Morse Theory and Supersymmetry

Jeremy van der Heijden

July 1, 2016

Bachelor Thesis Mathematics, Physics and Astronomy
Supervisors: prof. dr. Erik Verlinde, dr. Hessel Posthuma
Abstract

The remarkable result will be considered that using supersymmetric quantum mechanics one can derive the classical Morse inequalities. In particular, supersymmetry is used to define the Morse complex, which expresses the topology of a manifold in terms of the critical points of a real-valued function. First, the Morse complex is developed from the mathematics point of view, using gradient flow lines, stable and unstable manifolds and the Morse-Smale transversality condition. The Morse Homology Theorem, which says that the Morse complex is a complex having homology isomorphic to the singular homology, is stated without proof. Then, the Morse complex is developed from the physics point of view, which gives a beautiful interpretation of elementary particles as differential forms. The instanton calculation, which is used to define the boundary operator, is performed explicitly. Using considerations from supersymmetric quantum mechanics and the de Rham theory, a physicist’s proof of the Morse Homology Theorem is given.

Title: Morse Theory and Supersymmetry
Author: Jeremy van der Heijden, jj.vanderheijden@hotmail.com, 10580107
Supervisors: prof. dr. Erik Verlinde, dr. Hessel Posthuma
Second graders: prof. dr. Eric Opdam, prof. dr. Jan de Boer
Date: July 1, 2016

Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
Science Park 904, 1098 XH Amsterdam
http://www.science.uva.nl/math
## Contents

1 Introduction

2 Preliminaries in Differential Topology
   2.1 Differential Geometry ................................. 6
      2.1.1 Vector Fields, Integral Curves and Flows ........ 6
      2.1.2 The Gradient Vector Field ...................... 8
      2.1.3 Transversality and Orientation ................ 9
      2.1.4 Differential Forms and De Rham Cohomology ...... 11
      2.1.5 Hodge Theory ............................... 13
   2.2 Algebraic Topology ................................. 15
      2.2.1 Chain and Cochain Complexes ................. 15
      2.2.2 Singular Homology .......................... 16
   2.3 De Rham’s Theorem ............................... 17

3 Morse Theory ........................................ 20
   3.1 Morse Functions .................................. 20
   3.2 The Gradient Flow of a Morse Function ........... 22
   3.3 Stable and Unstable Manifolds .................... 24
      3.3.1 The Stable/Unstable Manifold Theorem ........ 24
      3.3.2 Proof of the Stable/Unstable Manifold Theorem . 26
   3.4 Morse-Smale Functions ............................ 30
   3.5 The Morse Homology Theorem ...................... 33

4 Supersymmetry ....................................... 38
   4.1 SUSY in 0+1 dimensions ........................... 38
   4.2 The Path Integral Formalism ...................... 42
   4.3 The Instanton Method ............................. 45
      4.3.1 Semiclassical Limit .......................... 45
      4.3.2 Euclidean Path Integral and Instantons ....... 47
   4.4 Grassmann Variables and the Fermionic Path Integral . 50

5 Witten on Morse Theory and Supersymmetry .......... 53
   5.1 Witten’s Proof of the Weak Morse Inequalities .... 53
   5.2 The Witten Complex .............................. 59

6 Populaire Samenvatting ................................ 67

Bibliography ........................................... 73
1 Introduction

“Regardless of any deviations, it was clear that I was to end up in math and physics” - Edward Witten

Over the past two centuries, mathematicians have developed a wondrous collection of conceptual machinery to peer into the invisible world of geometry in higher dimensions. Quite surprisingly, the advances in differential topology, which finds itself at the interface of topology and differential geometry, have not come solely from mathematics. More and more often, spectacular discoveries in mathematics have emerged from an interaction with theoretical physics. One such an important episode involved a 1982 paper on Morse theory and Supersymmetry by the physicist Edward Witten [30]. In his paper, Witten showed that supersymmetry, the fundamental connection between fermions and bosons, gave an elegant way of deriving important results in Morse theory. For a colorful history of Morse theory and the events leading up to Witten’s paper, we highly recommend reading Raoul Bott’s recollections in [8].

In this thesis, the profound connection between Morse theory and supersymmetry will be explored along the lines of Witten’s 1982 paper. In short, Morse theory provides a way of understanding the topology of a manifold in terms of the critical points of a real-valued function. The first part of this thesis will be an exposition of the ‘classical’ approach to finite dimensional Morse homology. The basic ideas surrounding Morse homology were developed during the first half of the twentieth century. We begin our discussion by introducing Morse functions and the gradient vector field. Solutions of this vector field, the gradient flows, give rise to stable and unstable manifolds which are well-studied objects in the theory of dynamical systems. Then, Morse functions will be considered which satisfy the Morse-Smale transversality condition, giving them all sorts of nice properties. For example, this allows one to define a sequence of groups, generated by the critical points of a Morse function, with corresponding boundary maps, induced by the gradient flow lines which connect one critical point to another. One can show that the above construction defines a complex, the so-called Morse complex, and that the corresponding homology is isomorphic to the singular homology. This result, also known as the Morse Homology Theorem, is one of the main results in Morse theory. Proving the Morse Homology Theorem is highly non-trivial and goes beyond the scope of this thesis.

Instead, we will move our attention to the interesting history of the Morse complex. The first time this complex appeared in the literature was, quite unexpectedly, in a paper on supersymmetry. Supersymmetry is a surprising subtle idea - the idea that the equations representing the basic laws of nature do not change if certain particles in the equations are interchanged with one another. The second part of this thesis aims to show that supersymmetry makes its appearance quite naturally in mathematics: in the study of the de Rham
cohomology. There is a beautiful interpretation of fermions and bosons as differential forms! Witten abused this identification to prove that the number of supersymmetric vacua is, in fact, a topological invariant. By deforming the de Rham complex the vacua become localized near the critical points of a Morse function. Now, finding the perturbative vacua comes down to counting critical points. In the exact spectrum, however, some of the perturbative vacua might be lifted. To say something about the topology we must accommodate for tunneling paths from one critical point to another. These tunneling paths, also known as instantons, are, in fact, the gradient flow lines. They can be used to define a suitable boundary operator. This is precisely the description of the Morse complex. From the physics point of view, it is not at all surprising that the Morse complex must be a complex which calculates the singular homology. Witten thus gave a physicist’s proof of the Morse Homology Theorem.

In Chapter 2, we have tried to give an accessible introduction into the concepts from differential topology necessary to understand this thesis. The reader who is already familiar with differential topology can skip most of this chapter. Chapter 3 will develop in great detail all the results needed from Morse theory to formulate the Morse Homology Theorem. In Chapter 4 the physical machinery will be given to understand Witten’s analysis: in section 4.1 we will explain the main features of supersymmetric quantum mechanics, while section 4.2 and 4.3 will be used to explain the instanton method. The instanton method, formulated in the path integral formalism, is used in [30] to calculate the relevant tunneling amplitudes. Chapter 5 will use the results from chapter 4 to explain Witten’s paper in great detail. More specifically, section 5.1 will use supersymmetry to derive the weak Morse inequalities and section 5.2 will go into the physical construction of the Morse complex. The relevant instanton calculation will be performed explicitly.

I would like to thank both prof. dr. Erik Verlinde and dr. Hessel Posthuma for supervising this bachelor thesis. In particular, I would like to thank Hessel Posthuma for pointing out Witten’s 1982 paper and for providing me with helpful feedback on my written work. A special thanks goes to Marcel Vonk, who has helped me attain physical intuition for the abstract theory and who was always available to answer my questions.

Jeremy van der Heijden
Amsterdam, July 1, 2016
2 Preliminaries in Differential Topology

In the category of topological spaces there is a number of functorial ways of associating to each space an algebraic object like a vector space or group in such a way that homeomorphic spaces have isomorphic objects. The simplest such a functor, which you are already familiar with, is the fundamental group. In this thesis, a different class of functors will be considered called homology groups. A great advantage of working with homology groups in differential topology is that we can approach the problem from two sides. On the one hand, we have the whole apparatus of differential geometry at our disposal to find the cohomology in terms of de Rham groups using closed and exact forms. On the other hand, there exists a purely topological way of calculating it through singular homology theory. In the 1930’s Georges de Rham was the first to prove the connection between de Rham groups and topology; the theorem that bears his name will be the main result of this chapter. To start things off we will present a short recap of some basic results in differential geometry and homology theory, which are relevant for our discussion of Morse theory. The de Rham’s Theorem is included to ensure that the results in chapter 3 and chapter 5 can be related. Most proofs are omitted.

2.1 Differential Geometry

This thesis is about smooth manifolds, so a proper introduction of these objects would be in order. In simple terms, they are spaces which locally look like some Euclidean space \( \mathbb{R}^n \), so that we can do calculus on them. The most familiar examples, apart from \( \mathbb{R}^n \) itself, are smooth curves, as circles or parabolas, and smooth surfaces, as spheres, tori, ellipsoids, etc. The aim of this section is to give an accessible introduction into the concepts from differential geometry necessary to understand this thesis. For a more formal introduction into manifold theory you can take up any book on basic differential geometry, for example [6] or [18].

2.1.1 Vector Fields, Integral Curves and Flows

In short, a manifold consists of a topological space \( M \) such that every point \( p \in M \) has a neighborhood \( U \), called a chart, that is homeomorphic to an open subset of \( \mathbb{R}^n \). Moreover, we can give \( M \) a smooth structure by demanding that the charts are smoothly compatible with each other. The smooth coordinate functions \( \phi : U \to \hat{U} \subset \mathbb{R}^n \) give us an identification of \( U \) and \( \hat{U} \). You can visualize this identification by thinking of a “grid” drawn on \( U \). Under this identification we can represent a point \( p \in U \) by its local coordinates \( \phi(p) = (x^1, ..., x^n) \).

A map \( f : M \to N \) from one manifold to another is called smooth, whenever its coordinate representation \( \hat{f} : \phi(U) \to \mathbb{R}^m \) is smooth. An important class of maps is given by the smooth function \( f : M \to \mathbb{R} \) for which we will write \( C^\infty(M) \).
In order to make sense of vector fields we need to introduce the concept of a tangent space, which you can think of as a ‘linear model’ of the manifold.

