Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Morse theory with examples

Bachelor Thesis

Yann Guggisberg

March-September 2018

Advisor: Dr. Ana Cannas da Silva
Department of Mathematics, ETH Zürich

Abstract

This text is a survey on the subject of differential topology, and in particular Morse theory. Morse theory uses critical points of functions to make statements about the homotopy type of given manifolds. The main theorems that we prove in the second chapter are quite remarkable, as they allow a precise characterization of topological spaces using just the critical points of a suggestively called Morse function.

The present work can be divided into three parts. In the first chapter, we define manifolds of Euclidean space, as well as useful tools to recognize when a given subset of $\mathbb{R}^n$ is indeed a manifold. Towards the end of the chapter, we define Morse functions and go over some of their properties. This first part can be seen as a preparation section where we describe the setting in which Morse theory will then be used.

The second chapter contains the statements and proofs of the main results of Morse theory, using the machinery developed in the previous chapter.

The last two chapters present applications of Morse theory as a way to characterize and classify the homotopy type of topological spaces. We present the Reeb theorem and a way to compute the homology groups of complex projective spaces as examples. We also give a classification of compact 1-manifolds using two different methods: Morse theory, and parametrization by arc-length.

Contents

Acknowledgements

I want to thank my supervisor Dr. Ana Cannas da Silva for proposing such an interesting subject and answering my numerous questions. She was able to explain every detail that was not clear and has helped me deepening my understanding of the topic.

Also, a big thank you to Emilie Epiney who has pointed out many typos in my text.

As this work is a survey based on the books and references cited later, none of the material presented here is the product of my own ideas, except for minor modifications and the proofs of a few auxiliary results. The sources from which the material comes from will be mentioned in the introduction, as well as cited throughout the text.

The LaTeX template used for this thesis is the one recommended by the CADMO at ETH.

Introduction

This introduction is meant to explain in detail from which sources each part of the text comes from. As said in the abstract, this work is a survey on differential topology and Morse theory. The text mainly follows the books " Differential Topology" by Guillemin and Pollack [9] and "Morse Theory" by Milnor [8].

The first chapter is mostly based on [9]. The definition 1.48 concerning the concept of the Hessian as a bilinear form, the statement and proof of the Morse Lemma 1.53 and everything else between definition 1.48 and remark 1.54, as well as the discussion about 1-parameter groups (section 1.8), however, come from [8]. A few of the proofs about the properties of Morse functions were inspired by [3], which is a compilation of solutions to exercise proposed in [9].

The second chapter is mostly based on [8]. The definition of a deformation retract as well as the notions of adjunction space and cell-complex were taken from lecture notes of an algebraic topology class given at ETH by Will Merry in 2017-2018 [6]. The brief explanation on Riemannian metrics comes from lecture notes by Andrews [2]. The precisions on the cellular approximation theorem were found in Hatcher's book "Algebraic Topology" [4]. In its proof of theorem 2.22, Milnor uses results from an article from Whitehead given here [10].

The third chapter deals with examples given in [8], whereas the forth chapter has two main sources. The proof by parametrization by arc-length comes from another of Milnor's books, namely "Topology from the differentiable Viewpoint" [7]. The proof using Morse theory is an Appendix of [9], except for the proof of the lemma 4.11, which was inspired from [1].

The Appendix A about the general definition A.12 of a smooth manifold and the necessary assumptions come from another set of lecture notes by Will Merry, namely his 2018 differential geometry course [5]. The precision on submanifolds comes from [11]. Appendix B was again taken from [9].

The illustrations were taken from [9], [8], and [7] and come directly from the proof that is followed in this text.

The objective of this text is to explain the ideas of the above sources and to arrange them in a way that is as coherent and natural as possible. As mathematics constitute a very interconnected field, it is difficult to write something that is entirely self-contain but I do hope that this survey is able to shed light on the elegance of the powerful results and applications of Morse theory.

Chapter 1

Manifolds of Euclidean space and smooth maps

One of the goals of differential geometry is the use of calculus to derive geometric properties of different spaces. To do this, it is sometimes necessary to generalize the basic tools of calculus, which are initially defined for subsets of Euclidean space. Such a generalization is most naturally made for objects that "locally look like" Euclidean space. Objects with this property are called manifolds and are the subject of this chapter. We need a few definitions to formalize what we mean by "locally looking like Euclidean space".

1.1 First definitions and examples

We start by giving a few very important definitions:

Definition 1.1 *Let $U \subseteq \mathbb{R}^n$ be an open set. A map $f : U \to \mathbb{R}^m$ is called* smooth *if its partial derivatives of all orders are continuous.*

This definition only works for functions with open domains, since the definition of partial derivative makes no sense otherwise. We can however generalize our definition as follows:

Definition 1.2 *A map $f : X \to \mathbb{R}^m$ for X an arbitrary subset of $\mathbb{R}^n$ is called* smooth *if for all x in X, there exists an open set $U \subseteq \mathbb{R}^n$ and a smooth map $F : U \to \mathbb{R}^m$ such that $F|_{U \cap X} = f$. We say that f can be locally extended to a smooth map F on an open set U.*

Definition 1.3 *A smooth map $f : X \to Y$ between subsets of two Euclidean spaces is a* diffeomorphism *if it is bijective and if the inverse map $f^{-1} : Y \to X$ is also smooth. We say that X and Y are* diffeomorphic *if there exists such a map.*

In the context of differential topology, diffeomorphic spaces are considered to be equivalent. They can be regarded as two instances of the same mathematical object.

We are now ready to formalize our definition of a manifold:

Definition 1.4 *Let X be a subset of $\mathbb{R}^n$. We say that X is a* k-dimensional manifold *if for all x in X there exists a neighbourhood $V \subset X$ of x that is diffeomorphic to an open set $U \subset \mathbb{R}^k$. A diffeomorphism $\Phi : U \to V$ is called a* parametrization *of the neighbourhood V. The inverse map $\Phi^{-1} : V \to U$ is called a* coordinate system *on V. When we write the map $\Phi^{-1} = (x_1, \ldots, x_k)$, the k smooth functions $x_1, \ldots, x_k$ are called* coordinate functions *on V. The number k is called* the dimension of the manifold *and we often write $dim(X) = k$.*

Remark 1.5 *Let $X \subset \mathbb{R}^n$. Recall that the (relatively) open sets of X are the sets that can be written as an intersection $U \cap X$, where U is an open subset of $\mathbb{R}^n$. This equips X with the* relative topology. *When not otherwise specified, it is with this topology that we will take subsets of Euclidean space.*

Remark 1.6 *This definition of a manifold is not the most general one. It is indeed not mandatory for a manifold X to be a subset of Euclidean space. The more abstract definition is however not particularly useful to the development of Morse theory, which is the main goal of this text. The definition given here also has the advantage to be more geometrically intuitive, as stressed out by Guillemin and Pollack in [9]. However, Milnor [8], which is the other main source for this text does not explicitly assume manifolds to be subsets of Euclidean space. The technical assumptions necessary for a topological space to be a manifold in whole generality are relegated to the Appendix A. For the following discussion, it is enough to think of manifolds as objects that can easily and smoothly be injected in Euclidean space. A property of Euclidean space that will be used several times in this text is the second axiom of countability. A recall of the definition can be found in Appendix A.*

Another closely related definition is that of a submanifold:

Definition 1.7 *Let X and Z be manifolds in $\mathbb{R}^n$ with $X \subseteq Z$, then we say that Z is a* submanifold *of X. In particular, X is a submanifold of $\mathbb{R}^n$, and any open subset of X is a submanifold of x.*

Remark 1.8 *This definition of a submanifold is a weak one, it is sometimes useful to make stronger assumptions. This is explained briefly in Appendix A.*

Remark 1.9 *At this point, it is worth noting that every property that we have defined so far are* local *properties. They are valid for a point x in a neighbourhood of this point only. For example, smoothness of a mapping, or being a manifold are local properties. This can be opposed to the term* global, *which refers to properties of mathematical objects as a whole. The distinction between local and global properties will be of importance in a few proofs later so it is important to keep it in mind.*

We now want to define the derivative of a smooth map between two manifolds. To do this, we will need to define some new concepts.

1.2 Tangent spaces and derivatives

We begin this section by recalling the definition of a derivative for a mapping between two Euclidean spaces.

Definition 1.10 *Let f be a smooth map from an open set of $\mathbb{R}^n$ into $\mathbb{R}^m$ and let $h \in \mathbb{R}^n$ be an arbitrary vector. Then, the derivative of f in direction h taken at a point x in the domain of the function is defined by:*

$$\mathrm{d}f_x(h) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$$

For x fixed, the map $\mathrm{d}f_x : \mathbb{R}^n \to \mathbb{R}^m$ sending h to $df_x(h)$ is a linear, which is a known fact from calculus. Another important result from analysis is the chain rule, which states that for open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ and smooth maps $f : U \to V$ and $g : V \to \mathbb{R}^l$ we have, for any x in U

$$\mathrm{d}(g \circ f)_x = \mathrm{d}g_{f(x)} \circ \mathrm{d}f_x$$

The derivative of a map is the best possible linear approximation of that function. In that spirit, we can use derivatives to investigate the linear space that best approximates a manifold X at a given point x. Suppose that X is a submanifold of $\mathbb{R}^n$ and that $\Phi : U \to X$ is a local parametrization around x, with U an open set in $\mathbb{R}^k$. By translation of the open set U, we can assume without loss of generality that $\Phi(0) = x$ We will use this fact frequently in order to ease the notation. The best linear approximation of $\Phi : U \to X$ at 0 is the map:

$$u \mapsto \Phi(0) + \mathrm{d}\Phi_0(u) = x + \mathrm{d}\Phi_0(u)$$

With that in mind, we can define the concept of a tangent space.

Definition 1.11 *The* tangent space *of X at the point x is the image of the map $\mathrm{d}\Phi_0 : \mathbb{R}^k \to \mathbb{R}^n$. It is denoted by $T_x(X)$. The elements of $T_x(X)$ are called* tangent vectors *to X at the point x.*

Remark 1.12 *(i) $T_x(X)$ is a vector subspace of $\mathbb{R}^n$*

(ii) The parallel translate $x + T_x(X)$ of the tangent space is the closest flat approximation of X at the point x

Before going any further, it is important to notice that our definition of tangent space is not quite finished yet. Indeed, what would happen if one chooses another parametrization around x? Would the tangent space be different? Luckily, this is not the case as we will show below:

Lemma 1.13 *The tangent space $T_x(X)$ is well-defined, i.e. it is independent of the choice of parametrization.*

Proof Assume that $\Psi : V \to X$ is another parametrization around x. Once again, we may suppose after translation that $\Psi(0) = x$. Now, if we shrink the sets U and V sufficiently, we may also assume that $\Phi(U) = \Psi(V)$. We then define the map $h := \Psi^{-1} \circ \Phi : U \to V$. As it is the composition of two diffeomorphims and because the images of Φ and Ψ coincide, the map h is a diffeomorphism as well. Rearranging the definition of h, we write $\Phi = \Psi \circ h$. Differentiating this equation yields : $\mathrm{d}\Phi_0 = \mathrm{d}\Psi_0 \circ \mathrm{d}h_0$. This equality implies that the image of $\mathrm{d}\Phi_0$ is contained in the image of $\mathrm{d}\Psi_0$. Doing the exact same thing, but reversing the roles of Φ and Ψ yields the converse inclusion. Thus, we obtain that $T_x(X) = \mathrm{d}\Phi_0(\mathbb{R}^k) = \mathrm{d}\Psi_0(\mathbb{R}^k)$, which shows that $T_x(X)$ is well-defined. □

Another fact worth noting is that the dimension of the tangent space as a vector space is the same as the dimension of its manifold.

Lemma 1.14 *Let X be a k-dimensional manifold embedded in $\mathbb{R}^n$ and x be a point in X. Then the vector space $T_x(X)$ has dimension k.*

Proof Let U be an open set in $\mathbb{R}^k$ and $\Phi : U \to V$ be a parametrization around x, with once again $\Phi(0) = x$ Let W be an open set in $\mathbb{R}^n$ and $F : W \to \mathbb{R}^k$ be a map extending Φ^{-1}. Then the composition $F \circ \Phi$ is the identity map on U. Since the derivative of a linear map is the same linear map (this is a fact from analysis), we get that the sequence

$$\mathbb{R}^k \xrightarrow{\mathrm{d}\Phi_0} T_x(X) \xrightarrow{\mathrm{d}F_x} \mathbb{R}^k$$

must be the identity on $\mathbb{R}^k$. This implies that $\mathrm{d}\Phi_0 : \mathbb{R}^k \to T_x(X)$ is an isomorphism and thus the dimension of $T_x(X)$ is k. □

We are now ready to construct the derivative of a smooth map $f : X \to Y$ between manifolds. We want this derivative to be a linear transformation of tangent spaces $\mathrm{d}f_x : T_x(X) \to T_{f(x)}(Y)$. Additionally, we would like this map to have two properties. First of all, it should extend our previous definition of derivative, i.e. if X and Y are open sets in Euclidean spaces, the two definitions of derivative should coincide. Moreover, the chain rule should hold for our new definition of derivative. Let $\Phi : U \to X$ be a parametrization around x and $\Psi : V \to Y$ be one around $y = f(x)$, with $U \subseteq \mathbb{R}^k$ and $V \subseteq \mathbb{R}^l$. Once again, assume that $\Phi(0) = x$ and $\Psi(0) = y$. By shrinking U if necessary, we have the following commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{\quad f \quad} & Y \\ {\scriptstyle \Phi}\uparrow & & \uparrow{\scriptstyle \Psi} \\ U & \xrightarrow[h = \Psi^{-1} \circ f \circ \Phi]{} & V \end{array}$$

By taking derivatives, we obtain a commutative square of linear maps:

$$\begin{array}{ccc} T_x(X) & \overset{\mathrm{d}f_x}{\dashrightarrow} & T_y(Y) \\ {\scriptstyle \mathrm{d}\Phi_0}\uparrow & & \uparrow{\scriptstyle \mathrm{d}\Psi_0} \\ \mathbb{R}^k & \underset{\mathrm{d}h_0}{\longrightarrow} & \mathbb{R}^l \end{array}$$

Since Φ is a diffeomorphism, its derivative $\mathrm{d}\Phi_0$ is an isomorphism, (this is easily shown) and thus the only possible definition for $\mathrm{d}f_x$ is the following:

$$\mathrm{d}f_x = \mathrm{d}\Psi_0 \circ \mathrm{d}h_0 \circ \mathrm{d}\Phi_0^{-1}$$

As with the definition of tangent space, it is important to show that this definition is independent of the choice of parametrization, and this is done in the exact same way as for the tangent space, i.e. in Lemma 1.13

Now that the concept of derivative of a map between manifolds is well defined, we need to check whether the chain rule holds as we wanted. Let $f : X \to Y$ and $g : Y \to Z$ be two smooth maps between manifolds. Let $\Phi : U \to X$ be a parametrization of X around x, $\Psi : V \to Y$ one of Y around $y = f(x)$ and finally, let $\eta : W \to Z$ parametrize Z around $z = g(y)$. Assume again that each parametrization sends 0 to x, y and z respectively. We obtain the following commutative diagram:

$$\begin{array}{ccccc} X & \overset{f}{\longrightarrow} & Y & \overset{g}{\longrightarrow} & Z \\ {\scriptstyle \Phi}\uparrow & & {\scriptstyle \Psi}\uparrow & & \uparrow{\scriptstyle \eta} \\ U & \underset{h=\Psi^{-1}\circ f\circ\Phi}{\longrightarrow} & V & \underset{j=\eta^{-1}\circ g\circ\Psi}{\longrightarrow} & W \end{array}$$

We observe the square:

$$\begin{array}{ccc} X & \overset{g\circ f}{\longrightarrow} & Z \\ {\scriptstyle \Phi}\uparrow & & \uparrow{\scriptstyle \eta} \\ U & \underset{j\circ h}{\longrightarrow} & W \end{array}$$

By definition, we have :

$$\mathrm{d}(g \circ f)_x = \mathrm{d}\eta_0 \circ \mathrm{d}(j \circ h)_0 \circ \mathrm{d}\Phi_0^{-1}$$

Using the chain rule for Euclidean sets, we know that $\mathrm{d}(j \circ h)_0 = (\mathrm{d}j)_0 \circ (\mathrm{d}h)_0$. This yields:

$$\mathrm{d}(g \circ f)_x = (\mathrm{d}\eta_0 \circ \mathrm{d}j_0 \circ \mathrm{d}\Psi_0^{-1}) \circ (\mathrm{d}\Psi_0 \circ \mathrm{d}h_0 \circ \mathrm{d}\Phi_0^{-1}) = \mathrm{d}g_y \circ \mathrm{d}f_x$$

This proves that the chain rule holds.

1.3 The inverse function theorem and immersions

Now that we have the tools needed to differentiate on manifolds, we can start examining how maps between manifolds behave depending on the dimension of the given manifolds. This will illustrate other ways to show whether an object is a manifold and give criteria to determine when a geometric object is well-behaved enough to be a manifold.

We start with a generalization of the inverse function theorem to manifolds.

Definition 1.15 *Let X and Y be smooth manifolds of the same dimension. If a map $f : X \to Y$ carries a neighbourhood of a point x diffeomorphically onto a neighbourhood of the point $y = f(x)$, we say that f is a* local diffeomorphism *at x. If f is a local diffeomorphism at every point, we simply call it a* local diffeomorphism*.*

Theorem 1.16 (The inverse function theorem) *Let $f : X \to Y$ be a smooth map such that the derivative $\mathrm{d}f_x$ at the point x is an isomorphism. Then f is a local isomorphism at x.*

Proof This result is already known for open subsets of Euclidean space. The proof is similar, it just needs to be reformulated in the language of manifolds using parametrizations. □

The inverse Function theorem 1.16 can only be applied when both manifolds have the same dimension. Now, let us assume that $dimX < dimY$. In this situation the best thing we can require for the derivative at a certain point is for it to be injective.

