Sard's Theorem, Morse Theory, and Applications

MAT367: Differential Geometry Essay

Jacob Chmura

This document gives a brief introduction to Morse Theory. We begin by motivating the study of specific types of functions on manifolds - called Morse functions, as a way to extract rich information about the space. After an illustrative example, we proceed to the definition of Morse Functions. We give a brief introduction to Measure Theory, culminating in Sard's Theorem which roughly says that the image of the set of critical points is "small". Then, after some definitions from algebraic topology, we return to Morse functions, and show their existence, and abundance. We work towards the Morse Lemma, and three important Theorems in Morse Theory, which describe intimate connections between the critical points of Morse Functions, and the topology of our manifold. We omit some proofs and focus on the intuition. Finally, we discuss interesting application of these ideas from a computational perspective, drawing on the disciplines of topological data analysis, and computer graphics.

Contents

1 Motivation **2**

2 Morse Functions **3**

2.1 Derivatives and Critical Points 3

2.2 The Hessian 3

2.3 Morse Functions 5

3 A Small Detour To Measure Theory **6**

3.1 Measure on $\mathbb{R}^n$ 6

3.2 Measure on Smooth Manifolds 7

3.3 Sard's Theorem 8

4 A Small Detour To Algebraic Topology **10**

4.1 CW Complexes 10

4.2 Homotopy Equivalence 10

5 The Existence, Density, and Local Form of Morse Functions **12**

5.1 Existence 12

5.2 Density 12

5.2.1 Morse Functions Are Generic 12

5.3 Local Form 13

5.3.1 Morse Lemma 13

6 The Morse Theorems **15**

6.1 Riemmanian Metrics and The Gradient 15

6.2 1-Parameter Group of Diffeomorphisms and Gradient Flow Lines 15

6.3 Stable and Unstable Manifold 16

6.3.1 Theorem 1 17

6.4 Critical Values and Topology 18

6.4.1 Theorem 2 18

6.4.2 Theorem 3 18

7 Applications **21**

7.1 Topological Data Analysis 21

7.2 Computer Graphics 22

8 References **24**

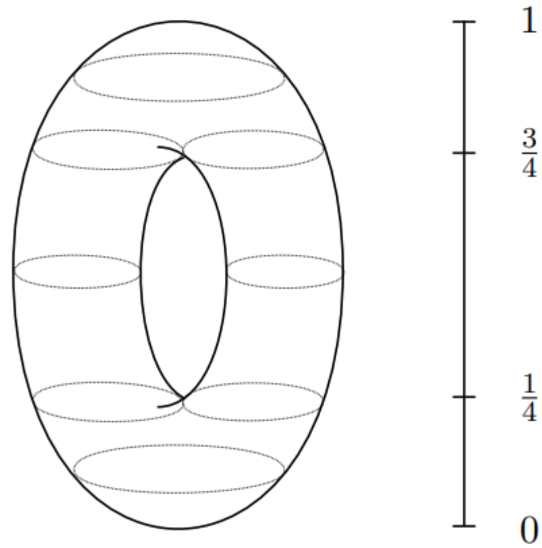
1 Motivation

The study of manifolds is extremely important. Perhaps we can summarize manifold theory as a tool that let's us do *flat* mathematics on *curved spaces*. And this is useful because *flat space*, namely $\mathbb{R}^n$ is *nice*, and because *curved spaces* are what one encounters in *practice* - in physics, chemistry, computer graphics, and optimization. How can we better understand what our curved space M looks like?

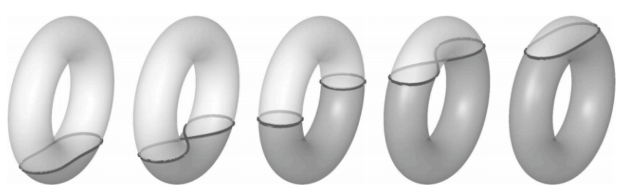
The central idea of *Morse Theory* is this:

A well chosen function $f : M \longrightarrow \mathbb{R}$ encodes a lot of information about M

Example 1. *Consider the torus $\mathbb{T}^2$ nicely embedded in $\mathbb{R}^3$: [10]*



and the height function $h : \mathbb{T}^2 \longrightarrow \mathbb{R}$. What happens at the critical values of h? Let's look at the sublevelsets of h as we sweep through it's values, giving the following horizontal slices: [10]



Notice how when we cross the critical points, our space changes in a fundamental way: $h^{-1}(\infty, a]$, $a < 0$ is the empty set, then $h^{-1}(\infty, a]$, $0 < a < \frac{1}{4}$ a disc, then $h^{-1}(\infty, a]$, $\frac{1}{4} < a < \frac{3}{4}$ a cylinder, then $\frac{3}{4} < a < 1$ changes again, and finally $h^{-1}(\infty, a]$, $a \geq 1$ recovers $\mathbb{T}^2$. As we cross through the critical points of h, the boundary and topology of the resultant space changes.

In the previous example, we saw how the height function h, through it's critical values, encoded topological information about our manifold $\mathbb{T}^2$. We will keep this example in mind, as we proceed to formalize and generalize the notions here, in the pursuit of studying rich properties that can be encoded via suitable functions on a manifold.

2 Morse Functions

2.1 Derivatives and Critical Points

We begin with familiar definitions.

Definition 1. Let M be a smooth manifold, and $f : M \longrightarrow N$ be a smooth map, and $p \in M$. The *differential of f at p* is the map $f_{*,p} : T_pM \longrightarrow T_{f(p)}N$ defined as follows. If $v \in T_pM$, then $f_{*,p}(v) \in T_{f(p)}N$ is the derivation:

$$f_{*,p}(v)(g) = v(g \circ f) \in \mathbb{R} \tag{1}$$

where $g \in C^{\infty}_{f(p)}(N)$

Proposition 2. *The differential of a map in local coordinates is the matrix of partial derivatives; the* Jacobian.

Remark. In what follows, we make the identification $T_p\mathbb{R}^n \sim \mathbb{R}^n$.

Definition 3. Let M be a smooth manifold and $f : M \longrightarrow \mathbb{R}$ a smooth function.

A $p \in M$ is called a *critical point of f* if the differential $f_{*,p} : T_pM \longrightarrow \mathbb{R}$ is the zero linear functional: $f_{*,p} \equiv 0$.

A $c \in \mathbb{R}$ is called a *critical value of f* if $\exists p \in M$ such that p is a critical point and $f(p) = c$.

On $\mathbb{R}^n$ we are familiar with the Hessian Matrix as the matrix of second order partial derivatives. We learn that algebraic properties of the Hessian give us further characterisation of the critical points of a C^2 function. Now we describe it's analog on smooth manifolds.

2.2 The Hessian

Definition 4. Let $f : M \longrightarrow \mathbb{R}$ be a smooth map, and $p \in M$ be a critical point. We define the *Hessian of f at p* as the map $f_{**,p} : T_pM \times T_pM \longrightarrow \mathbb{R}$ given by:

$$f_{**,p}(v, w) = v_p(\tilde{w}(f)) \tag{2}$$

where $v, w \in T_pM$ and $\tilde{w}$ is any local extension of w.

Lemma 5. *The Hessian is a well defined, symmetric bilinear form on T_pM*

Proof. [3] For symmetry, we note that if $\tilde{v}$ is an extension of v:

$$f_{**,p}(v, w) - f_{**,p}(w, v) = v_p(\tilde{w}(f)) - w_p(\tilde{v}(f)) = [\tilde{v}, \tilde{w}]_p(f) = f_{*,p}([\tilde{v}, \tilde{w}]) = 0 \tag{3}$$

since p is a critical point of f. Now we can see that $f_{**,p}$ is well-defined since:

$$\tilde{v}_p(\tilde{w}(f))) = v_p(\tilde{w}(f)) \tag{4}$$

is independent of our extension of v, and by symmetry:

$$\tilde{v}_p(\tilde{w}(f)) = \tilde{w}_p(\tilde{v}(f))) = w_p(\tilde{v}(f)) = \tilde{v}_p(w(f)) \tag{5}$$

shows our definition is independent of our extension of w. Therefore the hessian is well-defined.