**Definition 2.1.** Let $M$ be a smooth manifold and $p \in M$. A linear map $v : C^\infty(M) \to \mathbb{R}$ is called a derivation at $p$ if it satisfies

$$v(fg) = f(p)vg + g(p)vf,$$

for $f, g \in C^\infty(M)$.

The set of all derivations at $p$, denoted by $T_pM$, is a vector space called the tangent space of $M$ at $p$. An element of $T_pM$ is called a tangent vector at $p$.

For any smooth function $f : M \to \mathbb{R}$ the number $v(f)$ is thought of as the directional derivative of $f$ in the direction $v$. Geometrically we like to think about tangent vectors at a point as ‘arrows’ sticking out of that point which are tangent to $M$. When we choose local coordinates $\phi(p) = (x^1, \ldots, x^n)$ on the manifold we can define

$$\left. \frac{\partial}{\partial x^i} \right|_p f = \left. \frac{\partial (f \circ \phi^{-1})}{\partial x^i} \right|_{\phi(p)}(\phi(p)).$$

The $(\left. \frac{\partial}{\partial x^i} \right|_p)$ constitute a basis for $T_pM$ so that every tangent vector $v \in T_pM$ can be written as

$$v = v^i \left. \frac{\partial}{\partial x^i} \right|_p.$$

It will be useful to consider the set of all tangent vectors at all points of our manifold, called the tangent bundle of $M$

$$TM = \coprod_{p \in M} T_pM.$$  

The tangent bundle is itself a manifold [6]. We are now in the position to define the concept which will be of huge importance in our treatment to follow.

**Definition 2.2.** A smooth vector field on $M$ is a smooth section of the projection map $\pi : TM \to M$. More concretely, a smooth vector field is a smooth map $X : M \to TM$, usually written $p \mapsto X_p$, such that $\pi \circ X = \text{Id}_M$, or equivalently, $X_p \in T_pM$ for each $p \in M$.

The primary geometric objects associated with a smooth vector field are their integral curves, which are smooth curves whose velocity at each point is equal to the value of the vector field there. The collection of integral curves of a given vector field on $M$ determines a family of diffeomorphisms, called the flow.

**Definition 2.3.** If $X$ is a smooth vector field on $M$, an integral curve of $X$ is a smooth curve $\gamma : J \to M$ such that

$$\gamma'(t) = X_{\gamma(t)}$$

for all $t \in J$.

Finding integral curves boils down to solving a system of first order ordinary differential equations in a smooth chart. A vector field therefore always has local integral curves which are unique given the starting point. Suppose that for each point $p \in M$ the vector field $X$ has
a unique integral curve starting at \( p \) which is defined for all \( t \in \mathbb{R} \) (it will not always be the case that the integral curve is defined for all \( t \in \mathbb{R} \)). We write \( \varphi^{(p)} : \mathbb{R} \rightarrow M \) for the integral curve. Now for each \( t \in \mathbb{R} \), we can define a map \( \varphi_t : M \rightarrow M \) by setting:

\[
\varphi_t(p) = \varphi^{(p)}(t).
\]

Each \( \varphi_t \) ‘slides’ the manifold along the integral curves for a period \( t \). This family of diffeomorphisms defines a **global flow on** \( M \) as a continuous left \( \mathbb{R} \)-action on \( M \); in other words, a continuous map \( \varphi : \mathbb{R} \times M \rightarrow M \) satisfying

\[
\varphi_t \circ \varphi_s = \varphi_{t+s} \text{ for all } t, s \in \mathbb{R}, \quad \varphi_0 = \text{Id}_M.
\]

As already mentioned, global flows need not always exist for a given vector field. Whenever they do exist we say that a smooth vector field is **complete**.

**Theorem 2.4.** On a compact smooth manifold, every smooth vector field is complete.

The proof of this theorem is relatively easy and can be found in [18]. The only flows we will encounter are global flows generated by the smooth gradient vector field.

### 2.1.2 The Gradient Vector Field

A natural question to ask at this point is how we can make sense of the gradient of a function in a coordinate-independent sense. It turns out that we need the concept of a cotangent space.

**Definition 2.5.** Let \( M \) be a smooth manifold. For each \( p \in M \) we define the **cotangent space at** \( p \), denoted by \( T^*_pM \), as the dual space to \( T_pM \):

\[
T^*_pM = (T_pM)^*.
\]

Elements of \( T^*_pM \) are called **covectors at** \( p \).

We define the **cotangent bundle of** \( M \) as

\[
T^*M = \coprod_{p \in M} T^*_pM.
\]

A section of \( T^*M \) is called a **covector field or differential 1-form**. A very important covector field \( df \), called the **differential of** \( f \), is given by

\[
df_p(X_p) = X_p f \text{ for } X_p \in T_pM,
\]

where \( f \in C^\infty(M) \). Choosing local coordinates \( (x^i) \) on an open subset \( U \subset M \) we have the representation

\[
df = \frac{\partial f}{\partial x^i} dx^i,
\]

where the \( dx^i \) form a **frame** for the cotangent bundle \( T^*U \) such that \( dx^i \left( \frac{\partial}{\partial x^j} \right) = \delta_{ij} \).
In manifold theory the appropriate structure to do geometry is a Riemannian metric, which
is essentially a choice of inner product on every tangent space, varying smoothly from point to
point. More formally, a Riemannian metric on $M$ is a smooth symmetric covariant 2-tensor
field on $M$ that is positive definite at each point. A Riemannian manifold is a pair $(M, g)$,
where $M$ is a smooth manifold and $g$ is a Riemannian metric on $M$. The tensor $g$ takes as
input two tangent vectors $v, w \in T_p M$ and spits out a real number $g_p(v, w)$. In any local
coordinates $(x^i)$, a Riemannian metric can be written as

$$ g = g_{ij} dx^i dx^j, $$

where $g_{ij} = g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$. The Riemannian metric $g$ provides a natural isomorphism between the tangent and cotangent bundles $\hat{g} : TM \to T^* M$ as follows. For $p \in M$ and each $v \in T_p M$ let $\hat{g}(v) \in T^*_p M$ be the covector defined by $\hat{g}(v)(w) = g_p(v, w)$ for all $w \in T_p M$. This bundle isomorphism induces an isomorphism of sections which enables us to lower and raise indices. These musical isomorphisms have local coordinate expressions

$$ X^\flat = \hat{g}(X) = g_{ij} X^i dx^j $$

for a vector field $X = X^i \frac{\partial}{\partial x^i}$, and

$$ \omega^\flat = \hat{g}^{-1}(\omega) = g^{ij} \omega_i \frac{\partial}{\partial x^j} $$

for a covector field $\omega = \omega_i dx^i$.

The gradient of a function $f$ is the following smooth vector field on Riemannian manifolds:

$$ \nabla f = (df)^\flat = \hat{g}^{-1}(df). $$

Writing out the above definition, we see that $\nabla f$ is the unique vector field satisfying

$$ g(\nabla f, X) = X f $$

for all vector fields $X$.

2.1.3 Transversality and Orientation

In this section we will present some results on transversality and orientations which will be
used in subsequent chapters. First, we need to define what we mean by a submanifold.

**Definition 2.6.** If $f : M \to N$ is a smooth map between smooth manifolds then $f$ is called
an immersion if the differential

$$ df_x : T_x M \to T_{f(x)} N, \quad v \mapsto (g \to v(g \circ f)) $$

is injective for all $x \in M$. Whenever $f$ is an injective immersion $M$ is called an immersed
submanifold of $N$.

One special kind of immersion is of particular importance:

**Definition 2.7.** A smooth injective immersion $f : M \to N$ between smooth manifolds is
called an embedding if $f$ is a homeomorphism onto its image. In that case $M$ is called an
embedded submanifold of $N$. 

9
Definition 2.8. Let $M$ and $Z$ be embedded submanifolds of $N$. We say that $M$ intersects $Z$ transversally, $M \pitchfork Z$, if whenever $x \in M \cap Z$ we have
\[ T_x M \oplus T_x Z = T_x N. \]

In other words, two submanifolds intersect transversally when their tangent spaces span the tangent space of the full manifold.

Lemma 2.9. Let $M$ and $Z$ be immersed submanifolds of $N$ and suppose $M \pitchfork Z$. Then $M \cap Z$ is an immersed submanifold of $N$ of dimension
\[ \dim(M \cap Z) = \dim M + \dim Z - \dim N. \]

Lemma 2.9 will be crucial in the construction of Morse-Smale functions in section 3.4. There, one also finds examples of submanifolds which do or do not intersect transversally.

Now orientations on a manifold will be considered. We all know the informal rules for picking a preferred ordered basis in $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$. A basis for $\mathbb{R}^1$ is usually chosen to point to the right (i.e. in the positive direction) and a natural choice of ordered bases in $\mathbb{R}^2$ consists of those for which the rotation from the first basis vector to the second is counterclockwise. Also, everyone who has done some vector calculus has encountered the “right-handed” bases in $\mathbb{R}^3$. In order to make sense of the terms “to the right”, “counterclockwise” and “right-handed” in arbitrary vector spaces we need the following definition.

Definition 2.10. Let $V$ be a real vector space of finite dimension $n > 0$. On the set of ordered bases of $V$ we define an equivalence relation as follows. We say that two ordered bases $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$ are equivalent whenever the transition matrix $C = (c_{ij})$, defined by
\[ v_i = \sum_{j=1}^{n} c_{ij} w_j \]
has positive determinant. There are two equivalence classes and we define an orientation for $V$ as a choice of one of the equivalence classes. An ordered basis that is in the orientation is called positively oriented. Any basis that is not in the orientation is called negatively oriented.

The orientation $[e_1, \ldots, e_n]$ of $\mathbb{R}^n$ determined by the standard basis is called the standard orientation.

Definition 2.11. Let $M$ be a smooth manifold. An orientation of $M$ is a choice of orientation for each tangent space $T_x M$ such that the orientations of nearby tangent spaces are consistent in the following sense. Around every point $M$ there exists a coordinate chart $\phi : U \to \mathbb{R}^n$ which is orientation preserving, i.e. for every point $x \in U$ the linear isomorphism $d\phi_x : T_x M \to \mathbb{R}^n$ is orientation preserving where $\mathbb{R}^n$ is given the standard orientation. A smooth manifold which possesses an orientation is called oriented.

