UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2018:1

Morse Theory and Handle Decomposition

Kawah Rasolzadah

Examensarbete i matematik, 15 hp
Handledare: Thomas Kragh & Georgios Dimitroglou Rizell
Examinator: Martin Herschend
Februari 2018

Department of Mathematics
Uppsala University

Abstract

Cellular decomposition of a topological space is a useful technique for understanding its homotopy type. Here we describe how a generic smooth real-valued function on a manifold, a so called Morse function, gives rise to a cellular decomposition of the manifold.

Contents

1 Introduction **2**

1.1 The construction of attaching a k-cell 2

1.2 An explicit example of cellular decomposition 3

2 Background on smooth manifolds **8**

2.1 Basic definitions and theorems 8

2.2 The Hessian and the Morse lemma 10

2.3 One-parameter subgroups 17

3 Proof of the main theorem **21**

4 Acknowledgment **35**

5 References **36**

1 Introduction

Cellular decomposition of a topological space is a useful technique we apply to compute its homotopy type. A smooth Morse function (a Morse function is a function which has only non-degenerate critical points) on a smooth manifold gives rise to a cellular decomposition of the manifold. There always exists a Morse function on a manifold. (See [2], page 43). The aim of this paper is to utilize the technique of cellular decomposition and to provide some tools, among them, the lemma of Morse, in order to study the homotopy type of smooth manifolds where the manifolds are domains of smooth Morse functions. This decomposition is also the starting point of the classification of smooth manifolds up to homotopy type. More precisely, we here prove a result which provides the fundamental building block in the construction of the cellular decomposition. We show that:

Theorem*: Let $f : M \longrightarrow \mathbb{R}$ be a smooth function on a smooth manifold with boundary, and let $p \in M$ be a non-degenerate critical point of index λ. Setting $f(p) = c$, suppose that*

$$f^{-1}[c-\epsilon, c+\epsilon] = \{q \in M : c-\epsilon \leq f(q) \leq c+\epsilon\}$$

is compact and contains no critical point of f other than p nor any boundary points, for some $\epsilon > 0$. Then for all sufficiently small ϵ, the set $f^{-1}(-\infty, c+\epsilon]$ has the homotopy type of $f^{-1}(-\infty, c-\epsilon]$ with a λ-cell attached.

We proceed to give definitions needed for the above theorem.

1.1 The construction of attaching a k-cell

Definition 1.1 *The set of all points $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ whose coordinates satisfy the condition that $|x| \leq 1$ is called a **k-cell** and is denoted by $\mathbf{e^k}$ throughout the text. Shortly, $e^k = \{x = (x_1, \ldots, x_k) \in \mathbb{R}^k : |x| \leq 1\}$.*

We have already defined a k-cell but we will also be needing to attach k-cells to topological spaces. The precise definition of "attaching a k-cell" can be given as follows.

Let $e^k = \{x \in \mathbb{R}^k : |x| \leq 1\}$ be a k-cell and Y any topological space. Denote the boundary of e^k by $\dot{e}^k$ ($\dot{e}^k = \{x \in \mathbb{R}^k : |x| = 1\}$).

Let $g : \dot{e}^k \longrightarrow Y$ be a continuous map. Attaching a k-cell to Y via g is denoted by $Y \cup_g e^k$ and is obtained as follows. First take the topological sum of Y and e^k (disjoint union of Y and e^k) and then identify each $x \in \dot{e}^k$ with $g(x) \in Y$, and endow the resulting space with the quotient topology.

If we let e^0 be a point and let its boundary $\dot{e}^0$ be empty, then attaching a 0-cell, e^0 to Y is by definition just the union of Y and a disjoint point.

Definition 1.2 *Let X and Y be topological spaces and $f, g : X \longrightarrow Y$, continuous maps. We say that f and g are **homotopic** if there exists a continuous map $F : [0,1] \times X \longrightarrow Y$ such that $F(0,x) = f(x)$ and $F(1,x) = g(x)\ \forall\ x \in X$. the map F is called a **homotopy** between f and g.*

Definition 1.3 *Two topological spaces X and Y are said to be of the **same homotopy type** or **homotopy equivalent** if there exists continuous maps $f : X \longrightarrow Y$ and $g : Y \longrightarrow X$ such that $g \circ f$ is homotopic to $id : X \longrightarrow X$ and $f \circ g$ is homotopic to $id : Y \longrightarrow Y$. The maps f and g are called **homotopy equivalences** and each one of them is said to be the **homotopy inverse** to the other one.*

We illustrate the technique of cellular decomposition via an explicit example and then by the provided tools, throughout the text, we will finally be able to generalize this explicit example by the theorem above.

This exposition and the results in it, closely follow the first eighteen pages of Milnor's Morse Theory. Most of the figures are also taken from there.

1.2 An explicit example of cellular decomposition

Here we give an explicit example of how a Morse function can be used to construct a cellular decomposition of a the torus.

Definition 1.4 *let X and Y be topological spaces. A bijective map $f : X \longrightarrow Y$ is a **homeomorphism** if it is continuous and has a continuous inverse. We call X and Y **homeomorphic** if there exists a homeomorphism between them.*

For demonstrating our example, we consider a particular embedded torus $M \subseteq \mathbb{R}^3$ and a tangent plane V which is tangent to M at the lowest point of M as indicated in figure 1[1].

Let $f : M \longrightarrow \mathbb{R}$ ($\mathbb{R}$ is representing the real numbers) be the ***height function*** (the function that measures the height of the torus M above the plane V), and let M^a be the set of all points x in the torus M such that

[1]Taken from: J. Milnor, 1968, p.1

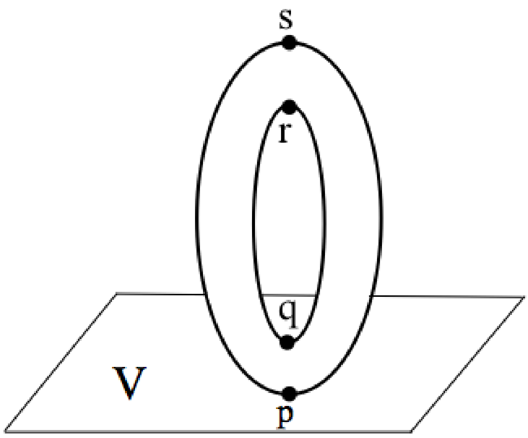

Figure 1

$f(x) \leq a$. We assume that f is Morse with exactly four critical points p, q, r, s, as shown in Figure 1. Now, the following statements are true:

(1) M^a is empty for any $a < 0 = f(p)$

(2) M^a is homeomorphic to a 2-cell for $f(p) < a < f(q)$

(3) M^a is homeomorphic to a cylinder for $f(q) < a < f(r)$. See figure 2[2].

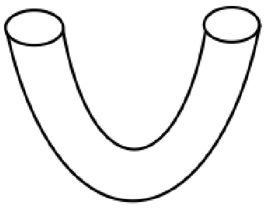

Figure 2

(4) M^a is homeomorphic to a genus one compact manifold having a circle as boundary for $f(r) < a < f(s)$. See figure 3[3].

(5) M^a is the full torus for $f(s) < a$.

It is possible to describe the change of the homeomorphism class at each moment (2) - (5), but it is not the point of our discussion here. Instead, we look at the change of the homotopy equivalence class of M^a. This is a weaker notion of equivalence, but it is more practical for computation.

[2]Taken from: J. Milnor, 1968, p.1
[3]Source: J. Milnor, 1968, p.2

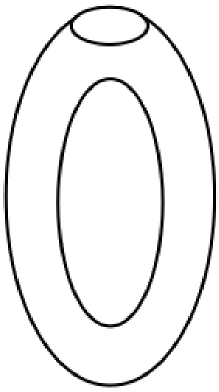

Figure 3

Now we are ready to describe the change of M^a as a passes through one of the points $f(p), f(q), f(r), f(s)$ in terms of homotopy type rather than homeomorphism type. So in terms of homotopy type:

(1) $\longrightarrow$ (2) (going from situation (1) to situation (2)) is just the operation of attaching a 0-cell to M^a where $a < f(p) = 0$. Because as far as we consider homotopy type, we cannot distinguish a point ($M^a \cup e^0$, $a < f(p) = 0$) from the space M^a for $f(p) < a < f(q)$. Hence we think of M^a, $f(p) < a < f(q)$ as a space, homotopy equivalent to a 0-cell instead of a space, homeomorphic to a 2-cell. See figure 4[4].

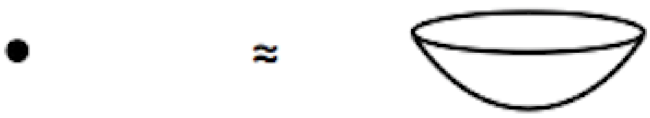

Figure 4

"$\approx$" means "is homotopy equivalent to" or "is of the same homotopy type as".

(2) $\longrightarrow$ (3) is the operation of attaching a 1-cell[5]. See figure 5[6].

(3) $\longrightarrow$ (4) is also the operation of attaching a 1-cell[7]. See figure 6[8].