Definition 1.17 *If the derivative $\mathrm{d}f_x : T_x(X) \to T_y(Y)$ at the point x of a map $f : X \to Y$ between two manifolds X and$Y$ is injective, the map f is called an* immersion *at x. If f is an immersion at every point $x \in X$, we simply say that f is an immersion.*

Example 1.18 *The simplest example of an immersion is the inclusion of $\mathbb{R}^k$ into $\mathbb{R}^l$ for $k \leq l$ where a point $(x_1, \dots, x_k)$ is sent to $(x_1, \dots, x_k, 0, \dots, 0)$. We call this map the* canonical immersion*.*

An interesting result is that the canonical immersion is actually, up to diffeomorphism, the only immersion there is. This means that with the right choice of parametrization, i.e. the right choice of diffeomorphism, we can write any immersion as the canonical immersion. This is formalized in the following theorem:

Theorem 1.19 (Local immersion theorem) *Let $f : X \to Y$ be an immersion at x, and let $y = f(x)$. Then there exist local coordinates around x and y such that f is locally equivalent to the canonical immersion near x, i.e.*

$$f(x_1, \dots, x_k) = (x_1, \dots, x_k, 0, \dots, 0).$$

Proof We start by choosing local parametrizations with $\Phi(0) = x$, $\Psi(0) = y$ and such that the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle\Phi}\uparrow & & \uparrow{\scriptstyle\Psi} \\ U & \xrightarrow[g]{} & V \end{array}$$

The function g above is uniquely defined by commutativity of the diagram. Our goal is to augment g in such a way that we may use the inverse function theorem 1.16. Since f is an immersion, g is one as well and thus $\mathrm{d}g_0 : \mathbb{R}^k \to \mathbb{R}^l$ is injective. By a well-suited change of basis in $\mathbb{R}^l$ we may assume that the linear function $\mathrm{d}g_0$ has an $l \times k$ matrix of the form:

$$\begin{pmatrix} I_k \\ 0 \end{pmatrix}$$

Here, I_k denotes the identity matrix in $\mathbb{R}^k$. Now, let us define a map $G : U \times \mathbb{R}^{l-k} \to \mathbb{R}^l$ by

$$G(x, z) = g(x) + (0, z)$$

The function G maps an open set of $\mathbb{R}^l$ into $\mathbb{R}^l$ and the derivative $\mathrm{d}G_0$ is the identity matrix I_l. Thus, we may use the inverse Function theorem 1.16, which tells us that G is a local diffeomorphism of $\mathbb{R}^l$ at 0. Now, the map Ψ is a diffeomorphism and thus also a local diffeomorphism at 0. Therefore, the map $\Psi \circ G$ is a local diffeomorphism at 0 as well, and we can use it as a local parametrization of Y around the point y. Moreover, by shrinking U and Y if necessary, we get the following diagram to commute:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle\Phi}\uparrow & & \uparrow{\scriptstyle\Psi} \\ U & \xrightarrow[\iota]{} & V \end{array}$$

Where ι denotes the canonical immersion. This completes the proof. $\square$

An immediate corollary of this theorem is the fact that if f is an immersion at the point x, it is also an immersion in a neighbourhood of x.

Now that we have characterized immersions a bit more precisely, let us examine the image of an immersion with the following question in mind: is the image of an immersion a submanifold? It is easy to find examples where that is not the case. What are the missing conditions that will allow us to make images of immersions nice submanifolds? Let us define some new concepts:

Definition 1.20 *A map $f : X \to Y$ is called* proper *if the preimage of every compact set in Y is compact in X.*

Definition 1.21 *An immersion that is injective and proper is called an* embedding.

With these new definitions we can now answer our question with the following theorem:

Theorem 1.22 *An embedding $f : X \to Y$ maps X diffeomorphically onto a submanifold of Y.*

Proof From the Local Immersion Theorem 1.19, we know that f maps any sufficiently small neighbourhood W of a point x diffeomorphically onto its image $f(W)$. If we can show that the set $f(W)$ is open for any open set W, any point of $f(X)$ will lie within a parametrizable neighbourhood, and thus $f(X)$ will be a submanifold. For a contradiction, let us assume that $f(W)$ is not open in $f(X)$. Then there exists a sequence of points y_i in $f(X)$ that do not lie in $f(W)$ but that converge to a point $y \in f(W)$. The set $\{y_i, y\}$ is compact because the y_i's converge. Therefore, because f is proper, its preimage in X is also compact. By injectivity, every y_i has exactly one preimage point x_i, and y has one preimage point x which must lie in W. Since the preimage set $\{x_i, x\}$ is compact, we have a converging subsequence (that we will still denote by x_i) with a limit point $z \in X$. Then by continuity we have $f(x_i) \to f(z)$. Since we also have $f(x_i) \to f(x)$ and since f is injective, we must have $x = z$, which means, that for i big enough, we have $x_i \in W$. This contradicts our construction of the y_i's that do not belong to W. Thus, $f(X)$ is a manifold. It is now easy to check that the map $f : X \to f(X)$ is a diffeomorphism, since it is already known that it is a local diffeomorphism and that it is bijective with f^{-1} also locally known to be smooth. □

1.4 Submersions

Let us now tackle the other dimensional case, namely, when $k = dim(X) \geq dim(Y) = l$. Let $f : X \to Y$ with $f(x) = y$. The strongest condition we can demand for the derivative $\mathrm{d}f_x : T_x(X) \to T_y(Y)$ is surjectivity.

Definition 1.23 *In the setting above, if $\mathrm{d}f_x$ is surjective, we call f a* submersion *at x. If f is a submersion at every point, we simply call it a submersion.*

Example 1.24 *As with immersions, the simplest example of a submersion is called the* canonical submersion *and is the standard projection of $\mathbb{R}^k$ onto $\mathbb{R}^l$ for $k \geq l$, i.e. the maps that sends a point $(x_1, \ldots, x_k)$ to the point $(x_1, \ldots, x_l)$.*

A similar statement as the Local Immersion Theorem also holds for submersions:

Theorem 1.25 (Local Submersion Theorem) *Let $f : X \to Y$ be a submersion at x, and $y = f(x)$. Then there exist local coordinates around x and y such that $f(x_1, \dots, x_k) = (x_1, \dots, x_l)$ for $k \geq l$, i.e f is locally equivalent to the canonical submersion near x.*

Proof Just as in the proof of the of the Local Immersion Theorem 1.19, we choose two parametrizations Φ and Ψ mapping x and y to 0 and such that the usual diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{\quad f \quad} & Y \\ {\scriptstyle\Phi}\uparrow & & \uparrow{\scriptstyle\Psi} \\ U & \xrightarrow[\quad g \quad]{} & V \end{array}$$

Again, we seek to modify g in order to apply the Inverse Function theorem 1.16. As $dg_0 : \mathbb{R}^k \to \mathbb{R}^l$ is surjective, by a change of basis in $\mathbb{R}^k$, we may assume that dg_0 has the $l \times k$ matrix $(I_l, 0)$. We now define $G : U \to \mathbb{R}^k$ as follows:

$$G(a) = (g(a), a_{l+1}, \dots, a_k),$$

for $a = (a_1, \dots, a_k)$. The matrix representation of dG_0 is the identity matrix I_k and thus G is a local diffeomorphism at 0. We can therefore define an inverse G^{-1} from a neighbourhood U' of 0 into U. By construction, g is the canonical submersion composed with G and thus we can write the canonical submersion as the function $g \circ G^{-1}$. We then obtain the following commutative square, where s is the canonical submersion:

$$\begin{array}{ccc} X & \xrightarrow{\quad f \quad} & Y \\ {\scriptstyle\Phi \circ G^{-1}}\uparrow & & \uparrow{\scriptstyle\Psi} \\ U' & \xrightarrow[\quad s \quad]{} & V \end{array}$$

□

As with immersions, an immediate corollary of this theorem is that a submersion at x is also a submersion in a whole neighbourhood of x.

We will now use the tools developed above to investigate the geometric nature of preimage sets such as $f^{-1}(y)$ for a smooth map $f : X \to Y$ between manifolds.

Definition 1.26 *Let $f : X \to Y$ be a smooth map of manifolds. A point $y \in Y$ is called a* regular value *for f if $df_x : T_x(X) \to T_y(Y)$ is surjective for every x in X such that $f(x) = y$. A point $y \in Y$ that is not a regular value of f is called a* critical value.

Theorem 1.27 (Preimage theorem) *Let y be a regular value of $f : X \to Y$, then the preimage set $f^{-1}(y)$ is a submanifold of X with dimension $dim(f^{-1}(y)) = dim\, X - dim\, Y$.*

Proof Since y is a regular value of f, f is a submersion at every point of the preimage, and thus for any x in $f^{-1}(y)$, by the local Submersion theorem 1.25, we can select local coordinates such that

$$f(x_1, \ldots, x_k) = (x_1, \ldots x_l)$$

and such that y corresponds to $(0, \ldots, 0)$. This allows us to write the preimage set as the set of points of the form $(0, \ldots, 0, x_{l+1}, \ldots, x_k)$. In other words, let U be the neighbourhood of x on which the coordinates x_i are defined. Then, $f^{-1}(y) \cap U$ is the set of points where $x_1 = 0, \ldots x_l = 0$. The remaining functions $x_{l+1}, \ldots x_k$ constitute a coordinate system on $f^{-1}(y) \cap U$, which is a relatively open subset of $f^{-1}(y)$. This exactly shows that $f^{-1}(y)$ is a submanifold of dimension $k - l$ and thus concludes the proof. □

The theorem above gives us another characterization of manifolds, which is often very useful when one wants to show that a particular object is indeed a manifold.

We may now consider a variant of the preceding argument to derive some results that will be of importance later. Let us take $g_1, \ldots, g_l$ l smooth, real-valued functions on a manifold X with $dim X = k \geq l$. Each g_i is smooth and thus the maps $\mathrm{d}(g_i)_x$ are linear functionals on the tangent space $T_x(X)$. We now define the function g as follows:

$$g = (g_1, \ldots, g_l) : \mathrm{X} \to \mathbb{R}^l$$

It is easily verified that $\mathrm{d}g_x : T_x(X) \to \mathbb{R}^l$ is surjective if and only if the linear functionals $\mathrm{d}(g_i)_x$ are linearly independent. This leads us to the following definition.

Definition 1.28 *In the setting described above, we say that the functions $g_1, \ldots, g_l$ are* independent *at x if the functionals $\mathrm{d}(g_i)_x$ for $1 \leq i \leq l$ are linearly independent.*

Proposition 1.29 *If the smooth, real-valued functions $g_1, \ldots, g_l$ on X are independent at each point where they all vanish, then the set Z of commons zeros is a submanifold of X with dimension $dim(X) - l$*

Proof If we set $g = (g_1, \ldots, g_l)$, the fact that the functions g_i are independent is equivalent to saying that $\mathrm{d}g_x$ is surjective for every x in Z, which means that $(0, \ldots, 0) \in \mathbb{R}^l$ is a regular value of g. The result then follows directly from the Preimage theorem 1.27. □

Definition 1.30 *Let X be a smooth manifold and $Z \subseteq X$ be a submanifold of X. We define the* codimension *of the submanifold Z by the formula $codim(Z) = dim(X) - dim(Z)$*

With this definition, the statement of the previous proposition can be restated in the following terms: "l independent functions on X cut out a submanifold of codimension l."

We now state and prove two partial converses to proposition 1.29.

Proposition 1.31 (partial converse 1) *If y is a regular value of the smooth map $f : X \to Y$ then the preimage submanifold $f^{-1}(y)$ can be cut out by independent functions*

Proof It is enough to choose a diffeomorphism h of a neighbourhood V of y onto a neighbourhood of the origin in $\mathbb{R}^l$, such that $h(y) = 0$. Set $g = h \circ f$. It is easily checked that 0 is a regular of g since y is a regular value of f. Therefore, we can take the l functions $g_1, \dots, g_l$. □

Proposition 1.32 (partial converse 2) *Every submanifold of X is* locally *cut out by independent functions. More specifically, for Z a submanifold of codimension l and for an arbitrary point $z \in Z$, we claim that there exist l independent functions $g_1, \dots, g_l$ defined on an open neighbourhood W of z in X such that $Z \cap W$ is the set of the common zeros of the g_i's.*

Proof The result follows directly from the Local Immersion theorem 1.19 for the immersion $Z \to W$. □

We end this section with another way to characterize the tangent space.

Proposition 1.33 *Let Z be the preimage set of a regular value $y \in Y$ under the smooth map $f : X \to Y$. Then the kernel of the derivative* $\mathrm{d}f_x : T_x(X) \to T_y(Y)$ *at any point x is exactly the tangent space to Z, i.e the space $T_x(Z)$.*

Proof The map f is constant on Z, and thus the map $\mathrm{d}f_x$ must be zero on $T_x(Z)$, which means that $T_x(Z)$ must be included in the kernel. But at the same time $\mathrm{d}f_x : T_x(X) \to T_y(Y)$ is surjective for any x in Z, which means that we must have, for $x \in Z$

$$dim(ker(\mathrm{d}f_x)) = dim(T_x(X)) - dim(T_y(Y)) = dim(X) - dim(Y) = dim(Z)$$

Thus, $T_x(Z)$ is a subspace of the kernel with the same dimension. Therefore, the two must be equal. □

1.5 Transversality

In this section, we will try to generalize the results that we have previously shown for preimage sets of the form $f^{-1}(y)$. Let $f : X \to Y$ be a smooth map of manifolds. Assume Z to be a submanifold of Y. We will examine the set of solutions to the condition $f(x) \in Z$ instead of considering the preimage of a single point. What assumptions are now necessary to make sure that $f^{-1}(Z)$ is a reasonable geometric object, i.e. a manifold? Being a manifold is a local property. That is, $f^{-1}(Z)$ is a manifold if and only if every point $x \in f^{-1}(Z)$ has a neighbourhood U in X such that $f^{-1}(Z) \cap U$ is a manifold. This observation allows us to reduce the relation $f(x) \in Z$ to the simpler case we have already considered previously, i.e. when Z is a single point, say $y \in Y$. In this case, using the results from the previous section (partial converse 1 1.31), we may write Z in a neighbourhood of y as the zero set of l independent smooth functions $g_1, \ldots, g_l$, with l the codimension of Z in Y. Then, around any point x such that $f(x) = y$, the preimage $f^{-1}(Z)$ is the zero set of the functions $g_1 \circ f, \ldots g_l \circ f$. We can then define $g = (g_1, \ldots, g_l)$ and apply our former results to the map $g \circ f : W \to \mathbb{R}^l$ where W is a neighbourhood around our point x. Hence, $(g \circ f)^{-1}(0)$ will be a submanifold if 0 is a regular value of $g \circ f$.

The condition we have obtained seems rather arbitrary, since the maps g_i are not necessarily unique. However, it is possible to reformulate the condition in terms of f only, which will insure that everything we have done is well-defined. By the chain rule, we have that:

$$\mathrm{d}(g \circ f)_x = \mathrm{d}g_y \circ \mathrm{d}f_x,$$

and thus the linear map $\mathrm{d}(g \circ f)_x : T_x(X) \to \mathbb{R}^l$ is surjective if and only if $\mathrm{d}g_y$ carries the image of $\mathrm{d}f_x$ onto the whole of $\mathbb{R}^l$. But g is a submersion by construction and thus $\mathrm{d}g_y : T_y(Y) \to \mathbb{R}^l$ is surjective linear map, whose kernel is by proposition 1.33, $T_y(Z)$. Thus, $\mathrm{d}g_y$ carries a subspace of $T_y(Y)$ onto $\mathbb{R}^l$ exactly when if that subspace and $T_y(Z)$ span the whole of $T_y(Y)$. We therefore conclude that $g \circ f$ is a submersion at the point $x \in f^{-1}(Z)$ if and only if

$$Image(\mathrm{d}f_x) + T_y(Z) = T_y(Y) \tag{1.1}$$

Definition 1.34 *If the condition above is fulfilled, we say that the map f is* transversal *to the submanifold Z, and we write $f \pitchfork Z$*

This whole discussion is summed up in the following theorem, that we incidentally proved above:

Theorem 1.35 *If the smooth map $f : X \to Y$ is transversal to a submanifold $Z \subseteq Y$, then the preimage $f^{-1}(Z)$ is a submanifold of X. Moreover, the codimension of $f^{-1}(Z)$ in X is equal to the codimension of Z in Y.*

Remark 1.36 *When Z is just a point y, its tangent space is the zero subspace of $T_y(Y)$. Thus, f is transversal to y if $\mathrm{d}f_x(T_x(X)) = T_y(Y)$ for all $x \in f^{-1}(y)$. This exactly means that y is a regular value of f. Hence, regularity is actually a special case a transversality.*

We now examine an important special case of the theorem above. This deals with the transversality of the inclusion map i of one submanifold $X \subseteq Y$

with another submanifold $Z \subseteq Y$. First, we notice that $i^{-1}(Z) = X \cap Z$ and that the derivative $\mathrm{d}i_x : T_x(X) \to T_x(Y)$ is simply the inclusion map of $T_x(X)$ into $T_x(Y)$. This yields that $i \pitchfork Z$ if and only if, for every $x \in X \cap Z$:

$$T_x(X) + T_x(Z) = T_x(Y). \tag{1.2}$$

It is worth noting that this equation is symmetric in X and Z.

Definition 1.37 *When the equation 1.2 above holds, we say that the two submanifolds X and Z are* transversal, *and we write $X \pitchfork Z$.*

We obtain the following theorem:

Theorem 1.38 *The intersection of two transversal submanifolds X and Z of Y is again a submanifold. Moreover, the codimension is additive with respect to intersection, i.e. :*

$$codim(X \cap Z) = codim X + codim Z$$

Proof By definition 1.37 of the transversality of two submanifolds, we have that $i^{-1}(Z) = X \cap Z$ is a submanifold of Y. Moreover, by the previous theorem (theorem 1.35), we know that the codimension of $i^{-1}(Z) = X \cap Z$ in X is equal to the codimension of Z in Y, i.e.

$$dim(X) - dim(X \cap Z) = dim(Y) - dim(Z)$$

$$\Leftrightarrow dim(Y) - dim(X \cap Z) = dim(X) + dim(Z) - 2dim(X \cap Z) \tag{1.3}$$

By symmetry of the transversality condition, we can also write:

$$dim(Z) - dim(X \cap Z) = dim(Y) - dim(X)$$

$$\Leftrightarrow -2dim(X \cap Z) = 2dim(Y) - 2dim(X) - 2dim(Z) \tag{1.4}$$

Plugging the second equation (1.3) into to first one (1.4) yields

$$dim(Y) - dim(X \cap Z) = 2dim(Y) - dim(X) - dim(Z)$$

which is exactly what we wanted. □

Remark 1.39 *The transversality of X and Z also depends on the ambient space Y in which they lie.*

1.6 Homotopy and Stability

We begin this section by defining one of the fundamental terms in topology, namely homotopy (for smooth maps). Let us denote by I the unit interval $[0, 1]$

Definition 1.40 (Homotopy of smooth maps) *Let $f_0, f_1 : X \to Y$ be two smooth maps of manifolds. We say that f_0 and f_1 are* homotopic *and we write $f_0 \simeq f_1$ if there exists a smooth map $F : X \times I \to Y$ such that $F(x, 0) = f_0(x)$ and $F(x, 1) = f_1(x)$ for all $x \in X$. We call the map F a* homotopy *between f_0 and f_1. We can think of homotopy as a smoothly evolving family of maps $f_t : X \to Y$ connecting f_0 to f_1. To recover this family of maps, simply set $f_t(x) = F(t, x)$.*

Definition 1.41 *One can show that being homotopic is an equivalence relation on smooth maps, and the resulting equivalence classes are called* homotopy classes.