Remark 2.12. If $V = W \oplus \tilde{W}$, then orientations $[w_1, \ldots, w_n]$ and $[\tilde{w}_1, \ldots, \tilde{w}_m]$ on $W$ and $\tilde{W}$ determine a unique orientation $[w_1, \ldots, w_n, \tilde{w}_1, \ldots, \tilde{w}_m]$ on $V$. Similarly, if orientations on $V$ and $W$ are given this determines a unique orientation on $\tilde{W}$. 

10
2.1.4 Differential Forms and De Rham Cohomology

The theory of differential forms can be viewed as a generalization of the theory of covector fields. Differential forms allow us to define the de Rham cohomology which will be the most important mathematical structure in this thesis.

Let us quickly come to business by defining the following sequence of vector spaces

\[ \cdots \xrightarrow{d} \Omega^{k-1}(M) \xrightarrow{d} \Omega^k(M) \xrightarrow{d} \Omega^{k+1}(M) \xrightarrow{d} \cdots \]

Here \( \Omega^k(M) \) is the vector space of differential \( k \)-forms on \( M \) and the map \( d : \Omega^k(M) \to \Omega^{k+1}(M) \) is the exterior derivative. The reader who is unfamiliar with differential forms is advised to read the corresponding chapters of [18] for a formal introduction into the subject.

In short, a \( k \)-form is a tensor field on \( M \) whose value at each point is an alternating covariant \( k \)-tensor. Thus, a \( k \)-form \( \omega \) consists of an alternating tensor \( \omega_{i_1 \ldots i_k} \) for every \( p \in M \) which takes in \( k \) tangent vectors in \( T_p M \), say \( v_1, \ldots, v_k \), and spits out a real number \( \omega_{i_1 \ldots i_k}(v_1, \ldots, v_k) \). Locally, every \( k \)-form \( \omega \in \Omega^k(M) \) can be written as

\[
\omega = \sum_{i_1 < \ldots < i_k} f_{i_1 \ldots i_k} dx^{i_1} \wedge \ldots \wedge dx^{i_k}. \tag{2.1}
\]

Here, \( dx^1, \ldots, dx^n \) is the usual orthogonal frame for the cotangent bundle, \( f_{i_1 \ldots i_k} \) a smooth real-valued function on \( M \) and the anticommutative wedge product satisfies

\[
d x^i \wedge dx^j = -dx^j \wedge dx^i \quad \text{and} \quad dx^i \wedge dx^i = 0.
\]

This gives the following property:

\[
d x^{i_1} \wedge \ldots \wedge x^{i_k} \left( \frac{\partial}{\partial x^{j_1}}, \ldots, \frac{\partial}{\partial x^{j_k}} \right) = \begin{cases} 
\text{sgn} \sigma & \text{if } (i_1, \ldots, i_k) \text{ is a permutation } \sigma \text{ of } (j_1, \ldots, j_k) \\
0 & \text{and neither have repeated indices} \\
0 & \text{otherwise}
\end{cases}
\]

From (2.1) it is clear that \( \Omega^0(M) \) is the space of all real-valued smooth functions and \( \Omega^1(M) \) is just the vector space of covector fields on \( M \). It is easy to check that for \( \omega \in \Omega^k(M), \eta \in \Omega^l(M) \) the wedge product satisfies the anticommutative property \( \omega \wedge \eta = (-1)^{kl} \eta \wedge \omega \).

**Definition 2.13.** The wedge product turns

\[
\Omega^*(M) := \bigoplus_{k=0}^n \Omega^k(M),
\]

into an associative, anticommutative graded algebra called the exterior algebra of \( M \).

Recall that for \( f \in C^\infty(M) \) and \( X \) a smooth vector field we have \( df(X) = X f \). If \( \omega \in \Omega^k(M) \) then we define its exterior derivative \( d\omega \in \Omega^{k+1}(M) \) as follows. If \( \omega = f dx^{i_1} \wedge \ldots \wedge dx^{i_k} \) then we put \( d\omega = df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \). This is extended linearly to all forms. Of course, one has to check that this definition is independent of the local coordinates used to define it. Note that the exterior derivative gives a linear map \( d : \Omega^k(M) \to \Omega^{k+1}(M) \), which satisfies

\[
d \circ d = 0.
\]
Moreover, if \( \omega \) is a \( k \)-form and \( \eta \) is a \( p \)-form then
\[
d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.
\]
Both properties follow from straightforward calculations. The vector space \( \Omega^\bullet(M) \) with differential operator \( d \) is called the de Rham complex on \( M \). The kernel of \( d \) are the closed forms, i.e. \( d\omega = 0 \), and the image of \( d \) the exact forms, i.e. \( \omega = d\eta \). In the words of Bott and Tu [7], the de Rham complex may be viewed as a God-given set of differential equations, whose solutions are the closed forms. Because \( d \circ d = 0 \), the exact forms are automatically closed. They are the trivial or “uninteresting” solutions. A measure of the size of the space of “interesting solutions” is the definition of the de Rham cohomology.

**Definition 2.14.** The \( p \)-th de Rham cohomology group of \( M \) is the real vector space
\[
H^p_{dR}(M) := \frac{\ker(d : \Omega^p \to \Omega^{p+1})}{\im(d : \Omega^{p-1} \to \Omega^p)} = \text{closed } p \text{-forms}/\text{exact } p \text{-forms},
\]
By convention, we consider \( \Omega^p(M) \) to be the zero vector space when \( p < 0 \) or \( p > \dim M \), so that \( H^p_{dR}(M) = 0 \) in those cases. For a closed \( p \)-form \( \omega \) in \( M \), we write \([\omega]\) for the equivalence class of \( \omega \) in \( H^p_{dR}(M) \), called the cohomology class of \( \omega \). If \([\omega] = [\omega']\), i.e., \( \omega - \omega' = d\eta \) for certain \( \eta \in \Omega^{p-1}(M) \), we say that \( \omega \) and \( \omega' \) are cohomologous.

**Definition 2.15.** If \( f : M \to N \) is a smooth map we define the pullback of \( f \) as the map sending \( \omega \), a differential form on \( N \), to the differential form \( f^*\omega \) on \( M \), satisfying
\[
(f^*\omega)_p(v_1, \ldots, v_k) = \omega_{f(p)}(df_p(v_1), \ldots, df_p(v_k)).
\]
Note that \( f^* : \Omega^k(N) \to \Omega^k(M) \) is linear over \( \mathbb{R} \). Moreover, \( f^*(d\omega) = d(f^*(\omega)) \). This ensures that \( f^* \) descends to a linear map \( f^* : H^p_{dR}(N) \to H^p_{dR}(M) \) on cohomologies satisfying \((f \circ g)^* = g^* \circ f^* \) and \((\Id_M)^* = \Id_{H^p_{dR}}\). In the language of category theory, the de Rham cohomology is a contravariant functor from the category of smooth manifolds \( \text{Man}^\infty \) to the category of abelian groups \( \text{Ab} \).

There is an important operation which relates vector fields to differential forms:

**Definition 2.16.** Let \( \omega \) be a \( k \)-form on \( M \). For each \( p \in M \) and \( v \in T_pM \), we define a linear map \( i_v \), called interior multiplication by \( v \), mapping the alternating \( k \)-tensor \( \omega_p \) to the alternating \((k-1)\)-tensor \( i_v\omega_p \) given by
\[
i_v\omega_p(w_1, \ldots, w_{k-1}) = \omega_p(v, w_1, \ldots, w_{k-1}).
\]
In other words \( i_v\omega_p \) is obtained by plugging in \( v \) in the first slot. We often write
\[
v \cdot \omega_p = i_v\omega_p.
\]
Interior multiplication extends naturally to vector fields and differential forms, by letting it act pointwise. If \( X \) is a vector field on \( M \) we define the linear map \( i_X : \Omega^k(M) \to \Omega^{k-1}(M) \) by sending a \( k \)-form \( \omega \) to a \((k-1)\)-form \( X \cdot \omega = i_X\omega \) given by
\[
(X \cdot \omega)_p = X_p \cdot \omega_p.
\]
2.1.5 Hodge Theory

The Hodge star operator is a linear map defined on the exterior algebra of differential forms of an oriented manifold $M$ of dimension $n$. It establishes a mapping from the space of $k$-forms to the space of $(n-k)$-forms. Using an inner product on differential forms, which comes from the Hodge star operator, we define the Laplace operator, which is a suitable generalization of the usual Laplacian. Then, Hodge’s Theorem is stated which gives a one-to-one correspondence between cohomology classes and harmonic forms.

Suppose $(M,g)$ is an oriented Riemannian manifold of dimension $n$. Note that we can extend the metric $g$ to work as an ‘inner product’ on the cotangent bundle $T^*M$ by defining

$$\langle \omega, \eta \rangle = g_p(\omega^i, \eta^i),$$

where $\omega, \eta \in T^*_pM$ are covectors and the sharp notation indicates the tangent-cotangent isomorphism. In local coordinates $(x^i)$ we can write

$$\langle \omega, \eta \rangle = g^{ij} \omega^i \eta^j.$$

Recall that $(g^{ij}) = (g_{ij})^{-1}$. One can show that the above inner product generalizes to alternating $k$-tensors $\Lambda^k(T^*_pM)$ on $T^*_pM$ by setting

$$\langle \omega^1 \wedge ... \wedge \omega^k, \eta^1 \wedge ... \wedge \eta^k \rangle = \det(\langle \omega^i, \eta^j \rangle),$$

whenever $\omega^1, ..., \omega^k, \eta^1, ..., \eta^k$ are covectors at $p$.

**Proposition 2.17.** Let $M$ be an oriented Riemannian manifold. There exists a unique smooth bundle homomorphism $\ast : \Lambda^k(T^*M) \rightarrow \Lambda^{n-k}(T^*M)$ satisfying

$$\omega \wedge \ast \eta = \langle \omega, \eta \rangle \ast (1)$$

for $\omega, \eta \in \Lambda^k(T^*M)$. The map $\ast$ is called the Hodge star operator.