[4] Taken from: J. Milnor, 1968, p.2
[5] J. Milnor, 1968, p.2
[6] Source: J. Milnor, 1968, p.2
[7] J. Milnor, 1968, p.2
[8] Source: J. Milnor, 1968, p.2

Figure 5

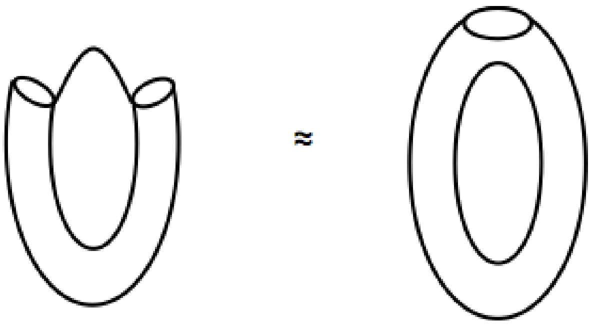

Figure 6

(4) $\longrightarrow$ (5) is the operation of attaching a 2-cell[9].

Remark 1.1 *Every homeomorphism $f : X \longrightarrow Y$ is a homotopy equivalence, simply by taking $g = f^{-1}$. But the converse is not true in general. An example could be the following: A 2-cell (a disc) and a 0-cell (a point) are homotopy equivalent (i.e. $\exists$ homotopy equivalence between them by shrinking the disc along the radial lines continuously to a single point) but there is no homeomorphism between them (there is clearly no bijection between them).*

Considering a homotopy type, M^a changes only when a passes through one of the points $f(p), f(q), f(r)$ and $f(s)$. So one may expect that the points p, q, r and s have a specific characterization. That characterization is that, they are critical points of $f : M \longrightarrow \mathbb{R}$. (See definition 2.4 for "critical point" of a map).

If we choose any coordinate system (x, y) near these points (See page 8 for a precise definition of a "coordinate system".), the derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ will both become zero at these critical points.

In the local coordinates (x, y) near the critical points, the function f can be expressed by $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$, i.e. $f = g \circ (x, y)$ near the critical points.

[9] J. Milnor, 1968, p.2

$$\begin{array}{ccc} M & \xrightarrow{(x,y)} & \mathbb{R}^2 \\ & {\scriptstyle f}\searrow \quad \swarrow{\scriptstyle g} & \\ & \mathbb{R} & \end{array}$$

Suppose that the height of the point q is a, i.e. $f(q) = a$. Choose a coordinate system (x, y) near q such that $(x, y)(q) = 0$. By the lemma of Morse (Lemma 2.3 which we will prove later), there are coordinates in which $f = g(x, y) = a - x^2 + y^2$.

Note that in the expression of g near each critical point the number of negative quadratic terms is the dimension of the cell that must be attached to be able to go from M^a to M^b, where $a < f(\text{critical point}) < b$.

For example near q, $f = c + x^2 - y^2$ has one negative quadratic term and hence, we need a 1-cell to go from M^a to M^b, where $a < f(q) < b$.

This exposition will entirely be about generalization of these facts for any differentiable function (not only $f : torus \longrightarrow \mathbb{R}$) on a manifold. (Definition of a "manifold" will be given later.)

2 Background on smooth manifolds

2.1 Basic definitions and theorems

Our results concern smooth manifolds, to which Morse theory can be applied. In the following we give the basic definitions.

Definition 2.1 *Let $U \subset \mathbb{R}^m$ be an open set, i.e. $\forall\ p \in U, \exists\ N_r(p)$ (ball centered at p of radius $r > 0$) such that $N_r(p) \subset U$. A map $f : U \longrightarrow \mathbb{R}^n$ is* ***smooth*** *(or differentiable of class C^∞) if it has continuous partial derivatives of all orders on U. For any arbitrary subset $X \subset \mathbb{R}^m$, a map $f : X \longrightarrow \mathbb{R}^n$ defined on X is called* ***smooth*** *if for each point $x \in X$, there exists an open set $U \subset \mathbb{R}^m$ and a smooth map $F : U \longrightarrow \mathbb{R}^n$ such that f is equal to F on $X \cap U$.*

Definition 2.2 *Let $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$. A bijective map $f : X \longrightarrow Y$ is a* ***diffeomorphism*** *if both f and its inverse $f^{-1} : Y \longrightarrow X$ are smooth. X and Y are* ***diffeomorphic*** *if such a map exists.*

The Inverse Function Theorem is used in the proof of Morse lemma. We recall the theorem here.

Theorem 2.1 *([2], p.13) Suppose that $f : X \longrightarrow Y$ is a smooth map whose derivative ∂f_x at the point x is an isomorphism. Then f is a local diffeomorphism at x.*

Definition 2.3 *Let X be a subset of the Euclidean space $\mathbb{R}^m$. X is called a k-dimensional* ***manifold*** *if it is locally diffeomorphic to $\mathbb{R}^k$, that is, for every point $p \in X$, there exists a neighborhood of x, $V \subset X$, such that V is diffeomorphic to an open set U of $\mathbb{R}^k$.*

A diffeomorphism $\varphi : U \longrightarrow V$ (U and V defined as above) is called a ***parameterization*** of V and the inverse diffeomorphism $\varphi^{-1} : V \longrightarrow U$ is called a ***coordinate system*** on V, and is often expressed as $(x_1, \ldots, x_k) : V \longrightarrow U$ to identify each point $p \in V$ as the point $(x_1(p), \ldots, x_k(p)) \in U$. The maps $x_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \ldots, k$ are called ***coordinate functions***.

Let $f : U \longrightarrow \mathbb{R}^n$ be a smooth map, where U is an open set of $\mathbb{R}^m$. Then the derivative of f at a point $p \in U$, denoted by df_p is a linear map $df_p : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ and is given by the matrix:

$$df_p = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_m}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(p) & \cdots & \frac{\partial f_n}{\partial x_m}(p) \end{pmatrix} \tag{1}$$

Let $M \subset \mathbb{R}^n$ be a k-dimensional manifold and p be a point in M. Let $\varphi : U \subset \mathbb{R}^k \longrightarrow V$ be a local parameterization of the neighborhood of p, where $\varphi(x) = p$. Then the tangent space of M at p, denoted by T_pM, is defined to be the image of $d\varphi_x : \mathbb{R}^k \longrightarrow \mathbb{R}^n$. T_pM is a subset of $\mathbb{R}^n$ and has the same dimension as the manifold M. (proof omitted.)

If $f : M \longrightarrow N$ is a smooth map where M and N are manifolds, then $df_p : T_pM \longrightarrow T_{f(p)}N$.

Definition 2.4 *Let $f : M \longrightarrow \mathbb{R}$ be a smooth real valued function on a manifold M. A point $p \in M$ is a **critical point** of f if the map $df_p : T_pM \longrightarrow T_{f(p)}\mathbb{R}$ is zero. By choosing a local coordinate system $(x_1, \ldots, x_n)$ on a neighborhood V of the critical point p, the definition means that*

$$\frac{\partial f}{\partial x_1}(p) = \cdots = \frac{\partial f}{\partial x_n}(p) = 0. \tag{2}$$

*A point p is called **regular** if df_p is surjective. The point $f(p) \in \mathbb{R}$ is called a **critical value** of f if p is a critical point of f and is called a **regular** value of f if p is a regular point of f.*

Recall that if $f : M \longrightarrow \mathbb{R}$ is the height function, then $M^a := \{x \in M : f(x) \leq a\}$.

Definition 2.5 *Let H^k be the upper half space i.e.*

$$H^k := \left\{(x_1, \ldots, x_k) \middle| \, x_j \in \mathbb{R}^k, x_k \geq 0\right\}. \tag{3}$$

*A subset $M \subset \mathbb{R}^n$ is a k-dimensional **manifold with boundary** if for every point $p \in M$, there exists a neighborhood V of p such that V is diffeomorphic to an open neighborhood U in H^k.*

Note that the property of being open is relative, i.e. it depends on the space in which U is imbedded. So openness in H^k is not the same as openness in $\mathbb{R}^k$. See figure 7.

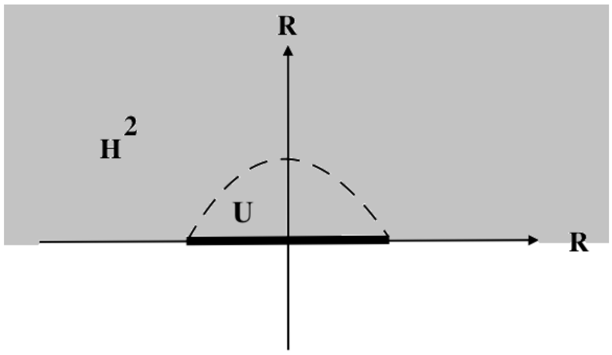

Figure 7: U is open in H^2 but not in $\mathbb{R}^2$

The following theorem (Preimage theorem), which will be given without proof, shows that M^a is a manifold with boundary if $a \in \mathbb{R}$ is a regular value of the function $f : M \longrightarrow \mathbb{R}$. The proof is based upon the Submersion theorem, which we omit.