Then by linearity of derivations, for $c \in \mathbb{R}$ and $v_1, v_2, w \in T_pM$:

$$f_{**,p}(cv_1 + v_2, w) = c \cdot v_{1,p}(\tilde{w}(f)) + \cdot v_{2,p}(\tilde{w}(f)) = c \cdot f_{**,p}(v_1, w) + f_{**,p}(v_2, w) \tag{6}$$

and bi-linearity of f_{**} follows from symmetry. □

We saw earlier that the coordinate description of the differential of a map reduces to the Jacobian of that map: the matrix of partial derivatives. We find an analogous result for the hessian.

Proposition 6. *The hessian of a map in local coordinates is the matrix of second order partial derivatives.*

Proof. Let $f : M \longrightarrow \mathbb{R}$ be a smooth map, and $p \in M$ be a critical point of f. Let $(U, \phi = x^1, .., x^m))$ be coordinates around p on M, and write:

$$v = \sum v_i \frac{\partial}{\partial x^i}|_p \tag{7}$$

$$w = \sum w_j \frac{\partial}{\partial x^j}|_p \tag{8}$$

$$\tag{9}$$

where $v_i, w_j \in C^\infty(U)$. In fact, since p is a critical point of f, we can take $w_j \in \mathbb{R}$ to be constants. Then:

$$f_{**,p}(v, w) = v_p(\sum w_j \frac{\partial f}{\partial x^j}) = \sum_{i,j} v_i w_j \frac{\partial^2 f}{\partial x^i x^j}(p) \tag{10}$$

Hence the matrix $\frac{\partial^2 f}{\partial x^i x^j}(p)$ of second order partial derivatives with respect to the local coordinates around p yield the hessian of f at p. □

Our intrinsic definition of the hessian shows that the it's rank in the local coordinate description is well-defined. This leads us to the following definitions:

Definition 7. A critical point p of $f : M \longrightarrow \mathbb{R}$ is called *non-degenerate* if the matrix

$$\frac{\partial^2 f}{\partial x^i x^j}(p) \tag{11}$$

is non-singular, for coordinates $x^1, .., .x^n$ about p.

Definition 8. Let $p \in M$ be a critical point of $f : M \longrightarrow \mathbb{R}$. The *index of p* is defined as the number of negative eigenvalues of if the matrix

$$\frac{\partial^2 f}{\partial x^i x^j}(p) \tag{12}$$

for coordinates $x^1, .., .x^n$ about p.

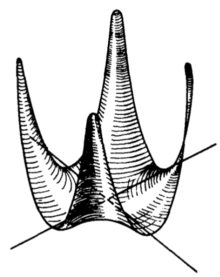
Equivalently, it is the maximal dimension of a subspace of T_pM on which $f_{**,p}$ is negative-definite. Intuitively, this is the number of independent directions we can travel on our manifold, such that going in that direction decreases the value of f.

2.3 Morse Functions

Any smooth function on a compact manifold will have critical points by virtue of the extreme value theorem. A condition we might like to impose is that each of these critical points be non-degenerate. In turn, this will mean that at each critical point, the quadratic term in the Taylor expansion of our function is non-degenerate. We give such functions a name.

Definition 9. Let $f : M \longrightarrow \mathbb{R}$ be a smooth function. We say that *f is a Morse function* if every critical point of f is non-degenerate.

Example 10. *[3] Let* $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *be given by* $f(x, y) = x^2y^2$



The set of critical points, all of which are degenerate consists of the union of the x and y axis. In any case, f is not *a morse function.*

Example 11. *This example shows one way in which we can from new Morse Functions out of existing ones. If* $f : M \longrightarrow \mathbb{R}$ *and* $g : M \longrightarrow \mathbb{R}$ *are Morse, then* $f + g : M \times N \longrightarrow \mathbb{R}$ *defined as:*

$$(f + g)(x, y) = f(x) + g(x) \tag{13}$$

is Morse. And the Critical points are pairs of critical points of f and g.

Proof. By C^∞-linearity of the differential, we have that $d(f + g) = df + dg$ so (p, q) is a critical point of $f + g$ if and only if p is a critical point of f and q is a critical point of g. Similarly the Hessian of $f + g$ is the sum of the hessian of f and the hessian of g, so at a critical point (p, q) the fact that f and g are non-degenerate at p and q, imply that $f + g$ is non-degenerate at (p, q). $\square$

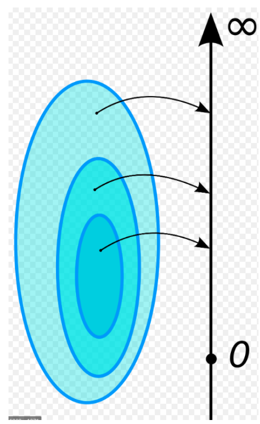
At face value, it is unclear how weak or how strong this condition actually is. After a small detour to measure theory, and algebraic topology, we will get an answer to this question.

3 A Small Detour To Measure Theory

The goal of this section is to say something about how *large* the set of critical values of a smooth map is, in the measure-theoretic sense, and this will come up when return to studying Morse Functions. We begin by recalling a few basic properties of zero-measure sets in $\mathbb{R}^n$, and then define the same on smooth manifolds using charts.

3.1 Measure on $\mathbb{R}^n$

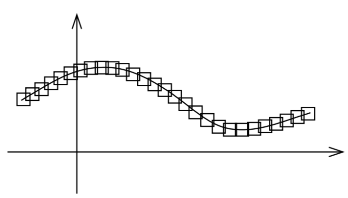
An indispensable tool in integration theory is the notion of *measure*. Broadly speaking, the idea is to assign a meaningful real number to specific classes of subsets, that determine how *big* a set might be.



Of particular important are those sets which we deem *negligible*: the sets of measure *zero*. Intuitively, these sets can be ignored by integration methods, for they provide no substance to the integration area in question. Of course, abstract measure theory is involved, so for this exposition, we will stick to a particular construction in $\mathbb{R}^n$.

Definition 1. A subset $A \subset \mathbb{R}^n$ is said to have *measure zero* if for any $\delta > 0$, A can be covered by a countable collection of open rectangles (or open balls), whose volume is less than δ.

[1]



Remark. We say a property ϕ on points $p \in \mathbb{R}^n$ holds *almost everywhere* if the set of points where it fails has measure zero:

$$\phi \ a.e. \quad \iff \{p \in \mathbb{R}^n : \neg\phi(p)\} \ has\ measure\ zero \tag{14}$$

We know state some basic properties of sets with measure zero, omitting proof, but giving intuition for why they are true.

Proposition 2. *We have:*

- *If $A \subset \mathbb{R}^n$ has measure zero, then the translated set $A + p = \{a + p : a \in A\}$ also has measure zero.*

- *Every subset of a set with measure zero has measure zero*
- *A countable union of sets of measure zero has measure zero*

Remark. If we can cover A by rectangles whose collective volume is arbitrarily small, then translating those rectangles to cover $A + p$ will remain a cover, and also have arbitrarily small volume.

If $B \subset A$ and A can be covered by rectangles whose collective volume is arbitrarily small, then this will also be a cover of B, and will also have arbitrarily small volume.

If $A := \{A_i\}_{i \in \mathbb{N}}$ is a collections of sets, we can cover A_i by volume $\frac{\varepsilon}{2^i}$. Then the union A can be covered by rectangles of volume $\sum_{i \in \mathbb{N}} \frac{\varepsilon}{2^i} < \varepsilon$.

