac{\langle HV, V \rangle}{\langle V, V \rangle}$$

increases continuously for every vector $V$ with $v_k \neq 0$. As a consequence of this, we see that $\lambda_i$ must increase continuously as $\ell$ increases.

**Lemma 3.10** Given two points $p, q \in \mathbb{R}^N$ having no tangential generalized billiard paths connecting them and an $\epsilon > 0$, there exists a $q' \in B_\epsilon(q)$ such that $-L_{\ell}^{(p,q')}$ is a Morse function.
Proof Since $\mathbb{R}^{2N} - T_k$ is open, there is an $\epsilon > 0$ such that $p \times B_k(q) \cap T_k = \emptyset$. First consider the case where there is only one degenerate billiard path, $P_1$. Let $v_1$ be the unit vector pointing from the last reflection to $q$. As $q'$ is moved in the direction of $v_1$ from $q$, all the eigenvalues of $P_1$ and all the non-degenerate paths vary continuously (Lemma 3.8 and Lemma 3.9). Choose a $0 < \delta < \epsilon$ so that all of the non-zero eigenvalues are bounded away from zero between $q$ and $q + \delta v_1$.

By Lemma 3.9, the eigenvalues of the path $P_1$ will increase monotonically. Hence all of the zero eigenvalues will have increased to positive values. Since no new zero eigenvalues have been created, we can set $q' = q + \delta v_1$.

Now suppose there is more than one degenerate billiard path (perhaps infinitely many). First choose a path $P_1$, as above, and $0 < \delta_i < \epsilon/2$ so that the Hessian $H(P_1)$ of the length function $L_{k}^{(p,q+\delta_i v_1)}$ is non-degenerate and no new zero eigenvalues have been created.

Repeat this process, each time choosing a degenerate billiard path $P_i$ and $0 < \delta_i < \epsilon/2^i$. After a finite number of iterations, all of the degenerate billiard paths will have been eliminated. Otherwise, we would find infinitely many non-degenerate critical points. This is impossible, because $X_k$ is compact.

Let $m$ be the number of times we must repeat the procedure. Since

$$d(q, q + \delta_1 v_1 + \cdots + \delta_m v_m) \leq \sum_{i=1}^{m} \delta_i < \epsilon$$

for all $m > 0$, we can set $q' = q + \delta_1 v_1 + \cdots + \delta_m v_m$. □

Lemma 3.11 For a given cone-transverse embedding $M \hookrightarrow \mathbb{R}^N$ the set of pairs $(p, q) \in \mathbb{R}^N \times \mathbb{R}^N$ such that $-L^{(p,q)}$ is a Morse function is open and dense in $\mathbb{R}^{2N}$. 


Proof Lemma 3.7 shows the set is open and Lemma 3.10 shows that it is dense.

\[
\]

3.5 Application of the Morse Inequalities

In this section, we finally apply the results of Chapter 2 to the case of \(-L_k : X_k \to \mathbb{R}\).

**Theorem 3.1** Suppose \( M \hookrightarrow \mathbb{R}^N \) is a smooth embedding of an \( n \)-manifold, and \( p, q \in \mathbb{R}^N \). Then for every \( \epsilon > 0 \), there is a \( p' \in B_{\epsilon}(p) \) and a \( q' \in B_{\epsilon}(q) \) such that if \( N_k \) is the number of billiard paths with \( k \) reflections connecting \( p' \) to \( q' \). Then

\[
N_k \geq \sum_{i=0}^{kn} b_i(X_k).
\]

**Proof** Choose \( p' \) and \( q' \) so that \(-L_k\) is a Morse function. Then

\[
N_k = \sum_{i=0}^{kn} m_i(-L_k).
\]

Since \( m_i(-L_k) \geq b_i(X_k) \), the result follows.