The notion of homotopy allows us to mathematically formalize the concept of stability. For a given property of a function to stable, it should still be valid if the function is slightly deformed. This intuitive picture leads us to the next definition.

Definition 1.42 *A property of the map f_0 is called* stable *if whenever a map $f_0 : X \to Y$ has this property and $f_t : X \to Y$ is a homotopy of f_0, then there exists an $\varepsilon > 0$ such that each f_t with $t < \varepsilon$ also has the property. The collection of maps that possesses a specific stable property is called a* stable class *of maps.*

The next theorem shows that all the properties that we have defined so far are stable, if the domain manifold is compact.

Theorem 1.43 (Stability theorem) *The following classes of smooth maps of a compact manifold X into a manifold Y are stable classes:*

1. *local diffeomorphisms*
2. *immersions*
3. *submersions*
4. *maps transversal to any specified closed submanifold $Z \subset Y$*
5. *embeddings*
6. *diffeomorphisms*

Proof The proof for the first four classes is done in the same way. Local diffeomorphisms are actually immersions in the special case $dim(X) = dim(Y)$, so we start with the proof for immersions. Let f_0 be an immersion and let f_t be a homotopy of f_0. We want to find an $\varepsilon > 0$ such that the map $\mathrm{d}(f_t)_x$ is injective for all $(x, t) \in X \times [0, \varepsilon) \subseteq X \times I$. Because X is compact, any open neighbourhood of $X \times 0$ will contain $X \times [0, \varepsilon]$ for ε small enough. Thus, it is enough to show that each point $(x_0, 0)$ has a neighbourhood $U \subseteq X \times I$ in which $\mathrm{d}(f_t)_x$ is injective for all $(x, t) \in U$. This is a local assertion, and thus we may assume that X is an open set in $\mathbb{R}^l$ and Y an open set in $\mathbb{R}^k$. Since $\mathrm{d}(f_0)_{x_0}$ is injective, the Jacobian

$$\left(\frac{\partial (f_0)_i}{\partial x_j} (x_0) \right)$$

is an $l \times k$ matrix containing a $k \times k$ submatrix with non-zero determinant. But by definition 1.40 of a homotopy, each partial derivative

$$\frac{\partial (f_t)_i}{\partial x_j}(x)$$

is continuous as a function on $X \times I$. Since the determinant is also a continuous function, we must have a neighbourhood around $(x_0, 0)$ where the Jacobian still contains a $k \times k$ submatrix with non-zero determinant, which is exactly what we wanted to prove. The proof for submersions is identical. For the fourth class, it is enough to recall that transversality can locally be expressed as a submersion condition and then use the previous part.

Let us now prove the claim for the class of embeddings. Since X is compact, the only thing left to show is the stability of the injectivity condition. For this, we define the smooth map $G : X \times I \to Y \times I$ by $G(x,t) = (f_t(x), t)$. Towards a contradiction, we assume that embeddings do not form a stable class, and so there exist a sequence $(t_i) \to 0$ and distinct points x_i and y_i such that $G(x_i, t_i) = G(y_i, t_i)$. By compactness of X, we may take a subsequence to get $x_i \to x_0$, and similarly $y_i \to y_0$. Then, by continuity:

$$G(x_0, 0) = \lim G(x_i, t_i) = \lim G(y_i, t_i) = G(y_0, 0)$$

But $G(x_0, 0) = f_0(x_0)$ and $G(y_0, 0) = f_0(y_0)$ and hence by injectivity of f_0 we must have $x_0 = y_0$. Locally, we may work in Euclidean space and thus consider the matrix $\mathrm{d}G_{(x_0,0)}$, which is just

$$\begin{pmatrix} & a_1 \\ \mathrm{d}(f_0)_{x_0} & \vdots \\ & a_l \\ 0 \dots 0 & 1 \end{pmatrix}$$

where the numbers a_j are not important. Since $\mathrm{d}(f_0)_{x_0}$ is injective, its matrix must have k independent rows, which implies that $\mathrm{d}G_{(x_0,0)}$ has $k+1$ independent rows, so $\mathrm{d}G_{(x_0,0)}$ is an injective linear map. This implies that G is an immersion around $(x_0, 0)$ and thus by the local immersion theorem, it must be injective in some neighbourhood of of $(x_0, 0)$. But for i large enough we have that both (x_i, t_i) and (y_i, t_i) belong to this neighbourhood, which leads to a contradiction.

The last remaining class of stable functions is the class of diffeomorphisms. Since a homotopy is a smooth deformation of a smooth function, any f_t will be a smooth map as well. We have already proven that injectivity was stable so it remains to show the same for surjectivity. Since a diffeomorphism is continuous by definition, points that belong to the same connected component cannot be mapped to two different connected components. We can thus without loss of generality assume that X is connected. Incidentally, because f_0 is a diffeomorphism, Y must be connected as well. Additionally, we know that f_t is local diffeomorphism for t small enough, which implies that it maps open sets onto open sets. Hence, $Im(f_t) = f_t(X)$ is an open set. But since X is compact and f_t continuous, we also have that $f_t(X)$ is compact. Since we are looking at manifolds of Euclidean space (or more generally because we require our topological spaces to be Hausdorff), any compact set is closed, which means that $f_t(X)$ is closed. Since Y is connected and $f_t(X)$ is both open and closed, we must have $f_t(X) = Y$, for t small enough which shows surjectivity. The inverse f_t^{-1} is already locally known to be smooth so the proof is complete. □

1.7 Sard's Theorem and Morse Functions

What we have seen so far is that regular values are useful to characterize preimage sets as nice geometric objects, namely manifolds. The next question that might come to mind is the following: "How common are regular values?" Maybe such a condition is too strong to be of any use once we are dealing with concrete examples of maps of manifolds. Luckily for us, this is not the case, as the next result shows:

Theorem 1.44 (Sard's theorem) *The set of critical values of a smooth map of manifolds $f : X \to Y$ has measure zero.*

Proof The proof can be found in the Appendix, since it has little to do with the topological considerations which interest us the most in this text. □

Corollary 1.45 *The regular values of any smooth map $f : X \to Y$ are dense in Y. Moreover, for any countable number of smooth maps $f_i : X \to Y$ the points of Y that are simultaneously regular values for all of the f_i's are dense.*

Proof This follows directly from Sard's theorem B.1 above and from the fact that the union of a countable number of sets of measure zero still has measure zero. □

We now give a new terminology for concepts that we have already encountered:

Definition 1.46 *Let $f : X \to Y$ be a smooth map of manifolds. A point x in X is called a* regular point *of f if $\mathrm{d}f_x : T_x(X) \to T_y(Y)$ is surjective. One sometimes says that f* is regular *at x. This is exactly the condition that f is a submersion at x. We now just have a new name for it. If $\mathrm{d}f_x$ is not surjective, we call x a* critical point *of f.*

Let us, for a moment, consider smooth functions on a manifold X, that is functions $f : X \to \mathbb{R}$. Then f is either regular at the point x, or $\mathrm{d}f_x = 0$.

The local submersion theorem 1.25 gives us the local behaviour of f at all its regular points. The goal is now to develop the machinery required to characterize the behaviour of f around critical points as well. It is worth noting that many maps are often compelled to have critical points by the topology of the manifolds on which they are defined. For example, a function f on a compact manifold X (for X not a single-point space) must take on a maximum and a minimum. If the value $f(x)$ is extremal, then f cannot be a coordinate system around x, which implies that the derivative $\mathrm{d}f_x$ is 0. However, there is a class of critical points that is "best-behaved". It is known from analysis that an easy way to check whether a critical point is a maximum, a minimum, or a saddle point is to compute the *Hessian matrix* $H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$. The definiteness of H then gives information about the nature of the critical point.

Definition 1.47 *If the Hessian matrix is non-singular at a critical point x, we say that x is a* non-degenerate critical point of f.

The non-degeneracy property is independent from the coordinate system that we choose, and this will be clear from the intrinsic definition of the Hessian that we will give now:

Definition 1.48 *If x is a critical point of f we define a symmetric bilinear functional f_{**} on $T_x(X)$, X being the manifold on which f is defined. The bilinear form f_{**} is called the* Hessian of f at the point x, *and is constructed as follows: if $v, w \in T_x(X)$, then we can extend those two vectors to smooth vector fields $\bar{v}$ and $\bar{w}$ on X and we set*

$$f_{**}(v, w) = \bar{v}_x(\bar{w}(f))$$

The notation might seem unfamiliar, so let us break it down a bit: $\bar{v}_x$ is the value of the vector field at the point x, which is of course just the tangent vector v. Furthermore, if g is a smooth function on X, we interpret $v(g)$ as the directional derivative of g at the point x in the direction v, which is a real number. In our case, our function g is $\bar{w}(f)$, which is a function on X defined by $\bar{w}(f)(p) = \bar{w}_p(f)$.

Let us now check that f_{**} is indeed symmetric:

Proposition 1.49 *The Hessian bilinear functional is symmetric.*

Proof We have:

$$\bar{v}_x(\bar{w}(f)) - \bar{w}_x(\bar{v}(f)) = [\bar{v}, \bar{w}]_x(f) = 0$$

The vector field $[\bar{v}, \bar{w}]$ is called the Poisson bracket of $\bar{v}$ and $\bar{w}$. It is essentially defined by the subtraction on the left-hand side, but it is not very important for our discussion. What is important is that the Poisson bracket is a vector field and thus $[\bar{v}, \bar{w}]_x(f)$ is a directional derivative of f at the point x. Since x is a critical point of f, all directional derivatives must be zero, which justifies the last equality and thus proves symmetry. $\square$

The symmetry of f_{**} allows us to check that it is well-defined, i.e. independent of the choice of extensions $\bar{v}$ and $\bar{w}$. By symmetry, we can write

$$v(\bar{w}(f)) = \bar{v}_x(\bar{w}(f)) = \bar{w}_x(\bar{v}(f)) = w(\bar{v}(f))$$

Since the left-hand side is independent of $\bar{v}$ and the right-hand side independent of $\bar{w}$, the Hessian is indeed well-defined.

Remark 1.50 *If $(x_1, \ldots, x_n)$ is a local coordinate system and $v = \sum_{i=1}^{n} a_i \frac{\partial f}{\partial x_i}\Big|_x$, $w = \sum_{j=1}^{n} b_j \frac{\partial f}{\partial x_j}\Big|_x$, we can choose $\bar{w} = \sum_{j=1}^{n} b_j \frac{\partial}{\partial x_j}$ where the b_j's are constant functions. This yields:*

$$f_{**}(v, w) = v(\bar{w}(f))(x) = v\left(\sum_{j=1}^{n} bj \frac{\partial f}{\partial x_j}\right) = \sum_{i,j=1}^{n} a_i b_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

so the Hessian matrix defined previously and known from analysis represents f_{**} with respect to the basis $\frac{\partial}{\partial x_1}\Big| x, i = 1, \ldots, n$

We can now introduce some useful terminology.

Definition 1.51 *The* index *of a bilinear functional H on a vector space V is defined to be the maximal dimension of a subspace $W \subseteq V$ on which H is negative definite. The* nullity *of H is the dimension of the null-space, i.e. the space $N = \{v \in V | H(v, w) = 0 \, \forall w \in V\}$.*

A critical point p in a manifold X can thus be non-degenerate if and only if f_{**} on T_pX has nullity 0. We will refer to the index of f_{**} as the *index of f at p*.

We will shortly state and prove the Morse Lemma 1.53, which shows that the behaviour of a function at critical points is entirely determined by the index of the given function. First, we need a technical lemma.

Lemma 1.52 *Let f be a smooth function in a convex neighbourhood V of 0 in $\mathbb{R}^n$ such that $f(0) = 0$. Then*

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

for some smooth functions g_i defined in V and such that $g_i(0) = \frac{\partial f}{\partial x_i}(0)$.

Proof

$$f(x_1,\ldots,x_n) = \int_0^1 \frac{\mathrm{d}f(tx_1,\ldots,tx_n)}{dt}\,dt = \int_0^1 \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(tx_1,\ldots,tx_n)\cdot x_i\,dt.$$

Therefore we can simply define $g_i(x_1,\ldots,x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1,\ldots,t_n)dt$ □

Lemma 1.53 (Morse Lemma) *Let f be a smooth map on a manifold and p be a non-degenerate critical point of f. Then there is a local coordinate system $(y_1,\ldots,y_n)$ in a neighbourhood U of p with $y_i(p) = 0$ for all i and such that, for all x in U, we have:*

$$f(x) = f(p) - (y_1(x))^2 - \cdots - (y_\lambda(x))^2 + (y_{\lambda+1}(x))^2 + \cdots + (y_n(x))^2 \quad (1.5)$$

where λ is the index of f at p.

Proof We first show that if f can be written as above inside the neighbourhood U, then the number λ must be the index of f at p. Let $(z_1,\ldots,z_n)$ be an arbitrary coordinate system such that

$$f(x) = f(p) - (z_1(x))^2 - \cdots - (z_\lambda(x))^2 + (z_{\lambda+1}(x))^2 + \cdots + (z_n(x))^2.$$

Then we must have:

$$\frac{\partial^2 f}{\partial z_i \partial z_j}(p) = \begin{cases} -2 & \text{if } i = j \leq \lambda, \\ 2 & \text{if } i = j > \lambda, \\ 0 & \text{otherwise ,} \end{cases}$$

Thus, the matrix representing the Hessian f_{**} with respect to the basis $\frac{\partial}{\partial z_1}\Big|_p, \ldots, \frac{\partial}{\partial z_n}\Big|_p$ is given by the diagonal matrix

$$\begin{pmatrix} -2 & & & & & \\ & \ddots & & & & \\ & & -2 & & & \\ & & & 2 & & \\ & & & & \ddots & \\ & & & & & 2 \end{pmatrix}$$

Therefore, there is a subspace of T_pX of dimension λ where f_{**} is negative definite, and a subspace V of dimension $n - \lambda$ where f_{**} is positive definite. If there were a subspace of dimension strictly greater than λ where f_{**} were negative definite, then this subspace would intersect with V, which is not possible, since f_{**} is positive definite on V. This shows that λ is the index of f at p. Now, we need to show that such a suitable coordinate system $(y_1,\ldots,y_n)$ exists. Without loss of generality, we can assume that p is the origin of $\mathbb{R}^n$ and that $f(p) = f(0) = 0$. Using the previous lemma 1.52, we may write

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

for $(x_1, \ldots, x_n)$ in some convex neighbourhood of 0. Since 0 is assumed to be a critical point of f, lemma 1.52 gives us

$$g_j(0) = \frac{\partial f}{\partial x_j}(0) = 0$$

Therefore we can apply lemma 1.52 once again to the functions g_j, which yields

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

for some smooth functions h_{ij}. Putting everything together, we can write

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

We may assume that $h_{ij} = h_{ji}$ by using the following trick. We write $\bar{h}_{ij} = \frac{1}{2}(h_{ij} + h_{ji})$. Then, we have indeed that $\bar{h}_{ij} = \bar{h}_{ji}$ and $f = \sum x_i x_j \bar{h}_{ij}$. Furthermore, lemma 1.52 also gives that the matrix $(\bar{h}_{ij}(0))$ is actually equal to $(\frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(0))$, which is non-singular by assumption. We claim there is a non-singular transformation of the coordinate system that yields the desired from for f, by shrinking the neighbourhood around 0 if necessary. This process is similar the diagonalization proof for quadratic forms. The proof of this claim uses induction. Let us assume by induction that there exist coordinates $u_1, \ldots, u_n$ in a neighbourhood U_1 of 0 such that

$$f = \pm(u_1)^2 \pm \cdots \pm (u_{r-1}) + \sum_{i,j \geq r} u_i u_j H_{ij}(u_1, \ldots, u_n)$$

in U_1 and such that the matrices $(H_{ij}(u_1, \ldots, u_n))$ are symmetric. After a linear coordinate transformation in the last $n - r + 1$ variables, we may assume that $H_{rr}(0) \neq 0$. Let $g(u_1, \ldots, u_n)$ denote the square root of $|H_{rr}(u_1, \ldots, u_n)|$. The function g is a smooth, non-zero function of $u_1, \ldots, u_n$ at least on some smaller neighbourhood $U_2 \subseteq U_1$ of 0. We can now introduce new coordinates $v_1, \ldots, v_n$ by $v_i = u_i$ for $i \neq r$. For the r-th coordinate function, we set

$$v_r(u_1, \ldots, u_n) = g(u_1, \ldots, u_n) \left[u_r + \sum_{i > r} u_i H_{ir}(u1, \ldots, u_n) / H_{rr}(u_1, \ldots, u_n) \right].$$

The inverse function theorem 1.16 ensures that $v_1, \ldots, v_n$ will be valid coordinate functions on a sufficiently small neighbourhood U_3 of 0. Then it is easy to check that

$$f = \sum_{i \leq r} \pm (v_i)^2 + \sum_{i,j>r} v_i v_j H'_{ij}(v_1, \ldots, v_n)$$

By iterating this process as many times as needed, namely n times, we get the desired coordinate system, which completes the proof. □

Remark 1.54 *The Morse lemma 1.53 thus tells us that we can locally describe the behaviour of a function around non-degenerate critical points. Two direct consequences of the lemma is that non-degenerate critical points are isolated, and thus in particular, a function on a compact manifold only has finitely many non-degenerate critical points. Indeed for each point x take an open set U_x such that U_x contains at most one critical point. This constitutes an open cover X and compactness gives us a finite subcover and thus finitely many critical points.*

It is now useful to define a new class of functions.

Definition 1.55 *A function whose critical points are all non-degenerate is called a* Morse function.

Morse functions are really useful, because they give a lot of information about the topology of their domain manifold. This is the main topic of this text and will be discussed at length in the next chapter. Another interesting point is that Morse functions are actually quite common, in a sense that will be specified in the theorem below. The occurrence of critical points that are degenerate is thus a rare phenomenon.