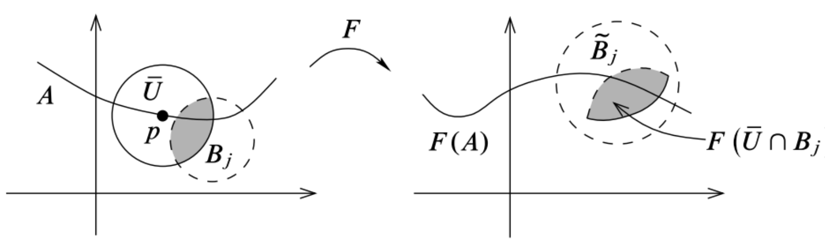
We now give a series of results that we will need to get to our goal of showing that critical points of smooth functions have measure zero. A short remark will be given for why each result is true.

Proposition 3. *Suppose A is an open or closed subset of* $\mathbb{R}^{n-1}$*, and* $f : A \longrightarrow \mathbb{R}$ *is continuous. Then the graph of f has measure zero in* $\mathbb{R}^n$

Remark. By the third property of the previous proposition, it's sufficient to consider the case where A is compact, since any open or closed set can be written as a countable union of compact subsets. Then a continuous function on a compact set will be uniformly continuous. This means we can partition our domain such that the output of our function is uniformly bounded by some small constant. On each interval, we can cover the graph by a rectangle of arbitrarily small volume.

Proposition 4. *Suppose* $A \subset \mathbb{R}^n$ *has measure zero and* $F : A \longrightarrow \mathbb{R}^n$ *is smooth. Then the image* $F(A)$ *has measure zero.*

[1]



Remark. We can cover A by countable many pre-compact open balls U, using the fact that $\mathbb{R}^n$ is second countable. It follows that the image $F(A)$ is the countable union of sets of the form $F(A \cap \overline{U})$. Again by the third property that we have for sets of measure zero, it's sufficient to show that each such that $F(A \cap \overline{U})$ has measure zero. Now, since $\overline{U}$ is compact, we can bound the variation of F for all pairs of points $x, y \in \overline{U}$ by a constant C. Then since we can cover A, and therefore $A \cap \overline{U}$ with rectangles whose volume is arbitrarily small, each image rectangle will have area expand by at most a factor of C. Collectively then, the image of our zero-volume cover of $A \cap \overline{U}$ will also have arbitrarily small volume; and thus measure zero.

3.2 Measure on Smooth Manifolds

With the tools from above, we can now discuss sets of measure zero on manifolds. As one would expect, the definition transfers through charts into $\mathbb{R}^n$.

Definition 5. If M is a smooth n-manifold, we say that a subset $A \subset M$ has <u>*measure zero in M*</u> if for every smooth chart (U, ϕ) for M, the subset $\phi(A \cap U) \subset \mathbb{R}^n$ has n-dimensional measure zero.

3.3 Sard's Theorem

Theorem 3.6. *Suppose M and N are smooth manifolds and $F : M \longrightarrow N$ is a smooth map. Then the set of critical values of F has measure zero in N.*

Proof. We give a rough outline of the proof given in [1]

Let $m := dim(M)$ and $n := dim(N)$. The proof is by induction on m.

If $m = 0$, then the result is immediate because if $n = 0$ then F has no critical points, while if $n > 0$ the entire image of F has measure zero because it is countable.

Now suppose $m \geq 1$ and assume the theorem holds for maps whose domains have dimension less than m. Since we can cover our manifolds M and N with countably many coordinates charts, we can reduce our work by assuming at the outset that F is a smooth map from an open subset $U \subset \mathbb{R}^n$ to $\mathbb{R}^n$, and we use coordinates $x^1, ..., x^m$ on U and $y^1, ..., y^n$ on $\mathbb{R}^n$.

Let $C \subset U$ denote the critical points of F. We define the decreasing sequence of subsets $C \supset C_1 \supset C_2 \supset ...$ as follows:

$$C_k = \{x \in C : \text{ for } 1 \leq i \leq k, \text{ all the ith derivatives of } F \text{ vanish at } x\} \tag{15}$$

By continuity, C and all of the C_k are closed in U. We take three steps to proving that $F(C)$ has measure zero. The high level goal is that we can write:

$$F(C) = \bigcup_i F(C_i - C_{i+1}) \tag{16}$$

where here $C_0 := C$. By the third proposition above, it's sufficient to show that each element of this union has measure zero.

<u>Step 1: $F(C - C_1)$ has measure zero</u>

The idea is to note that if $a \in C - C_1$, then some partial derivative of F is not zero at a. After rearranging coordinates, we can assume it is the first. Then the fact that $\frac{\partial F^1}{\partial x^1}|_a \neq 0$ means by the inverse function theorem that we can define new smooth coordinates in some neighborhood $V_a \subset U$ by:

$$(u, v^2, v^3, ..., v^m) = (F^1, x^2, ..., x^m) \tag{17}$$

and after shrinking V_a, we can assume $\overline{V_a}$ is compact, and then F has coordinates representation and Jacobian like:

$$F(u, v^2, .., v^m) = (u, F^2(u, v), ..., F^n(u, v)) \tag{18}$$

$$DF(u, v) = \begin{pmatrix} 1 & 0 \\ * & \frac{\partial F^i}{\partial v^j} \end{pmatrix} \tag{19}$$

Then $C \cap \overline{V_a}$ is the set of points where the $(n-1) \times (m-1)$ sub-matrix in the bottom right has rank less than n-1.

Again, using the fact that countable union of sets of measure zero have measure zero, it's sufficient to show that $F(U \cap \overline{V_a})$ has measure zero in $\mathbb{R}^n$.

The key step is to realize that it's sufficient to show that intersection with each hyper-plane $y^1 = c$ has (n-1)-dimensional measure zero. Looking at the last (n-1) coordinates of our map, we apply the induction hypothesis for maps from manifolds of dimension $m - 1$ and this completes the first step.

<u>Step 2: for each k, $F(C_k - C_{k+1})$ has measure zero</u>

The idea here, is that if we are at a point $a \in C_k - C_{k+1}$ then some $(k+1)$st partial derivative of F does not vanish. Let $y : U \longrightarrow \mathbb{R}$ be some kth partial of F and in this step, assume it has at least one non-vanishing first partial derivative at a.

Then a is a regular point of the smooth map y, so if we define Y to be the zero set of y, it will be a smooth hyper-plane by the regular level set theorem, defined on a set of regular points $V_a \ni a$. By definition of C_k all kth partials of F (including y) vanish on C_k so $C_k \cap V_a \subset Y$. Since the differential is not surjective at any $p \in C_k \cap V_a$, this means $F(C_k \cap V_a)$ is contained in the set of critical values of $F|_Y : Y \longrightarrow \mathbb{R}^n$ which has measure zero by the induction hypothesis. Covering U by countably many neighborhoods V_a completes this step.

Step 3: for $k > m/n - 1$, $F(C_k)$ has measure zero

Now we treat the case of points where all partial derivatives of F vanish, includes the first.

The idea here is a bit more complicated, but essentially, we for each $a \in U$ pick a closed cube $a \in E \subset U$, and since U can be covered by countably many such cubes we reduce to showing that $F(C_k \cap E)$ has measure zero. Using uniform continuity, we bound the variation of F on this cube, and partition the cube into small enough pieces, based on the variation of F such that the image of each sub-cube is contained in a union of balls whose collective radii are arbitrarily small. □

With this powerful theorem at our disposal, we return to the study of Morse functions, and discuss some of their properties, but first we will need some definitions from algebraic topology.