Theorem 2.2 *(Preimage Theorem, [2], p. 21) Let X and Y be smooth manifolds and $f : X \longrightarrow Y$ be a smooth map. If $y \in Y$ is a regular value of f, then $f^{-1}(y)$ is a smooth manifold and $dim(f^{-1}(y)) = dim(X) - dim(Y)$.*

Note that the dimension equality makes sense as far as $dim(X) \geq dim(Y)$.

Now if $a \in \mathbb{R}$ is a regular value of $f : M \longrightarrow \mathbb{R}$, then $f^{-1}(a)$ by the Preimage theorem is a smooth manifold of a dimension which is equal to

$$dim(M) - 1 \qquad (dim(\mathbb{R}) = 1) \tag{4}$$

and $M^a = f^{-1}(-\infty, a]$ is a manifold with the boundary $f^{-1}(a)$ which also is a submanifold of M.

2.2 The Hessian and the Morse lemma

Definition 2.6 *A critical point p, of a map f, is called* ***non-degenerate*** *if there exists a coordinate system $(x_1, \ldots, x_n)$ on a neighborhood of p in which the Hessian matrix $H = (\frac{\partial^2 f}{\partial x_i x_j}|_p)$, $i, j = 1, \ldots, n$ is non-singular (i.e. H has maximal rank $\Longleftrightarrow det(H) \neq 0$)*

Lemma 2.1 *Non-degeneracy of a critical point does not depend on the choice of a coordinate system.*

Proof:

Let $x = (x_1, \ldots, x_n) : V \longrightarrow U \subset \mathbb{R}^n$ and $x' = (x'_1, \ldots, x'_n) : V \longrightarrow U' \subset \mathbb{R}^n$ be two coordinate systems on a neighborhood V of a critical point p. x and x' differ from each other by a local diffeomorphism $\psi : U' \longrightarrow U$.

Since both x and x' are diffeomorphisms, the following diagram:

$$\begin{array}{ccc} & V & \\ {}^{x'}\swarrow & & \searrow^{x} \\ U' & \xrightarrow[\psi]{} & U \end{array}$$

commutes, i.e. $\psi = x \circ x'^{-1} : U' \longrightarrow U$. We have

$$\begin{cases} x_1 = \psi_1(x'_1, \ldots, x'_n) \\ \vdots \\ x_n = \psi_n(x'_1, \ldots, x'_n) \end{cases}$$

and for $q \in V$, $(x_1, \ldots, x_n)(q) = \Big(\psi_1(x'_1(q), \ldots, x'_n(q)), \ldots, \psi_n(x'_1(q), \ldots, x'_n(q))\Big)$.

(In the below computation, the Einstein summation convention is used.)

Now

$$\begin{aligned} H' = \frac{\partial^2 f}{\partial x'_i \partial x'_j} &= \frac{\partial}{\partial x'_i}(\frac{\partial f}{\partial x'_j}) = \frac{\partial}{\partial \psi_l}\frac{\partial \psi_l}{\partial x'_i}(\frac{\partial f}{\partial \psi_k}\frac{\partial \psi_k}{\partial x'_j}) = \\ &\overset{\bullet}{=} \frac{\partial}{\partial x_l}\frac{\partial x_l}{\partial x'_i}(\frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial x'_j}) \\ &= \frac{\partial x_l}{\partial x'_i}(\frac{\partial^2 f}{\partial x_l \partial x_k} \cdot \frac{\partial x_k}{\partial x'_j} + \frac{\partial f}{\partial x_k} \cdot \frac{\partial^2 x_k}{\partial x_l \partial x'_j}) = \\ &= \frac{\partial^2 f}{\partial x_l \partial x_k} \cdot \frac{\partial x_l}{\partial x'_i}\frac{\partial x_k}{\partial x'_j} + \frac{\partial f}{\partial x_k}(\frac{\partial x_l}{\partial x'_i} \cdot \frac{\partial^2 x_k}{\partial x_l \partial x'_j}) \\ &\overset{\star}{=} \frac{\partial^2 f}{\partial x_l \partial x_k} \cdot \frac{\partial x_l}{\partial x'_i}\frac{\partial x_k}{\partial x'_j}. \end{aligned}$$

$\bullet : \frac{\partial x_i}{\partial x'_j} = \frac{\partial \psi_i}{\partial x'_j}$

$\star : \frac{\partial f}{\partial x'_k}|_p = 0$ since p is a critical point.

$\frac{\partial^2 f}{\partial x_l \partial x_k}$ is the Hessian of f in x_i-coordinates and $H' = J \cdot H \cdot J^t$ where $H = \frac{\partial^2 f}{\partial x_l \partial x_k}$, J is the Jacobian and J^t is the transpose of the Jacobian. The lemma follows since $det(H') = det(J)det(H)det(J^t)$ where $det(J) = det(J^t) \neq 0$

□

Before giving the next definition, We need to recall the definition of positive and negative definite respectively. (It is given in elementary linear algebra.)

If $\mathbf{v} = (v_1, \ldots, v_n)$ is a column vector in $\mathbb{R}^n$ and $A = (a_{ij})$ is an $n \times n$, real, symmetric matrix (i.e. $a_{ij} = a_{ji}$ for $1 \leq i, j \leq n$). Then the expression

$$Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v} = \sum_{i,j=1}^{n} a_{ij} v_i v_j$$

is called a ***quadratic form*** on $\mathbb{R}^n$ corresponding to the matrix A. Observe that $Q(\mathbf{v})$ is a real number for every n-vector $\mathbf{v}$.

We say that A is positive definite if $Q(\mathbf{v}) > 0$ for every nonzero vector $\mathbf{v}$ and similarly negative definite if $Q(\mathbf{v}) < 0$ for every non zero vector $\mathbf{v}$.

Definition 2.7 *The **index of the Hessian** H, at a non-degenerate critical point, of a smooth map $f : V \longrightarrow \mathbb{R}$, is defined to be the maximal dimension of a subspace of V on which H is negative definite.*

Observe that the Hessian of $f : M \longrightarrow \mathbb{R}$ at p is a symmetric matrix and its index is well-defined by lemma 2.1.

For simplicity, we say the index of f at p instead of the index of the Hessian of f on T_pM.

Now, we are almost ready for the lemma of Morse which describes the behavior of f at p by the index of f. But before stating Morse lemma, we need the following lemma.

Lemma 2.2 *Let f be a smooth function on a convex neighborhood V of 0 in $\mathbb{R}^n$, with $f(0) = 0$. Then*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i g_i(x_1, \ldots, x_n) \tag{5}$$

for some suitable smooth functions g_i defined on V, with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

Proof:

Let $x_i(t) = tx_i$. By the fundamental theorem of calculus and the assumption that $f(0) = 0$ as well as the chain rule we have

$$
\begin{aligned}
f(x_1,\dots,x_n) &= \int_0^1 \frac{df(tx_1,\dots,tx_n)}{dt}dt = \\
&= \int_0^1 \sum_{i=1}^{n} \frac{\partial f(tx_1,\dots,tx_n)}{\partial x_i} \cdot x_i dt \\
&= \sum_{i=1}^{n} \int_0^1 \frac{\partial f(tx_1,\dots,tx_n)}{\partial x_i} \cdot x_i dt.
\end{aligned} \tag{6}
$$

We can simply let $g_i(x_1,\dots,x_n)$ to be $\int_0^1 \frac{\partial f(tx_1,\dots,tx_n)}{\partial x_i}dt$ to establish the lemma.

□

Lemma 2.3 *(Lemma of Morse) Let p be a non-degenerate critical point of a smooth function $f : M \longrightarrow \mathbb{R}$ where M is a n-manifold. Then there is a local coordinate system $(y_1,\dots,y_n)$ in a neighborhood U of p with $y_i(p) = 0$ for all i and such that the equality*

$$
f = f(p) - y_1^2 - \cdots - y_\lambda^2 + y_{\lambda+1}^2 + \cdots + y_n^2 \tag{7}
$$

holds throughout U, where λ is the index of f at p.

Proof:

We first prove that if f is expressed as above then λ is the index of f at p.

If

$$
f(q) = f(p) - (x_1(q))^2 - \cdots - (x_\lambda(q))^2 + (x_{\lambda+1}(q))^2 + \cdots + (x_n(q))^2 \tag{8}
$$

for some coordinate system $(x_1,\dots,x_n)$, then

$$
\frac{\partial^2 f}{\partial x_i \partial x_j}(p) = \begin{cases} -2 & if \quad i = j \leq \lambda, \\ 2 & if \quad i = j > \lambda, \\ 0 & if \quad i \neq j. \end{cases} \tag{9}
$$

Hence the Hessian of f with respect to the basis $\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}|_p$ is

$$H = \begin{bmatrix} -2 & & & & & \\ & \ddots & & & & \\ & & -2 & & & \\ & & & 2 & & \\ & & & & \ddots & \\ & & & & & 2 \end{bmatrix}.$$

Now, for any column vector $\mathbf{v} = (v_1, \ldots, v_\lambda, 0_{\lambda+1}, \ldots, 0_n) \in \mathbb{R}^n$, the quadratic form $Q(\mathbf{v})$ on $\mathbb{R}^n$ corresponding to the matrix H has the form:

$$Q(\mathbf{v}) := \mathbf{v}^T H \mathbf{v} = -2v_1^2 - \cdots - 2v_\lambda^2 < 0 \tag{10}$$

and similarly for any column vector $\mathbf{v} = (0_1, \ldots, 0_\lambda, v_{\lambda+1}, \ldots, v_n)$,

$$Q(\mathbf{v}) = 2v_{\lambda+1}^2 + \cdots + 2v_n^2 > 0. \tag{11}$$

Therefore there is a subspace of T_pM (recall that the Hessian of f at p is defined on T_pM) of dimension λ on which H is negative definite and similarly a subspace V of T_pM of dimension $n - \lambda$, on which H is positive definite.