Theorem 1.56 *Let us take a manifold $X \subseteq \mathbb{R}^n$ of dimension k, with $x_1, \ldots, x_n$ the usual coordinate functions on $\mathbb{R}^n$. Let $f : X \to \mathbb{R}$ be an arbitrary smooth function and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. We define a new function f_a on X by setting*

$$f_a = f + a_1 x_1 + \cdots + a_n x_n$$

Then the function f_a is a Morse function for almost every $a \in \mathbb{R}^n$.

We divide the proof in two steps by first stating the same result in $\mathbb{R}^k$ as a lemma.

Lemma 1.57 *Let f be a smooth function on an open set $U \subseteq \mathbb{R}^k$. Then for almost every $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$, the function f_a defined as above is a Morse function on U.*

Proof We again define the map $g : U \to \mathbb{R}^k$ as follows:

$$g = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_k} \right)$$

Now, we can write the derivative of f_a at a point p by

$$(\mathrm{d}f_a)_p = \left(\frac{\partial f_a}{\partial x_1}(p), \dots, \frac{\partial f_a}{\partial x_k}(p)\right) = g(p) + a$$

Thus p is a critical point of f_a if and only if $g(p) = -a$. Now, we observe that since f_a and f have the same second partial derivatives, they also have the same Hessian, namely the matrix $(\mathrm{d}g)_p$. Now let us assume that $-a$ is a regular value of g. Then, whenever $g(p) = -a$, $(\mathrm{d}g)_p$ is non-singular, which exactly means that every critical point of f_a is non-degenerate. Finally, using Sard's theorem B.1, we know that our assumption that $-a$ is a regular value of g holds for almost every $a \in \mathbb{R}^k$, which completes the proof. □

Proof (proof of Theorem 1.56) Let x be any point in X and $x_1, \dots, x_n$ the usual coordinate functions on $\mathbb{R}^n$. We first claim and prove the following: the restrictions of some k of these coordinate functions $x_{i_1}, \dots, x_{i_k}$ to X constitute a coordinate system in a neighbourhood of x. Let $\Phi_1, \dots, \Phi_n$ be the standard basis of linear functionals on $\mathbb{R}^n$. Then some k of these functionals $\Phi_{i_1}, \dots, \Phi_{i_k}$ are linearly independent when we restrict them to the tangent space $T_x(X)$. Now, the derivative of the function $x_i : \mathbb{R}^n \to \mathbb{R}$ is exactly the functional $\Phi_i : \mathbb{R}^n \to \mathbb{R}$, and so the derivative of the restriction of x_i to X must be the restriction of Φ_i to $T_x(X)$. The linear independence of the functionals Φ_{i_k} implies that the map $(x_{i_1}, \dots, x_{i_k}) : X \to \mathbb{R}^k$ is a local diffeomorphism at x. Since the point x we had was arbitrary, we can cover X with open subsets U_α such that for each α, some k of the coordinate functions form a coordinate system. Furthermore, since we are in the Euclidean space $\mathbb{R}^n$, the second axiom of countability holds and we thus may assume that there are countably many U_α's. Let us take one of these sets U_α and assume without loss of generality and to ease the notation that $(x_1, \dots, x_n)$ is a coordinate system on U_α. For any $n-k$-tuple $c = (c_{k+1}, \dots, c_n)$ we define the function

$$f_{(0,c)} = f + c_{k+1}x_{k+1} + \cdots + c_n x_n$$

By lemma 1.57, we know that for almost every $b \in \mathbb{R}^k$, the function

$$f_{(b,c)} = f_{(0,c)} + b_1 x_1 + \cdots + b_k x_k$$

is a Morse function on the open set U_α. Now, let S_α be the set of all $a \in \mathbb{R}^n$ such that f_a is not a Morse function on U_α. What we have shown with our previous construction is that every "horizontal slice" $S_\alpha \cap \mathbb{R}^k \times c$ has measure zero, when considered as a subset of $\mathbb{R}^k$. It is a consequence of Fubini's theorem that every S_α consequently has measure 0 in $\mathbb{R}^n$. It is clear that a function has a degenerate critical point in X if and only if it has one in some set U_α. Thus the set of $a \in \mathbb{R}^n$ for which f_a is not a Morse function is the union of all the sets S_α. We know that there are countably many U_α

and thus countably many S_α. Since the union of a countable number of sets of measure zero still has measure zero, the proof is complete. □

We finish this section with a few interesting facts about Morse functions.

Example 1.58 *Let $f : S^{k-1} \to \mathbb{R}$ be the height function $(x_1, \dots, x_k) \mapsto x_k$ on the sphere. We can view f as the restriction of the projection $\pi : \mathbb{R}^k \to \mathbb{R}$ to the sphere. This implies that for $x \in S^{k-1}$, the derivative $\mathrm{d}f_x : T_x(S^{k-1}) \to \mathbb{R}$ is the restriction of the derivative of the projection π, which is $\mathrm{d}\pi_x = \pi$ because the projection is a linear map. Thus, x will be a critical point if and only if we have $T_x(S^{k-1}) = \mathbb{R}^{k-1} \times \{0\}$. To characterize the tangent space in another way, let us examine the function $g : S^{k-1} \to \mathbb{R}$ defined by $g(x) = ||x||^2$. It is then clear that 1 is a regular value of g and that we have $g^{-1}(1) = S^{k-1}$. Using proposition 1.33, we may write $T_a(S^{k-1}) = ker(x \mapsto 2a^t x)$ where a^t is the transpose of a. Then, we have that $T_x(S^{k-1}) = \mathbb{R}^{k-1} \times \{0\}$ if and only if $a = (0, \dots, 0, 1) = N$ or $a = (0, \dots, 0, -1) = S$. The critical points are thus the two poles N and S. One can compute the Hessian matrix of the function composed with the inverse of the stereographic projections to verify that those critical points are indeed non-degenerate. Thus f is a Morse function.*

Proposition 1.59 *Let f be a smooth function on an open set $U \subseteq \mathbb{R}^k$. For each $x \in U$, we denote by $H(x)$ the Hessian matrix of f at the point x. Then f is a Morse function if and only if the following inequality holds across the set U.*

$$det(H)^2 + \sum_{i=1}^{k} \left(\frac{\partial f}{\partial x_i} \right)^2 > 0 \tag{1.6}$$

Proof "⇒" Let us take an arbitrary point $x \in U$. If x is a regular point, then we must have $\frac{\partial f}{\partial x_i} \neq 0$ for some i, which is enough to give us the desired inequality 1.6. If x is a critical point, it must be non-degenerate because f is a Morse function, and thus we must have $det(H) \neq 0$, which again directly yields 1.6.

"⇐" Let $x \in U$ be a critical point of f. Then we have that $\frac{\partial f}{\partial x_i}(x) = 0$ for $i = 1, \dots, k$. But then we must have by assumption that $det(H)^2 > 0$, which implies that H is non-singular. Since we chose an arbitrary critical point, we can conclude that f is a Morse function, as desired. □

Proposition 1.60 *Let $f_t : \mathbb{R}^k \to \mathbb{R}$ be a family of homotopic functions. Suppose additionally that f_0 is a Morse function in some neighbourhood U of a compact set K. Then each f_t is also a Morse function on U for t sufficiently small.*

Proof The argument is fairly similar to the proof of the stability theorem. The compactness of K implies that any open neighbourhood of $K \times \{0\}$ in $\mathbb{R}^k \times [0, 1]$ contains $K \times [0, \varepsilon]$ for some $\varepsilon > 0$ small enough.

Then, because every f_t is smooth and the homotopy is smooth as well, each partial $\frac{\partial f_t}{\partial x_i}(x)$ is smooth as a function of $\mathbb{R}^k \times [0,1]$, and similarly for $\frac{\partial^2 (f_t)_i}{\partial x_i^2}(x)$, and hence also for $det(H_t)(x)$, where H_t denotes the Hessian matrix of f_t. This implies that

$$det(H)^2 + \sum_{i=1}^{k} \left(\frac{\partial f}{\partial x_i} \right)^2 > 0$$

in some neighbourhood L of $K \times [0,1]$. Thus by the first remark of this proof, there exists some neighbourhood $V \times [0, \varepsilon] \subset L$ where this inequality holds. Then, using proposition 1.59 concludes the proof. □

Proposition 1.61 (Stability of Morse functions) *Let f be a Morse function on the compact manifold X, and let f_t be a homotopic family of functions with $f_0 = f$. Then each function f_t is a Morse function for t sufficiently small. In other words, the property of being a Morse function is a stable property.*

Proof For an arbitrary point $x \in X$, let $\Phi : \mathbb{R}^k \supset V_x \to X$ be a local parametrization around x such that $\Phi(0) = x$. Now choose r such that $\overline{B}_r(0) \subset V_x$ where $B_r(0)$ denotes the Euclidean ball of radius r around 0. We define $U_x = \Phi(B_r(0))$, which is an open neighbourhood of the compact set $\Phi(\overline{B}_{\frac{r}{2}}(0)$ (Compactness follows from Φ being a diffeomorphism of Euclidean spaces). Using the previous theorem 1.60 we know that there exists $\varepsilon_x > 0$ such that f_t is a Morse function on U_x for all t in $[0, \varepsilon_x)$.

Since X is a compact manifold, and $\{U_x\}_{x \in X}$ is an open cover of X, we can find a finite subcover $U_{x_1}, \ldots, U_{x_N}$. By setting $\varepsilon = min\{\varepsilon_{x_1}, \ldots, \varepsilon_{x_n}\}$, we get that f_t is a Morse function on the whole manifold X for t in $[0, \varepsilon)$. This completes the proof. □

Proposition 1.62 *Let X be a compact manifold, then there exist Morse functions on X which take distinct values for each of the critical points.*

Proof The idea of the proof is to use the stability of Morse functions to modify the values of the functions slightly at critical points where it is necessary. More precisely, let f be a Morse function on the compact manifold X. One can construct f using theorem 1.56. Let $x_1, \ldots x_N$ be its critical points. For every critical point, we take a smooth cut-off function ρ_i that is identically 1 on a small neighbourhood of x_i and 0 outside a slightly bigger neighbourhood. We then choose constants $a_1, \ldots, a_N$ such that for $i \neq j$

$$f(x_i) + a_i \neq f(x_j) + a_j.$$

We then define the function $g = f + \sum_{i=1}^{N} a_i \rho_i$. Since the ρ_i's are constant around the critical points of f, this new function has the same critical points, and they are non-degenerate again because the ρ_i's are locally constant. We can choose the coefficients a_i to be very small so that g get arbitrarily close to f. The function g is then a Morse function taking different values at every critical point. □

1.8 1-parameter group of diffeomorphisms

We conclude this chapter by a technical discussion that will be of use later.

Definition 1.63 *A* 1-parameter group of diffeomorphisms *of a manifold M is a C^∞ map*

$$\varphi : \mathbb{R} \times M \to M$$

such that the following holds:

1. *for each $t \in \mathbb{R}$, the map $\varphi_t : M \to M$ defined by $\varphi_t(q) = \varphi(t,q)$ is a diffeomorphism of M onto itself.*
2. *for all $t,s \in \mathbb{R}$, we have $\varphi_{t+s} = \varphi_t \circ \varphi_s$*

Given such a 1-parameter group φ of diffeomorphisms of M, we define the vector field X on M as follows. For every smooth real valued function f, set

$$X_q(f) = lim_{h\to 0}\frac{f(\varphi_h(q)) - f(q)}{h}$$

This vector field is said to generate *the group φ*

Lemma 1.64 *A smooth vector field X on M which vanishes outside of a compact set $K \subseteq M$ generates a unique 1-parameter group of diffeomorphims of M.*

Proof Given any smooth curve $t \to c(t) \in M$ we define the velocity vector $\frac{\mathrm{d}c}{\mathrm{d}t} \in T_{c(t)}M$ by the identity $\frac{\mathrm{d}c}{\mathrm{d}t}(f) = lim_{h\to 0}\frac{fc(t+h)-fc(t)}{h}$. Now, let φ be a 1-parameter group of diffeomorphisms generated by the vector field X. Then, we obtain the following equalities:

$$\frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t}(f) = \lim_{h\to 0}\frac{f(\varphi_{t+h}(q)) - f(\varphi_t(q))}{h} = \lim_{h\to 0}\frac{f(\varphi_h(p)) - f(p)}{h} = X_p(f),$$

where $p = \varphi_t(q)$. Thus, the curve $t \mapsto \varphi_t(q)$ satisfies the differential equation

$$\frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t} = X_{\varphi_t(q)} \tag{1.7}$$

with the initial condition $\varphi_0(q) = q$. Note that if we use local coordinates $u_1, \ldots, u_n$, the differential equation has the more familiar form $\frac{\mathrm{d}u_i}{\mathrm{d}t} = x_i(u_1, \ldots, u_n)$, for $i = 1, \ldots, n$. It is known form analysis that such a differential equation, at least locally, has a unique solution depending smoothly on the initial condition. Hence, for every $q \in M$, there exists a neighbourhood U and a number $\varepsilon > 0$ such that the differential equation 1.7 has a unique smooth solution for $q \in U$ and $|t| < \varepsilon$.

Using compactness, we can cover K with finitely many of these sets U. We now define $\varepsilon_0 > 0$ to be the smallest ε among all the finitely many pairs (U, ε). For, $q \notin K$, we set $\varphi_t(q) = q$. This yields a solution to 1.7 for all $q \in M$, as long as $|t| < \varepsilon_0$. This solution is smooth as a function of both variables, and granted $|t|, |s|, |t+s| < \varepsilon_0$, it is clear that $\varphi_{t+s} = \varphi_t \circ \varphi_s$.

We now need to define φ_t for $|t| \geq \varepsilon_0$. Since any number t can be written as k times $\varepsilon_0/2$ plus a remainder $r < \varepsilon_0$. Then, for $t = k(\varepsilon_0/2) + r$ with $k \geq 0$, we an write

$$\varphi_t = \varphi_{\varepsilon_0/2} \circ \varphi_{\varepsilon_0/2} \circ \cdots \circ \varphi_{\varepsilon_0/2} \circ \varphi_r$$

where $\varphi_{\varepsilon_0/2}$ is iterated k times. If $k < 0$, we can just replace $\varepsilon_0/2$ by $-\varepsilon_0/2$ iterated $-k$ times. Thus, φ_t is well defined for all values of t. As a composition of smooth functions, it is smooth and it is clear by construction that it satisfies the condition $\varphi_{s+t} = \varphi_t \circ \varphi_s$. This completes the proof. $\square$

Chapter 2

Morse theory

2.1 An important example

The reason for all the terminology we introduced in the first chapter is that critical points of Morse functions can be used to describe the topology of the manifold on which the Morse functions are defined. This is actually what is meant with Morse theory. We will shortly illustrate this with a very insightful example that actually serves as an introduction in Milnor's book [8]. However, we first need a few new concepts.

Definition 2.1 *Let X and Y be topological spaces, with $X' \subset X$ a closed subset. Let $f : X' \to Y$ be a continuous map. The* adjunction space *$X \cup_f Y$ is defined to be the disjoint union $X \sqcup Y$ where we identify x with $f(x)$ for all x in X'. More formally, we have that $X \cup_f Y = (X \sqcup Y)/\sim$, where $\sim$ is the smallest equivalence relation, such that $x \sim f(x), \forall x \in X'$*

An informal way of thinking about this space is by taking Y and attaching $X \setminus X'$ to it. We will now focus on a particular instance of adjunction space.

Definition 2.2 *Let B^n be the closed unit ball of dimension n with $n \geq 1$, and S^{n-1} be the $n-1$-unit sphere. We call the space $E^n := B^n \setminus S^{n-1}$* the standard n-cell. *This is actually just the open unit ball. For the case $n = 0$, we define E^0 to simply be a point. If X is a topological space, and $E \subseteq X$ is a subset homeomorphic to E^n, we call E* an n-cell in *X. If Y is another topological space, and $f : S^{n-1} \to Y$ is continuous, the space $B^n \cup_f Y$ is said to be obtained from Y by* attaching an n-cell.

We now have the necessary vocabulary to tackle Milnor's example.

Example 2.3 *Let M be a torus tangent to the plane V, and let $f : M \to \mathbb{R}$ denote the height function above the plane. We define M^a to be the set of all $x \in M$ such that $f(x) \leq a$. Then, by choosing different values for a, we observe the following:*

1. *If $a < 0 = f(p)$ then M^a is the empty set.*

2. *If $f(p) < a < f(q)$, then M^a is homeomorphic to a 2-cell.*
3. *If $f(q) < a < f(r)$, then M^a is homeomorphic to a cylinder.*
4. *If $f(r) < a < f(s)$, then M^a is homeomorphic to a compact manifold of genus one having a circle as boundary.*
5. *If $f(s) < a$, then M^a is the full torus.*

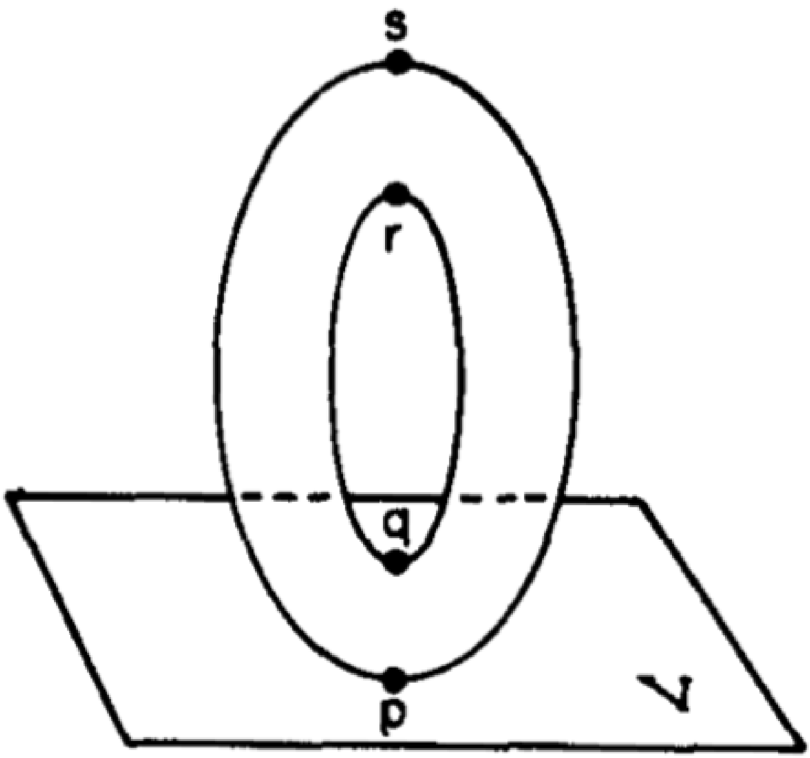


Figure 2.1: A torus tangent to a plane (from [8])

We would like to describe the change of the set M^a as a passes through the points $f(p), f(q), f(r), and f(s)$. For that we define yet another new concept.