4 A Small Detour To Algebraic Topology

4.1 CW Complexes

Definition 1. A <u>*k-cell*</u> for $k \in N$ is the closed ball $\{y \in \mathbb{R}^k : ||y|| \leq 1\}$

Example 2. *A 0-cell is a point. A 1-cell is a line segment. A 2-cell is a disk. Etc.*

Definition 3. A <u>*CW complex W of dimension k*</u> is a topological space formed by the following inductive procedure:

The *0-skeleton* W^0 of W is a set of 0-cells

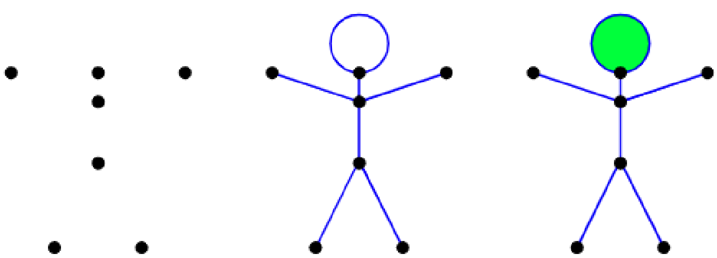
The *1-skeleton* W^1 of W is formed by gluing the endpoints of 1-cells to the 0-skeleton via the quotient topology.

The *k-skeleton* W^k of W is formed by gluing the endpoints of k-cells to the $(k-1)$ skeleton via the quotient topology.

Then $W = W^k$

Example 4. *A stick figure represented as a CW complex containing seven 0-cells, seven 1-cells, and a single 2-cell.*

The left most figure shows the 0-skeleton, the middle figure shows the 1-skeleton, and the full 2-skeleton is on the right. [9]

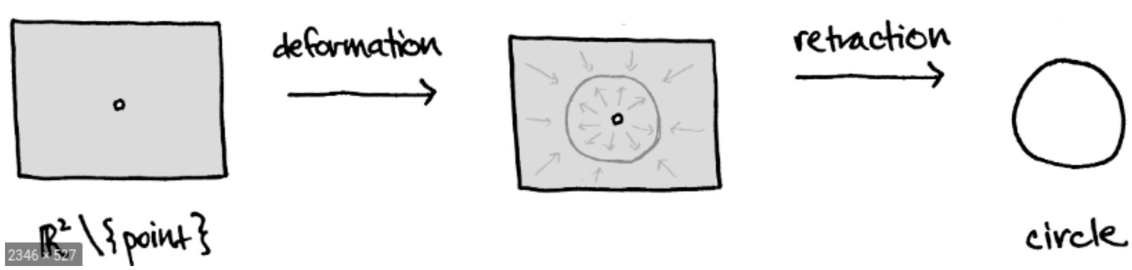


4.2 Homotopy Equivalence

In what way are the topological space $\mathbb{R}^2 - \{point\}$ and S^1 the same?

They cannot be homeomorphic, because S^1 is compact, while $\mathbb{R}^2 - \{point\}$ is not.

And yet, we can warp one into the other as the diagram shows: [11]

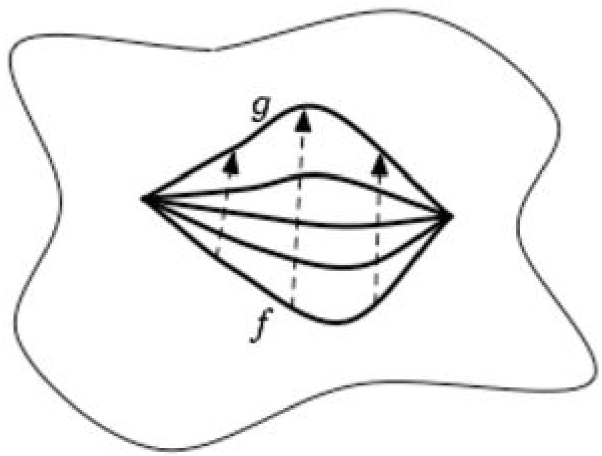


The notion that this figure captures is the one of *homotopy equivalence*. This is a rather involved area, so we will give the definition, provide some intuition, but will not go much further than we need for our Morse Theorems.

Definition 5. A homotopy between two continuous functions $f, g : X \longrightarrow Y$ where X and Y are topological spaces, is a continuous function $H : X \times [0, 1] \longrightarrow Y$ such that $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$

Intuitively, H parameterizes a set of functions $h_t : X \longrightarrow Y$, such that at time 0, $h_0 = f$ and at time 1, $h_1 = g$. The full map H can be thought of as a slider control that allows us to nicely transition from f to g as the slider t transition from 0 to 1.

If there is such an H, we say that f and g are *homotopic*.



Definition 6. Let X, Y be topological spaces. We have that they are homotopy equivalent if there exist a pair of continuous maps $f, g : X \longrightarrow Y$ such that $g \circ f$ is homotopic to the identity map on X.

Remark. Intuitively, spaces X and Y are homotopically equivalent if they can be transformed into one another by bending, shrinking and expanding operations.

Remark. Being homeomorphic is stronger than being homotopic. If X, Y are homeomorphic, then $g \circ f$ can be the identity (not only homotopic to it).

5 The Existence, Density, and Local Form of Morse Functions

5.1 Existence

The following example shows the existence of a large simple class of Morse functions.

Example 1. *Let $M \subset \mathbb{R}^n$ be a submanifold. For almost every $p \in \mathbb{R}^n$ the map:*

$$f_p : M \longrightarrow \mathbb{R} \tag{20}$$
$$x \mapsto ||x-p||^2 \tag{21}$$

is a Morse function.

Proof. [10] Let f_p as above, Then the derivative of f_p is:

$$df_{p,x}(v) = 2(x-p, v) \tag{22}$$

Therefore the critical points occur exactly when T_xM is normal to (x-p). Choose local coordinates $(u_1, ..., u_d)$ for M such that:

$$\frac{\partial f_p}{\partial u_i} = 2(x-p) \cdot \frac{\partial x}{\partial u_i} \tag{23}$$
$$\frac{\partial^2 f_p}{\partial u_i u_j} = 2(\frac{\partial x}{\partial u_i}\frac{\partial x}{\partial u_j} + (x-p) \cdot \frac{\partial^2 x}{\partial u_i u_j}) \tag{24}$$

Then by definition, x is a non-degenerate critical point if and only if x-p is normal to T_xM and the matrix on the right has rank d.

By *Sard's Theorem* it suffices to show that the $p \in \mathbb{R}^n$ such that x-p is normal to T_xM and the matrix on the right is singular, are the critical points of a smooth map.

Let $NM = \{(x,v) \in T_x\mathbb{R}^n : v \in T_xM^\perp\}$

Let $E : NM \longrightarrow \mathbb{R}^n$ by $E(x,v) = x+v$. Then $p = x+v$ is a critical point of E if and only if :

$$2(\frac{\partial x}{\partial u_i}\frac{\partial x}{\partial u_j} + v \cdot \frac{\partial^2 x}{\partial u_i u_j}) \tag{25}$$

is singular. Therefore the set of all f_p which are not Morse Functions corresponds to a subset of the critical points of E, which by Sard's Theorem has measure zero in $\mathbb{R}^n$. Then for almost all p, f_p is Morse

□

Remark. By Whitney's Embedding theorem, it follows that Morse functions exist on all smooth manifolds.

5.2 Density

Now that we know that Morse functions actually exist, we return to the question posed in an earlier section: How general or how strict are Morse Functions?

5.2.1 Morse Functions Are Generic

In some sense, it is not a strict condition at all. We will find that the class of Morse Functions is dense in $C^\infty(M)$ if M is compact. This says, that the property of being Morse is *generic*: *most* functions are Morse, and any smooth function can be approximated by a Morse function in C^k-norm.