Fact 2.1 *λ is the maximal dimension, on which H is negative definite.*

Proof of 2.1

Any subspace of T_pM with dimension greater than λ, on which H is negative definite will be a subspace that intersects V, which is impossible by the fact that H is positive definite on V. Therefore λ is the index of f at p.

□

Now, we prove the remaining part of the lemma.

We can, without loss of generality, choose p to be the origin of $\mathbb{R}^n$ such that $f(p) = f(0) = 0$.

By lemma 2.2 we have

$$f(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j g_j(x_1, \ldots, x_n) \tag{12}$$

for $(x_1, \ldots, x_n)$, in some neighborhood of 0, where the smooth functions $g_j = \frac{\partial f}{\partial x_j}$.

Since we have assumed that 0 is a critical point,

$$g_j(0) = \frac{\partial f}{\partial x_j}|_0 = 0 \tag{13}$$

Therefore, lemma 2.2 can be applied to g_j and hence

$$g_j(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i h_{ij}(x_1, \ldots, x_n) \tag{14}$$

for certain smooth functions h_{ij}. By plugging equation (14) into equation (12), we get

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} x_i x_j h_{ij}(x_1, \ldots, x_n). \tag{15}$$

We can assume that $h_{ij} = h_{ji}$ by letting

$$H_{ij} = \frac{1}{2}(h_{ij} + h_{ji}). \tag{16}$$

Then in this form,

$$\begin{aligned} f &= \sum_{i,j=1}^{n} x_i x_j H_{ij} = \sum_{i=1}^{n}(x_i x_1 H_{i1} + \sum_{j=2}^{n} x_i x_j H_{ij}) = \\ &= \sum_{i=1}^{n} x_i x_1 H_{i1} + \sum_{i=1}^{n}\sum_{j=2}^{n} x_i x_j H_{ij} = \\ &= x_1^2 H_{11} + \sum_{i=2}^{n} x_i x_1 H_{i1} + \sum_{j=2}^{n} x_1 x_j H_{1j} + \sum_{i,j=2}^{n} x_i x_j H_{ij} = \\ &= x_1^2 H_{11} + 2\sum_{i=2}^{n} x_i x_1 H_{i1} + \sum_{i,j=2}^{n} x_i x_j H_{ij}, \end{aligned} \tag{17}$$

where $H_{ij} = H_{ji}$. Now

$$\frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_0 = h_{ij}(0) + h_{ji}(0) = 2H_{ij}(0). \tag{18}$$

So by the fact that $\frac{1}{2}(\frac{\partial^2 f}{\partial x_i \partial x_j})\Big|_0 = H_{ij}(0)$ and the assumption that 0 is a non-degenerate critical point of f, $H_{ij}(0)$ is non-singular (i.e. $det(H_{ij}(0) \neq 0)$, and therefore, we can apply a linear transformation to the coordinate function $x_1, \dots, x_n$ so that

$$\frac{1}{2}\frac{\partial^2 f}{\partial x_1^2} \neq 0. \tag{19}$$

Hence $H_{11}(0) \neq 0$.

We know that H_{11} is a smooth function (h_{ij}'s are smooth) and hence by having the property of continuity, $H_{11}(0) \neq 0$ in a neighborhood of 0.

Now let $\mathbf{y} = (y_1, \dots, y_n)$ be a new coordinate in the neighborhood of 0 such that

$$y_i = x_i \qquad \text{if} \quad i \neq 1 \tag{20}$$

and

$$y_1 = \sqrt{|H_{11}|}\Big(x_1 + \sum_{j=2}^{n} x_j \frac{H_{1j}}{H_{11}}\Big) \tag{21}$$

Thus by the Inverse Function Theorem (See Theorem 2.1) y is a diffeomorphism. (A coordinate system in our case.)

What remains to be shown is that f is expressible as equation (7).

Now, we have

$$\begin{aligned} y_1^2 &= |H_{11}|\Big(x_1 + \sum_{j=2}^{n} x_j \frac{H_{1j}}{H_{11}}\Big)^2 \\ &= \pm H_{11}x_1^2 \pm 2\sum_{j=2}^{n} x_1 x_j H_{1j} \pm \sum_{j=2}^{n} x_j^2 \frac{H_{1j}^2}{H_{11}}, \end{aligned} \tag{22}$$

where the signs of first and last term are positive if $H_{11} > 0$ and negative if $H_{11} < 0$.

By using equation (17)

$$y_1^2 = \pm f \pm \sum_{i,j=2}^{n} x_i x_j H_{ij} \pm \frac{1}{H_{11}}\Big(\sum_{j=2}^{n} x_j H_{1j}\Big)^2 \tag{23}$$

$\Longleftrightarrow$

$$f = \pm y_1^2 \pm \sum_{i,j=2}^{n} x_i x_j H_{ij} \pm \frac{1}{H_{11}} \Big(\sum_{j=2}^{2} x_j H_{1j} \Big)^2 \tag{24}$$

Now, by isolating the terms after $\pm y_1^2$ which depend on local coordinates x_i, $i \geq 2$, we reach the desired expression for f except that f has one term $\pm y_1^2$. This can easily be solved by induction on the number of y_i until we get

$$f = \pm y_1^2 \pm \cdots \pm y_\lambda^2 \pm \cdots \pm y_n^2. \tag{25}$$

□

Corollary 2.2.1 *Non-degenerate critical points are isolated.*

2.3 One-parameter subgroups

We conclude this section by a lemma which will be needed for proving the first theorem in the next section, but before giving the lemma, we need the definition of "1-parameter group of diffeomorphisms" and also some discussion about it.

Definition 2.8 *A **1-parameter group of diffeomorphisms** of a manifold M is defined to be a smooth (C^∞) map $\varphi : \mathbb{R} \times M \longrightarrow M$ with the following two properties* [10]*:*

i) For each $t \in \mathbb{R}$, the map $\varphi_t : M \longrightarrow M$ defined by $\varphi_t(q) = \varphi(t, q)$ is a diffeomorphism of M onto itself.

ii) For all $t, s \in \mathbb{R}$, $\varphi_{t+s} = \varphi_t \circ \varphi_s$

Recall that a tangent vector field can be equivalently seen as a so-called smooth derivation of functions (in the direction of the vector field).

Now let φ be a 1-parameter group of diffeomorphisms and let f be any smooth real-valued function. There exists an induced vector field X on M defined by the following identity:

$$X_q(f) := \lim_{h \longrightarrow 0} \frac{f(\varphi_h(q)) - f(\varphi_0(q))}{h}, \tag{26}$$

[10] J. Milnor, 1968, p.9 and p.10

where φ_0 is the identity and hence $\varphi_0(q) = q$.

In other words, this is a smooth choice of one vector for each tangent plane T_qM for $q \in M$.

We say that the vector field X generates the group φ. (φ is also called the flow of the vector field X.)

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

Proof:

Note that for any point $\gamma(t_0) = p$ with coordinate $(x_1, \ldots, x_n) \in \mathbb{R}^n$, the velocity vector is defined to be

$$\frac{d\gamma}{dt} = \sum_{i=1}^{n} \frac{\partial\gamma}{\partial x_j}\frac{dx_i}{dt}. \tag{27}$$

We first prove that if φ is a 1-parameter group of diffeomorphisms, generated by the vector field X, then φ is unique. For that, consider the curve $\varphi_t(q)$ in M for each constant $q \in M$, instead of $\gamma(t)$. Now we have

$$\begin{aligned} \frac{d(f \circ \varphi_t(q))}{dt} &= \lim_{h \longrightarrow 0} \frac{f(\varphi_{t+h}(q)) - f(\varphi_t(q))}{h} \\ &= \lim_{h \longrightarrow 0} \frac{f\big(\varphi_h(\varphi_t(q))\big) - f(\varphi_t(q))}{h} \\ &= \lim_{h \longrightarrow 0} \frac{f(\varphi_h(p)) - f(p)}{h} = X_{\varphi_t(q)}(f), \end{aligned} \tag{28}$$

where $\varphi_t(q) = p$. Thus the ordinary differential equation

$$\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)}, \tag{29}$$

is satisfied by the curve $\varphi_t(q)$ for each fixed q with the initial condition $\varphi_0(q) = q$.

By the theory of ordinary differential equations, there exists a unique locally defined solution depending smoothly on the initial condition.