This shows that the more complicated the topology of the path space \( X_k \), the greater the number of generalized paths there must be. In Chapter 3 and Chapter 4, we will relate the number of billiard paths directly to the topology of the underlying manifold \( M \).
Chapter 4

Morse Theory for Some Other Stratified Spaces

In Chapter 3, we introduced the blow up space $X_k$, a compactification of $M^k - \Delta'$, where $\Delta'$ is the union of all the diagonals $\{\alpha_i = \alpha_{i+1}\}$. We then studied the length function $L_k$ on this manifold with corners. We followed this procedure because $L_k$ is singular on the set $\Delta'$. The Morse theory for manifolds with corners then allowed us to relate the number of generalized billiard paths to the topology of $X_k$.

In this chapter, we will show how one may study the function $L_k : M \times \cdots \times M \to \mathbb{R}$ directly, without blowing up anything. We still face problems on the set $\Delta'$, of course. Our approach now is to treat the product $M^k$ as a stratified space, define another modified gradient flow with respect to this stratification, and then prove the Morse theorems in this case. We can then use these theorems to relate the number of generalized billiard paths to the topology of the manifold $M$.

4.1 $M \times \cdots \times M$ as a Stratified Space

The first step is to think of $M^k$ as a stratified space. When we looked at manifolds with corners, the stratifications were a natural feature of the spaces in question. The space $M^k$ is naturally a smooth manifold. We must impose on this smooth space a stratified structure.

A point $P$ in $M^k$ is an ordered $k$-tuple of points in $M$, $P = (\alpha_1, \ldots, \alpha_k)$. Define $F_j(M^k) = \{(\alpha_1, \ldots, \alpha_k) : \alpha_i \neq \alpha_{i+1} \text{ for exactly } j + 1 \text{ choices of } i \in \{0, \ldots, k\}\}$. Here, as usual, $\alpha_0 = p$ and $\alpha_{k+1} = q$. $F_j(M^k)$ is the set of paths with $j$ distinct reflections. The components of $F_j(M^k)$ will be the strata of $M^k$. 
We need to show that these components satisfy the conditions in Definition 2.4. The following three lemmas take care of this task.

**Lemma 4.1** Each component of $F_j(M^k)$ is a manifold.

**Proof** There is a natural isomorphism of the tangent space at a point of $M^k$:

$$T_PM^k \cong T_{\alpha_1}M \oplus \cdots \oplus T_{\alpha_k}M.$$ 

Here we choose a coordinate chart $x_i$ for $M$ at each point $\alpha_i$. Then a coordinate chart for $M^k$ at $P$ is given by $(x_1; \ldots; x_k)$. A tangent vector in $M^k$ is a choice of a tangent vector at each point $\alpha_i \in M$. A basis for the full tangent space is given by a basis for the tangent space at each reflection.

At a point in $F_{k-1}(M^k)$ there is a single pair $\alpha_i = \alpha_{i+1}$. Instead of the coordinate chart above, we can think about the coordinate chart given by

$$(x_1; \ldots; x_i; x_{i+2}; \ldots; x_k; x_{i+1} - x_i).$$

This coordinate system takes a neighborhood $U \cap F_{k-1}(M^k)$ to $\mathbb{R}^{(k-1)n} \subset \mathbb{R}^{kn}$. This shows that $F_{k-1}(M^k)$ is a submanifold of $M^k$ with dimension $(k-1)n$.

In general, a component of $F_j(M^k)$ may be identified by a partition $I_1, \ldots, I_j$ of $\{1, \ldots, k\}$ such that $\ell$ and $m$ are in the same $I_i$ if and only if $\alpha_\ell = \alpha_m$. If we choose a representative $i_\ell \in I_\ell$, then the map

$$(x_1; \ldots; x_{i_\ell})$$

gives a coordinate chart from $U_P \cap F_j(M^k)$ to $\mathbb{R}^{jn}$. So $F_j(M^k)$ is a manifold, and hence each component is a manifold. □

**Lemma 4.2** $M^k$ is the union of the connected components of all the $F_j(M^k)$. 

Proof Every point \( P \in M^k \) is in some \( F_j(M^k) \), so
\[
M^k = \bigcup_j F_j(M^k).
\]
Moreover, each \( F_j(M^k) \) is the union of its connected components. The lemma follows directly.