**Proof.** We construct it by considering an oriented orthonormal frame $(\varepsilon^i)$ for $T^*M$ and setting

$$\ast(\varepsilon^{i_1} \wedge ... \wedge \varepsilon^{i_k}) = \varepsilon^{j_1} \wedge ... \wedge \varepsilon^{j_{n-k}},$$

where $j_1, ..., j_{n-k}$ is selected such that $(\varepsilon^{i_1}, ..., \varepsilon^{i_k}, \varepsilon^{j_1}, ..., \varepsilon^{j_{n-k}})$ is positive for the cotangent bundle orientation. In particular, we have

$$\ast(1) = \varepsilon^1 \wedge ... \wedge \varepsilon^n,$$

$$\ast(\varepsilon^1 \wedge ... \wedge \varepsilon^n) = 1.$$

One easily checks that this definition of $\ast$ satisfies (2.2). \qed

Note that the Hodge star operator induces a map of sections $\ast : \Omega^k(M) \rightarrow \Omega^{n-k}(M)$. Moreover, by a straightforward calculation we have

$$\ast \ast = (-1)^{k(n-k)} : \Omega^k(M) \rightarrow \Omega^k(M),$$

where the factor $(-1)^{k(n-k)}$ arises as a determinant of the base change in this calculation.

13
Definition 2.18. We define an inner product on $\Omega^k(M)$ by setting
\[
(\omega, \eta) := \int_M \langle \omega, \eta \rangle \ast (1) = \int_M \omega \wedge \ast \eta,
\]
where we used the definition of $\ast$ at the second equality.

The fact that $(\cdot, \cdot)$ is an inner product follows from the fact that $\langle \cdot, \cdot \rangle$ is an inner product and that $\ast(1) \geq 0$ is a positive form. For $1 \leq k \leq n$ define a map $d^* : \Omega^k(M) \to \Omega^{k-1}(M)$ by $d^*\omega = (-1)^{n(k+1)+1} \ast d \ast \omega$, where $\ast$ is the Hodge star operator. Extend this definition to 0-forms by setting $d^*\omega = 0$ for $\omega \in \Omega^0(M)$. Obviously $d^* \circ d^* = 0$. Now a short calculation shows that
\[
(d^*\omega, \eta) = (\omega, d\eta),
\]
for all $\omega \in \Omega^k(M)$ and $\eta \in \Omega^{k-1}(M)$. The codifferential $d^*$ is the formal adjoint of $d$ with respect to the inner product in (2.3). A combination of $d$ and $d^*$ allows us to define the most important operator for our analysis.

Definition 2.19. The Laplace operator on $\Omega^k(M)$ is
\[
\Delta := dd^* + d^* d : \Omega^k(M) \to \Omega^k(M).
\]
A k-form $\omega$ is called harmonic if $\Delta \omega = 0$. We will write $\ker(\Delta|_{\Omega^k})$ for the harmonic k-forms.

Proposition 2.20. We have
\[
\Delta \omega = 0 \iff d\omega = 0 \text{ and } d^* \omega = 0.
\]
In particular harmonic forms are closed.

Proof. By rewriting
\[
(\Delta \omega, \omega) = (dd^* \omega, \omega) + (d^* d\omega, \omega) = (d^* \omega, d^* \omega) + (d\omega, d\omega) = \|d^* \omega\|^2 + \|d\omega\|^2
\]
the statement follows immediately. \qed

We end this section with a famous result due to Hodge, which is at the core of Witten’s analysis in section 5.1.

Theorem 2.21 (Hodge). Let $M$ be a compact Riemannian manifold. Then every cohomology class in $H^k_{\text{dr}}(M)$ contains precisely one harmonic form: the inclusion $\ker(\Delta|_{\Omega^k}) \hookrightarrow \Omega^k(M)$ induces an isomorphism in cohomology.

A proof of this statement using Dirichlet’s principle can be found in [16].

Corollary 2.22. Let $M$ be a compact oriented smooth manifold. Then all cohomology groups $H^p_{\text{dr}}(M)$ are finite dimensional.

This allows us to give the following definition.

Definition 2.23. The $p$-th Betti number of $M$ is defined to be $\beta_p(M) = \dim H^p_{\text{dr}}(M)$. 

14
2.2 Algebraic Topology

We begin this section with a basic notion from homological algebra: chain and cochain complexes. Then, we will look at a very particular chain complex which finds its inspiration in topology: singular homology theory. Only the basics will be treated here so for a more detailed account you can consult a standard book on algebraic topology as [6].

2.2.1 Chain and Cochain Complexes

With the de Rham complex we have seen the first instance of a more general concept in homological algebra: a cochain complex.

Definition 2.24. Let $R$ be a ring and consider the sequence of $R$-modules

\[ \ldots \rightarrow A^{k-1} \xrightarrow{d} A^k \xrightarrow{d} A^{k+1} \xrightarrow{d} \ldots \]

where $d : A^k \rightarrow A^{k+1}$ is an $R$-linear map. Such a sequence is a cochain complex if the composition of any two successive application of $d$ is the zero map

\[ d \circ d = 0. \]

For the above cochain complex we often write $(A^\bullet, d)$ or just $A^\bullet$.

In all of our applications, the ring will be either $\mathbb{Z}$ in which case we are looking at a sequence of abelian groups with homomorphisms, or $\mathbb{R}$, in which case we have vector spaces and linear maps. The concept of modules is just a nice way of combining the above two cases. Now such a sequence of modules is called exact if the image of each $d$ is equal to the kernel of the next. It is easy to see that exact sequences are cochain complexes. However, in general the converse is not true.

Definition 2.25. The $p$-th cohomology group of $A^\bullet$ is the quotient module

\[ H^p(A^\bullet) := \frac{\ker (d : A^p \rightarrow A^{p+1})}{\operatorname{im} (d : A^{p-1} \rightarrow A^p)}. \]

It can be thought of as a measure of the failure of exactness at $A^p$.

The obvious example of a cochain complex is the de Rham complex on a manifold $M$. If $A^\bullet$ and $B^\bullet$ are complexes, a cochain map from $A^\bullet$ to $B^\bullet$, $f : A^\bullet \rightarrow B^\bullet$, is a collection of linear maps $f : A^p \rightarrow B^p$, where we write $f$ for every map for simplicity, such that the diagram

\[ \cdots \rightarrow A^{k-1} \xrightarrow{d} A^k \xrightarrow{d} A^{k+1} \xrightarrow{d} \cdots \]

\[ \cdots \rightarrow B^{k-1} \xrightarrow{d} B^k \xrightarrow{d} B^{k+1} \xrightarrow{d} \cdots \]

commutes. The fact that $f \circ d = d \circ f$ means that any cochain map induces a linear map on cohomology $f^* : H^p(A^\bullet) \rightarrow H^p(B^\bullet)$ for each $p$ defined by $f^*[a] = [f(a)]$. The reader
who is somewhat familiar with category theory will now recognize the cohomology $H^p$ to be covariant functor sending a chain complex $A^\bullet$ to the abelian group $H^p(A^\bullet)$ and the cochain map $f : A^\bullet \to B^\bullet$ to the the induced map $f^* : H^p(A^\bullet) \to H^p(B^\bullet)$.

As with all constructions in category theory we can reverse the arrows in our diagram to obtain a dual construction. In our case, this corresponds to the concept of a chain complex...

... \[ \cdots \xrightarrow{\partial} A_{k+1} \xrightarrow{\partial} A_k \xrightarrow{\partial} A_{k-1} \xrightarrow{\partial} \cdots \]

where the $R$-linear maps $\partial$ go in the direction of decreasing indices. We often write $(A_\bullet, \partial)$ or just $A_\bullet$ for the above chain complex. Since $\partial \circ \partial = 0$, we can define a concept for chain complexes which is analogous to cohomology, called homology.

**Definition 2.26.** The $p$-th homology group of $A_\bullet$ is quotient module

$$H_p(A_\bullet) := \frac{\ker(\partial : A^p \to A^{p-1})}{\text{im}(\partial : A^{p+1} \to A^p)}.$$

One often says complex for both chain an cochain complexes when it is clear from the context which of the two is meant.

### 2.2.2 Singular Homology

We now give a brief summary of singular homology theory.

**Definition 2.27.** Let $\mathbb{R}^\infty$ have the basis $e_0, e_1, \ldots$. Then the standard $p$-simplex is

$$\Delta_p = \left\{ \sum_{i=0}^p \lambda_i e_i \left| \sum_{i=0}^p \lambda_i = 1, \ 0 \leq \lambda_i \leq 1 \right. \right\}.$$

For example, $\Delta_0 = \{1\}$, $\Delta_1 \cong [0,1]$, $\Delta_2$ is the triangle with vertices $(0,0), (1,0)$ and $(0,1)$ with interior, and $\Delta_3$ is the solid tetrahedron. Given points $v_0, \ldots, v_k \in \mathbb{R}^n$, let $[v_0, \ldots, v_k]$ denote the map $\Delta_p \to \mathbb{R}^n$ taking $\sum \lambda_i e_i \mapsto \sum \lambda_i v_i$. For each $i = 0, \ldots, p$ we define the $i$-th face map in $\Delta_p$ to be the singular $(p-1)$-simplex $F_i : \Delta_{p-1} \to \Delta_p$ defined by

$$F_i = [e_0, \ldots, \hat{e}_i, \ldots, e_p].$$

Here, the notation of putting a hat over a symbol indicates that this symbol is omitted. So $F_i$ maps $\Delta_{p-1}$ homeomorphically onto a boundary face of $\Delta_p$, by dropping the $i$-th vertex.

**Definition 2.28.** Let $M$ be a topological space. A continuous map $\sigma : \Delta_p \to M$ is called a singular $p$-simplex. The $p$-th singular chain group of $M$, denoted by $\Delta_p(M)$ is the free abelian group generated by all singular $p$-simplices. An element of this group, called a singular $p$-chain, is a finite formal linear combination of singular $p$-simplices with integer coefficients $\sum_{\sigma} n_\sigma \sigma$. For convenience we put $\Delta_p(M) = 0$ for $p < 0$.