Local solutions for such a differential equation allow us to conclude that for any point $p \in M$, there exists a neighborhood N of p and a number $\epsilon > 0$ such that the differential equation

$$\frac{d\varphi_t(q)}{dt} = X_{\varphi_t(q)} \qquad , \ \varphi_0(q) = q \tag{30}$$

has a unique smooth solution for q inside N and for $|t| < \epsilon$.

Choose $\epsilon > 0$ to be the smallest of such ϵ. Since K is compact, a finite number of the neighborhoods N will cover K. Hence φ_t is defined on K. Now, to see that φ_t is defined on $M \backslash K$, let $\varphi_t(q) = c$, $\forall q \in M \backslash K$ where c is a constant. It follows that $\varphi_t(q)$ will be a smooth solution for the differential equation for $|t| < \epsilon_0$, $\forall q \in M$.

For fixed $q \in M, \varphi_t(q)$ is a local solution of (30). If one starts walking from $q = \varphi_0(q)$, after time s, one will end up at the point $\varphi_s(q)$ and after walking furthermore a time t, one reaches the point $\varphi_{t+s}(q)$. But one reaches the same point (i.e. $\varphi_{t+s}(q)$) if one starts from $\varphi_s(q)$ and walks a time t. This shows that if $|t|, |s|, |t+s| < \epsilon$, then

$$\varphi_{t+s} = \varphi_t \circ \varphi_s. \tag{31}$$

Furthermore for $|t| < \epsilon_0$,

$$\varphi_{-t+t}(q) = \varphi_{-t} \circ \varphi_t(q) = \varphi_0(q) = q, \tag{32}$$

which shows that $\varphi_{-t} = \varphi_t^{-1}$. Hence φ_t^{-1} is a smooth local solution of the ordinary differential equation and the diffeomorphism property follows by the fact that φ_t is bijective.

What remains to be done to complete the proof, is to define φ_t for $|t| \geq \epsilon_0$.

Let $|t| > \epsilon_0$ for any number t. Then t is expressible as a multiple of $\epsilon_0/2$ with a reminder r such that $|r| < \epsilon_0/2$. If

$$t = m\Big(\frac{t_0}{2}\Big) + r, \ m \in \mathbb{Z}, \tag{33}$$

define φ_t by the following identities:

$$\varphi_t = \begin{cases} \varphi_{\epsilon_0/2} \circ \cdots \varphi_{\epsilon_0/2} \circ \varphi_r & \text{if} \quad m \geq 0 \\ \varphi_{-\epsilon_0/2} \circ \cdots \varphi_{-\epsilon_0/2} \circ \varphi_r & \text{if} \quad m < 0 \end{cases} \tag{34}$$

where the transformation is repeated m times.

φ_t is well-defined and smooth (it is the composition of smooth maps) and it also satisfies the identity $\varphi_{t+s} = \varphi_t \circ \varphi_s$.

So φ_t is defined for any number $t \in \mathbb{R}$.

$\square$

3 Proof of the main theorem

In this section, we will proof two theorems which generalize the ideas discussed in the introduction. Recall that if $f : M \longrightarrow \mathbb{R}$, where M is a manifold, then we defined

$$M^a = f^{-1}(-\infty, a] = \{x \in M : f(x) \leq a\}. \tag{35}$$

We will let M^a to be the above subset throughout this section also.

Definition 3.1 *Let X be a subset of a set Y. The injection $\iota : X \longrightarrow Y$, defined by $\iota(x) = x, \forall x \in X$ is called the* ***inclusion map****.*

Note that the inclusion map is the identity map id_X if $X = Y$.

Definition 3.2 *For every point $p \in M$ let $\langle \cdot, \cdot \rangle_p$ be a defined inner product on the tangent space T_pM of a manifold M at the point p which depends smoothly on p[11]. The collection of all these inner products is called a* ***Riemannian metric*** *on M.*

There always exists a Riemannian metric induced from $\mathbb{R}^n$ since $M \subseteq \mathbb{R}^n$.

Theorem 3.1 *Let $f : M \longrightarrow \mathbb{R}$ be a smooth (C^∞) function, where M is a manifold and $\mathbb{R}$ is the space of real numbers, and let also that $a < b$ be two points in $\mathbb{R}$. Assume that the set*

$$f^{-1}[a, b] = \{x \in M : a \leq f(x) \leq b\} \tag{36}$$

is compact, and it neither contains any critical points of f nor any boundary points of M. Then the following two statements hold:

1) There exists a diffeomorphism from M^a onto M^b.

2) M^a is a deformation retract (See definition 3.3 of deformation retract.) of M^b and in particular the inclusion map $\iota : M^a \longrightarrow M^b$ is a homotopy equivalence.

Note that $f^{-1}[a, b] \cong f^{-1}[a] \times [a, b]$.

Proof of the theorem:

[11] The dependence is in the sense that $\langle \cdot, \cdot \rangle_p$ changes as p changes and the smoothness means that for any two vector fields X and Y on M, the map $p \mapsto \langle X(p), Y(p) \rangle$ is a smooth function.

Pushing M^b down to M^a along orthogonal trajectories of hypersurfaces $f = c$ (c is a constant) is the general idea of the proof.

(Note/recall that hypersurfaces, in their embedding spaces, have codimension 1. A picture in mind could be as figure 8[12].)

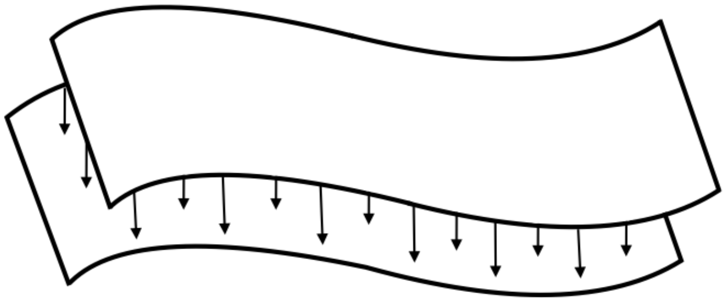

Figure 8

Let $\langle u, v\rangle$ be the inner product of two tangent vectors u and v, where $\langle \cdot, \cdot\rangle$ is defined by a chosen Riemannian metric on M. Recall from elementary calculus that the directional derivative of f along the vector u is given by $\langle u, grad\ f\rangle$, by definition of gradient. Hence the directional derivative of f along the vectors of any vector field X (or simply along X) is given by

$$\langle X, grad\ f\rangle, \quad X,\ grad\ f \in TM \tag{37}$$

and the gradient of f is characterized by

$$\langle X, grad\ f\rangle = X(f). \tag{38}$$

Note that $grad\ f\big|_p = 0$ for any critical point of f. Thus the vector field, $grad\ f$, vanishes at critical points of f. Recall also that if $\gamma : \mathbb{R} \longrightarrow M$ is a parameterized curve with the velocity vector $\frac{d\gamma}{dt}$ and $(x_1, \ldots, x_n)$ is a coordinate system on M, then

$$\langle \frac{d\gamma}{dt}, grad\ f\rangle = \sum_{i=1}^{n} \frac{d\gamma_i}{dt}\frac{\partial f}{\partial x_i} = \frac{d}{dt}(f \circ \gamma). \tag{39}$$

Now we need to introduce a new function. So let β be a smooth real-valued function on M which vanishes outside of some neighborhood of the compact

[12]Taken from: J. Milnor, 1968, p.12

subset $f^{-1}[a,b]$. Throughout the compact set $f^{-1}[a,b]$ (recall that $f^{-1}[a,b]$ does not contain any critical points of f as it is stated in the theorem and thus $grad\ f\Big|_{f^{-1}[a,b]} \neq 0$), we require β to satisfy

$$\beta = \frac{1}{\langle grad\ f, grad\ f\rangle}. \tag{40}$$

Let X be a vector field, defined by

$$X_q = \beta(q)(grad\ f)_q \in T_qM. \tag{41}$$

Since both β and $grad\ f$ are smooth and we have assumed that β vanishes outside of the compact set $f^{-1}[a,b]$, X is the vector field that satisfies the conditions that are sufficient (according to lemma 2.4) for a vector field to generate a 1-parameter group of diffeomorphisms. Thus X generates a 1-parameter group of diffeomorphisms.

$$\varphi_t : M \longrightarrow M. \tag{42}$$

Let $t \mapsto f(\varphi_t(q))$ ($f(\varphi_t(q) : \mathbb{R} \longrightarrow \mathbb{R}$) be a function for any fixed $q \in M$. If $\varphi_t(q)$ lies in the set $f^{-1}[a,b]$, then

$$\frac{d(f\circ\varphi_t(q))}{dt} = \langle\frac{d\varphi_t(q)}{dt}, grad\ f\rangle \tag{43}$$

follows by equation (39) and since $\varphi_t : M \longrightarrow M$ is the 1-parameter group of diffeomorphisms, generated by the vector field X on the compact set $f^{-1}[a,b]$, and at the same time $\varphi_t(q)$, by assumption lies in $f^{-1}[a,b]$, we have

$$\langle\frac{d\varphi_t(q)}{dt}, grad\ f\rangle = \langle X_q, grad\ f\rangle. \tag{44}$$

By equations (40) and (41) and that $\langle u,u\rangle = |u|^2$ for any vector u,

$$\begin{aligned}\langle X, grad\ f\rangle &= \frac{1}{|grad\ f|^2}\langle grad\ f, grad\ f\rangle \\ &= \frac{1}{|grad\ f|^2}\cdot |grad\ f|^2 = 1.\end{aligned} \tag{45}$$

Hence as far as the map $\varphi_t(q)$ lies in the set $f^{-1}[a,b]$ (and consequently $f(\varphi_t(q))$ has values between a and b) the map

$$t \mapsto f(\varphi_t(q)) \tag{46}$$

is linear and has derivative $+1$.