Definition 2.4 *A continuous map $f : X \to Y$ between two topological spaces is called a* homotopy equivalence *if there exists a continuous map $f : Y \to X$ such that $g \circ f \simeq Id_X$ and $f \circ g \simeq Id_Y$, where, as in Chapter 1, "$\simeq$" denotes the relation of being homotopic. If there exists a homotopy equivalence between two spaces X and Y, we say that those two spaces have* the same homotopy type.

Remark 2.5 • *It is clear that homeomorphic spaces have the same homotopy type, however, the converse is in general not true.*

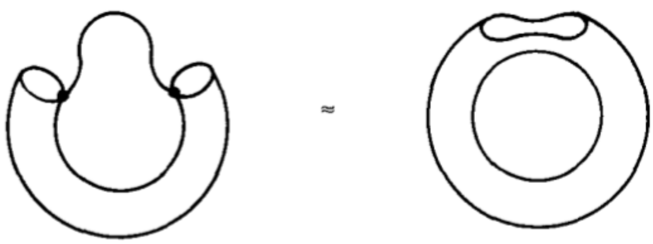
- *If a topological space has the same homotopy type as a point, it is often called* contractible.

Let us now examine how the homotopy type of M^a changes. (All the illustrations below were taken from [8]).

- $(1) \to (2)$ *is the operation of attaching a 0-cell, since a 2-cell is contractible and thus has the same homotopy type as a 0-cell, which is just a point.*

- $(2) \to (3)$ *is the operation of attaching a 1-cell.*

- $(3) \to (4)$ *again consists in attaching a 1-cell.*



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

Furthermore, we can observe that the points p, q, r, s at which the homotopy type of M^a changes are the critical points of f. At p, we can choose coordinates (x, y) such that $f = x^2 + y^2$, at s such that $f =$ constant $- x^2 - y^2$, and at q and r

such that $f = constant + x^2 - y^2$. An important thing to note at this point is that the number of minus signs in the expression of f (which we already defined as the index of f) is the dimension of the cell we need to attach to go from one homotopy type of M^a to the next. In the rest of this chapter, we will generalize this idea for differentiable functions on manifolds.

2.2 Homotopy type in terms of critical values

For the following section for a smooth map f on a manifold M, we denote by M^a the preimage set $f^{-1}(-\infty, a] = \{p \in M : f(p) \leq a\}$.

We begin with a topological definition.

Definition 2.6 *Let X be a topological space and let $X' \subseteq X$ be a subspace. Let $\iota : X' \to X$ denote the inclusion. We call X' a* deformation retract *of X if there exists a smooth map $r : X \to X'$ such that $r \circ \iota = id_{X'}$ and $\iota \circ r \simeq id_X$.*

The proof of the next theorem uses tools from differential geometry that are not the main focus of this text but will be briefly described here. The concept we want to define here is that of a Riemannian metric, which allows us to define lengths and angles on smooth manifolds. Every smooth manifolds admits a Riemannian metric and we will outline a proof of this without getting into the details too much. The main ingredient is a partition of unity, a concept that may be familiar from analysis. For this description of manifolds, we use the more general setup of appendix A and not just the definition we had of manifolds of Euclidean space.

Definition 2.7 *A* Riemannian metric *g on a smooth manifold M is a smoothly chosen inner product $g_x : T_xM \times T_xM \to \mathbb{R}$ on every tangent space T_xM of M. More precisely, $\forall x \in M, g_x$ has the following properties:*

1. $g_x(u, v) = g_x(v, u), \forall u, v \in T_xM$
2. $g_x(u, u) \geq 0, \forall u \in T_xM$
3. $g(u, u) = 0 \iff u = 0$

Moreover, g is smooth in the sense that for any smooth vector fields X and Y, the map $x \mapsto g_x(X_x, Y_x)$ is smooth.

Remark 2.8 *Locally, we can describe a metric in terms of its coefficient in a local chart, defined by $g_{ij} = g(\partial_i, \partial_j)$. The smoothness of g is equivalent to the smoothness of all the coefficients g_{ij}.*

Example 2.9 $\mathbb{R}^n$ *can be given a Riemannian metric in many ways. For example, let f_{ij} be bounded smooth functions such that $f_{ij} = fji$. Then, for C big enough, the functions $g_{ij} = C\delta_{ij} + f_{ij}$ are positive definite everywhere, and so define a Riemannian metric. Here δ_{ij} denotes the Kronecker delta function.*

Proposition 2.10 *Every smooth manifold M admits a Riemannian metric.*

Proof (sketch) Let us pick an atlas $\{\varphi_\alpha : U_\alpha \to V_\alpha\}$ and a subordinate partition of unity $\{\rho_\alpha\}$. On each of the sets $V_\alpha \subset \mathbb{R}^n$ take a Riemannian metric $g^{(\alpha)}$ as in example 2.9. We then define

$$g(u,v) = \sum_\alpha \rho_\alpha g^{(\alpha)}(D\varphi_\alpha(u), D\varphi_\alpha(v))$$

We can then verify that all the desired properties hold. □

Theorem 2.11 *Let $f : M \to \mathbb{R}$ be a smooth function on a manifold M. Let $a < b$ and suppose that $f^{-1}[a,b]$ is compact and contains no critical points of f. Then M^a is diffeomorphic to M^b. Moreover, M^a is a deformation retract of M^b, so that the inclusion map $\iota : M^a \to M^b$ is a homotopy equivalence.*

Proof The idea of the proof is to push M^b onto M^a along the orthogonal trajectories of the hypersurfaces $f = constant$. To do this, choose a Riemannian metric on M, and let $\langle X, Y\rangle$ denote the inner product of two tangent vectors determined by this choice of metric. The gradient of f is the vector field ∇f on M, which fulfills the identity

$$\langle X, \nabla f\rangle = X(f)$$

for any vector field X, and where $X(f)$ denotes the directional derivative of f along X. Now, ∇f vanishes precisely at the critical points of f. If $c : \mathbb{R} \to M$ is a curve with velocity vector $\frac{\mathrm{d}c}{\mathrm{d}t}$, note the identity

$$\left\langle \frac{\mathrm{d}c}{\mathrm{d}t}, \nabla f \right\rangle = \frac{\mathrm{d}(f \circ c)}{\mathrm{d}t}.$$

Define the map $\rho : M \to \mathbb{R}$ to be smooth and equal to $1/\langle \nabla f, \nabla f\rangle$ on the compact set $f^{-1}[a,b]$ and vanishing outside a compact neighbourhood of this set. Then, we may define a vector field X satisfying the condition of lemma 1.64 as follows:

$$X_q = \rho(q)(\nabla f)_q$$

Thus, X generates a 1-parameter group of diffeomorphisms $\varphi_t : M \to M$. For a fixed point q in M, let us now consider the function $t \mapsto f(\varphi_t(q))$. If $\varphi_t(q)$ belongs to $f^{-1}[a,b]$, then we have

$$\frac{\mathrm{d}f(\varphi_t(q))}{\mathrm{d}t} = \left\langle \frac{\mathrm{d}\varphi_t(q)}{\mathrm{d}t}, \nabla f \right\rangle = \langle X, \nabla f\rangle = 1$$

Hence, the correspondence $t \mapsto \varphi_t(q)$ is linear with derivative 1 as long as $f(\varphi_t(q))$ lies between a and b. Now, the diffeomorphism $\varphi_{b-a} : M \to M$

clearly carries M^a diffeomorphically onto M^b, which proves the first statement of the theorem. Now let us define a 1-parameter family of maps $r_t : M^b \to M^b$ as follows:

$$r_t(q) = \begin{cases} q & \text{if } f(q) \leq a \\ \varphi_{t(a-f(q))}(q) & \text{if } a \leq f(q) \leq b \end{cases}$$

Then r_0 is the identity, and r_1 is a retraction from M^b to M^a. Hence M^a is a deformation retract of M^b. This completes the proof. $\square$

Remark 2.12 *The compactness of $f^{-1}[a,b]$ is a necessary condition.*

We have seen that the homotopy type does not change between two preimage sets if there are no critical points involved. We now formalize what we have seen in our example 2.3 with the torus; namely that when we have a critical point, "passing through it" with the level set changes the homotopy by attaching an n-cell. Here is the precise statement:

Theorem 2.13 *Let $f : M \to \mathbb{R}$ be a smooth function on a manifold M and let $p \in M$ be a non-degenerate critical point of f with index λ. Write $c = f(p)$. Assume that for some $\varepsilon < 0$, the set $f^{-1}[c-\varepsilon, c+\varepsilon]$ is compact and contains no critical point of f except for p. Then the set $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon}$ with a λ-cell attached*

Proof The proof of this is quite lengthy and necessitates a few intermediate steps. To make the process as clear as possible, we first describe the idea of the proof and then get into the details. All the figures in this proof come from [8].

We will define a new function $F : M \to \mathbb{R}$ which coincides with f except that we demand $F < f$ in a small neighbourhood of p. Thus, the preimage set $F^{-1}(-\infty, c-\varepsilon]$ will be equal to $M^{c-\varepsilon}$ together with a small region H near the point p. Then, a rather direct argument will show that if we take the cell e^λ in a suitable way, $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c-\varepsilon} \cup H$. Finally, using the previous theorem on F and $F^{-1}[c-\varepsilon, c+\varepsilon]$ we will show that $M^{c-\varepsilon} \cup H$ is a deformation retract of $M^{c+\varepsilon}$, which will conclude the proof.

The following figure (2.2) illustrates this idea. H is the horizontally shaded region.

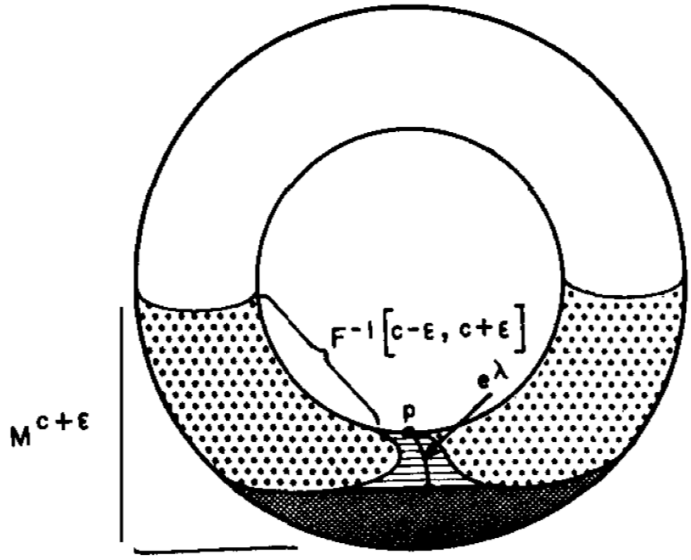


Figure 2.2: idea of the proof

Choose a coordinate system $u_1, \ldots, u_n$ in a neighbourhood U of p such that

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

holds in U. This is possible by the Morse Lemma 1.53. Thus, the critical point p will have coordinates $u_1(p) = \cdots = u_n(p) = 0$. Take $\varepsilon > 0$ sufficiently small so that

1. The set $f^{-1}[c - \varepsilon, c + \varepsilon]$ is compact and contains no critical point other than p.

2. The image of the neighbourhood U under the diffeomorphic embedding $(u_1, \ldots, u_n) \to \mathbb{R}^n$ contains the ball $(u_1, \ldots, u_n) : \sum_{i=1}^{n} (u_i)^2 \leq 2\varepsilon$.

Now, we define the λ-cell that we will need as follows:

$$e^\lambda = (u_1)^2 + \cdots + (u_\lambda)^2 \leq \varepsilon \text{ and } u_{\lambda+1} = \cdots = u_n = 0.$$

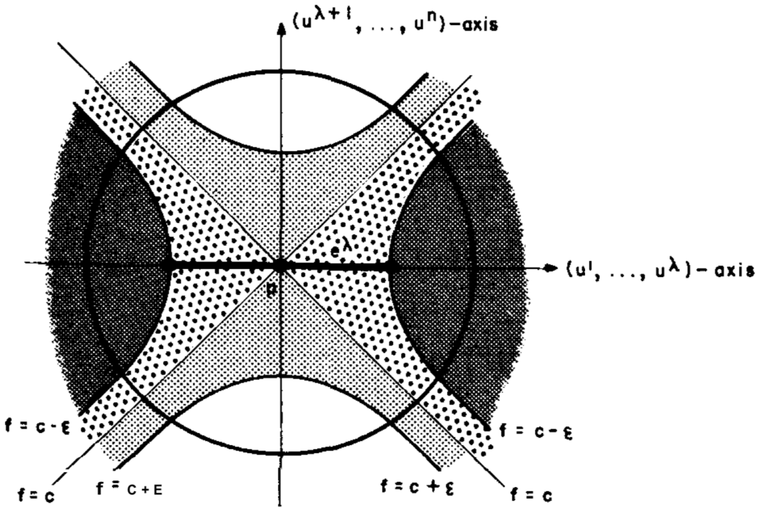


Figure 2.3: The cell e^λ

The figure (2.3) above illustrates the situation. The axes represent the hyperplanes $u^{\lambda+1} = \cdots = u^n = 0$ and $u^1 = \ldots u^\lambda = 0$ respectively. The circle represents the boundary of the ball of radius $\sqrt{2\varepsilon}$ and the hyperbolas are the hypersurfaces $f^{-1}(c-\varepsilon)$ and $f^{-1}(c+\varepsilon)$. The region $M^{c-\varepsilon}$ is shaded, $f^{-1}[c-\varepsilon, c]$ is filled with big dots and $f^{-1}[c, c+\varepsilon]$ is the region with the small dots.

Looking at the intersection $e^\lambda \cap M^{c-\varepsilon}$, we see that it is precisely the boundary of e^λ, denoted by $\dot{e}^\lambda$. Thus, e^λ is attached to $M^{c-\varepsilon}$, as we wanted.

We now want to prove that $M^{c-\varepsilon} \cup e^\lambda$ is a deformation retract of $M^{c+\varepsilon}$. In order to do this, we construct a new smooth function $F : M \to \mathbb{R}$ as follows. First, let $\mu : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying the following three conditions:

1. $\mu(0) > \varepsilon$
2. $\mu(r) = 0$ for $r \geq 2\varepsilon$
3. $-1 < \mu'(r) \leq 0$ for all r,

where $\mu'(r)$ denotes the derivative of μ. Now let F coincide with f outside of the coordinate neighbourhood U and let us define F inside U as

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

It is easy to see that F is smooth and well-defined on the whole of M. At this point, it is convenient to introduce the following notations. We define $\xi, \eta : U \to [0, \infty)$ by

$$\xi = (u_1)^2 + \cdots + (u_\lambda)^2 \text{ and } \eta = (u_{\lambda+1})^2 + \cdots + (u_n)^2$$

Then we have $f = c - \xi + \eta$ and $F(q) = c - \xi(q) + \eta(q) - \mu(\xi(q) + 2\eta(q))$, for all q in U.

Claim 1 $F^{-1}(-\infty, c+\varepsilon] = M^{c+\varepsilon}$

Proof Outside of the ellipsoid given by $\xi + 2\eta \leq 2\varepsilon$, the functions f and F agree. Inside the ellipsoid, we have

$$F \leq f = c - \xi + \eta \leq c + \frac{1}{2}\xi + \eta \leq c + \varepsilon$$

which is what we wanted. □

Claim 2 *The critical points of F are the same as those of f.*

Proof Note that we have:

$$\frac{\partial F}{\partial \xi} = -1 - \mu'(\xi + 2\eta) < 0 \quad \text{and} \quad \frac{\partial F}{\partial \eta} = 1 - 2\mu'(\xi + 2\eta) \geq 1.$$

Since $\mathrm{d}F = \frac{\partial F}{\partial \xi}\mathrm{d}\xi + \frac{\partial F}{\partial \eta}\mathrm{d}\eta$, where $\mathrm{d}\xi$ and $\mathrm{d}\eta$ are simultaneously 0 only at the point p, it follows that F has no other critical point than p. □

Claim 3 *The region $F^{-1}(-\infty, c-\varepsilon]$ is a deformation retract of $M^{c+\varepsilon}$.*

Proof Consider the set $F^{-1}[c-\varepsilon, c+\varepsilon]$. By the first claim, together with the inequality $F \leq f$, we have that $F^{-1}[c-\varepsilon, c+\varepsilon] \subseteq f^{-1}[c-\varepsilon, c+\varepsilon]$, therefore $F^{-1}[c-\varepsilon, c+\varepsilon]$ is compact. The only possible critical point contained in this region is p, but we have

$$F(p) = c - \mu(0) < c - \varepsilon,$$

which means that $F^{-1}[c-\varepsilon, c+\varepsilon]$ has no critical points. Using theorem 2.11 now yields the assertion. □

It will be convenient to write $F^{-1}(-\infty, c-\varepsilon]$ as $M^{c-\varepsilon} \cup H$, where H denotes the closure of $F^{-1}(-\infty, c-\varepsilon] \setminus M^{c-\varepsilon}$.

We now return to our cell e^λ, which we defined as containing all the points q where $\xi(q) \leq \varepsilon$ and $\eta(q) = 0$. Since we have that $\frac{\partial F}{\partial \xi} < 0$, we obtain, for $q \in e^\lambda$,

$$F(q) \leq F(p) < c - \varepsilon.$$

At the same time however, any point q in e^λ fulfils $f(q) \geq c - \varepsilon$. This means that the cell e^λ is contained in the set H.