It is not hard to see that if \( i \leq j \) then \( F_i(M^k) \subset \overline{F_j(M^k)} \). We can label the connected components of the sets \( \{F_i : i = 0, \ldots, k\} \) with the indexing set \( A \).

Lemma 4.3 If \( \rho, \sigma \in \mathcal{A} \) and \( H_\rho \cap \overline{H_\sigma} \neq \emptyset \) then \( H_\rho \subset \overline{H_\sigma} \).

Proof Let \( P \in H_\rho \) we will show that for any \( \varepsilon > 0 \) there is a point \( Q \in H_\sigma \) such that \( d_{M^k}(P, Q) < \varepsilon \).

Let \( I_1, \ldots, I_r \) be the partition corresponding to \( H_\rho \) and \( J_1, \ldots, J_s \) the partition corresponding to \( H_\sigma \). Since \( H_\rho \cap \overline{H_\sigma} \neq \emptyset \), it must be true that for each \( J_q \) there is an \( I_p \) such that \( J_q \subset I_p \).

To prove the lemma, it is sufficient to show that given any \( P \in H_\alpha \) and \( \varepsilon > 0 \) there is a \( Q \in H_\beta \) such that \( Q \in B_\varepsilon(P) \subset M^k \). We can accomplish this by defining \( Q \) as follows.

Let \( \delta = \frac{\varepsilon}{\sqrt{k}} \). Choose \( \beta_1 \in B_\delta(\alpha_1) \). For \( i > 1 \), if \( \{i-1, i\} \in J_q \) for some \( q \), set \( \beta_i = \beta_{i-1} \). Otherwise, choose \( \beta_i \in B_\delta(\alpha_i) \) in such a way that \( \beta_i \notin \{\beta_1, \ldots, \beta_{i-1}\} \).

Then \( Q \in H_\beta \) and
\[
d_{M^k}(P, Q) = \left( \sum_{i=1}^k d_{M^k}(\alpha_i, \beta_i)^2 \right)^{1/2} < \left( \sum_{i=1}^k \delta^2 \right)^{1/2} = \varepsilon,
\]
proving the lemma.

When \( H_\rho \) and \( H_\sigma \) satisfy the hypotheses of Lemma 4.3 we write \( H_\rho \preceq H_\sigma \).
4.2 A Continuous Flow on $M^k$

A considerable portion of our treatment of Morse theory for Manifolds with corners was devoted to showing the existence of a continuous flow along which the value of the Morse function was decreasing. This is an important point again here, but because of the foundation we have laid it will not cause as much difficulty.

In Section 3.4 we showed that under certain conditions $-L_k : X_k \to \mathbb{R}$ satisfies Property (3). It follows as a corollary that $+L_k : X_k \to \mathbb{R}$ satisfies Property (3) as well. The results of Sections 2.4 and 2.5 then may be applied to this function.

Section 2.4 tells us that the vector field $-\nabla (+L_k)$ can be modified to produce a vector field $G^+$ on $X_k$. Section 2.5 says that the vector field $G^+$ induces a continuous flow $\varphi^+ : X_k \times [0, \infty) \to X_k$. At the end of Section 3.3 we mention a map $g : X_k \to M^k$ that 'puts back together' the blow up $X_k$.

We want to define a flow $\vartheta$ on $M^k$ in the following manner. Given a point $p \in M^k$ choose a point $p' \in g^{-1}(p)$. Then set

$$\vartheta(p, t) = g \circ \varphi^+(p', t).$$

We must show, however, that this does not depend on the choice of $p'$.