Now we know that $\varphi_{b-a} : M \longrightarrow M$ is a diffeomorphism. By assumption $f(\varphi_{b-a}(q)) \in [a,b] \ \forall \ q \in f^{-1}[a,b]$ and $f(\varphi_{b-a}(q))$ takes the value b at most and a at least. Now $\varphi_{b-a}(q) = \varphi_b \circ \varphi_{-a}(q)$ and $f(\varphi_{-a}(q)) \in [a,b]$. (Since $f(\varphi_t(q)) \in [a,b] \ \forall \ t \in \mathbb{R}$.)

Hence $\tilde{q} \coloneqq \varphi_{-a}(q) \in f^{-1}[a,b]$. φ_b is a diffeomorphism and $f(\varphi_b(\tilde{q}))$ is equal to b at most $\forall \ \tilde{q} \in f^{-1}[a,b]$. This shows that φ_{b-a} takes M^a diffeomorphically to M^b.

For proving the second part of the theorem, we need the following definitions.

Definition 3.3 *Let A be a subspace of a space (or a topological space) X. We say that A is a* ***retract*** *of X if there exists a continuous map $f : X \longrightarrow X$ with the following two properties:*

1) $f(x) \in A \ \forall x \in X$,

2) $f(x) = x \ \forall x \in A$.

We call f a ***retraction****.*

Equivalently, if we restrict the codomain of f to A i.e. $f : X \longrightarrow A$, then we say that A is a ***retract*** *of X (and f a* ***retraction****) if f satisfies the second condition above.*

Definition 3.4 *Let A be a subspace of a space (or a topological space) X. If there is a continuous map $F : [0,1] \times X \longrightarrow X$ such that*

1) $F(0,x) = x \quad \forall x \in X$,

2) $F(1,x) \in A \quad \forall x \in X$,

3) $F(1,x) = x \quad \forall x \in A$,

then F is called a ***deformation retraction*** *and A is called a* ***deformation retract*** *of X. If we substitute the third condition above with $F(t,x) = x \ \forall x \in A$, then we call F a* ***strong deformation retraction****.*

Observe that condition 1) says that $F(0,x) = id_X(x)$ and conditions 2) and 3) together say that $F(1,x) = f(x)$, where $f : X \longrightarrow X$ is a retraction from X to A.

Note that the deformation retraction $F : [0.1] \times X \longrightarrow X$ is a homotopy between id_X and the retraction (See previous definition.) $f : X \longrightarrow X$. Note also that viewing the retraction as $f : X \longrightarrow A$, then F is a homotopy between id_X and $\iota \circ f$, where ι is the inclusion map $\iota : A \longrightarrow X$. Combining this with the fact that $f \circ \iota = id_A$ (and hence $f \circ \iota$ and id_A are trivially homotopic) proves the following lemma.

Lemma 3.1 *Any deformation retraction gives a homotopy equivalence from X to A.*

For proving the second part of the theorem, let $\gamma_t : M^b \longrightarrow M^b$ be a 1-parameter family of maps, defined by

$$\gamma_t(q) = \begin{cases} q & \text{if} \quad f(q) < a \\ \varphi_{t(a-f(q))}(q) & \text{if} \quad a \le f(q) \le b. \end{cases} \tag{47}$$

So γ_0 is the identity map on M^b and $\gamma_1 : M^b \longrightarrow M^b$, by the proof of part 1) is a retraction from $M^{f(q) \le b}$ to M^a. Thus M^a is a deformation retract of $M^{f(q) \le b}$.

Now, let $\iota : M^a \longrightarrow M^b$ be the inclusion map and γ_1 as above but view $\gamma_1 : M^b \longrightarrow M^a$ instead. Then $\gamma_1 \circ \iota = id_{M^a}$ and hence $\gamma_1 \circ \iota \sim id_{M^a}$, ($\sim$ means "homotopic") and $\iota \circ \gamma_1 \sim id_{M^b}$ since γ is a homotopy between $\gamma_0 = id_{M^b}$ and $\iota \circ \gamma_1$. Hence the inclusion map $\iota : M^a \longrightarrow M^b$ is a homotopy equivalence and this completes the proof of theorem 3.1.

□

Remark 3.1 *The compactness of the set $f^{-1}[a,b]$ is important to be regarded in the theorem.*

If we for example remove a point p from the set $f^{-1}[a,b]$, then $f^{-1}[a,b] \setminus p$ is not compact (not closed anymore). See figure 9. Observe that in such a case, M^a is not a deformation retract of M^b, because in a deformation of a manifold, we are not allowed to puncture the manifold.

Theorem 3.2 *Let $f : M \longrightarrow \mathbb{R}$ be a smooth function on a smooth manifold with boundary, and let $p \in M$ be a non-degenerate critical point of index λ. Setting $f(p) = c$, suppose that*

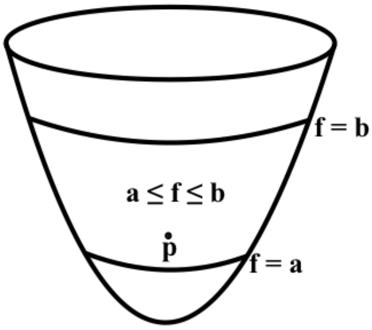

Figure 9

$$f^{-1}[c-\epsilon, c+\epsilon] = \{q \in M : c-\epsilon \le f(q) \le c+\epsilon\}$$

is compact and contains no critical point of f other than p nor any boundary points, for some $\epsilon > 0$. Then for all sufficiently small ϵ, the set $M^{c+\epsilon}$ has the homotopy type of $M^{c-\epsilon}$ with a λ-cell attached.

Before giving a rigorous proof of this theorem, giving the idea of the proof may make it easier to understand. We proceed as follows.

Let $f : M \longrightarrow \mathbb{R}$ be the height function on a torus. (See page 3 for definition of "height function".) We take a small neighborhood U of the critical point p and then define a new function $F : M \longrightarrow \mathbb{R}$ so that it is equal to the height function f outside of U and $F < f$ in the small neighborhood U. We define F in the way that the region $F^{-1}(-\infty, c-\epsilon]$ consists of

$$M^{c-\epsilon} = f^{-1}(-\infty, c-\epsilon] \tag{48}$$

as well as a region H near the critical point p.

In figure 10[13], the region $M^{c-\epsilon}$ is the red region and the region H is shown by horizontal lines.

We will see that the region $F^{-1}[c-\epsilon, c+\epsilon]$ is compact and contains no critical points and hence we can apply theorem 3.1 to $F^{-1}[c-\epsilon, c+\epsilon]$ and the function F, to show that $F^{-1}(-\infty, c-\epsilon] = M^{c-\epsilon} \cup H$ is a deformation retract of $F^{-1}(-\infty, c+\epsilon]$.

[13]Taken from: J. Milnor, 1968, p.14

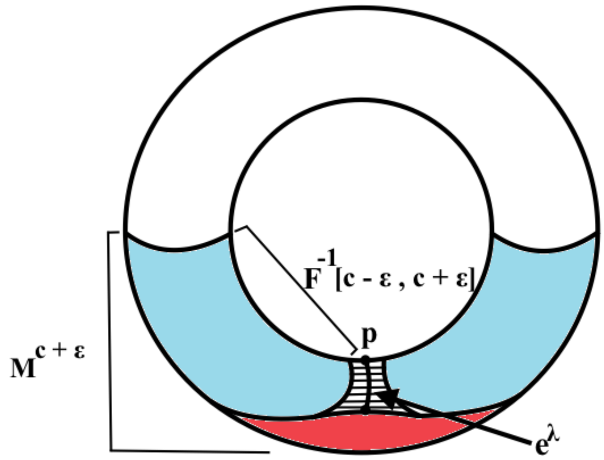

Figure 10

Choose a proper λ-cell, e^{λ} inside H (on the torus, it will be $e^1 \subset H$, and by proper λ-cell, we mean the one which allows us to reach our goal because, not every e^{λ} attached to $M^{c-\epsilon}$ is homotopy equivalent to $M^{c+\epsilon}$). Now what we need to do to complete the proof, is to show that $M^{c-\epsilon} \cup e^{\lambda}$ is a deformation retract of $M^{c-\epsilon} \cup H$ and this will simply be don by pushing H along the horizontal lines in H.