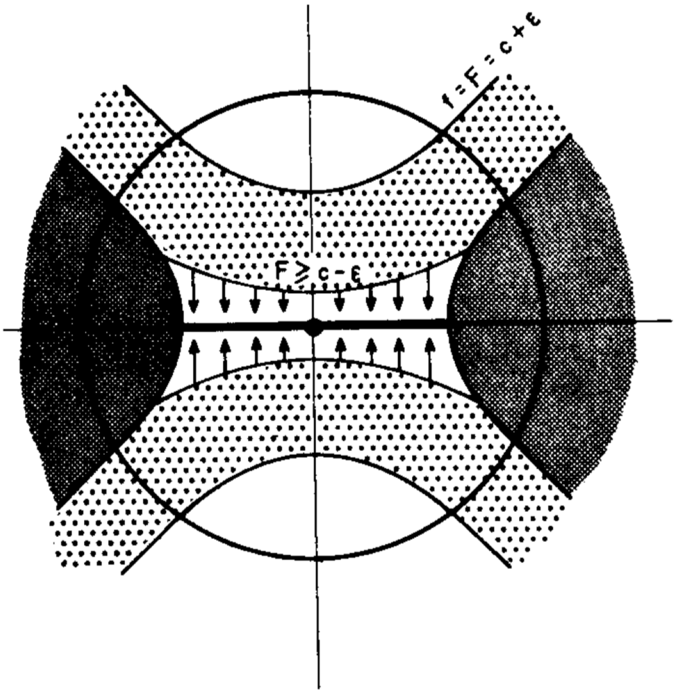


Figure 2.4: The region H

In Figure (2.4), $M^{c-\varepsilon}$ is the darkened region, H contains the vertical arrows and $F^{-1}[c-\varepsilon, c+\varepsilon]$ is the dotted area. The arrows inside H represent a deformation retract r_t that is described in the claim below.

Claim 4 *$M^{c-\varepsilon} \cup e^{\lambda}$ is a deformation retract of $M^{c-\varepsilon} \cup H$.*

Proof We want to construct a deformation retraction $r_t : M^{c-\varepsilon} \cup H \to M^{c-\varepsilon} \cup e^{\lambda}$. We define this map r_t to be the identity outside of the coordinate neighbourhood U. Inside of the region U, it is necessary to distinguish three cases.

Case 1 In the region where $\xi \leq \varepsilon$ let r_t be the following map

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

With this definition, r_1 is just the identity, and r_0 maps the whole region into the cell e^{λ}. Since we have that $\frac{\partial F}{\partial \eta} > 0$, each r_t will map $F^{-1}(-\infty, c-\varepsilon]$ into itself.

Case 2 Inside the region $\varepsilon \leq \xi \leq \eta + \varepsilon$, we define r_t to be

$$(u_1, \ldots, u_n) \mapsto (u_1, \ldots, u_\lambda, s_t u_{\lambda+1}, \ldots, s_t u_n)$$

where $s_t \in [0,1]$ is defined as follows

$$s_t = t + (1-t)((\xi - \varepsilon)/\eta)^{1/2}.$$

Then r_1 is again the identity, and r_0 maps the entire region into the hypersurface $f^{-1}(c-\varepsilon)$. One can verify that the functions $s_t u_i$ are still continuous as $\xi \to \varepsilon$ and $\eta \to 0$. This definition coincides with that of case 1 in the limit case when $\xi = \varepsilon$.

Case 3 For the region $\eta + \varepsilon \leq \xi$ (which is exactly $M^{c-\varepsilon}$), we simply define r_t to be the identity, which again coincides with case 2 for $\xi = \eta + \varepsilon$.

This proves Claim 4. □

With Claim 3 and 4, the proof of theorem 2.13 is complete. □

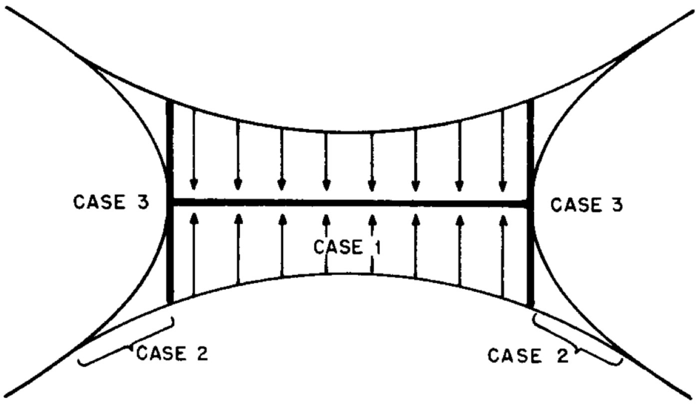


Figure 2.5: The deformation retract

Remark 2.14 *We can easily generalize the proof of the previous theorem to the following result. Suppose that there are k non-degenerate critical points* $p_1, \ldots, p_k$ *with indices* $\lambda_1, \ldots, \lambda_k$ *in* f^{-1}. *Then* $M^{c+\varepsilon}$ *has the same homotopy type as* $M^{c-\varepsilon} \cup e^{\lambda_1} \cup \cdots \cup e^{\lambda_k}$.

Remark 2.15 *A similar argument also allows us to show that* M^c *is a deformation retract of* $F^{-1}(-\infty, c]$, *which is a deformation retract of* $M^{c+\varepsilon}$. *Thus,* $M^{c-\varepsilon} \cup e^{\lambda}$ *is a deformation retract of* M^c.

We now define a new type of topological spaces, and later see that their homotopy type is the same as a lot of other topological spaces, which makes them really useful.

Definition 2.16 *Let* $X' \subseteq X$ *be topological spaces such that* X' *is closed in* X. *A* cellular decomposition *of the pair* (X, X') *is a sequence of subspaces*

$$X' = X^{-1} \subseteq X^0 \subseteq X^1 \subseteq \cdots \subseteq X$$

such that

1. *X carries the colimit topology, i.e a set $C \subseteq X$ is closed if and only if $C \cap X^n$ is closed for every $n \geq -1$*
2. *For each $n \geq 0$, X^n is obtained from X^{n-1} by attaching n-cells. (If $X' = \emptyset$ and $n = 0$, we mean that X^0 is a discrete set of points.)*

We call the pair (X', X) a relative cell complex. *If $X' = \emptyset$, we say that X is a cell complex, and we write just X instead of $(X, \emptyset)$. The topological space X^n is called the* n-skeleton *of (X, X') and the decomposition (X^n), $n \geq -1$ is the* skeleton filtration *of (X, X').*

Remark 2.17 *Cell complexes are also often called CW-complexes, where the C stands for "closure finite", and the W for "weak topology".*

Before stating our main result about cell complexes, we prove two topological lemmas about spaces with cells attached.

Lemma 2.18 (Whitehead) *Let φ_0 and φ_1 be homotopic maps from the sphere $\dot{e}^\lambda$ to the topological space X. Then the identity map of X extends to a homotopy equivalence*

$$k : X \cup_{\varphi_0} e^\lambda \to X \cup_{\varphi_1} e^\lambda$$

Proof We define k by the formulae

$$\begin{cases} k(x) = x & \text{for } x \in X \\ k(tu) = 2tu & \text{for } 0 \leq t \leq \frac{1}{2}, \quad u \in \dot{e}^\lambda \\ k(tu) = \varphi_{2-2tu}(u) & \text{for } \frac{1}{2} \leq t \leq 1, \quad u \in \dot{e}^\lambda \end{cases}$$

where φ_t is the homotopy between φ_0 and φ_1. We can define a corresponding map

$$l : X \cup_{\varphi_1} e^\lambda \to X \cup_{\varphi_0} e^\lambda$$

by similar formulae. It is then just a matter of checking that the compositions lk and kl are homotopic to the respective identity maps. □

Lemma 2.19 *Let $\varphi : \dot{e}^\lambda \to X$ be an attaching map. Any homotopy equivalence $f : X \to Y$ extends to a homotopy equivalence*

$$F : X \cup_\varphi e^\lambda \to Y \cup_{f\varphi} e^\lambda$$

Proof We define F as follows

$$\begin{cases} F|_X = f \\ F|_{e^\lambda} = \text{identity} \end{cases}$$

Let $g : Y \to X$ be a homotopy inverse to f and define $G : Y \cup_{f\varphi} e^\lambda \to X \cup_{gf\varphi} e^\lambda$ by the corresponding conditions $G|_Y = g$ and $G|_{e^\lambda} =$ identity.

Since $gf\varphi$ is homotopic to φ, it follows from lemma 2.18 that there is a homotopy equivalence

$$k : X \cup_{gf\varphi} e^\lambda \to X \cup_\varphi e^\lambda$$

We will first show that $kGF : x \cup_\varphi e^\lambda \to X \cup_\varphi e^\lambda$ is homotopic to the identity.

Let h_t be a homotopy between gf and the identity. Using the specific definition we had for k, G, and F, we have:

$$\begin{cases} kGF(x) = gf(x) & \text{for } x \in X \\ kGF(tu) = 2tu & \text{for } 0 \le t \le \frac{1}{2}, \quad u \in \dot{e}^\lambda, \\ kGF(tu) = h_{2-2t}\varphi(u) & \text{for } \frac{1}{2} \le t \le 1, \quad u \in \dot{e}^\lambda. \end{cases}$$

The required homotopy $q_\tau : X \cup_\varphi e^\lambda \to X \cup_\varphi e^\lambda$ is now given by

$$\begin{cases} q_\tau(x) = h_\tau(x) & \text{for } x \in X, \\ q_\tau(tu) = \frac{2}{1+\tau} tu & \text{for } 0 \le t \le \frac{1+\tau}{2}, \quad u \in \dot{e}^\lambda, \\ q_\tau(tu) = h_{(2-2t+\tau)}\varphi(u) & \text{for } \frac{1+\tau}{2} \le t \le 1, \quad u \in \dot{e}^\lambda. \end{cases}$$

Thus, F has a left homotopy inverse. The proof that F is a homotopy equivalence will now be purely formal, and will use the following

Claim 5 *If a map F has a left homotopy inverse L and a right homotopy inverse R, then F is a homotopy equivalence; and R (or L) is a two-sided homotopy inverse.*

Proof Let id denote the identity map. The relations $LF \simeq id$ and $FR \simeq id$ imply that

$$L \simeq L(FR) = (LF)R \simeq R.$$

This yields $RF \simeq LF \simeq id$, which proves that R is a two-sided inverse. □

We now conclude the proof of the lemma as follows. The relation $kGF \simeq id$ asserts that F has a left homotopy inverse, and we can conduct a similar proof for the map G.

1. Since $k(GF) \simeq id$, and k is known to have a left inverse, it follows that $(GF)k \simeq id$.

2. Since $G(Fk) \simeq id$ and G is known to have a left inverse, we get $(Fk)G \simeq id$.

3. Since $F(kG) \simeq id$ and kG is also the left inverse of F, F is a homotopy equivalence, exactly as we wanted. □

The proof of our next theorem requires an important result of algebraic topology called the cellular approximation theorem that we will only state here. The proof can be found in Hatcher's Algebraic Topology, from which the following definition and statement are taken.

Definition 2.20 *Let $f : X \to Y$ be a map between two cell complexes. We say that f is a* cellular map *if for all $n \geq -1$, $f(X^n) \subset Y^n$.*

Theorem 2.21 (cellular approximation theorem) *Every map $f : X \to Y$ between cell complexes is homotopic to a cellular map.*

We can now state another important result about the homotopy type of a topological space with respect to its critical points.

Theorem 2.22 *If f is a differentiable function on a manifold M with no degenerate critical points, and if each set M^a is compact, then M has the homotopy type of a cell-complex, with one cell of dimension λ for each critical point of index λ*

Proof Let $c_1 < c_2 < c_3 < \dots$ be the critical values of $f : M \to \mathbb{R}$. The sequence c_i has no cluster point since each M^a is compact. For $a < c_1$, we have $M^a = \emptyset$. Assume that $a \neq c_1, c_2, \dots$ and that M^a has the same homotopy type has a cell-complex. Let c be the smallest c_i such that $a < c_i$. By theorems 2.11 and 2.13, and remark 2.14, $M^{c+\varepsilon}$ has the homotopy type of $M^{c-\varepsilon} \cup_{\varphi_1} e^{\lambda_1} \cup \dots \cup e^{\lambda_{j(c)}}$ for some maps $\varphi_1, \dots, \varphi_{j(c)}$, for ε small enough. Furthermore, there is a homotopy equivalence $h : M^{c-\varepsilon} \to M^a$. By assumption, we also have a homotopy equivalence $h' : M^a \to K$, where K is a cell-complex. Then, by cellular approximation (theorem 2.21), each map $h' \circ h \circ \varphi_j$ is homotopic to a map

$$\psi_j : \dot{e}^\lambda \to (\lambda_j - 1) - \text{skeleton of K}.$$

Then $K \cup_{\psi_1} \cup e^{\lambda_1} \cup \dots \cup_{\psi_{j(c)}} e^{\lambda_{j(c)}}$ is a cell-complex, and has the same homotopy type as $M^{c+\varepsilon}$ by the lemmas 2.18 and 2.19.

By induction it follows that each of the M^a has the same homotopy type of a cell-complex. If M is compact, this completes the proof. If M is not compact, but all critical points lie in one of the compact sets M^a, a proof similar to the one of theorem 2.11 shows that M^a is a deformation retract of M, so the proof is complete in that case as well.

If there are infinitely many critical points, then the construction above yields the following infinite sequence of homotopy equivalences:

$$\begin{array}{ccccccc} M^{a_1} & \subset & M^{a_2} & \subset & M^{a_3} & \subset & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ K_1 & \subset & K_2 & \subset & K_3 & \subset & \dots \end{array}$$

Each of these homotopy equivalences extends the previous one. Let K denote the union of the K_i in the direct limit topology, i.e. the finest possible compatible topology, and let $g : M \to K$ be the limit map. Then g induces isomorphisms of homotopy groups in all dimensions. We only need to apply theorem 1 of "Combinatorial homotopy I" [10] by Whitehead to conclude that g is a homotopy equivalence. □

Remark 2.23 *We have also proved that each* M^a *has the homotopy type of a finite cell-complex, with one cell of dimension* λ *for each critical point of index* λ *in* M^a*. This is true even if a is a critical value.*

Chapter 3

Examples and applications

3.1 A few application of Morse theory

In this section, we use the machinery of Morse theory to obtain interesting topological facts. The following section uses concepts of algebraic topology that we unfortunately do not have time to explain in this text. If the reader is however familiar with homology, this chapter should be accessible.

We first prove a nice application of the big theorems of the previous section. In example 1.58, we saw that the height function on a sphere has exactly two critical points and that they are non-degenerate. Here is a much more general statement.

Theorem 3.1 (Reeb) *If M is a compact manifold and f is a differentiable function on M with only two critical points, both non-degenerate, then M is homeomorphic to a sphere.*

Proof The two critical points must be the minimum and the maximum points, since M is compact. Without loss of generality, let us say that $f(p) = 0$ is the minimum, and $f(q) = 1$ is the maximum. If ε is small enough, using the Morse lemma 1.53, the sets M^{ε} and $f^{-1}[1-\varepsilon, 1]$ are closed n-cells. But M^{ε} is homeomorphic to $M^{1-\varepsilon}$ by theorem 2.11. Thus M is the union of two closed $n - cells$, $M^{1-\varepsilon}$ and $f^{-1}[1-\varepsilon, 1]$, matched along their boundary. It is now easy to construct a homeomorphism between M and S^n. □

Remark 3.2 *The theorem still holds with degenerate critical points, but the proof is not as easy.*

Let us illustrate the use of the machinery we have developed with another example.

Definition 3.3 *The* complex projective space $\mathbb{C}P^n$ *is the space of lines in* $\mathbb{C}^{n+1}$ *that go through the origin. We can see it as a space of tuples* $(z_0, \ldots, z_n) \in \mathbb{C}^{n+1}$

such that $(z_0, \dots, z_n)$ *is identified with* $(\lambda z_0, \dots, \lambda z_n)$ *for* $\lambda \in \mathbb{C}$. *Since the factor* λ *is arbitrary, we may choose* $\sum_{j=0}^{n} |z_j|^2 = 1$ *for the representative of each equivalence class. We denote the equivalence class of* $(z_0, \dots, z_n)$ *by* $(z_0 : \dots : z_n)$.

We define a real valued function f on $\mathbb{C}P^n$ by $f(z_0 : \dots : z_n) = \sum_{j=0}^{n} c_j |z_j|^2$ where the c_j's are distinct real constants. We consider the following local coordinate system. Let U_0 be the set of $(z_0 : \dots : z_n)$ such that $z_0 \neq 0$, and set

$$|z_0| \frac{z_j}{z_0} = x_j + i y_j$$

The functions $x_1, y_1, \dots, x_n, y_n : U_0 \to \mathbb{R}$ are then the required coordinate functions, as they map U_0 diffeomorphically onto the open unit ball in $\mathbb{R}^{2n}$. Additionally, we have that $|z_j|^2 = x_j^2 + y_j^2$ for $1 \leq j \leq n$ and $|z_0|^2 = 1 - \sum_{j=1}^{n}(x_j^2 + y_j^2)$. Given these equations, we may write the function f in the neighbourhood U_0 as follows:

$$f = c_0 + \sum_{j=1}^{n} (c_j - c_0)(x_j^2 + y_j^2)$$

The only critical point in U_0 is thus the center point of the coordinate system, i.e. $p_0 = (1 : 0 : 0 : \dots : 0)$. The function f is non-degenerate at this point, and its index is twice the number of j with $c_j < c_0$. Reasoning in the exact same way, we can use coordinate systems centered around the points $p_1 = (0 : 1 : 0 : \dots : 0)$, $p_2 = (0 : 0 : 1 : 0 : \dots : 0), \dots, p_n = (0 : \dots : 0 : 1)$. Thus, the only critical points (all non-degenerate) of f are exactly the points $p_1, \dots, p_n$. The index of f at the point p_k is twice the number of j with $c_j < c_k$. Since all the c_j's are distinct, every possible even index between 0 and $2n$ occurs exactly once. We may now use theorem 2.22 to conclude that $\mathbb{C}P^n$ has the homotopy type of a cell-complex of the form

$$e^0 \cup e^2 \cup e^4 \cup \dots \cup e^{2n}.$$

We thus have been able the characterize the homotopy type of all the complex projective spaces with the use of Morse theory.

Remark 3.4 *The reader familiar with algebraic topology will see that the above computation allows us the calculate the homology groups of* $\mathbb{C}P^n$. *Indeed, using cellular homology, we have:*

$$H_i(\mathbb{C}P^n; \mathbb{Z}) = \begin{cases} \mathbb{Z} & \text{for } i = 0, 2, 4, \dots, 2n \\ 0 & \text{for other values of } i \end{cases}$$

3.2 Morse inequalities

When Morse treated this topic, the theorem 2.22 had not been proven yet. To go around this, Morse described the connections between the topology of a manifold M and the critical points of function f on M via a collection of inequalities. We will now describe Morse's original method. In the following pages, we will, as indicated before, use some concepts from algebraic topology such as homology groups that we will not be able to define in every details to avoid an important detour.

Definition 3.5 *Let S be a function from certain pairs of spaces to the integers. We call S* subadditive *if whenever $X \supset Y \supset Z$, we have $S(X,Z) \leq S(X,Y) + S(Y,Z)$. If we have equality, we call S* additive.

We now give an example of such a function with the following definitions.