**Lemma 4.4** If $p', p'' \in g^{-1}(p)$, then for every $t \in [0, \infty)$,

$$g \circ \varphi^+(p', t) = g \circ \varphi^+(p'', t).$$

**Proof** Remember that the point $p'$ in a stratum $S \subset X_k$ consists of a configuration of points in $M^k$, and a collection of infinitesimal tangent space diagrams. The map $g$ can be thought of as simply forgetting about the infinitesimal diagrams. Since the value of $L_k$ depends only on the configuration of reflections in $M^k$ and not on the infinitesimal diagrams, $L_k$ is constant on the set $g^{-1}(p)$. The vector $-\nabla L_k$ will be normal to $g^{-1}(p)$. 
It follows that the vector \( G(p') \) consists solely of a vector in \( T_pM^k \). Moreover, this vector depends only on the configuration of \( p \in M^k \). Correspondingly, the flow \( \varphi^+ \) changes the configuration \( P \) in \( M^k \), but not the infinitesimal diagrams. The result is that for \( \tau \) sufficiently small, \( g \circ \varphi(p', \tau) = g \circ \varphi(p'', \tau) \). Here \( \tau \) must be chosen small enough that neither \( \varphi(p', \tau) \) nor \( \varphi(p'', \tau) \) leave the stratum \( S \).

Suppose that \( \tau \) is large enough that, say, \( \varphi(p', t) \) enters a new spectrum. Then let \( \tau_0 = \sup \{ t \in (0, \infty) : \varphi(p', t) \text{ is in the stratum } S \} \). Let \( \{ t_i \}_{i=1}^\infty \subset (0, \tau_0) \) be a sequence that approaches \( \tau_0 \). Since \( \varphi(p', t) \) and \( \varphi(p'', t) \) are both continuous for \( t \in (0, \tau_0) \), we can say

\[
\varphi(p'', \tau_0) = \lim_{i \to \infty} \varphi(p'', t_i) = \lim_{i \to \infty} \varphi(p', t_i) \varphi(p', \tau_0).
\]

The process can then begin again from this point. \( \square \)

Thus the flow is well defined. The result is a flow that, starting from a point in \( E_i(M^k) \), flows along the vector field \( -\nabla(L_k|_{E_i}) \) until it runs into a lower dimensional stratum. It then flows in the new stratum by the negative gradient of the function restricted to that stratum. The flow will never enter a higher dimensional stratum, since that would increase the length of the path.

### 4.3 The Morse Theorems

The proofs of the Morse theorems follow very much the same outline as the proofs in Section 2.7. Before we may proceed we need to make a

**Definition 4.1** An essential critical point of \( L_k : M^k \to \mathbb{R} \) is a critical point of any of the functions \( L_k|_{E_i} : E_i \to \mathbb{R} \).

Because any path entering a higher dimensional stratum will increase length, every critical point is an essential critical point. We will use the following notation:

\( (M^k)_a = L_k^{-1}((-\infty, a]) \).
Lemma 4.5 If there are no essential critical points in $L_k^{-1}([a, b])$, then for each point $\vec{x}$ in $(M^k)_b$ there is a time $t$ such that $\phi(\vec{x}, t) \in (M^k)_a$.

Proof The set $L_k^{-1}([a, b])$ is compact, and contains no essential critical points. It follows that

$$\mu = \min_i \inf_{P \in L_k^{-1}([a, b])} \|\nabla(L_k|E_i)\| > 0$$

Since the directional derivative $(-\nabla(L_k|E_i))[f]$ is given by

$$(-\nabla(L_k|E_i))[L_k] = -\left(\sum_i \frac{\partial L_k}{\partial x_i} e_i\right)[L_k]$$

$$= -\sum_i \frac{\partial L_k}{\partial x_i} (e_i)[L_k]$$

$$= -\sum_i \frac{\partial L_k}{\partial x_i} \frac{\partial L_k}{\partial x_i}$$

$$= -\|\nabla(L_k|E_i)\|.$$ 

and $\frac{\partial}{\partial t} \phi = (-\nabla(L_k|E_i))[L_k]$, it follows that along flow lines, the value of $L_k$ is decreasing at a rate bounded away from zero. So for $t > \frac{b-a}{\mu}$, $\phi(P, t) \in (M^k)_a$ for all $P \in (M^k)_b$. \qed

Theorem 4.1 If $a < b$ and $L_k^{-1}([a, b])$ contains no essential critical points, then $(M^k)_a$ is a deformation retract of $(M^k)_b$, so the inclusion map $(M^k)_a \hookrightarrow (M^k)_b$ is a homotopy equivalence.