The boundary of a singular $p$-simplex $\sigma : \Delta_p \to M$ is the singular $(p-1)$-chain $\partial \sigma$ defined by

$$\partial \sigma := \sum_{i=0}^p (-1)^i \sigma \circ F_i.$$
This extends uniquely to a group homomorphism $\partial : \Delta_p(M) \to \Delta_{p-1}(M)$, called the singular boundary operator. The important fact about the boundary operator is that for any singular chain $c$, $\partial(\partial c) = 0$. Thus, we have a sequence of abelian groups and homomorphisms

$$... \xrightarrow{\partial} \Delta_{k+1}(M) \xrightarrow{\partial} \Delta_{k}(M) \xrightarrow{\partial} \Delta_{k-1}(M) \xrightarrow{\partial} ...$$

which is a complex, called the singular chain complex. The $p$-th homology group of this complex $H_p(\Delta\cdot(M))$ is called the $p$-th singular homology group for which we write $H_p(M)$. A singular $p$-chain $c$ is called a cycle if $\partial c = 0$ and a boundary if $c = \partial b$ for some singular $(p+1)$-chain $b$.

A continuous map $f : M \to N$ induces a homomorphism $f_\# : \Delta_p(M) \to \Delta_p(N)$ on each singular chain group by $f_\#(\sigma) = f \circ \sigma$ on singular simplices $\sigma$ and extended linearly to chains. Since $f_\# \circ \partial = \partial \circ f_\#$, $f_\#$ is a chain map, an therefore induces a homomorphism on the singular homology groups, denoted by $f_* : H_p(M) \to H_p(N)$. It is clear that $(g \circ f)_* = g_* \circ f_*$ and $(\text{Id}_M)_* = \text{Id}_{H_p(M)}$, so the $p$-th singular homology defines a covariant functor from the category of topological spaces $\text{Top}$ to the category of abelian groups $\text{Ab}$. In particular, homeomorphic spaces have isomorphic singular homology groups.

### 2.3 De Rham’s Theorem

In this section we wish to relate de Rham cohomology to singular homology. The connection will be established by integrating differential forms over singular chains. In order to do this we must make a slight modification to singular homology when considering a manifold $M$. We will restrict ourselves to smooth singular simplices $\sigma : \Delta_p \to M$. Such a simplex is smooth in the sense that it has a smooth extension to a neighborhood of each point. Beginning with smooth singular simplices we can define smooth chain groups $\Delta^\infty_p(M)$ and smooth singular homology groups $H^\infty_p(M)$ similarly to the non-smooth case. We have the following result.

**Theorem 2.29.** For any smooth manifold $M$ the inclusion map $i : \Delta^\infty_p(M) \subseteq \Delta_p(M)$ induces an isomorphism $i_* : H^\infty_p(M) \to H_p(M)$.

The proof is rather technical and can be found in [18]. The important conclusion one can draw from this theorem is that by moving to smooth simplices we do not lose or gain any essential information about the homology of the manifold. From now on we will just write $\Delta_p(M)$ and $H_p(M)$ for the smooth analogs of the chain and homology groups.

Suppose we have a $p$-form $\omega$ on a manifold $M$ and a smooth singular $p$-simplex $\sigma : \Delta_p \to M$. We define

$$\int_{\sigma} \omega := \int_{\Delta_p} \sigma^* \omega.$$  

This makes sense because $\Delta_p$, as a smooth $p$-submanifold in $\mathbb{R}^p$, inherits the orientation of $\mathbb{R}^p$. For a $p$-chain $c = \sum n_\sigma \sigma$ we define

$$\int_c \omega = \sum n_\sigma \int_{\sigma} \omega.$$
This gives us a homomorphism

\[ \Psi(\omega) : \Delta_p(M) \to \mathbb{R}, \quad c \mapsto \int_c \omega. \]

Note that this is linear in \( \omega \) as well. Hence, we have a linear map

\[ \Psi : \Omega^p(M) \to \text{Hom}(\Delta_p(M), \mathbb{R}), \]

where \( \text{Hom}(\Delta_p(M), \mathbb{R}) \) denotes the space of homomorphisms from \( \Delta_p(M) \) to \( \mathbb{R} \). Let \( \omega \) be a \((p-1)\)-form and \( \sigma \) a smooth singular \( p \)-simplex in \( M \). Applying Stokes’ Theorem [18] at (*), we obtain

\[ \Psi(d\omega)(\sigma) = \int_{\sigma} d\omega = \int_{\Delta_p} \sigma^*(d\omega) = \int_{\Delta_p} d(\sigma^*\omega) = \int_{\partial \Delta_p} \sigma^*\omega \]

\[ = \sum_i (-1)^i \int_{\Delta_{p-1}} F_i^* \circ \sigma^*\omega = \sum_i (-1)^i \int_{\Delta_{p-1}} (\sigma \circ F_i)^*\omega \]

\[ = \int_{\partial \sigma} \omega = \Psi(\omega)(\partial \sigma) = \delta(\Psi(\omega))(\sigma), \]

where

\[ \delta : \text{Hom}(\Delta_{p-1}(M), \mathbb{R}) \to \text{Hom}(\Delta_p(M), \mathbb{R}) \]

is defined as the transpose of \( \partial \), i.e. \( \delta(f)(c) = f(\partial c) \). Thus, we have the commutative diagram

\[ \begin{array}{ccc}
\Omega^{p-1}(M) & \xrightarrow{\Psi} & \text{Hom}(\Delta_{p-1}(M), \mathbb{R}) \\
d \downarrow & & \delta \downarrow \\
\Omega^p(M) & \xrightarrow{\Psi} & \text{Hom}(\Delta_p(M), \mathbb{R})
\end{array} \]

That is, \( \Psi \) is a cochain map. The groups on the right (the duals of the chain groups) together with the map \( \delta \) form a cochain complex. We write

\[ \Delta^p(M; \mathbb{R}) := \text{Hom}(\Delta_p(M), \mathbb{R}). \]

for the smooth singular cochain complex. The singular cohomology of \( M \) (with coefficients in \( \mathbb{R} \)) is

\[ H^p(M; \mathbb{R}) := H^p(\Delta^*(M; \mathbb{R})). \]

The cochain map \( \Psi \), above, then induces a homomorphism

\[ \Psi^* : H^p_{dR}(M) \to H^p(M; \mathbb{R}). \]

The de Rham’s Theorem states the following:

**Theorem 2.30** (de Rham). The homomorphism

\[ \Psi^* : H^p_{dR}(M) \to H^p(M; \mathbb{R}) \]

is an isomorphism for all smooth manifolds \( M \).
For a proof of this theorem one should consult [6]. We would now like to relate singular cohomology $H^p(M; \mathbb{R})$ which has coefficients in $\mathbb{R}$ to singular homology $H_p(M)$ which has coefficients in $\mathbb{Z}$. The Universal Coefficient Theorem makes this relation precise. In its most general form the Universal Coefficient Theorem is a statement relating homologies and cohomologies of an arbitrary chain complex with coefficients in $\mathbb{Z}$ and an abelian group $G$ respectively. It involves some intense techniques from homological algebra, for example Ext-functors, which go beyond the scope of this thesis. In [6] one finds several variations of the Universal Coefficient Theorem with corresponding proofs. The main conclusion we draw from the Universal Coefficient Theorem is that if we know the singular homology $H_p(M)$ of a space it is a purely algebraic problem to describe the singular cohomology groups $H^p(M; G)$ with arbitrary coefficients. If we take real coefficients matters simplify substantially.

**Corollary 2.31.** For every smooth manifold $M$ there is an isomorphism

$$H^p(M; \mathbb{R}) \cong \text{Hom}(H_p(M), \mathbb{R}).$$

Thus, singular cohomology with real coefficients corresponds to the dual of singular homology. Combining this with Theorem 2.30 we obtain an explicit relation between singular homology groups $H_p(M)$ and the de Rham groups $H^p_{dR}(M)$. Moreover, this ensures that there is an equivalent definition of the $p$-th Betti number $\beta_p(M)$ of $M$ as the rank of $H_p(M)$.

The important message one should take home from the above considerations is that singular homology and the de Rham cohomology are in some sense dual notions. They are fundamentally different in their approach, but calculate the same topological invariant. So why have we bothered to establish this relationship? On the one hand, this thesis develops the Morse complex in the homology framework. On the other hand, following [30], supersymmetry is used to develop an equivalent complex in the cohomology framework. The results in this section ensure that both approaches give the same result.
3 Morse Theory

3.1 Morse Functions

A critical point of a smooth function $f : M \to \mathbb{R}$ is a point $p \in M$ such that $df_p = 0$.

**Definition 3.1.** A Morse function $f : M \to \mathbb{R}$ on a smooth manifold $M$ is a smooth function whose critical points are all non-degenerate, i.e. $H_p(f)$, the Hessian of $f$ at $p$, satisfies $\det(H_p(f)) \neq 0$ for all critical points $p$.

The Hessian $H_p(f)$ of a smooth function $f : M \to \mathbb{R}$ at a critical point $p$ is the symmetric bilinear map $H_p(f) : T_pM \times T_pM \to \mathbb{R}$ defined as follows. For tangent vectors $v, w \in T_pM$ choose extensions $\bar{v}$ and $\bar{w}$ to smooth vector fields on an open neighborhood of $p$. Then $\bar{w}(f)$ is again a smooth function, whose value at $p \in M$ is $\bar{w}(f)(p) = \bar{w}_p(f)$. So we can set $H_p(f)(v, w) = \bar{v}_p(\bar{w}(f)) = v(\bar{w}(f))$.

By definition the above is independent of the extension $\bar{v}$ of $v$. To show that the value is also independent of the extension of $\bar{w}$ we use the identity

$$\bar{v}_p(\bar{w}(f)) - \bar{w}_p(\bar{v}(f)) = [\bar{v}, \bar{w}]_p(f).$$

Since $p$ is a critical point of $f$ we have $[\bar{v}, \bar{w}]_p(f) = df_p([\bar{v}, \bar{w}]_p) = 0$ so (3.1) shows that $\bar{v}_p(\bar{w}(f)) = \bar{w}_p(\bar{v}(f))$, i.e. the Hessian is symmetric. It also shows that the value of the Hessian is independent of the choice of extension for $w$. Thus, at a critical point $p$ the Hessian of $f$ is a well-defined symmetric bilinear form on $T_pM$. In a local coordinate chart $\phi(p) = (x^1, \ldots, x^n)$ the matrix of $H_p(f)$ is expressed by the matrix of second partial derivatives

$$H_p(f) = \left( \frac{\partial^2 (f \circ \phi^{-1})}{\partial x^i \partial x^j}(\phi(p)) \right).$$

Since this matrix is symmetric, it is diagonalizable with real eigenvalues. The signs of the eigenvalues are uniquely determined by $H_p(f)$, while the magnitudes of the eigenvalues depends on the choice of coordinate chart.