Theorem 5.2. *Let M be a compact smooth manifold. Then the set of Morse functions is C^∞ dense in $C^\infty(M)$*

Proof. The idea is this. Use Whitney's embedding theorem to embed M into $\mathbb{R}^N$ via an embedding ϕ.

Since M is compact, for any $\varepsilon > 0$ we can choose a $u \in \mathbb{R}^N$ such that $||\phi(p) - u||^2 < \varepsilon$ for every $p \in M$. Then take f_u as defined in the previous example. This will be a Morse Function, and $|f(\phi(p)) - f_u(\phi(p)) < \varepsilon$ for every $p \in M$. □

5.3 Local Form

While Morse functions are plentiful, one should not get the impression that they are not powerful. Here we show that Morse Functions have a specific type of local form:

5.3.1 Morse Lemma

The Morse Lemma is a powerful statement about the behavior of functions around a non-degenerate critical point. It states that the function behaves quadratic-ally, regardless of the class of function, and that the kind of quadratic function is determined by the index of the critical point. The Morse Lemma states that the neighborhood about a non-degenerate critical point can be deformed into the neighborhood of the non-degenerate critical point of a quadratic function

Lemma 3. *Let $f : M \longrightarrow \mathbb{R}$ be smooth, and let p be a non-degenerate critical point for f.*

Then there exists a local coordinate system $(U, y^1, ..., y^n)$ around p called a Morse Chart for f at p *such that $y^i(p) = 0$ for all i, and such that $f(q) = f(p) - (y^1(q))^2 - ... - (y^\lambda(q))^2 + (y^{\lambda+1}(q))^2 + ... + (y^n(q))^2$ for all $q \in U$. And λ is the index of f at p.*

Proof. We give a rough outline of the proof in [3]

The proof relies on this familiar lemma from calculus:

Lemma 4. *If f is real-valued smooth function in a convex neighborhood V of $0 \in \mathbb{R}^n$, then there exist C^∞ functions $g_1, .., g_n$ such that:*

$$f(x^1, .., x^n) = \sum_{i=1}^{n} x_i g_i(x_1, ..., x_n) \tag{26}$$

which one proves using the fundamental theorem of calculus, and finds:

$$g_i(x^1, ..., x^n) = \int_0^1 \frac{\partial f}{\partial x^i} t(x^1, ..., x^n) dt \tag{27}$$

The proof of Morse Lemma goes like this.

Let $(U, x^1, ..., x^n)$ be any coordinate system about p, which, after a translation we can suppose that $f(p) = 0$.

By the above lemma, we write $f(x^1, .., x^n) = \sum_{i=1}^{n} x_i g_i(x_1, ..., x_n)$.

Since 0 is assumed to be a critical point:

$$g_j(0) = \frac{\partial f}{\partial x^j}(0) = 0 \tag{28}$$

So we can apply the lemma to g_j and find that:

$$g_j(x^1,..,x^n) = \sum_{i=1}^{n} x_i h_{i,j}(x_1,...,x_n) \tag{29}$$

fir certain smooth functions $h_{i,j}$.

Thus, we get:

$$f(x^1,..,x^n) = \sum_{i,j} x_i x_j h_{i,j}(x^1,...,x^n) \tag{30}$$

We can suppose $h_{i,j} = h_{j,i}$ by writing $\overline{h_{i,j}} = \frac{1}{2}(h_{i,j} + h_{j,i})$.

Then the matrix $\overline{h_{i,j}(0)} = \frac{1}{2}\frac{\partial^2 f}{\partial x^i x^j}(0))$ and hence is non-singular since we assumed p is a non-degenerate critical point.

The final step relies on a famous diagonalization theorem from linear algebra:

Theorem 5.5. *Every quadratic form Γ can be diagonalized, and therefore written in the form:*

$$\Gamma(x) = \lambda_1 y_1^2 + ... + \lambda_n y_n^2 \tag{31}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of the matrix A such that $\Gamma(x) = x^T A x$

If we apply this theorem to our matrix $\overline{h_{i,j}(0)}$, we find that after potential restricting to a smaller neighborhood of 0,

$$f(q) = f(p) - (y^1(q))^2 - ... - (y^\lambda(q))^2 + (y^{\lambda+1}(q))^2 + ... + (y^n(q))^2 \tag{32}$$

where λ defines the number of negative-eigenvalues of $h_{i,j} = h_{j,i}$ or equivalently, the index of f at p. This completes the proof. $\square$

Corollary 6. *Critical points of Morse functions are isolated, and Morse functions on compact manifolds have finitely many critical points.*

Proof. In a Morse chart about a point p, the differential only vanishes at p.

If a topological space is compact, there are finitely many isolated points. This proves the corollary. $\square$

6 The Morse Theorems

6.1 Riemmanian Metrics and The Gradient

To define geometric concepts such as lengths and angles on a vector space, one uses an inner product. For manifolds, the appropriate structure is a Riemannian metric, which is essentially a choice of inner product on each tangent space, varying smoothly from point to point. A choice of Riemannian metric allows us to define geometric concepts such as lengths, angles, and distances on smooth manifolds

Definition 1. [1] A *Riemannian metric g on a smooth manifold M* is a smooth symmetric 2-tensor field on M that is positive definite at each point. At each point $p \in M$, the metric assigns a positive-definite inner product:

$$g_p : T_pM \times T_pM \longrightarrow \mathbb{R} \tag{33}$$

A smooth manifold endowed with this metric g is called *a riemannian manifold* and is denoted (M, g).

Example 2. *On $\mathbb{R}^n$, we have the canonical metric with respect to the standard coordinates $x^1, ..., x^n$ given by:*

$$g_p^{canonical} : T_p\mathbb{R}^n \times T_p\mathbb{R}^n \longrightarrow \mathbb{R} \tag{34}$$

$$(\sum_i a_i \frac{\partial}{\partial x^i}, \sum_J b_j \frac{\partial}{\partial x^j}) \mapsto \sum_i a_i b_i \tag{35}$$

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a smooth function, then we are familiar with the gradient vector field*:* grad f *defined by:*

$$\nabla_x f = (\frac{\partial f}{\partial x^1}(x), .., \frac{\partial f}{\partial x^n}(x)) \tag{36}$$

In calculus, we learn that ∇f is perpendicular to the level sets of f, and $|\nabla f|$ is proportional to the rate of increase of f.

We know that the differential df takes a vector field X and returns a measurement of how quickly f is increasing along the flow line of X. So we can equivalently define the gradient vector field as the vector field for which:

$$g_p^{canonical}(\nabla_x(f), v) = df_p(v) \tag{37}$$

On Riemmanian Manifolds, we can immediately generalize, using the equipped riemmanian metric g.

Definition 3. [1] Let (M, g) be a riemmnian manifold. The gradient vector field of f is the vector field on M such that:

$$g(\nabla_p(f), v) = df_p(v) \tag{38}$$

where $p \in M$ and $v \in T_pM$

6.2 1-Parameter Group of Diffeomorphisms and Gradient Flow Lines

We know by *The Fundamental Theorem of Flows* that on manifolds, a smooth vector field X determines a 1-parameter group of diffeomorphisms induced by the the flow of X.

We know describe a similar construction, using the gradient vector field.

Definition 4. Let (M, g) be a riemmian manifold. A *gradient flow line* is an integral curve:

$$\gamma : \mathbb{R} \longrightarrow M \tag{39}$$

that satisfies the following flow equation:

$$\frac{d\gamma}{dt} = -\nabla_{\gamma}(f) \tag{40}$$

Using our intuition for the gradient vector field as determining the direction of ascent of a function f, we can think about the gradient flow trajectories as paths of steepest descent: at each point $p \in M$, following γ pushes us along the manifold M, in the steepest descent of our function f.