The proof of this theorem is in no way different from the proof of Theorem 2.1.

Theorem 4.2 Let $P$ be an essential critical point of $L_k$ with index $\lambda$. Set $L_k(p) = c$. Suppose that for some $\epsilon > 0$, $L_k^{-1}([c - \epsilon, c + \epsilon])$ contains no essential critical points other than $P$. Then $(M^k)_{c+\epsilon}$ is homotopy equivalent to $(M^k)_{c-\epsilon}$ with a $\lambda$-cell attached.
Proof Choose a coordinate system $x : U_P \to \mathbb{R}^{kn}$ in which we can write

$$f = f(P) - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_{n(k-j)}^2 + |x_{n(k-j)+1}| + \cdots + |x_{kn}|.$$ 

Then choose $\epsilon > 0$ sufficiently small so that $L_k^{-1}[c - \epsilon, c + \epsilon]$ contains no essential critical points other than $P$, and the image $x(U_P)$ contains the closed ‘ball’

$$\{(x_1, \ldots, x_n) : \sum_{i=1}^{n(k-j)} x_i^2 + \sum_{i=n(k-j)+1}^{kn} |x_i| \leq 2\epsilon\}$$

The proof from here will consist of the following three steps:

1. Define a region $H$.
2. Show $(M^k)_{c-\epsilon} \cup H \simeq (M^k)_{c+\epsilon}$.
3. Show $(M^k)_{c-\epsilon} \cup e^\lambda \simeq (M^k)_{c-\epsilon} \cup H$.

We begin by tweaking the function $L_k$ a bit. Choose a $C^\infty$ function $\mu : \mathbb{R} \to \mathbb{R}$ that satisfies

$$\mu(0) > \epsilon$$
$$\mu(r) = 0, \text{ for } r > 2\epsilon$$
$$-1 < \mu' \leq 0.$$ 

If we write

$$\xi = x_1^2 + \cdots + x_\lambda^2$$
$$\eta = x_{\lambda+1}^2 + \cdots + x_{n-j}^2$$
$$\zeta = |x_{n-j+1}| + \cdots + |x_n|,$$

then we can write $L_k = c - \xi + \eta + \zeta$.

Define a new function $\Gamma$ by

$$\Gamma = L_k - \mu(\xi + 2\eta + 2\zeta)$$
$$c - \xi + \eta + \zeta - \mu(\xi + 2\eta + 2\zeta).$$ 

We will use this function (and its level sets) to define the region $H$. 
Claim 4.1  The essential critical points of $\Gamma$ and $L_k$ are identical.

Outside our 'ball' of 'radius' $2\epsilon$, $\Gamma = L_k$ and so any critical points there must coincide. Inside, the function $L_k$ has a single essential critical point at $P$. To find the essential critical points of $\Gamma$ we must compute $d\Gamma$.

$$d\Gamma = (-1 - \mu')d\xi + (1 - 2\mu')d\eta + (1 - 2\mu')d\zeta,$$

where $d\zeta(v) = ||v||$. The coefficients $(-1 - \mu')$ and $(1 - 2\mu')$ are nowhere zero and $d\xi$ and $d\eta$ are simultaneously zero only at $p$. Thus $p$ is a critical point of $L_k$ and hence an essential critical point.

Claim 4.2  $\Gamma^{-1}(-\infty, c + \epsilon) = L_k^{-1}(-\infty, c + \epsilon)$.