**Definition 3.2.** Let $p$ be a critical point of the Morse function $f : M \to \mathbb{R}$. The index of $p$ is defined as the dimension of the subspace of $T_pM$ on which $H_p(f)$ is negative definite, i.e. the number of negative eigenvalues of the Hessian matrix at $p$, and is denoted by $\lambda_p$. 

20
To get a feeling of what is going on one can think of the index of a critical point $p$ as the number of ‘directions’ in which you can walk down along your manifold. We should look at a few examples to get accustomed with the above definitions.

**Example 3.3.** Let

$$S^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} | x_1^2 + \ldots + x_{n+1}^2 = 1\}$$

be the $n$-sphere and define the height function $f : S^n \to \mathbb{R}$ by $f(x_1, \ldots, x_{n+1}) = x_{n+1}$. It is quite easy to see that $f$ is a smooth Morse function on $S^n$ with two critical points, the north pole $N = (0, \ldots, 0, 1)$ and the south pole $S = (0, \ldots, 0, -1)$. The critical points $N$ and $S$ are of index $\lambda_N = n$ and $\lambda_S = 0$ respectively.

![Figure 3.1: The sphere $S^n$ with standard height function $f$.](image)

**Example 3.4.** We consider an all-time favorite, the torus $T^2$ resting vertically on the plane $z = 0$ in $\mathbb{R}^3$. The height function $f : T^2 \to \mathbb{R}$ is a Morse function. There are four critical points: the maximum $p$, the saddle points $q$ and $r$, and the minimum $s$. The indices are given by $\lambda_p = 2, \lambda_q = \lambda_r = 1, \lambda_s = 0$ respectively. We will revisit this example on multiple occasions.

![Figure 3.2: The torus $T^2$ with standard height function $f$.](image)

Recall that we can always embed a smooth manifold $M$ of dimension $n$ into $\mathbb{R}^k$ for some $k > n$ [6]. Denoting by $(x_1, \ldots, x_k)$ the coordinates of a point $x \in M$, we have the following surprising result on the abundance of Morse functions:
Theorem 3.5. For almost all $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ (with respect to the Lebesgue measure on $\mathbb{R}^k$), the function $f : M \to \mathbb{R}$ given by $f(x) = a_1 x_1 + \ldots + a_k x_k$, is a Morse function.

For a proof of this statement we refer to [4]. One more result will be given, since it turns out to be important in our future treatment.

Lemma 3.6. Non-degenerate critical points are isolated.

Proof. Let $p \in M$ be a non-degenerate critical point of $f : M \to \mathbb{R}$, and let $\phi : U \to \mathbb{R}^m$ be a chart around $p$ such that $\phi(p) = 0$. Consider the map $g : \phi(U) \to \mathbb{R}^m$ given by

$$g(x) = \left( \frac{\partial}{\partial x^1}(f \circ \phi^{-1})(x), \ldots, \frac{\partial}{\partial x^n}(f \circ \phi^{-1})(x) \right).$$

Note that $g(0) = 0$ and $dg_0 = H_0(f)$ is non-singular. By the Inverse Function Theorem $g$ is a diffeomorphism of some neighborhood $U_0$ of 0 to another neighborhood $\tilde{U}_0$ of 0. In particular, $g$ is injective on $U_0$, that is, for all $x \in U_0 \setminus \{0\}$ we have $g(x) \neq g(0) = 0$. Thus, $x$ is not a critical point of $f$.

The following corollary is an immediate consequence.

Corollary 3.7. A Morse function on a finite dimensional compact smooth manifold has a finite number of critical points.

3.2 The Gradient Flow of a Morse Function

In this section we will look at the vector field induced by the gradient of a Morse function. Recall that the gradient vector field of $f$ is the unique smooth vector field $\nabla f$ such that $g(\nabla f, X) = df(X) = Xf$ for all smooth vector fields $X$ on $M$. We have a local 1-parameter group of diffeomorphisms $\varphi_t : M \to M$ generated by the negative gradient $-\nabla f$, i.e.

$$\frac{d}{dt} \varphi_t(x) = -(\nabla f)(\varphi_t(x)), \quad \varphi_0(x) = x.$$

The integral curve $\gamma_x : \mathbb{R} \to M$ given by $\gamma_x(t) = \varphi_t(x)$ is called a gradient flow line.

Proposition 3.8. Every smooth function $f : M \to \mathbb{R}$ on a finite dimensional smooth Riemannian manifold $(M, g)$ decreases along its gradient flow lines.

Proof. This is a short computation:

$$\frac{d}{dt} f(\gamma_x(t)) = \frac{d}{dt} (f \circ \varphi_t(x))$$

$$= df_{\varphi_t(x)} \circ \frac{d}{dt} \varphi_t(x)$$

$$= df_{\varphi_t(x)}(-\nabla f)(\varphi_t(x))$$

$$= -\|(\nabla f)(\varphi_t(x))\|^2 \leq 0,$$

which proves that $f$ is decreasing along the gradient flow line $\gamma_x$. 

\[\square\]
Proposition 3.9. Let $f : M \to \mathbb{R}$ be a Morse function on a finite dimensional compact smooth Riemannian manifold $(M,g)$. Then every gradient flow line of $f$ begins and ends at a critical point, i.e. for any $x \in M$, $\lim_{t \to \infty} \gamma_x(t)$ and $\lim_{t \to -\infty} \gamma_x(t)$ exist and are both critical points of $f$.

Proof. Let $x \in M$ and let $\gamma_x(t)$ be the gradient flow line through $x$. Because $M$ is compact, $\gamma_x(t)$ is defined for all $t \in \mathbb{R}$ by Theorem 2.4 and the image of $f \circ \gamma_x : \mathbb{R} \to \mathbb{R}$ is a bounded subset of $\mathbb{R}$. Using Proposition 3.8 we must therefore require that

$$\lim_{t \to \pm \infty} \frac{d}{dt} f(\gamma_x(t)) = -\lim_{t \to \pm \infty} \|\nabla f(\varphi_t(x))\|^2 = 0.$$

Take $t_n \in \mathbb{R}$ to be the sequence with $\lim_{n \to \infty} t_n = -\infty$. Since $\{\gamma_x(t_n)\} \subset M$ is an infinite set of points in a compact manifold, it has an accumulation point $p$. We have $\|\nabla f(\gamma_x(t_n))\| \to 0$ for $n \to \infty$ so $p$ must be a critical point of $f$. By Lemma 3.6 we can choose a closed neighborhood $U$ of $p$ with no other critical points contained in it. Now suppose $\lim_{t \to -\infty} \gamma_x(t) \neq p$. Then there is an open neighborhood $V \subset U$ of $p$ and a sequence $\tilde{t}_n \in \mathbb{R}$ with $\lim_{n \to \infty} \tilde{t}_n = -\infty$ such that $\gamma_x(\tilde{t}_n) \in U - V$. Thus, the sequence $\{\gamma_x(\tilde{t}_n)\}$ has an accumulation point in the compact set $U - V$, which, by the above argument, must be a critical point of $f$. This contradicts with our choice of $U$. We conclude that $\lim_{t \to -\infty} \gamma_x(t) = p$. A similar argument shows that $\lim_{t \to \infty} \gamma_x(t) = q$ for some critical point $q \in M$. \qed

One of our first significant results in Morse theory is a consequence of Proposition 3.8.

Theorem 3.10. Let $f : M \to \mathbb{R}$ be a smooth function on a finite dimensional smooth manifold with boundary. For all $a \in \mathbb{R}$, let

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

Let $a < b$ and assume that $f^{-1}([a, b])$ is compact and contains no critical points of $f$. Then $M^a$ is diffeomorphic to $M^b$.

A proof of this beautiful statement is given in [4]. The idea of the proof is the following: Since $f$ has no critical points on $f^{-1}([a, b])$, the gradient vector field $\nabla f$ does not vanish there. The vector field which takes the value $\nabla f / \|\nabla f\|^2$ on $f^{-1}([a, b])$ and vanishes outside a suitable neighborhood of $f^{-1}([a, b])$ has compact support and defines a global flow $\Psi_t : M \to M$. One can show that $\Psi_{b-a} : M \to M$ is, in fact, a diffeomorphism from $M^a$ to $M^b$.

Historically, the Morse inequalities are the next important result in Morse theory. We will only consider the inequalities in their weak form. A proof using supersymmetric quantum mechanics can be found in section 5.1.

Proposition 3.11 (Weak Morse Inequalities). Let $M_p$ be the number of critical points of a Morse function $f$ whose Morse index is $p$. Then

$$M_p \geq \beta_p,$$

where $\beta_p$ is the $p$-th Betti number.
3.3 Stable and Unstable Manifolds

In this section our main goal is to prove the Stable/Unstable Manifold Theorem for a Morse function. The idea is that we can define two submanifolds associated to a critical point: the stable and unstable manifolds. Intuitively, these are all the points which respectively come from or eventually reach the critical point after acting on our manifold with the gradient flow. The Stable/Unstable Manifold Theorem for Morse functions tells us that the tangent space at a critical point splits in two parts, identifying the stable and unstable manifold as smooth embeddings into the manifold. The proof of this important statement is quite involved and requires multiple non-trivial theorems and lemmas, for which we will not always write down explicit proofs to enhance the readability. Since [4] is followed closely, we will refer in those cases.

3.3.1 The Stable/Unstable Manifold Theorem

Let \( f : M \rightarrow \mathbb{R} \) be a smooth function on a finite dimensional compact smooth Riemannian manifold \((M,g)\). Recall that the gradient vector field determines a smooth flow \( \varphi : \mathbb{R} \times M \rightarrow M \) by \( \varphi_t(x) = \gamma_x(t) \) where \( \frac{d}{dt} \gamma_x(t) = -\nabla f|_{\gamma_x(t)} \) and \( \gamma_x(0) = x \). Since \( M \) is compact, \( \varphi_t \) is a 1-parameter group of diffeomorphisms defined on \( \mathbb{R} \times M \).