Definition 3.6 *Let (X,Y) be a pair of topological spaces (as defined in the setting of homology groups) and let F be any field serving as coefficient group. We define $R_\lambda(X,Y)$ to be the rank over F of $H_\lambda(X,Y;F)$ so long as this rank is finite. We call this number the λ-th* Betti number *of (X,Y).*

R_λ is a subadditive function, which we can see from the long exact sequence for a triple of spaces (X,Y,Z):

$$\ldots \longrightarrow H_\lambda(Y,Z) \longrightarrow H_\lambda(X,Z) \longrightarrow H_\lambda(X,Y) \longrightarrow \ldots$$

An example of additive function is the so-called *Euler characteristic* $\chi(X,Y)$ defined as follow:

$$\chi(X,Y) = \sum(-1)^\lambda R_\lambda(X,Y)$$

Lemma 3.7 *Let S be subadditive and let $X_0 \subset \cdots \subset X_n$. Then $S(X_n,X_0) \leq \sum_{i=1}^n S(X_i,X_{i-1})$. If S is additive then equality holds.*

Proof We prove the claim by induction on n. For $n = 1$, the inequality is trivial and for $n = 2$, this is the definition of subadditivity (respectively of additivity). Now let's assume that the claim holds for $n-1$, then $S(X_{n-1},X_0) \leq \sum_{i=1}^{n-1} S(X_i,X_{i-1})$. Therefore, we have

$$S(X_n,X_0) \leq S(X_n,X_{n-1}) + S(X_{n-1},X_0) \leq \sum_{i=1}^{n} S(X_i,X_{i-1}) \tag{3.1}$$

and the result holds for n. This completes the proof. □

Remark 3.8 *Let us write $S(X,\emptyset) = S(X)$. Then, taking $X_0 = \emptyset$ in Lemma 3.7 yields*

$$S(X_n) \leq \sum_{i=1}^{n} S(X_i,X_{i-1}) \tag{3.2}$$

with equality if S is an additive function.

We can now make a connection between the Betti number of a compact manifold M and the number of critical points of a function f on M with the following result.

Theorem 3.9 (Weak Morse inequalities) *If C_λ denotes the number of critical points of index λ on the compact manifold M, then*

1. $R_\lambda(M) \leq C_\lambda$ *, and*
2. $\sum(-1)^\lambda R_\lambda(M) = \sum(-1)^\lambda C_\lambda$

Proof Let M be a compact manifold and f a differentiable function on M with isolated, non-degenerate critical points. Let $a_1 < \cdots < a_k$ be such that M^{a_i} contains exactly i critical points, and $M^{a_k} = M$. Then, if we denote by λ_i the index of the i-th critical point, we have:

$$H_*(M^{a_i}, M^{a_{i-1}}) = H_*(M^{a_{i-1}} \cup e^{\lambda_i}, M^{a_{i-1}}) = H_*(e^{\lambda_i}, \dot{e}^{\lambda_i}) = \begin{cases} \text{coefficient group if } * = \lambda_i \\ 0 \text{ otherwise} \end{cases}$$

where the second equality is obtained by excision. We can then apply the inequality 3.2 to $\emptyset = M^{a_0} \subset \cdots \subset M^{a_k} = M$ with $S = R_\lambda$ to get

$$R_\lambda(M) \leq \sum_{i=1}^{k} R_\lambda(M^{a_i}, M^{a_{i-1}}) = C_\lambda$$

where C_λ denotes the number of critical points of index λ. Applying the same formula to $S = \chi$, we get

$$\chi(M) = \sum_{i=1}^{k} \chi(M^{a_i}, M^{a_{i-1}}) = C_0 - C_1 + \cdots \pm C_n \qquad \square$$

We can get sharper inequalities than what we have just demonstrated. To do this, we need the following lemma:

Lemma 3.10 *The function S_λ is subadditive, where*

$$S_\lambda(X,Y) = R_\lambda(X,Y) - R_{\lambda-1}(X,Y) + R_{\lambda-2}(X,Y) - \cdots \pm R_0(X,Y)$$

Proof Given an exact sequence

$$\ldots \xrightarrow{h} A \xrightarrow{i} B \xrightarrow{j} C \xrightarrow{k} \ldots \longrightarrow D \longrightarrow 0$$

of vector spaces note that the rank of the homomorphism h plus the rank of i is equal to the rank of A. Iterating this, we get:

$$\begin{aligned} rank\ h &= rank\ A - rank\ i \\ &= rank\ A - rank\ B + rank\ j \\ &= rank\ A - rank\ B + rank\ C - rank\ k \\ &= \ldots \\ &= rank\ A - rank\ B + rank\ C - \cdots \pm rank\ D \end{aligned}$$

Hence, the right-hand side of the last equality is non-negative. Now, consider the long exact sequence of a triple $X \supset Y \supset Z$. We use the above computation on the connecting homomorphism

$$H_{\lambda+1}(X,Y) \xrightarrow{\partial} H_\lambda(Y,Z).$$

This yields

$$rank\,\partial = R_\lambda(Y,Z) - R_\lambda(X,Z) + R_\lambda(X,Y) - R_{\lambda-1}(Y,Z) + \cdots \geq 0$$

Collecting the terms appropriately, we can rewrite this as

$$S_\lambda(Y,Z) - S_\lambda(X,Z) + S_\lambda(X,Y) \geq 0$$

which is what we wanted to prove. □

We can now use the subadditive function S_λ on the spaces $\emptyset \subset M^{a_1} \subset M^{a_2} \subset \cdots \subset M^{a_k}$ to get the Morse inequalities:

$$S_\lambda(M) \leq \sum_{i=1}^{k} S_\lambda(M^{a_i}, M^{a_{i-1}}) = C_\lambda - C_{\lambda-1} + \cdots \pm C_0 \tag{3.3}$$

or equivalently

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots \pm R_0(M) \leq C_\lambda - C_{\lambda-1} + \cdots \pm C_0 \tag{3.4}$$

Remark 3.11 *The weak Morse inequalities from lemma 3.9 are a consequence of equation 3.4. Indeed adding the equations 3.4 for λ and $\lambda - 1$ together yields the first weak inequality, and comparing the equations 3.4 for λ and $\lambda - 1$ with $\lambda > n$ gives us the weak equality.*

We now give an illustration of the use of the Morse inequalities.

Corollary 3.12 *If $C_{\lambda+1} = C_{\lambda-1} = 0$ then $R_\lambda = C_\lambda$ and $R_{\lambda+1} = R_{\lambda-1} = 0$*

Proof Suppose that $C_{\lambda+1} = 0$. We have

$$R_{\lambda+1}(M) - R_\lambda(M) + \cdots \pm R_0(M) \leq C_{\lambda+1} - C_\lambda + \cdots \pm C_0$$

and

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots \pm R_0(M) \leq C_\lambda - C_{\lambda-1} + \cdots \pm C_0.$$

Adding the two together yields $R_{\lambda+1} \leq C_{\lambda+1}$ and thus $R_{\lambda+1} = 0$. Now, comparing equation 3.4 for λ and $\lambda + 1$, we have

$$R_\lambda(M) - R_{\lambda-1}(M) + \cdots \pm R_0(M) \leq C_\lambda - C_{\lambda-1} + \cdots \pm C_0$$

and

$$-R_\lambda(M) + R_{\lambda-1}(M) - \cdots \mp R_0(M) \leq -C_\lambda + C_{\lambda-1} - \cdots \mp C_0.$$

Multiplying the second inequality by -1 and comparing the two gives us an equality:

$$R_\lambda - R_{\lambda-1} + \cdots \pm R_0 = C_\lambda - C_{\lambda-1} + \cdots \pm C_0.$$

If we now suppose that $C_{\lambda-1} = 0$, the exact same argument yields $R_{\lambda-1} = 0$ and

$$R_{\lambda-2} - R_{\lambda-3} + \cdots \pm R_0 = C_{\lambda-2} - C_{\lambda-3} + \cdots \pm C_0.$$

Subtracting this equality from the one above completes the proof. □

Chapter 4

Classification of compact 1-manifolds

In this chapter, we will show two alternative ways of classifying compact 1-manifolds. The first one will use parametrizations by arc-length and will follow what is done in Milnor's "Topology from the differentiable Viewpoint" [7]. The other method will use Morse theory as in Guillemin and Pollack's "Differential Topology" [9].

4.1 Proof via parametrization by arc-length

Here is the theorem that we want to prove, first by using parametrization by arc-length:

Theorem 4.1 *Any smooth, connected 1-dimensional manifold is diffeomorphic either to the circle S^1 or to some interval of real numbers.*

Since any interval is diffeomorphic to either $(0, 1)$, $(0, 1]$, or $[0, 1]$ it follows that there are only four distinct connected 1-manifolds up to diffeomorphism.

For the first proof, we will need the concept of arc-length. In the following, I denotes an interval.

Definition 4.2 *A map $f : I \to M$ is a* parametrization by arc-length *if it maps I diffeomorphically onto any open subset of M, and if the vector $\mathrm{d}f_s(1) \in T_{f(s)}M$ has length 1 for all $s \in I$.*

Remark 4.3 *Any given local parametrization $I' \to M$ can be transformed into a parametrization by arc-length by a change of variables.*

Remark 4.4 *Since a parametrization by arc-length maps I onto an open subset of M, I can only have boundary points if M has boundary points as well.*

Lemma 4.5 *Let $f : I \to M$ and $g : J \to M$ be parametrizations by arc-length. Then $f(I) \cap g(J)$ has at most two components. If it has one component, then f can be extended to a parametrization by arc-length of $f(I) \cup g(J)$. If the intersection has two components, then M is diffeomorphic to S^1.*

Proof The composition $g^{-1} \circ f$ maps some relatively open subset of I diffeomorphically onto a relatively open subset of J because f and g are diffeomorphisms. Moreover, because we are dealing with parametrization by arc-length, the derivative of $g^{-1} \circ f$ must be equal to ± 1 everywhere.

Now, consider the graph $\Gamma \subset I \times J$, consisting of all (s, t) such that $f(s) = g(t)$. Since $g^{-1} \circ f$ is continuous, its graph Γ must be closed. (This is the closed graph theorem in its point-set topology version.) Thus, Γ is closed subset of $I \times J$, and since it is has derivative equal to ± 1, it is made up of line segments with exactly that slope. Now, by combining the fact that Γ is closed and that $g^{-1} \circ f$ is locally a diffeomorphism, we know that these line segments must extend to the boundary of $I \times J$. Bijectivity of $g^{-1} \circ f$ then assures us that there is at most one segment on each of the four edges of $I \times J$. This proves that the graph Γ has at most two components, and if there are two, they must have the same slope.

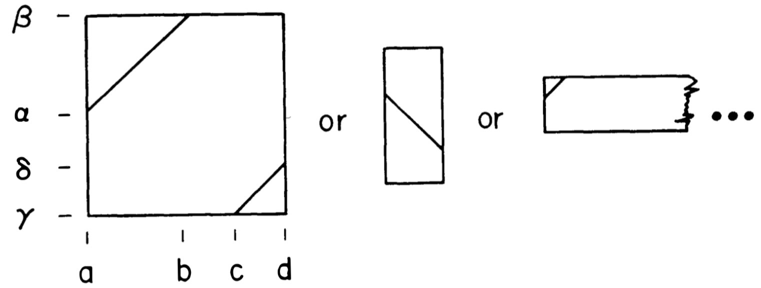


Figure 4.1: Three possible scenarios for the graph Γ (from [7]

If Γ has only one connected component, then the map $g^{-1} \circ f$ extends to a linear map $L : \mathbb{R} \to \mathbb{R}$. By piecing f and $g \circ L$ together, we get the extension that we wanted, namely

$$F : I \cup L^{-1}(J) \to f(I) \cup g(J).$$

If Γ has two connected components, we may assume without loss of generality that they have slope 1 and so must be arranged as in the figure (4.1). Let us write $J = (\gamma, \beta)$. Translating J if needed, we may, again without loss of generality, assume that $\gamma = c$ and $\delta = d$, so that

$$a < b \leq c < d \leq \alpha < \beta$$

Let now $\theta = 2\pi/(\alpha - a)$. We define the diffeomorphism $h : S^1 \to M$ as follows:

$$h(cos\theta, sin\theta) = \begin{cases} f(t) \text{ for } a < t < d, \\ g(t) \text{ for } c < t < \beta. \end{cases}$$

Since S^1 is compact and h is a diffeomorphism, the image $h(S^1)$ is compact. But since it is also open, it must be the entire manifold M, which completes the proof. □

Proof (Classification Theorem) Any arbitrary parametrization by arc-length can be extended to a parametrization by arc-length $f : I \to M$ that we call maximal. It is maximal in the sense that we cannot get a parametrization by arc-length over a larger interval than I. Let us assume that M is not diffeomorphic to S^1. We will show that in that case, f is surjective and thus a diffeomorphism. If the open set $f(I)$ were not the whole of M, there would be a limit point of $f(I)$ in $M \setminus f(I)$. If we parametrize a neighbourhood of x by arc-length and then apply lemma 4.5, we see that f can be extended over a larger interval, which contradicts our hypothesis of f being maximal. □

4.2 Proof via Morse functions

In this section, we prove a slightly different version of theorem 4.1, and in order to do this, we first want to generalize our definition of manifold. Indeed, objects such as a closed ball or a compact cylinder are not manifolds as in definition 1.4 because neighbourhoods of points on the boundaries are not diffeomorphic to open subsets of $\mathbb{R}^n$. The simplest example of this is the upper half space $H^k \subset \mathbb{R}^k$ consisting of all points where the last coordinate is non-negative. The boundary of H^k is $\mathbb{R}^{k-1}$ under its usual embedding in $\mathbb{R}^k$

Definition 4.6 *A subset $X \subseteq \mathbb{R}^n$ is a k-*dimensional manifold with boundary *if every point of X possesses a neighbourhood diffeomorphic to an open set in the space H^k (of course with the relative topology). As in the first chapter, such a diffeomorphism is called a* local parametrization *of X. The* boundary *of X, which we denote by ∂X, is the set of all the points that belong to the image of the boundary of H^k under some local parametrization. We define the* interior *of X to be $Int(X) = X \setminus \partial X$.*

Remark 4.7 *One should not confuse the boundary and the interior of a manifold as we have just defined them with the topological definitions of boundary and interior. The two notions agree for $dim(X) = n$, but not in the case $dim(X) < n$. It is also worth noting that the smooth manifolds as in definition 1.4 are also manifolds with boundary. However, their boundary is empty.*

The main theorem that we want to prove is now the following:

Theorem 4.8 *Every compact connected one-dimensional manifold with boundary is diffeomorphic either to a circle or a closed interval.*

As in the previous section, we first need to prove intermediate steps. All the illustrations in the proofs below are taken from [9].

Lemma 4.9 (smoothing lemma) *Let g be a smooth function on $[a, b]$ with positive derivative everywhere except at one interior point c. Then there exists a globally smooth function $\tilde{g}$ that agrees with g near the endpoints and has positive derivative everywhere*

Proof Let $\rho : \mathbb{R} \to \mathbb{R}$ be a smooth non-negative function vanishing outside of a compact subset of the interval (a, b), and such that it is equal to 1 near c. Additionally, assume we have $\int_a^b \rho = 1$. We define

$$\tilde{g}(x) = g(a) + \int_a^x [k\rho(s) + g'(s)(1 - \rho(s))]\mathrm{d}s,$$

where k is the positive constant given by

$$k = g(b) - g(a) - \int_a^b g'(s)(1 - \rho(s))\mathrm{d}s$$

It is now easy to verify that $\tilde{g}$ has all the given properties. □

In order to prove the theorem, we choose a Morse function f on the compact one dimensional manifold X. We denote by S the finite (see remark 1.54) set of critical points of f united with the boundary points of X. Then, the set $X \setminus S$ consists of a finite number of connected 1-manifolds $L_1, \ldots, L_n$.

We now take a few steps towards our full proof, with first the following proposition.

Proposition 4.10 *f maps each L_i diffeomorphically onto an open interval of $\mathbb{R}$.*

Proof Let L denote an arbitrary L_i. Since f is a local diffeomorphism, and L is connected, $f(L)$ is open and connected in $\mathbb{R}$. Moreover, $f(L)$ is contained in $f(X)$, which is compact since X is compact, and thus we must have $f(L) = (a, b)$. It remains to show that $f : L \to (a, b)$ is injective. The proof will then be complete because smoothness follows from the fact that f is a local diffeomorphism. Let p be any point of L and set $c = f(p)$. We want to show that every other point $q \in L$ can be joined to p be a curve $\gamma : [c, d] \to L$ (or a curve $\gamma : [d, c] \to L$) such that $f \circ \gamma =$ identity and $\gamma(d) = q$. Since $f(q) = d \neq c = f(p)$, this yields injectivity. Since f is a local diffeomorphism, it is clear that the set Q of all the points q that can be joined by a such a curve γ is both open and closed. Therefore, $Q = L$ □

Here is another statement that will be of use.

Lemma 4.11 *Let L be a subset of the one-dimension manifold with boundary X such that L is diffeomorphic to an open interval of $\mathbb{R}$. Then its closure $\bar{L}$ contains at most two points that are not in L*

Proof Given a diffeomorphism $g : (a,b) \to L$, let x, y, and z be three distinct points in $\bar{L} \setminus L$. We want to reach a contradiction. Let us take sequences $(x_n), (y_n)$ and (z_n) in L converging to those three points respectively. Since g is a diffeomorphism, the sequences $g(x_n), g(y_n)$ and $g(z_n)$ can only converge towards a or b. Without loss of generality, we may assume that $g(x_n) \to a$ as well as $g(y_n) \to a$. But since x and y are distinct, there exist open neighbourhoods U_x of x and U_y of y in X such that $U_x \cap U_y = \emptyset$. By definition of convergence, for N big enough, we also must have $x_k \in L \cap U_x$ and $y_k \in L \cap U_y$ as soon as $k > N$. Then, because g is a diffeomorphism, $g(L \cap U_x)$ and $g(L \cap U_y)$ must be disjoint relatively open subsets of the interval (a,b). However they also have the property that for any open subset V of a, we have $V \cap g(L \cap U_x) \neq \emptyset$ as well as $V \cap g(L \cap U_y) \neq \emptyset$. The second property contradicts the fact that the two sets are disjoint. The proof is thus complete. □

Here, the diffeomorphism f from L_i to an open interval extends to the closure $\bar{L}_i$ This implies that each of the L_i's has precisely two boundary points and the figure (4.2) below cannot occur.