Remark. If M is compact, then the domain of any such integral curve γ is $\mathbb{R}$, and induces a group of diffeomorphisms via the $\mathbb{R}$-action:

$$p \mapsto t \cdot p \tag{41}$$
$$\gamma(t) = t \cdot m \tag{42}$$
$$\gamma'(t) = -(\nabla_{\gamma(t)} f) \tag{43}$$
$$\gamma(0) = p \tag{44}$$

We see from this that each integral curve γ starts and ends at a critical point in the sense that:

$$lim_{t \longrightarrow \pm\infty} \gamma(t) =: C \tag{45}$$
$$C \subset \ \textit{Critical points of f} \tag{46}$$

6.3 Stable and Unstable Manifold

As noted above, gradient flow lines connect critical points. We now define the following objects:

Definition 5. Let M be a compact manifold, $f : M \longrightarrow \mathbb{R}$ be a Morse function and $p \in M$ a critical point of f in M.

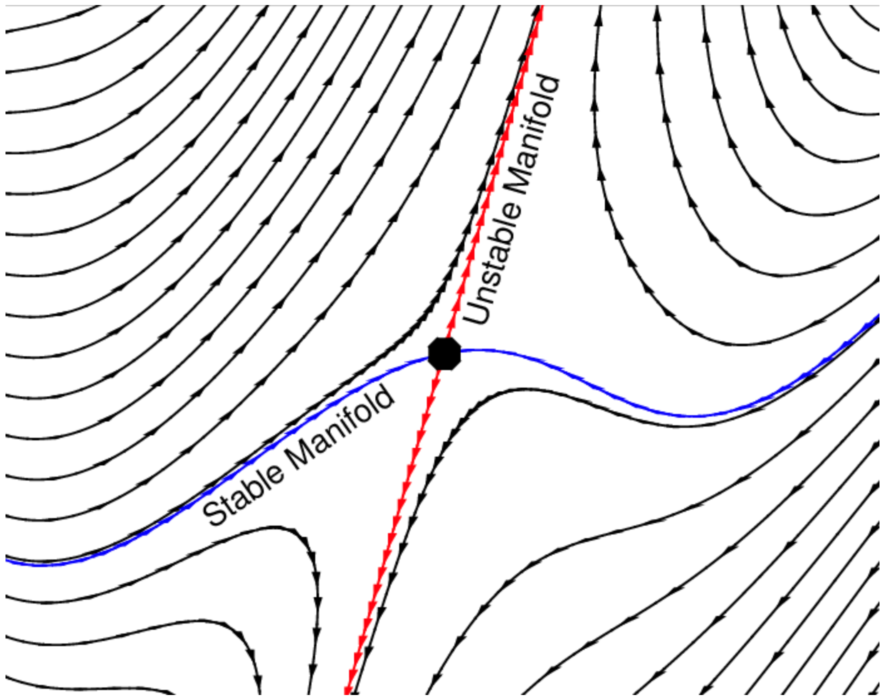
The stable manifold S(p) is the set of points which *flow down* to p:

$$S(p) := \{x \in M : lim_{t \longrightarrow \infty} t \cdot x = p\} \tag{47}$$

The unstable manifold U(p) is the set of points which *flow up* to p:

$$U(p) := \{x \in M : lim_{t \longrightarrow -\infty} t \cdot x = p\} \tag{48}$$

Example 6. *[7] Below are the gradient flow lines of a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, along with the stable and unstable manifold for a point p shaded dark. The stable manifold is the set of points shaded in blue, and the gradient flow line flows down to p. Conversely, the unstable manifold is shaded red, and the gradient flow line flows up to p.*



Proposition 7. *Let the index of $p \in M$ be λ where M is compact. Then $S(p)$ is homeomorphic to $\mathbb{R}^{n-\lambda}$ and $U(p)$ is homeomorphic to $\mathbb{R}^{\lambda}$.*

Proof. Intuitively: Clearly, the stable and unstable manifold partition the space. $U(p)$ will be given by travelling along smooth gradient flows, that decrease the value of f, and $S(p)$ will be given by travelling along smooth gradient flows, that increase the value of f. By definition of the index λ of the critical point p, there are λ independent directions we can travel from p and remain in $U(p)$ and thus $n - \lambda$ direction we can travel and remain in $S(p)$. □

As a consequence, it follows that a Morse function f on M provides a nice decomposition of M:

$$M = \bigcup_{p:p \text{ is critical}} S(p) = \bigcup_{p:p \text{ is critical}} U(p) \tag{49}$$

This comes from the fact that every point of M lies on a flow line γ, and we can always find $lim_{t \longrightarrow \pm\infty} \gamma(t)$. Each element of the union is disjoint by the uniqueness of integral curves, which stems from the uniqueness of solutions to differential equations.

6.3.1 Theorem 1

The next theorem, explains how each piece of our decomposition $M = \bigcup_{p:p \text{ is critical}} U(p)$ fit together. We will not have the tools from algebraic geomtry to give a proof, however we will see how this decomposition works for the height function on the torus, at the end of this section.

Theorem 6.8. *Let M be a compact manifold, and let $f : M \longrightarrow \mathbb{R}$ be a Morse function on M. Then M has the homotopy type of a cell complex, with one cell of dimension λ for each critical point of index λ*

Remark. This says that M is homotopy equivalent to a topological space of the form:

$$X_r = D^{\lambda_1} + D^{\lambda_2} + D^{\lambda_3} + ... \tag{50}$$

where $0 = \lambda_1, \lambda_2, ..., \lambda_r = m$ are the indices of the critical points of f, and the + signs denoted a topological gluing of the i-cells D^i.

6.4 Critical Values and Topology

Now we give some answers to the original question posed in the introductory section. We explain how the topology changes as we take sub level sets at different points.

As a first step, we make the follow remark:

Remark. Let M be a manifold and $f : M \longrightarrow \mathbb{R}$ a smooth function. The regular level set theorem asserts that if a is a regular value of f, then the level set $f^{-1}(a)$ is an embedded submanifold.

The same is true for sublevel sets:

Denote $M^a := f^{-1}((-\infty, a])$. Then M^a is submanifold (with boundary).

6.4.1 Theorem 2

What happens when we traverse sublevel sets without crossing any critical points?

Theorem 6.9. *Let $f : M \longrightarrow \mathbb{R}$. Suppose $a, b \in \mathbb{R}$ and $f^{-1}([a, b])$ is compact and f has no critical point in $f^{-1}([a, b])$. Then M^a is diffeomorphic to M^b.*

Proof. [8] The idea of the proof is that without any critical point in $f^{-1}([a, b])$ the flow lines generated by ∇f are *nice*, and we can follow them from M^a to M^b. Then, the 1-parameter group of diffeomorphism induced by the flow will give us our diffeomorphism.

By definition, ∇f is the vector field characterized by the equality $g(X, \nabla f) = X(f)$ for any vector field X, where g is a riemannian metric on M.

By assumption ∇f does not vanish on $f^{-1}([a, b])$ since we have no critical points there.

Consider a smooth non-negative function $\rho : M \longrightarrow \mathbb{R}$ satisfying:

- $\rho(q) = \frac{1}{g(\nabla f, \nabla f)_q}$ for $q \in f^{-1}([a, b])$
- $\rho = 0$ outside of a compact neighborhood of $f^{-1}([a, b])$

Then define the vector field $X : q \mapsto \rho(q)(\nabla f)_q$ so that X is the normalized gradient in $f^{-1}([a, b])$ and vanishes outside of a compact set. By the fundamental theorem of flows, it generates a 1-parameter group of diffeomorphisms $\phi : \mathbb{R} \times M \longrightarrow \mathbb{R}$.

For a fixed $q \in M$, we have a map $t \mapsto f(\phi_t(q))$ which gives the value of f after following the point q along its gradient flow line for time t.