**Definition 3.12.** Let \( p \in M \) be a non-degenerate critical point of \( f \). The **stable manifold** of \( p \) is defined to be
\[
W_s(p) = \{ x \in M | \lim_{t \to \infty} \varphi_t(x) = p \}.
\]
The **unstable manifold** of \( p \) is defined to be
\[
W_u(p) = \{ x \in M | \lim_{t \to -\infty} \varphi_t(x) = p \}.
\]

We now formulate the main result of this section.

**Theorem 3.13** (Stable/Unstable Manifold Theorem for a Morse function). Let \( f : M \rightarrow \mathbb{R} \) be a Morse function on a compact smooth Riemannian manifold \((M,g)\) of dimension \( m < \infty \). If \( p \in M \) is a critical point of \( f \), then the tangent space at \( p \) splits as
\[
T_pM = T_p^sM \oplus T_p^uM
\]
where the Hessian is positive definite on \( T_p^sM \) and negative definite on \( T_p^uM \). Moreover, the stable and unstable manifolds are surjective images of smooth embeddings
\[
E^s : T_p^sM \rightarrow W^s(p) \subset M,
E^u : T_p^uM \rightarrow W^u(p) \subset M.
\]
Hence, \( W^s(p) \) is a smoothly embedded open disk of dimension \( m - \lambda_p \), and \( W^u(p) \) is a smoothly embedded open disk of dimension \( \lambda_p \), where \( \lambda_p \) is the index of the critical point.

A detailed proof of the above statement will be given in the following subsection. We first consider an important proposition and work out some examples.
Proposition 3.14. Let $f : M \to \mathbb{R}$ be a Morse function on a compact smooth closed Riemannian manifold $(M, g)$, then $M$ is a disjoint union of stable manifolds of $f$, i.e.

$$M = \bigsqcup_{p \in \text{Cr}(f)} W^s(p).$$

Analogously

$$M = \bigsqcup_{q \in \text{Cr}(f)} W^u(q).$$

Proof. By the existence and uniqueness theorem for ODE’s, every point $x \in M$ lies on a unique gradient flow line $\gamma_x$. By Proposition 3.9 every gradient flow line begins and ends at a critical point. This proves the statement.

Let us now consider some examples.

Example 3.15. Let

$$S^n = \{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} | x_1^2 + ... + x_{n+1}^2 = 1\}$$

be the $n$-sphere. Recall that the height function $f : S^n \to \mathbb{R}$ is Morse with two critical points, the north pole $N = (0, ..., 0, 1)$ and the south pole $S = (0, ..., 0, -1)$. The critical points $N$ and $S$ are of index $\lambda_N = n$ and $\lambda_S = 0$ respectively. Now with respect to the standard metric we have

$$W^u(N) = S^n - \{S\}, \quad W^s(N) = \{N\} \quad \text{and} \quad W^u(S) = \{S\}, \quad W^s(S) = S^n - \{N\}.$$  

Note that $W^u(N)$ is indeed diffeomorphic to an open disk of dimension $\lambda_N$, and $W^u(S)$ is diffeomorphic to an open disk of dimension $\lambda_S$.

Example 3.16. Now we will return to the torus $\mathbb{T}^2$ resting vertically on the plane $z = 0$ in $\mathbb{R}^3$. The height function $f : \mathbb{T}^2 \to \mathbb{R}$ is a Morse function. There are four critical points: the maximum $p$, the saddle points $q$ and $r$, and the minimum $s$ with indices $\lambda_p = 2$, $\lambda_q = \lambda_r = 1$, $\lambda_s = 0$. The unstable manifold $W^u(s)$ is just the critical point $s$, the unstable manifold
$W^u(r)$ is the circle through $r$ and $s$ minus the critical point $s$, the unstable manifold $W^u(q)$ is the circle around the hole in the middle minus the critical point $r$. Every other point lies in $W^u(p)$. Therefore, the torus can be written as

$$\mathbb{T}^2 = W^u(p) \bigsqcup W^u(q) \bigsqcup W^u(r) \bigsqcup W^u(s).$$

3.3.2 Proof of the Stable/Unstable Manifold Theorem

The proof of Theorem 3.13 will be given in this subsection. First we have to find local formulas for $\nabla f$, $d\nabla f|_p$, and $d\varphi|_p$.

**Lemma 3.17.** In the local coordinates $x_1, ..., x_m$ on $U \subset M$ we have

$$\nabla f = \sum_{i,j} g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j}.$$  

**Proof.** Choosing local coordinates $x_1, ..., x_m$ on $U$ we can write $\nabla f = \sum_j X_j \frac{\partial}{\partial x^j}$. Then, for any $j = 1, ..., m$ we have

$$\frac{\partial f}{\partial x^j} = g(\nabla f, \frac{\partial}{\partial x^j}) = \sum_i g_{ij} X_i.$$  

Thus,

$$(X_1 \cdots X_m)(g_{ij}) = \left(\frac{\partial f}{\partial x^1} \cdots \frac{\partial f}{\partial x^m}\right)$$

which gives

$$(X_1 \cdots X_m) = \left(\frac{\partial f}{\partial x^1} \cdots \frac{\partial f}{\partial x^m}\right)(g^{ij}).$$

Hence,

$$X_j = \sum_i \frac{\partial f}{\partial x^i} g^{ij}.$$  

\[\square\]
Lemma 3.18. If $x_1, \ldots, x_m$ is a local coordinate system around a critical point $p \in U \subset M$ such that $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^m}$ is an orthonormal basis for $T_p U$ with respect to the metric $g$, then the matrix for the differential of $\nabla f : U \to \mathbb{R}^m$ is equal to the matrix of the Hessian at $p$, i.e.
\[
\frac{\partial}{\partial \vec{x}} \nabla f \big|_p = H_p(f).
\]

Proof. First note that we have written $\frac{\partial}{\partial \vec{x}} \nabla f$ for the matrix representation of $d(\nabla f)$. In local coordinates $x_1, \ldots, x_m$ on $U$ we have
\[
\nabla f = \left( g^{ki} \frac{\partial f}{\partial x^k} \right)
\]
by Lemma 3.17. The matrix of the differential of $\nabla f$ can be computed as follows
\[
\frac{\partial}{\partial \vec{x}} \nabla f \big|_p = \left( \frac{\partial}{\partial x^j} g^{ki} \frac{\partial f}{\partial x^k} \right) = \left( \frac{\partial g^{ki}}{\partial x^j} \frac{\partial f}{\partial x^k} + g^{ki} \frac{\partial^2 f}{\partial x^j \partial x^k} \right).
\]
Therefore, at a critical point $p \in U$ we have
\[
\frac{\partial}{\partial \vec{x}} \nabla f \big|_p = \left( g^{ki} \frac{\partial^2 f}{\partial x^j \partial x^k} \right),
\]
so if $(\frac{\partial}{\partial x^i})_p$ is orthonormal at $p$, then
\[
\frac{\partial}{\partial \vec{x}} \nabla f \big|_p = \left( \frac{\partial^2 f}{\partial x^j \partial x^i} \right) = H_p(f).
\]

Lemma 3.19. If $x_1, \ldots, x_m$ is a local coordinate system around a critical point $p \in U \subset M$ such that $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^m}$ is an orthonormal basis for $T_p U$ with respect to the metric $g$, then for any $t \in \mathbb{R}$ the matrix for the differential of $\varphi_t$ at $p$ is equal to the exponential of minus the matrix of the Hessian at $p$, i.e.
\[
\frac{\partial}{\partial \vec{x}} \varphi_t \big|_p = e^{-H_p(f)t}.
\]

Proof. By the existence and uniqueness theorem for ODE’s we know that
\[
\varphi : (-\varepsilon, \varepsilon) \times U \to U
\]
is smooth in both coordinates for some neighborhood $U$ of $p$ and small $\varepsilon > 0$. Because $\varphi$ satisfies
\[
\frac{d}{dt} \varphi(t, x) = -(\nabla f)(\varphi(t, x))
\]
for any $x \in U$ we can interchange the order of differentiation giving us
\[
\frac{d}{dt} \frac{\partial}{\partial \vec{x}} \varphi(t, x) = -\frac{\partial}{\partial \vec{x}} (\nabla f)(\varphi(t, x)) = -\left( \frac{\partial}{\partial \vec{x}} \nabla f \right)(\frac{\partial}{\partial \vec{x}} \varphi(t, x)).
\]
Therefore, \( \Phi : \mathbb{R} \times TU \to TU \) defined by \( \Phi(t, x) = \frac{\partial}{\partial x} \phi(t, x) \) is a solution to the linear system of ODE’s given by
\[
\frac{d}{dt} \Phi(t, x) = -\left( \frac{\partial}{\partial x} \nabla f \right)(\Phi(t, x)), \quad \Phi(0, x) = I_{m \times m}.
\]
Because \( e^{-\left( \frac{\partial}{\partial x} \nabla f \right)t} \) is also a solution to the system, we have
\[
\Phi(t, x) = e^{-\left( \frac{\partial}{\partial x} \nabla f \right)t},
\]
since the solution is unique. Hence, at the critical point \( p \)
\[
\left. \frac{\partial}{\partial x} \phi(t, x) \right|_p = e^{-H_p(f)t},
\]
by Lemma 3.18.