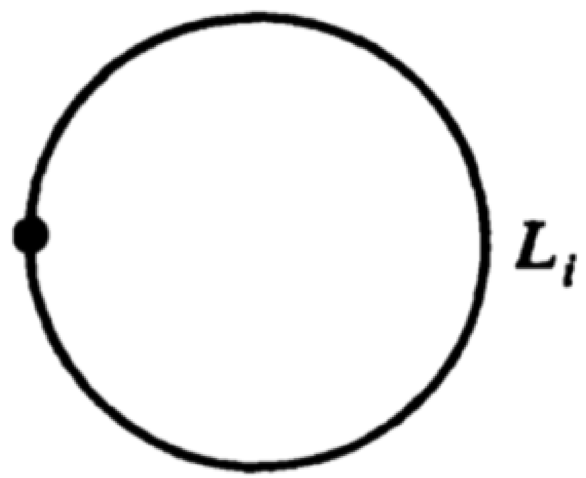


Figure 4.2: L_i cannot have exactly one boundary point.

Then, since X is a manifold, each point $p \in S$ belongs to the boundary of one or two $\bar{L}_i$. Indeed, if p belonged to the boundary of more than two of the sets $\bar{L}_i$ we would no longer have a manifold, as figure (4.3) below illustrates. In the case where p belongs to exactly one of the $\bar{L}_i$, we have $p \in \partial X$.

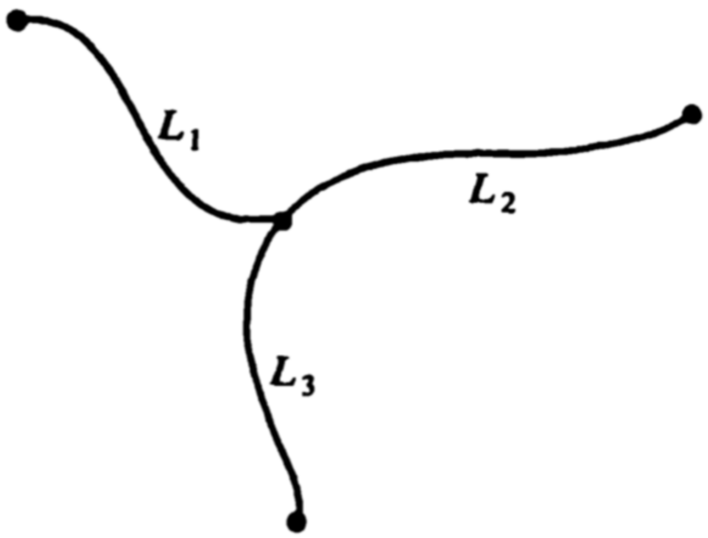


Figure 4.3: This cannot happen because we are dealing with a manifold.

Let us now take a sequence $L_1, \ldots, L_k$ such that each consecutive pair $\bar{L}_i$, $\bar{L}_{i+1}$ has a common boundary point that we call p_i for $i = 1, \ldots, k-1$. We call such a sequence of L_i's a chain. Let us denote by p_0 the other boundary point of $\bar{L}_1$ and by p_k the other boundary point of $\bar{L}_k$. There are only finitely many L_i's and thus there exists a maximal chain, i.e. a chain that we cannot extend by appending another L_j after the last set of the sequence (L_k).

We are now ready for the last claim towards the full proof of the classification theorem.

Proposition 4.12 *If $L_1, \ldots, L_k$ is a maximal chain, then it contains every L_i. If $\bar{L}_1$ and $\bar{L}_k$ have a common boundary point X is diffeomorphic to a circle. Otherwise it is diffeomorphic to a closed interval.*

Proof Let us assume that one of the L_i's (that we now call L to simplify the notation) is not part of the maximal chain that we chose. Then $\bar{L}$ cannot contain the points p_0 or p_k, since we could otherwise extend the chain. But since X is a manifold, as discussed before, $\bar{L}$ does not contain any other point p_j. Hence, the union $\bigcup_{i=1}^k \bar{L}_i$ does not intersect with any of the $\bar{L}$ excluded from the chain. From this fact, we conclude that $\bigcup_{i=1}^k \bar{L}_i$ is both open and closed in X, and thus by connectivity:

$$X = \bigcup_{i=1}^{k} \bar{L}_i$$

We know from proposition 4.10 that f behaves nicely on every L_i, but it may reverse directions as it goes through some of the boundary points as shown in figure (4.4) Let $a_i = f(p_i)$ so that f maps the manifold L_i diffeomorphically onto the interval (a_{i-1}, a_i) or (a_i, a_{i-1}) depending on which interval makes sense. Now, for every index i between 1 and k, we pick a function $\tau_i : \mathbb{R} \to \mathbb{R}$ carrying a_{i-1} to $i-1$ and a_i to i. Note that τ_i is an affine function, i.e. is of the form $t \mapsto \alpha t + \beta$. We now define the function $f_i : \bar{L}_i \to [a_{i-1}, i]$ to be the composition $\tau_i \circ f$.

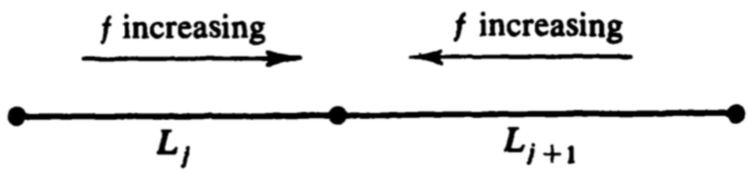


Figure 4.4: f may reverse direction when going through a critical point.

We now have two cases. Firstly, if $a_0 \neq a_k$, then the f_i agree on points where the domains of definition overlap, and we can thus define a map $F : X \to [0, k]$ by just setting $F = f_i$ on $\bar{L}_i$. By definition, F is continuous and a diffeomorphism at every point except $p_1, \ldots, p_{k-1}$. Fortunately we can use the smoothing lemma 4.9 to obtain a global diffeomorphism. In the other scenario, we have $a_0 = a_k$. We define $g_j = exp[i(2\pi/k)j]$ where i is the imaginary root. In a similar way as above, we can then define a map $G : X \to S^1$ by simply setting $G = g_j$ on $\bar{L}_j$, which is continuous and a diffeomorphism at every point except for, again, $p_1, \ldots, p_{k-1}$. Another use of the smoothing lemma 4.9 makes G a global diffeomorphism, and completes the proof of the classification theorem. □

Appendix A

A More general setup for manifolds

We use this appendix to present a more general definition of a smooth manifold. We also specify which assumptions are to be made on a topological space to avoid pathological cases when defining manifolds. It will then be clear that manifolds of Euclidean space as we have defined them are a special case.

We start with a few topological definitions.

Definition A.1 *Let X be a topological space. We say that X is* Hausdorff *if for any pair of points $x \neq y$ there exist open subsets U, V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$*

Definition A.2 *Let X be a topological space with an open cover $\{U_a\}_{a \in A}$. A* refinement *of this cover is another open cover $\{V_b\}_{b \in B}$ such that for every $b \in B$ there exists an $a \in A$ with $V_b \subset U_a$. Moreover, an open cover $\{U_a\}_{a \in A}$ is called* locally finite *if for every $x \in X$ there exists an open neighbourhood W of x such that the set $\{a \in A | U_a \cap W \neq \emptyset\}$ is a finite set. In other words, the W intersects with only finitely many sets of the open cover.*

Definition A.3 *A topological space is called* paracompact *if every open cover admits a locally finite refinement.*

Definition A.4 *A topological space X is called* locally Euclidean of dimension n *if for every point $x \in X$, there exists an open U of x and an open set $V \subset \mathbb{R}^n$ and a homeomorphism $\sigma : U \to V$.*

We are now ready to define topological manifolds, which we will than equip with a so-called smooth structure to get smooth manifolds.

Definition A.5 *A topological space M is called a* topological manifold of dimension n *if:*

1. *M is locally Euclidean of dimension n.*

2. *M is Hausdorff and has at most countably many connected components.*
3. *M is paracompact.*

Remark A.6 *It is a fact that manifolds in Euclidean space meet all the requirements to be topological manifolds.*

We now recall the definition of a concept that is used in the main text several times.

Definition A.7 *We say that a topological space X is* second countable *if it has a countable basis. One also says that the second axiom of countability holds.*

Remark A.8 *It is again a fact of point-set topology that any Hausdorff locally Euclidean topological space is second countable if and only if it is paracompact and has at most countably many connected components. This shows that our definition of topological manifold implies the second axiom of countability and that we used in the more specific setting of manifolds of Euclidean space.*

We now define the tools to make a topological manifold into a smooth manifold.

Definition A.9 *Let M be a topological manifold of dimension n. A* smooth atlas *on M is a collection*

$$\Sigma = \{\sigma_a : U_a \to V_a | a \in A\}$$

where $\{U_a\}_{a\in A}$ is an open cover of M, the V_a's are open sets in $\mathbb{R}^n$, and each $\sigma_a : U_a \to V_a$ is a homeomorphism such that the following compatibility condition holds: Suppose we have $a, b \in A$ such that $U_a \cap U_b \neq \emptyset$ then the composition (called transition map*) given by*

$$\sigma_b^{-1} \circ \sigma_a : \sigma_a(U_a \cap U_b) \to \sigma_b(U_a \cap U_b)$$

should be a diffeomorphism. We call the maps σ_a the charts *of the atlas.*

Definition A.10 *We say that two smooth atlases Σ_1 and Σ_2 are* equivalent *if their union is a smooth atlas as well. It is clear that this defines an equivalence relation on the set of smooth atlases on a given topological manifold.*

Definition A.11 *A* smooth structure *on a topological manifold is an equivalence class of smooth atlases.*

We can now define a smooth manifold in the broader sense.

Definition A.12 *A* smooth manifold of dimension *n is a pair (M, Σ) where M is a topological manifold of dimension n and Σ is a smooth structure on M.*

Remark A.13 *Again, smooth manifolds of Euclidean space as defined in the first chapter are a special case of the definition A.12.*

We conclude this section with an alternative and stronger definition of a submanifold that is sometimes necessary to avoid pathological examples.

Definition A.14 (embedded submanifold) *A subset N of an n-dimensional smooth manifold M is called an (embedded)* submanifold *of dimension k, for $k \leq n$ if for all $x \in N$ there exists a chart $\sigma : U \to V \subset \mathbb{R}^n$ of M with $x \in U$ such that $\sigma(N \cap U)$ is the intersection of a k-dimensional plane with $\sigma(U)$. The pairs $(N \cap U, \sigma|_{N \cap U})$ form an atlas on N.*

Remark A.15 *With this definition, the inclusion map from N to M is a topological embedding, which implies that the submanifold topology of N coincides with the relative topology. A weaker definition of a submanifold can be made when the inclusion map is only an injective immersion and not an embedding. In this case the submanifold topology and the relative topology may be different.*

Appendix B

Proof of Sard's theorem

We give a proof of Sard's theorem, which, although less topological than the rest of the text, is nonetheless important. Here is the statement of theorem 1.44 again.

Theorem B.1 *Let $f : X \to Y$ be a smooth map of manifolds, and let C be the set of critical points of f in X. Then the set of critical values $f(C)$ has measure zero in Y.*

We divide our proof into several steps. The first important observation is that we can use the second axiom of countability to find countable collections of open sets U_i and V_i such that the U_i's cover X and the V_i's cover Y with $f(U_i) \subseteq V_i$. Additionally, we want the U_i's and V_i's to be diffeomorphic to open euclidean sets. Therefore, the proof of Sard's theorem B.1 reduces to the following theorem:

Theorem B.2 *Let U be an open set in $\mathbb{R}^n$ and $f : U \to \mathbb{R}^p$ be a smooth map. Let C be the set of critical points of f. Then $f(C)$ has measure zero in $\mathbb{R}^p$.*

Proof The theorem is obviously true for $n = 0$, so we now prove the assertion by induction, assuming the theorem to hold for $n - 1$. We start by partitioning C into a sequence $C \supseteq C_1 \supseteq C_2 \supseteq \dots$, where C_1 is the set of all $x \in U$ such that $(\mathrm{d}f)_x = 0$ and, for $i \geq 1$, C_i is the set of all x such that all the partial derivatives up to order i vanish at the point x. The sets C_i are closed subsets of C.

We now prove the following

Claim 6 *The image $f(C \setminus C_1)$ has measure zero.*

Proof Our goal is to find an open set V around each $x \in C \setminus C_1$ such that $f(C \cap V)$ has measure zero. By the second axiom of countability again, $C \setminus C_1$ is covered by countably many of these neighbourhoods and this will yield our claim. For any $x \in C \setminus C_1$ there is a partial derivative, let us say

$\partial f / \partial x_1$ for simplicity, that does not vanish at the point x. We define the map $h : U \to \mathbb{R}^n$ as follows:

$$h(x) = (f_1(x), x_2, \ldots, x_n)$$

The derivative dh_x is non-singular by construction, so h maps a neighbourhood V of x diffeomorphically onto an open set V'. We can now examine the composition $g = f \circ h^{-1}$ which maps V' into $\mathbb{R}^p$ with the same critical values as the function f restricted to V. By construction, g has the property that it maps points of the form $(t, x_2, \ldots, x_n) \in V'$ to points of the form $(t, y_2, \ldots, y_p) \in \mathbb{R}^p$. In other words, it preserves the first coordinate. Thus, for every t, g induces a map $g^t : (t \times \mathbb{R}^{n-1}) \cup V' \to t \times \mathbb{R}^{p-1}$. We can now write the derivative of g in the form

$$\frac{\partial g_i}{\partial x_j} = \left(\begin{array}{c|c} 1 & 0 \\ \hline & \\ * & \frac{\partial g_i^t}{\partial x_j} \end{array} \right)$$

Since the determinant of the matrix on the right is just $det(\partial g_i^t / \partial x_j)$, a point of $r \times \mathbb{R}^n$ is critical for g^t if and only if it is critical for g. Using the induction hypothesis, Sard's theorem holds for $n-1$ and so the set of critical values of g^t has measure zero. Consequently, using Fubini's theorem, the set of critical values of g is of measure 0. □

Our second claim is the following

Claim 7 *$f(C_k \setminus C_{k+1})$ is of measure zero for $k \geq 1$.*

Proof This proof is similar to the one we conducted above, but a bit simpler. For $x \in C_k \setminus C_{k+1}$ there is a derivative of order $k+1$ that is not zero at the point x. We can therefore find a partial derivative of order k (let us call it ρ) that vanishes at x by definition of C_k but such that $\partial \rho / \partial x_1$ does not vanish at x. We chose the derivative in direction x_1 to ease the notation, and this can be done without loss of generality. We then define the function $h : U \to \mathbb{R}^n$ in the same fashion as before by $h(x) = (\rho(x), x_2, \ldots, x_n)$. As above, h maps a neighbourhood V of x diffeomorphically onto an open set V'. By construction, h maps $C_k \cap V$ into the hyperplane $0 \times \mathbb{R}^{n-1}$. This implies that the map $g = f \circ h^{-1}$ has all its critical points of type C^k in the hyperplane $0 \times \mathbb{R}^{n-1}$. Let $\bar{g} : (0 \times \mathbb{R}^{n-1}) \cap V' \to \mathbb{R}^p$ be the restriction of g. By induction, the set of critical values of $\bar{g}$ has measure zero. Additionally, since the critical points of g of type C^k are clearly critical points of $\bar{g}$ the image of these critical points is of measure zero, which implies that $f(C_k \cap V)$ has measure zero. $C_k \setminus C_{k+1}$ can be covered by countably many such sets V, which proves the claim. □

We conclude with this last claim

Claim 8 *For $k > n/p - 1$, $f(C_k)$ is of measure zero.*

Proof Let $S \subset U$ be a cube whose sides are of length δ. If $k > n/p - 1$, we will show that $f(C_k \cap S)$ has measure zero. Since C_k can be covered by countably many such cubes, this will prove the claim. By Taylor's theorem, the compactness of S and the definition of C_k, we observe that

$$f(x+h) = f(x) + R(x,h)$$

where we can bound the last term as follows

$$|R(x,h)| < a|h|^{k+1}$$

This holds for x in $C_k \cap S$ such that $x + h \in S$. The constant a only depends on f and S. To proceed, we divide S into r^n cubes with sides of length δ/r. Let us take S_1 to be one of these small cubes containing a point $x \in C_k$. Then, we can write any other point of S_1 as $x + h$, with the bound

$$|h| < \sqrt{n}\left(\frac{\delta}{r}\right)$$

Using the first estimation, we have that $f(S_1)$ lies in a cube with sides of length b/r^{k+1} centered about $\delta(x)$. The constant b is given by $b = 2a(\sqrt{n}\delta)^{k+1}$. From this follows that $f(C_k \cap S)$ is contained in the union of at most r^n cubes with a volume v bounded by

$$v \leq r^n \left(\frac{b}{r^{k+1}}\right)^p = b^p r^{n-(k+1)p}$$

With the assumption that $k + 1 > n/p$, we have $v \to 0$ as $r \to \infty$, so that $f(C_k \cap S)$ must have measure zero. □

Putting the three claims together completes the proof of Sard's theorem B.1. □

Bibliography

[1] ”classification of smooth 1-manifolds”. https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5122_2017S/16.pdf. Accessed: 2018-09-18.

[2] B. Andrews. Lecture notes: ”riemannian metrics”. https://maths-people.anu.edu.au/~andrews/DG/DG_chap9.pdf. Accessed: 2018-09-17.

[3] C. Eur. ”differential topology by guillemin & pollack solutions”. https://math.berkeley.edu/~ceur/notes_pdf/Eur_GPDiffTopSolns.pdf. Accessed: 2018-09-18.

[4] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[5] W. J. Merry. Lecture notes: ” smooth manifolds”. https://www.merry.io/differential-geometry-i-lecture-notes/1-smooth-manifolds. Accessed: 2018-09-26.

[6] W. J. Merry. ”lecture notes: Attaching spaces and cell complexes”. https://www.merry.io/algebraic-topology-i-lecture-notes/18-attaching-spaces-and-cell-complexes. Accessed: 2018-07-20.

[7] J. Milnor. *Topology from the differentiable Viewpoint*. the University Press of Virginia, 1965.

[8] J. Milnor. *Morse Theory*. Princeton University Press, 1973.

[9] A. Pollack V. Guillemin. *Differential Topology*. AMS Chelsea Publishing, 1974.

[10] J. H.C. Whitehead. Combinatorial homotopy i. *Bull. Amer. Math. Soc.*, 55:213–245, 1949.

[11] Wikipedia. ”submanifold”. https://en.wikipedia.org/wiki/Submanifold. Accessed: 2018-09-26.