Note that:

$$\frac{df(\phi_t(q))}{dt} = g(\frac{d\phi_t(q)}{dt}, (\nabla f)_q) = g(X_q, (\nabla f))q) = 1 \tag{51}$$

by the definition of X. This shows that $t \mapsto f(\phi_t(q))$ is linear with derivative 1. We will have that $\phi_{b-a} : M^a \longrightarrow M$ is a diffeomorphism onto M^b. Intuitively, it is a bijection from M^a to M^b because along the gradient flow lines, f is increasing linearly along the path $\phi_t(q)$ until it reaches b, and never decreases because ρ is non-negative.

□

6.4.2 Theorem 3

What happens when we traverse a sublevel set crossing a critical point of index λ?

It will turn out that that as we cross critical points, our topology will change in a very specific way, which is connected to the index of the critical point.

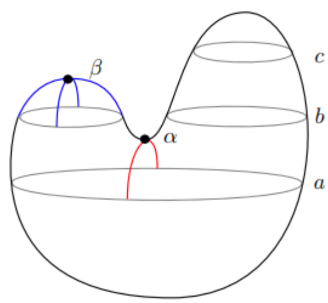
Now we can state the second Morse Theorem.

Theorem 6.10. *Let $f : M \longrightarrow \mathbb{R}$. Suppose $a, b \in \mathbb{R}$ and $f^{-1}([a, b])$ is compact and f has exactly one critical point α in $f^{-1}([a, b])$. Then M^b is homotopy equivalent to M^a with a λ-cell attached.*

Remark. What is this theorem saying? It says that if we are in a situation where f has a critical point in between two sublevel sets M^a and M^b, then these spaces will no longer be diffeomorphic. There is however a strong relationship between them, which is dependent on the index of that critical value. It turns out, they are homotopy equivalent, i.e. we can smoothly warp one into the other, but only after attaching a $\lambda - cell$, that is, after attaching a rectangle solid, whose dimension coincides with the index of the smaller critical value, which in turn is intuitively the number of independent directions that f is decreases at the point $f^{-1}(a)$!

The proof of this theorem is quite technical and involved. Instead, we will give an example of this theorem in action.

Example 11. *Consider the following surface, embedded in $\mathbb{R}^3$ where f is the height function. [8]*



In this figure, M^a looks like a deformed disc. From our detour on algebraic topology, we provided some intuition for why this is homotopic equivalent to a point.

On the other hand M^b is homotopic equivalent to a circle. We can imagine the two circular rings to the left and right of α forming a slinky*, and letting the slinky collapse onto S^1.*

Also, notice that α is the unique critical point in $f^{-1}([a, b])$. It has index 1: from α, there is a single negative eigenvalue, which we can visualize since there is just a single direction in which we can move from that point on the surface, and decrease the value of f, i.e. the height. The unstable manifold, depicted in red, shows that gradient flow line that decreases the value of f along this single direction.

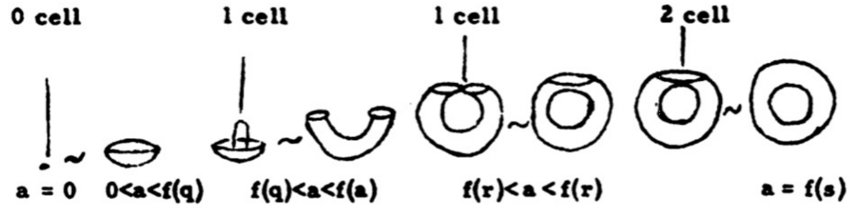
Finally, we have the following homotopy equivalences as the theorem suggests:

$$M^a \sim \{p\} \implies M^a + 1cell = \{p\} + S^1 \sim S^1 \sim M^b \tag{52}$$

As our final example, we come back to the torus T^2 with the Morse height function as introduced at the beginning of this document.

Example 12. *Let T^2 be the torus embedded in $\mathbb{R}^3$ as in the beginning of the document and let f be the height function.*

The following diagram shows how our topology changes as we cross the critical points of f in our sublevel sets [12]:



We begin by analyzing the index of each critical point.

The first critical point is the bottom most point on our torus. It has an index of 0, since there is no direction we can move on the torus that well decrease our elevation.

The second and third critical points occur as we cross the inner hole of our torus. In each case, there is a 1-dimensional direction we can move to decrease our elevation, so these critical points have index 1.

The final critical point is at the top most point on our torus. It has index 2, because there are two independent directions, given by the full two directions on the torus that we can travel to reduce our elevation: any way we go, our height decreases.

Now let's analyze the homotopy types of our sublevel sets.

At our first critical point, our sublevel set consists of a single point.

After this, but before we reach the second critical point, our sublevel sets look like a disc. Again, we know these are homotopic equivalent to a point. This is precisely what our theorem says, we attach a 0-cell to a point, but this doesn't change our topology.

As we pass our second critical point, we are in the same situation as in the previous example. Our sublevel set looks like a slinky, which can be retracted onto the circle by collapsing the slinky onto a circle. We have added a 1-cell, as predicted, because the index of this second critical point is 1.

As we pass our third critical point, our sublevel set looks like it added a handle onto the slinky. This handle is precisely the circle, the 1-cell, which got introduced due to the index 1 critical point.

And finally, as we leave the torus through our final critical point, our sublevel set is the full torus. This final transition is perhaps the most difficult to describe. The point is, if we attach a sphere to sit nicely on top of the result from our previous step, we recover the full torus. This can be somewhat visualized on the right most figure in the above diagram.

7 Applications

As a computer science student, I was pleased to find interesting application of Morse Theory in topological data analysis, and in computer graphics.

7.1 Topological Data Analysis

Topological data analysis deals with understanding data through methods of topology.

In many real world data sets, for example, in machine learning, the collected data is high-dimensional, noisy, sparse and non-linear.

It turns out, most optimization algorithms on high-dimensional data make the following assumption.

Proposition 1. *Suppose we have a dataset $\mathcal{D} = \{(x\}_{i=1}^{D}$ of points $x \in \mathbb{R}^N$.* [1]

Then there exists a submanifold $M \subset \mathbb{R}^N$ such that $\mathcal{D} \subset M$ and dim$(M) << N$.

At a high level, our assumption says something like:

Ignoring noise, the *informative part of our data* can be embedded into a much *smaller* space: the relevant features can be expressed in a much lower dimension.

The question then becomes: *how do we find this low dimensional manifold efficiently?*

This is an extremely difficult problem. However, for point clouds (finite set of points in some euclidean space), the area of topological data analysis aims to do just this, by finding simple combinatorial representation of the data, which capture the key patterns.

Morse Theory

Suppose our data set $\mathcal{D} \subset \mathbb{R}^N$ consists of a finite sampling from an unknown probability density function $g : \mathbb{R}^N \longrightarrow [0, \infty)$.

Then we are interested in reginos of high density because:

- If $g(x)$ is very smal for some $x \in \mathcal{D}$ then x can be regarded as a noisy datapoint. It does not represent the main features.

Therefore, we can focus on the *superlevel sets*:

$$Y^a = g^{-1}([a, \infty]) \tag{53}$$

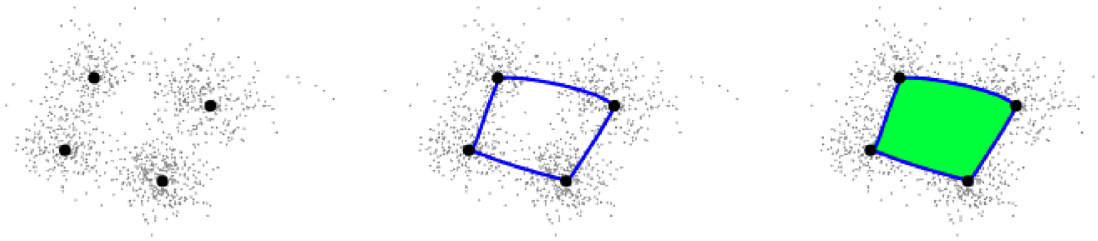
- these encode how the dense regions of the data $\mathcal{D}$ is organized

Just as we analyzed sublevel sets in this document, superlevel sets have analogs of the main theorems we covered.