Outside the set $B = \{\xi + 2\eta + 2\zeta \leq 2\epsilon\}$ we know that $\mu = 0$, so $\Gamma = L_k$. Inside this set, we see that

$$\Gamma \leq L_k = c - \xi + \eta + \zeta.$$ 

Equality holds on the boundary of $B$. Also,

$$c - \xi + \eta + \zeta \leq c + \left(\frac{1}{2}\xi + \eta + \zeta\right).$$

Here equality holds when $\xi = 0$. Finally, we note that

$$c + \left(\frac{1}{2}\xi + \eta + \zeta\right) \leq c + \epsilon.$$ 

Here again, equality holds on the boundary of $B$. So we see that within this set, $\Gamma \leq c + \epsilon$ and $L_k \leq c + \epsilon$, unless $\xi = 0$ and $\eta + \zeta = \epsilon$, in which case $\Gamma = L_k = c + \epsilon$.

Claim 4.3  $\Gamma^{-1}(-\infty, c - \epsilon)$ is a deformation retract of $(M^k)_{c+\epsilon}$. 
Consider the region $\Gamma^{-1}[c-\epsilon, c+\epsilon]$. It is compact, but does it contain any critical points? The only possibility is $P$, but

$$\Gamma(P) = c - \mu(0) < c - \epsilon,$$

so $P \notin \Gamma^{-1}[c-\epsilon, c+\epsilon]$ and Theorem 4.1 applies: $\Gamma^{-1}(-\infty, c - \epsilon]$ is a deformation retract of

$$\Gamma^{-1}(-\infty, c + \epsilon] = L_k^{-1}(-\infty, c + \epsilon] = (M^k)_{c+\epsilon}.$$

Define the region $H$ by

$$H = \Gamma^{-1}(-\infty, c - \epsilon] - (M^k)_{c-\epsilon}.$$

Recall that we have defined the $\lambda$-cell $e^\lambda$ such that $e^\lambda = \{Q : \xi(Q) < \epsilon, \eta(Q) = \zeta(Q) = 0\}$. Note that $e^\lambda \subseteq H$, since $\frac{\partial \xi}{\partial \xi} = -1 - \mu' < 0$ implies for $Q \in e^\lambda$,

$$\Gamma(Q) < \Gamma(P) < c - \epsilon.$$

Note also that $e^\lambda \cap (M^k)_{c-\epsilon} = \partial e^\lambda$.

**Claim 4.4** $(M^k)_{c-\epsilon} \cup e^\lambda$ is a deformation retract of $(M^k)_{c-\epsilon} \cup H$.

For each $t \in [0,1]$ we define a map $r_t : (M^k)_{c-\epsilon} \cup H \to (M^k)_{c-\epsilon} \cup H$ as follows:

**Case 1.** If $Q \in (M^k)_{c-\epsilon}$, set $r_t(Q) = Q$ for all $t$.

**Case 2.** If $Q \in H$ and $\xi(Q) < \epsilon$, then set

$$r_t(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, (1-t)x_{\lambda+1}, \ldots, (1-t)x_n).$$

Then $r_0$ is the identity map, and $r_0 : (M^k)_{c-\epsilon} \cup H \to (M^k)_{c-\epsilon} \cup e^\lambda$. Moreover, $r_t(Q) \in e^\lambda$ for each $t$, because $\frac{\partial \xi}{\partial \eta} > 0$ and $\frac{\partial \xi}{\partial \xi} > 0$. (Moving toward $e^\lambda$ decreases $\Gamma$.)

**Case 3.** If $\epsilon \leq \xi(Q) \leq \eta(Q) + \zeta(Q) + \epsilon$, then define $r_t$ by

$$r_t(x_1, \ldots, x_n) = (x_1, \ldots, x_\lambda, s_t x_{\lambda+1}, \ldots, s_t x_n).$$
where

\[ s_t = (1 - t) + t \left( \frac{\xi - \epsilon}{\eta + \zeta} \right)^{1/2}. \]

Here again, \( r_0 \) is the identity, and now \( r_1(Q) \in L_k^{-1}(c - \epsilon) \). It was shown in the proof of Theorem 2.2 that the functions \( s_t x_t \) are continuous as \( \xi \to \epsilon, \eta \to 0, \zeta \to 0 \).