Proof of Theorem 3.2

Since p is a non-degenerate critical point of f with index λ, there exists by Morse lemma a coordinate system $(u_1, \ldots, u_n)\colon U \longrightarrow \mathbb{R}^n$ in the neighborhood of p U such that

$$f = c - (u_1)^2 - \cdots - (u_\lambda)^2 + (u_{\lambda+1})^2 + \cdots + (u_n)^2 \tag{49}$$

in U and that

$$u_i(p) = 0 \quad \text{for} \quad i = 1, \ldots, n \tag{50}$$

where $u_i(p)$'s are coordinates of the critical point p.

We want the region $f^{-1}[c-\epsilon, c+\epsilon]$ to be compact and to contain p ($f(p) = c$) as the only critical point of f in this region. For this goal, we choose a sufficiently small $\epsilon > 0$ to satisfy our requirement. Also let the same $\epsilon > 0$ be sufficiently small so that the closed ball

$$\{(u_1, \ldots, u_n) : u_1^2 + \cdots + u_n^2 \leq 2\epsilon\} \subset \mathbb{R}^2 \tag{51}$$

is contained in the image of the neighborhood U, under the coordinate system $(u_1, \ldots, u_n) : U \longrightarrow \mathbb{R}^n$. Recall that the property of being smooth is included in the definition of a coordinate system. So $(u_1, \ldots, u_n)$ is differentiable.

Also note that the image of U under $(u_1, \ldots, u_n)$ is embedded in $\mathbb{R}^n$, and throughout the proof, instead of looking at a point $q \in M$, we look at the image of q in $\mathbb{R}^n$ under this coordinate system.

Now define the λ-cell e^λ to be the set of points in U with

$$(u_1)^2 + \cdots + (u_\lambda)^2 \leq \epsilon \quad \text{and} \quad u_{\lambda+1} = 0 = \cdots = u_n = 0. \tag{52}$$

In figure 11[14] the situation so far is illustrated.

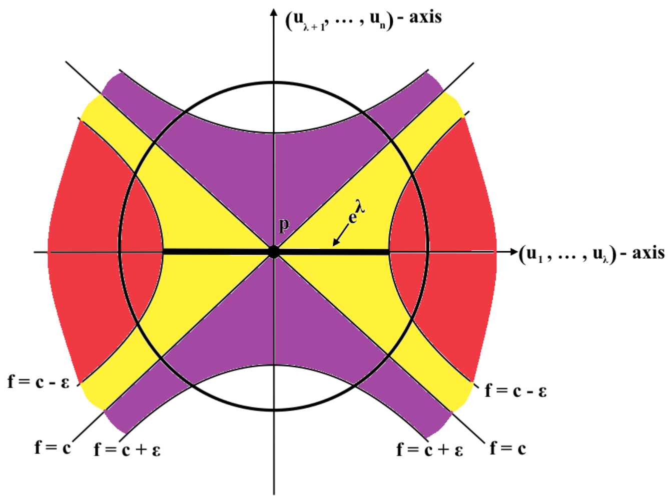

Figure 11

Note also that the circle which is the boundary of the ball, has radius $\sqrt{2\epsilon}$, and the hyperbolas (i.e. the curves $f = c - \epsilon$ and $f = c + \epsilon$) are the image of the hypersurfaces (i.e. $f^{-1}(c - \epsilon)$ and $f^{-1}(c + \epsilon)$) in M.

The red region represents the region $M^{c-\epsilon}$ and the regions marked with yellow and purple represent the regions $f^{-1}[c - \epsilon, c]$ and $f^{-1}[c, c + \epsilon]$ respectively.

[14]Taken from: J. Milnor, 1968, p.15

The horizontal line that goes through p and touches the hyperbolas $f = c-\epsilon$ represents the λ-cell e^λ and by (49) and (52) we see that

$$\dot{e}^\lambda = M^{c-\epsilon} \cap e^\lambda \tag{53}$$

where $\dot{e}^\lambda$ denotes the boundary of e^λ. Thus we certainly know that e^λ is attached to $M^{c-\epsilon}$ as we require in the theorem. (We must show that $M^{c-\epsilon} \cup e^\lambda$ is a deformation retract of $M^{c+\epsilon}$.)

Now we need to construct a smooth function $F : M \longrightarrow \mathbb{R}$. For that, let first

$$\mu : \mathbb{R} \longrightarrow \mathbb{R}$$

be a smooth function such that

$$\begin{cases} \mu(0) > \epsilon \\ \mu(r) = 0 \qquad \text{if} \ \ r \geq 2\epsilon \end{cases} \tag{54}$$

and

$$-1 < \mu'(r) = \frac{d\mu}{dr} \leq 0 \qquad \text{for} \ \ \forall\, r \in \mathbb{R}. \tag{55}$$

Let F be equal to f outside of the neighborhood $U \subset M$ and throughout U, define F as follows.

$$F = f - \mu((u_1)^2 + \cdots + (u_\lambda)^2 + 2(u_{\lambda+1})^2 + \cdots + 2(u_n)^2). \tag{56}$$

F is well-defined and the smoothness of F is inherited from the smoothness of f and μ.

For the sake of convenience, let

$$\begin{aligned} \xi &= (u_1)^2 + \cdots + (u_\lambda)^2 \qquad \text{and} \\ \eta &= (u_{\lambda+1})^2 + \cdots + (u_n)^2 \end{aligned} \tag{57}$$

where $\xi, \eta : U \longrightarrow [0, \infty)$.

Hence

$$f = c - \xi + \eta \tag{58}$$

and

$$F(q) = c - \xi(q) + \eta(q) - \mu(\xi(q) + 2\eta(q)) \tag{59}$$

for any q in the neighborhood U.

The proof will be established by four assertions.

<u>Assertion 1:</u> The regions $F^{-1}(-\infty, c+\epsilon]$ and $M^{c+\epsilon} = f^{-1}(-\infty, c+\epsilon]$ coincide with each other.

Proof: If $\xi + 2\eta > 2\epsilon$, then by (54) $\mu(\xi + 2\eta) = 0$ and hence $F = f$ outside of the region where $\xi + 2\eta > \epsilon$.

In the region $\xi + 2\eta \leq 2\epsilon$ (note that this is an ellipsoid) by (54) and (56) we have

$$F \leq f = c - \xi + \eta \leq c + \frac{1}{2}\xi + \eta \leq c + \epsilon. \tag{60}$$

The last inequality holds because $\xi + 2\eta \leq 2\epsilon \Longleftrightarrow \frac{1}{2}\xi + \eta \leq \epsilon$. It follows that the sets $F^{-1}(-\infty, c+\epsilon]$ and $f^{-1}(-\infty, c+\epsilon]$ coincide.

$\square$

<u>Assertion 2:</u> The functions F and f have the same critical points.

Proof: We have

$$dF = \frac{\partial F}{\partial \xi} d\xi + \frac{\partial F}{\partial \eta} d\eta \tag{61}$$

where

$$\begin{aligned} \frac{\partial F}{\partial \xi} &= -1 - \mu'(\xi + 2\eta) < 0 \\ \frac{\partial F}{\partial \eta} &= 1 - 2\mu'(\xi + 2\eta) \geq 1. \end{aligned} \tag{62}$$

$dF = 0$ if both $d\xi$ and $d\eta$ are zero and they are clearly zero at the critical point of f, p. So dF equals to zero only at p and hence p is the only critical point of F inside U.

□

By (54) and (56) and also the fact that $u_i(p) = 0$ for $i = 1, \ldots, n$ we see that

$$F(p) = c - \mu(0) < c - \epsilon \qquad (\mu(0) > \epsilon). \tag{63}$$

So in the level set, $F^{-1}(c - \epsilon)$ contains only points in U above p. Thus (63) combined with Assertion 1 implies that

$$F^{-1}[c - \epsilon, c + \epsilon] \subset f^{-1}[c - \epsilon, c + \epsilon]. \tag{64}$$

Therefore $F^{-1}[c - \epsilon, c + \epsilon]$ is compact. ($f^{-1}[c - \epsilon, c + \epsilon]$ is compact by assumption.)

By Assertion 2, p is the only critical point of F in U but

$$F(p) < c - \epsilon. \tag{65}$$

So p is not in the region $F^{-1}[c - \epsilon, c + \epsilon]$. Thus $F^{-1}[c - \epsilon, c + \epsilon]$ contains no critical points.

Assertion 3: The region $F^{-1}(-\infty, c - \epsilon]$ is a deformation retract of $M^{c+\epsilon} = f^{-1}(-\infty, c + \epsilon]$.

Proof: F is a smooth function and by above discussion $F^{-1}[c - \epsilon, c + \epsilon]$ is compact and does not contain any critical points. By theorem 3.1 the region $F^{-1}(-\infty, c - \epsilon]$ is a deformation retract of the region $F^{-1}(-\infty, c + \epsilon]$ and by Assertion 1, $F^{-1}(-\infty, c + \epsilon]$ coincides with $f^{-1}(-\infty, c + \epsilon] = M^{c+\epsilon}$. Hence $F^{-1}(-\infty, c - \epsilon]$ is a deformation retract of $M^{c+\epsilon}$.