Using discrete analogous of Morse Theory, researcher can find maximum probability 0-cells, 1-cells and so forth on the point cloud, and recover information about g, and hence extract the main features from data. The fact that the manifold can be decomposed into simpler cell spaces, whose dimension is determined by the algebraic properties of g, implies that an aggregation of these cells gives a relatively descriptive picture of the entire manifold.

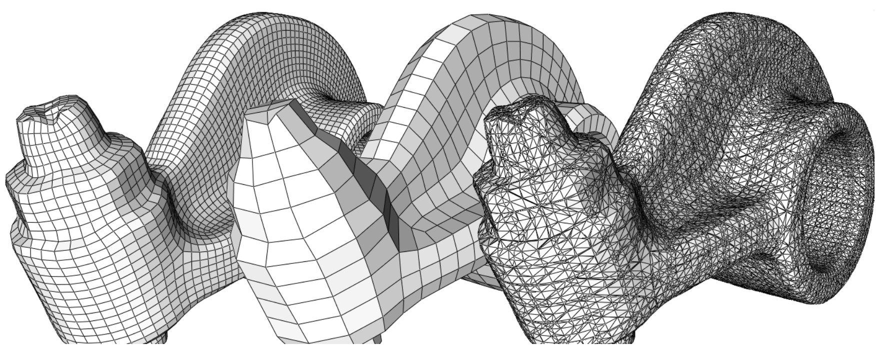
The following shows what such a method looks like when analyzing social network data. Each point is a set of features gathered on a user, projected to a plane using principal component analysis.

[1] Typically N is extremely large. For example, in image data, even small photos live in $x \in \mathbb{R}^{28\times28\times3}$, for a 28-28 pixel colored image.



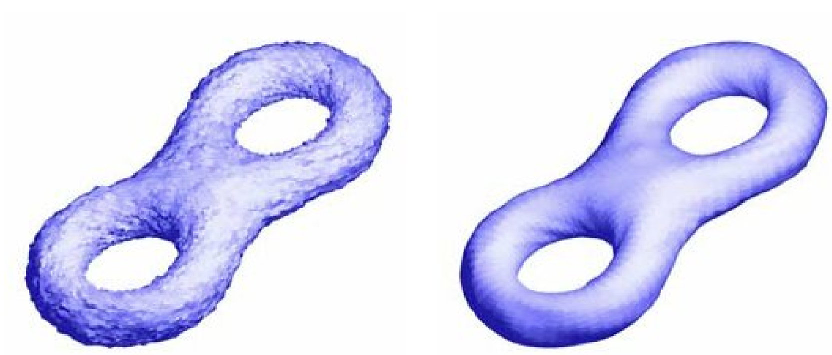
7.2 Computer Graphics

In computer graphics, discrete differential geometry is heavily used. One studies discrete analogues of smooth geometric objects, by making piece-wise linear approximations, yielding *surface meshes* that describe the surface of interest.



One frequent task in this field is the idea of *denoising*. For example, we may have a non-uniform sample of points on our surface. When applying smooth functions to our mesh [2], a non-uniform sample may lead to weird looking visuals artifacts.

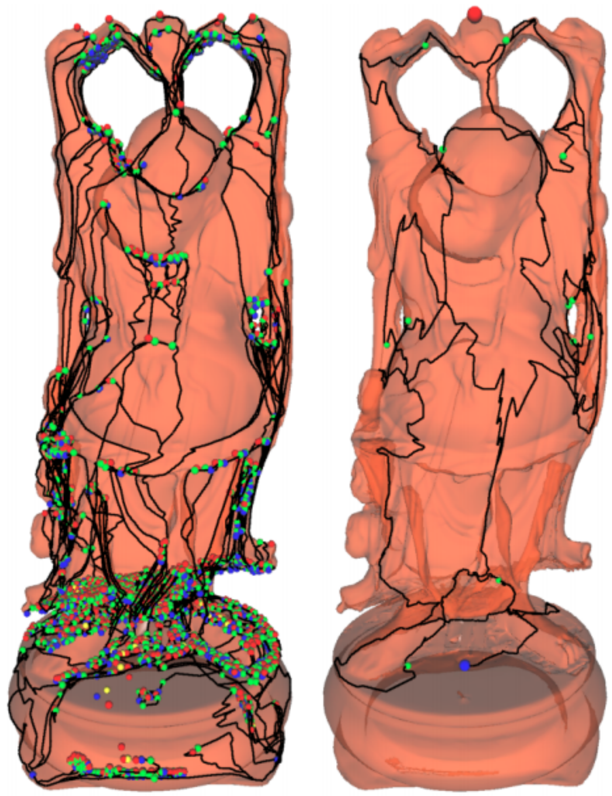
A frequent way to deal with this is to make small changes to our surface, that *smooths it*, such that these artifacts go away.



Interestingly, most artifacts are the result of critical points that are not well behaved. In light of Morse Theory, a dual approach to this problem is to *change the function instead of the surface.* We know we can approximate the function we want to apply to our surface with a Morse Function, and we know the critical

[2]such as an explosion animation in a video game, or a realistic human body movement in a video game, or a texture map in a video

points of Morse Functions are non-degenerate by assumption. This gives us a way to apply interesting graphics operations without needing to modify our surface at all (which can be quite large). An example of height function (left) shows many higher order degenerate critical points, which would ruin a transformation of the surface that depends on computing elevation. A modified height-function (right) on the same surface shows how the set of critical points becomes more well-behaved:



[5]

8 References

References

[1] John M. Lee. *Introduction to Smooth Manifolds.* https://math.berkeley.edu/~jchaidez/materials/reu/lee_smooth_manifolds.pdf

[2] Loring W. Tu *An Introduction to Manifolds.* https://im0.p.lodz.pl/~kubarski/AnalizaIV/Wyklady/L-Tu-1441973990.pdf

[3] J. Milnor *Morse Theory.* https://www.maths.ed.ac.uk/~v1ranick/papers/milnmors.pdf

[4] Allen Hatcher *Algebraic Topology.* https://pi.math.cornell.edu/~hatcher/AT/AT.pdf

[5] Xinlai Ni, Michael Garland, John C. Hart *Fair Morse Functions for Extracting the Topological Structure of a Surface Mesh.* https://www.cs.jhu.edu/~misha/Fall07/Papers/Ni04.pdf

[6] Zhongping Ji, Ligang Liu, Guojin Wang *Non-iterative Global Mesh Smoothing with Feature Preservation.* https://www.researchgate.net/publication/228988731_Non-iterative_Global_Mesh_Smoothing_with_Feature_Preservation

[7] Martin A. Guest *Morse Theory in the 1990's.* https://arxiv.org/pdf/math/0104155.pdf

[8] Shintaro Fushida-Hardy *Morse Theory.* https://stanford.edu/~sfh/morse.pdf

[9] Henry Adams, Atanas Atanasov, Gunnar Carlsson *Morse Theory in Topological Data Analysis.* https://www.researchgate.net/publication/51963152_Morse_theory_in_topological_data_analysis

[10] Arun Debray *Morse Theory: Notes UTexas.* https://web.ma.utexas.edu/users/a.debray/lecture_notes/m392c_Morse_notes.pdf

[11] Tai-Danae Bradley *Clever Homotopy Equivalences.* https://www.math3ma.com/blog/clever-homotopy-equivalencesf

[12] Raoul Bott *Morse Theory and Application to Homotopy Theory.* https://www.math.ucla.edu/~petersen/Bott_book.pdf