Note that this definition agrees with Case 1 when \( \xi = \epsilon \) and with Case 2 when \( \xi - \eta - \zeta = \epsilon \). Thus \( r \) provides a deformation retraction of \( (M^k)_{c-\epsilon} \cup H \) to \( (M^k)_{c-\epsilon} \cup e^h \).

This concludes the proof of Theorem 4.2. \( \square \)

Just as in Chapter 2, these two theorems may be used to show that \( M^k \) is homotopy equivalent to a CW-complex having one \( \lambda \) cell for each essential critical point of \( L_k : M^k \to \mathbb{R} \) with index \( \lambda \). The Morse Inequalities also follow as before.
Chapter 5

Billiard Paths Revisited

This short chapter brings to bear the results of Chapter 4 in counting the number of billiard paths connecting two points near a manifold.

5.1 A Comparison of Essential Critical Points in $X_k$ and $M^k$

Let us investigate the relationship between critical points in $X_k$ and $M^k$. Recall the stratification of $M^k$. If a stratum $H$ is in $F_j(M^k)$, then its dimension is $nj$. Consequently, every essential critical point in $H$ has index at most $nj$. Consequently, we get the following

Lemma 5.1 Any essential critical point of $L_k : M^k \to \mathbb{R}$ with index

\[ \lambda > n(k - 1) \]

is in $F_k(M^k)$.

For any path $P \in F_k(M^k)$, the preimage $g^{-1}(P)$ consists of a single point, $P'$. It is easy to see that $P$ is an essential critical point of $L_k$ if and only if $P'$ is an essential critical point of $-L_k$, since

\[ -\nabla(L_k) = 0 \iff -\nabla(-L_k) = 0. \]

We showed at the start of Section 3.4 that all of the essential critical points of $-L_k : X_k \to \mathbb{R}$ lie in the interior of $X_k$. As a result of this we have the following

Lemma 5.2 The essential critical points of $-L_k : X_k \to \mathbb{R}$ are in one-to-one correspondence with the essential critical points of $L_k : M^k \to \mathbb{R}$ that lie in $F_k(M^k)$. 
5.2 Application of the Morse Inequalities

So we see that counting the number of generalized billiard paths with exactly \( k \) reflections is equivalent to counting the number of essential critical points in \( E_k(M^k) \). There are at least as many of these essential critical points as there are essential critical points with index greater than \( nk \).

**Theorem 5.1** The number of generalized billiard paths connecting \( p \) to \( q \) in the vicinity of a manifold \( M \) satisfies

\[
N_k^{(p,q)} \geq \sum_{j=0}^{n-1} \sum_{i_1 + \cdots + i_k = j} b_{i_1}(M) \cdots b_{i_k}(M).
\]

**Proof** If \( m_j(L_k) \) denotes the number of essential critical points of \( L_k : M^k \to \mathbb{R} \), then

\[
N_k^{(p,q)} \geq \sum_{j=0}^{nk} m_j(L_k) \\
\geq \sum_{j=n(k-1)+1}^{nk} m_j(L_k) \\
\geq \sum_{j=n(k-1)+1}^{nk} b_j(M^k).
\]

Since \( M^k \) is a manifold, we can use Poincaré duality to say \( b_i(M^k) = b_{nk-i}(M^k) \). Then we see

\[
N_k^{(p,q)} \geq \sum_{j=0}^{n-1} b_j(M^k).
\]

Using the Künneth Theorem, it can be shown that

\[
b_j(M^k) = \sum_{i_1 + \cdots + i_k = j} b_{i_1}(M) \cdots b_{i_k}(M).
\]
Now we can finally write

\[ N_n^{(p,q)} \geq \sum_{j=0}^{n-1} \sum_{i_1 + \cdots + i_k = j} b_{i_1}(M) \cdots b_{i_k}(M) \]

proving the theorem.

Two things are evident from this expression. First, the more complicated the topology of \( M \), the more generalized billiard paths there will be. The second is that as the number of reflections \( k \) increases, the number of generalized billiard paths with \( k \) reflections increases, and rather quickly.
Bibliography