□

From now on, for convenience denote the region $F^{-1}(-\infty, c - \epsilon]$ by $M^{c-\epsilon} \cup H$, where H is the closure of $F^{-1}(-\infty, c - \epsilon] \setminus M^{c-\epsilon}$. Note that $M^{c-\epsilon} \cup H$ by theorem 3.1 is diffeomorphic to $M^{c+\epsilon}$. We have already defined e^λ to be the set of points q in U such that

$$\xi(q) \leq \epsilon \qquad \text{and} \qquad \eta(q) = 0. \tag{66}$$

Note that e^λ is in the region H. To see this, let $q \neq p$ be any point in e^λ. Then by the fact that $\frac{\partial F}{\partial \xi} < 0$ inside U and that p is the origin of U where $\frac{\partial F}{\partial \xi}\big|_p = 0$ we have

$$F(q) \leq F(p) < c - \epsilon. \tag{67}$$

But by (49) and (52) $\forall\ q \in e^\lambda$, $q \neq p$

$$f(q) \geq c - \epsilon. \tag{68}$$

So the region $F^{-1}(-\infty, c-\epsilon] \setminus M^{c-\epsilon}$ contains e^λ except points q where $f(q) = c-\epsilon$. But since H is the closure of this region, e^λ is entirely inside H. In figure 12[15], H is the region containing the vertical arrows and e^λ is the horizontal line inside H with ends at $f = c - \epsilon$. The red region is $M^{c-\epsilon}$ and the region $F^{-1}[c-\epsilon, c+\epsilon]$ is shown by blue color.

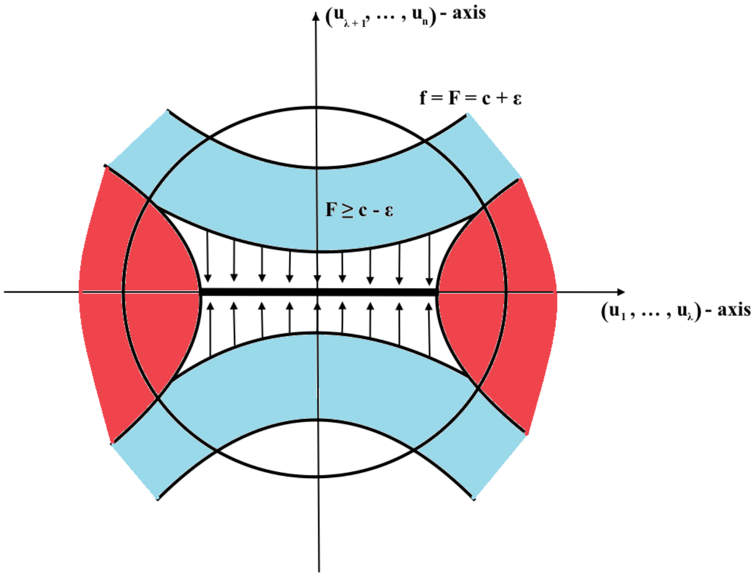

Figure 12

<u>Assertion 4:</u> $M^{c-\epsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\epsilon} \cup H$.

Proof: The idea of the proof is to push the region H toward e^λ along the vertical arrows in H as indicated in figure 12. But for a precise proof of the

[15]Taken from: J. Milnor, 1968, p.18

assertion we need to look at three cases in the neighborhood U and define a function

$$r_t : M^{c-\epsilon} \cup H \longrightarrow M^{c-\epsilon} \cup H \tag{69}$$

for each case. The three cases are illustrated in figure 13[16].

We let r_t be the identity function outside of U.

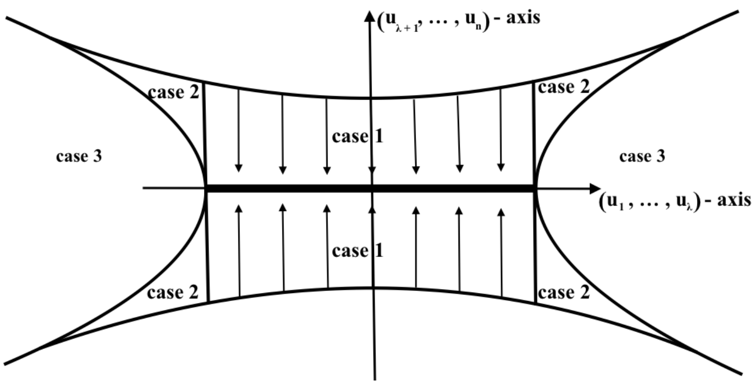

Figure 13

<u>Case 1:</u> Define r_t to be the transformation

$$(u_1, \ldots, u_n) \longrightarrow (u_1, \ldots, u_\lambda, tu_{\lambda+1}, \ldots, tu_n) \tag{70}$$

in the region $\xi \leq \epsilon$ (i.e. the set of points $q \in H$ such that $\xi(q) \leq \epsilon$).

Now r_1 is clearly the identity map and r_0 maps the entire region into e^λ because all points in this region after this transformation will have coordinates $\xi \leq \epsilon$ and $\eta = 0$ which coincides with e^λ. Note that for any $t \in \mathbb{R}$, the map r_t stays in the region $H^{c-\epsilon} \cup H$. This is simply because we have $\frac{\partial F}{\partial \eta} > 0$. So the value of F does not increase by change of t.

<u>Case 2:</u> Let s_t be a number in the interval $[0, 1]$ defined by

$$s_t = t + (1 - t)\sqrt{\frac{\xi - \epsilon}{\eta}}. \tag{71}$$

Now define r_t in the region $\epsilon \leq \xi \leq \eta + \epsilon$ to be the transformation

[16] Taken from: J. Milnor, 1968, p.19

$$(u_1, \dots, u_n) \longrightarrow (u_1, \dots, u_\lambda, s_t u_{\lambda+1}, \dots, s_t u_n). \tag{72}$$

Note that the area indicated by case 2 in figure 13 represents the region in which

$$\begin{cases} \xi \geq \epsilon \quad \text{and} \\ f = c - \xi + \eta \geq c - \epsilon \Longleftrightarrow -\xi + \eta \geq -\epsilon \end{cases} \tag{73}$$

i.e. $\epsilon \leq \xi \leq \eta + \epsilon$. Now $s_1 = 1$ and consequently r_1 is again the identity map. $s_0 = \sqrt{\frac{\xi - \epsilon}{\eta}}$, hence the region is entirely mapped into the hypersurface $f^{-1}(c - \epsilon)$. To see this, we have to show that for any point q in this region (i.e. where $\epsilon \leq \xi(q) \leq \eta(q) + \epsilon$), $f(q) = c - \epsilon$ after transforming q by r_0. $f(q) = (c - \xi + \eta)\Big|_q$. After transforming q by r_0 we have

$$\begin{aligned} (c - \xi + \eta)\Big|_q &= (c - \xi + \frac{\xi - \epsilon}{\eta} u_{\lambda+1}^2 + \dots + \frac{\xi - \epsilon}{\eta} u_n^2)\Big|_q \\ &= (c - \xi + (\frac{\xi - \epsilon}{\eta})\eta)\Big|_q = c - \epsilon. \end{aligned} \tag{74}$$

Note that in the region $\xi = \epsilon$, r_0 maps every point to $\dot{e}^\lambda$ (the boundary of e^λ) and on the hypersurface $f^{-1}(c - \epsilon)$ where $\xi - \epsilon = \eta$, r_t is the identity. Note also that the definitions of r_t in cases 1 and 2 are the same if $\xi = \epsilon$.

Case 3: In the region in which $f = c - \xi + \eta \leq c - \epsilon$ (i.e. where $\xi - \epsilon \geq \eta$ or simply $M^{c-\epsilon}$), let r_t again be the identity map which coincides with the definition of r_t in case 2 if $\xi - \epsilon = \eta$.

So $M^{c-\epsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\epsilon} \cup H$ and by Assertion 3, $M^{c-\epsilon} \cup H$ is a deformation retract of $M^{c+\epsilon}$. Hence $M^{c-\epsilon} \cup e^\lambda$ is a deformation retract of $M^{c+\epsilon}$. Thus by Lemma 3.1

$$M^{c+\epsilon} \approx M^{c-\epsilon} \cup e^\lambda. \tag{75}$$

Thereby the proof of Theorem 3.2 is completed.

4 Acknowledgment

I would like to thank both of my advisors Thomas Kragh and Georgios Dimitroglou Rizell for helping me throughout this work. I have learned from both of them and I am grateful for that.

5 References

[1] Milnor, John W., 1968. *Morse Theory.* Third Edition. Princeton, New Jersey: Princeton university press.

[2] Guillemin, Victor, & Pollack, Alan, 1974. *Differential Topology.* USA: American Mathematical Society.

[3] Milnor, John W., 1965 *Topology from the Differentiable Viewpoint.* Princeton university press.

[4] Do Carmo, Manferedo P., 1976. *Differential Geometry of Curves & Surfaces.* Revised & updated second edition. New Jersey: Prentice-Hall, Inc., Englewood Cliffs.

[5] Spivak, Michael, 1999. *A Comprehensive Introduction to Differential Geometry.* Volume one. Third Edition. Houston Texas.