If \( p \) is a non-degenerate critical point of index \( \lambda_p \), then there is a basis for \( T_p U \) such that \(-H_p(f)\) diagonal matrix with \( \lambda_p \) of the diagonal entries strictly negative, say \( \alpha_1, \ldots, \alpha_{\lambda_p} \), and \( m - \lambda_p \) of the diagonal entries strictly positive \( \beta_{\lambda_p+1}, \ldots, \beta_m \). Note that none of the entries are zero since the Hessian is non-degenerate. With respect to this basis we have the expression
\[
e^{-H_p(f)t} = \begin{pmatrix} e^{-\alpha_1 t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-\lambda_p t} \end{pmatrix}.
\]
We see that \( \lambda_p \) of the diagonal entries have length greater than one, and \( m - \lambda_p \) have length less than one when \( t > 0 \). Therefore, Lemma 3.19 implies that \( d\phi|_p : T_pM \to T_pM \) has no eigenvalues of length one for \( t \neq 0 \), i.e., \( p \) is a hyperbolic fixed point of \( \phi : M \to M \). Moreover, we have a splitting \( T_pM \cong T^s_pM \oplus T^u_pM \) where for \( t > 0 \)
\[
d\phi|_p : T^s_pM \to T^s_pM \text{ is contracting},
\]
\[
d\phi|_p : T^u_pM \to T^u_pM \text{ is expanding}.
\]
If \( t > 0 \) then the dimension of \( T^u_pM \) is \( \lambda_p \), and if \( t < 0 \) then the dimension of \( T^s_pM \) is \( \lambda_p \). The
next step is to show that a splitting as above gives rise to a submanifold structure on \( W^s(p) \) and \( W^u(p) \).

**Theorem 3.20** (Global Stable Manifold Theorem for a Diffeomorphism). If \( \varphi : M \to M \) is
a smooth diffeomorphism of a finite dimensional smooth manifold \( M \) and \( p \) a hyperbolic fixed
point of \( \varphi \), then
\[
W^s_p(\varphi) = \{ x \in M | \lim_{n \to \infty} \varphi^n(x) = p \}
\]
is an immersed submanifold of \( M \) with \( T_p W^s_p(\varphi) = T^s_pM \). Moreover, \( W^s_p(\varphi) \) is the surjective
image of a smooth injective immersion
\[
E^s : T^s_pM \to W^s_p(\varphi) \subset M.
\]
Hence, \( W^s_p(\varphi) \) is a smooth injectively immersed open disk in \( M \).
A proof of Theorem 3.20 which uses the Local Stable Manifold Theorem can be found in [4]. There one finds a proof of the Local Stable Manifold Theorem which is due to M.C. Irwin. Irwin’s proof is based on the Lipschitz Inverse Function Theorem and it describes the local stable manifold of a hyperbolic fixed point as the graph of a differentiable map between Banach spaces of Cauchy sequences [15]. It contains some fairly difficult real analysis. The Local Stable Manifold Theorem allows us to define the smooth immersion $E_s$ in the above statement whose construction is due to Smale [25]. The construction of $E_s$ is not obvious. If we replace $\phi$ by $\phi^{-1}$ in the preceding theorem, we arrive at an analogous result for unstable manifolds.

**Theorem 3.21** (Global Unstable Manifold Theorem for a Diffeomorphism). If $\phi : M \to M$ is a smooth diffeomorphism of a finite dimensional smooth manifold $M$ and $p$ a hyperbolic fixed point of $\phi$, then

$$W_u^p(\phi) = \{ x \in M \mid \lim_{n \to -\infty} \phi^n(x) = p \}$$

is an immersed submanifold of $M$ with $T_p W_u^p(\phi) = T_p^u M$. Moreover, $W_u^p(\phi)$ is the surjective image of a smooth injective immersion

$$E^u : T_p^u M \to W_u^p(\phi) \subset M.$$

Hence, $W_u^p(\phi)$ is a smooth injectively immersed open disk in $M$.

The Global Stable and Unstable Manifold Theorems show that the stable and unstable manifolds are the images of certain smooth injections. Since $\phi^n_t = \phi_{tn}$ for all $n \in \mathbb{N}$, it is clear that, when we take $\phi$ to be the gradient flow, for any critical point $p \in M$ and any fixed $t > 0$ we have $W_u^p(\phi_t) = \{ x \in M \mid \lim_{n \to -\infty} \phi^n_t(x) = p \} = \{ x \in M \mid \lim_{\lambda \to \infty} \phi_\lambda(x) = p \} = W^s(p)$. With the same reasoning $W_u^p(\phi_t) = W^u(p)$. Thus, Theorem 3.20 and 3.21 hold for the stable and unstable manifolds of the smooth gradient vector field. To finalize our proof we need the following lemma, which shows that if $M$ is compact the stable and unstable manifolds are smoothly embedded open disks.

**Lemma 3.22.** If $f : M \to \mathbb{R}$ is a Morse function on a finite dimensional compact smooth Riemannian manifold $(M, g)$ and $p$ is a critical point of $f$, then

$$E^s : T_p^s M \to W^s(p) \subset M$$

and

$$E^u : T_p^u M \to W^u(p) \subset M$$

are homeomorphisms onto their images.

**Proof.** We will only prove the lemma for $W^s(p)$ because the unstable manifold of $f$ is the stable manifold of $-f$. Moreover, $E^s$ is continuous because it is smooth. Hence, it suffices to prove $(E^s)^{-1} : W^s(p) \to T^s_p M$ is continuous where $W^s(p) \subset M$ has been given the subspace topology.

By Lemma 3.6 we can choose an open set $U \subset M$ containing $p$ such that it contains no other critical points. Since $E^s$ is a local diffeomorphism, we can choose an open set $V \subset T^s_p M$ around $0 \in T^s_p M$ such that $E^s(V) \subset U$ and $E^s|_V$ is a homeomorphism onto its image. Let
Proposition 3.24. Let \( x_j \in W^s(p) \) be a sequence converging to some point \( x \in W^s(p) \), and \( y_j = (E^s)^{-1}(x_j) \) and \( y = (E^s)^{-1}(x) \). Now suppose that \( y_j \) does not converge to \( y \) as \( j \to \infty \). For every \( t \in \mathbb{R} \), \( x_j' = \varphi_t(x_j) \) is a sequence that converges to \( x' = \varphi_t(x) \). For \( t \) sufficiently large, \( x^t \) and \( x_j^t \) are in \( U \) for all \( j \) sufficiently large. Write \( y_j^t = (E^s)^{-1}(x_j^t) \) and \( y^t = (E^s)^{-1}(x') \). The assumption that \( y_j \) does not converge to \( y \) implies that \( y_j^t \) does not converge to \( y^t \). Therefore, we must have \( \|y_j^t\| \to \infty \) if \( j \to \infty \). Otherwise, there would be a value of \( t > 0 \) such that \( y_j^t \in V \) for all \( j \) sufficiently large, and since \( E^s\big|_p \) is a homeomorphism we would have \( y_j^t \to y^t \) as \( j \to \infty \).

If \( \|y_j^t\| \to \infty \) if \( j \to \infty \), there is a subsequence of \( x_j^t \) that converges to a critical point \( q \) by Proposition 3.9 and Corollary 3.7. Since \( x_j^t \in U \) for all \( j \) sufficiently large an \( p \) is the only critical point in \( U \) we must have \( q = p \). This would give us a gradient flow line from \( p \) to itself, which contradicts Proposition 3.8. Hence, \( y_j^t \to y^t \) as \( j \to \infty \) and therefore, \( y_j \to y \) as \( j \to \infty \). \( \square \)

Now combining Theorem 3.20 and 3.21, the results on the hyperbolic nature of the critical points and the above lemma we have completed the proof of Theorem 3.13, the Stable/Unstable Manifold Theorem for a Morse function.

### 3.4 Morse-Smale Functions

Until now our main objects of study have been Morse functions on smooth Riemannian manifolds. Although they constitute a rich collection of functions for which we can define (un)stable manifolds, we need to impose a stronger requirement on the functions to make sure the (un)stable manifolds behave ‘nicely’. That is, we would like the whole of flow lines going from one critical point to another to behave like a submanifold. To achieve this we need some kind of transversality condition, which we have already mentioned in section 2.1.3. The statements below make these requirements precise.

**Definition 3.23.** A Morse function \( f : M \to \mathbb{R} \) on a finite dimensional smooth Riemannian manifold \((M, g)\) is said to satisfy the *Morse-Smale transversality condition* if and only if the stable and unstable manifolds of \( f \) intersect transversally, i.e.

\[
W^u(p) \cap W^s(q)
\]

for all \( p, q \in \text{Cr}(f) \). A Morse function satisfying the Morse-Smale transversality condition is called a *Morse-Smale function*.

From the Morse-Smale transversality condition we derive the following as an immediate consequence.

**Proposition 3.24.** Let \( f : M \to \mathbb{R} \) be a Morse-Smale function on a finite dimensional compact smooth Riemannian manifold \((M, g)\). If \( p, q \in \text{Cr}(f) \) such that \( W^u(p) \cap W^s(q) \neq \emptyset \), then \( W^u(p) \cap W^s(q) \) is an embedded submanifold of \( M \) of dimension \( \lambda_p - \lambda_q \).

**Proof.** By Theorem 3.13, \( W^u(p) \) and \( W^s(q) \) are smooth embedded submanifolds of \( M \) of dimension \( \lambda_p \) and \( m - \lambda_q \) respectively. Hence, by Lemma 2.9, \( W^u(p) \cap W^s(q) \) is an embedded
A submanifold of $M$ of dimension
\[ \dim W^u(p) + \dim W^s(q) - m = \lambda_p + (m - \lambda_q) - m = \lambda_p - \lambda_q. \]

This completes the proof.

For notational convenience we will denote $W^u(p) \cap W^s(q) = W(p, q)$. The corollary below tells us something important about the gradient flow lines connecting two critical points.

**Corollary 3.25.** Let $f : M \rightarrow \mathbb{R}$ be a Morse-Smale function on a finite dimensional compact smooth Riemannian manifold $(M, g)$. Then the index of the critical points is decreasing along the gradient flow lines, i.e., if $p, q \in \text{Cr}(f)$ with $W(p, q) \neq \emptyset$, then $\lambda_p > \lambda_q$.

**Proof.** If $W(p, q) \neq \emptyset$, then $W(p, q)$ contains at least one flow line from $p$ to $q$. Because gradient flow lines have dimension one we must have $\dim W(p, q) \geq 1$. 

Let us consider an example of a Morse-Smale function.

**Example 3.26 (The tilted torus).** Perversely enough, the main example to explain Morse Theory to the uninitiated, namely the torus $T^2$ resting vertically on the plane $z = 0$ in $\mathbb{R}^3$ with the standard height function $f : T^2 \rightarrow \mathbb{R}$, is not Morse-Smale! This can be easily seen from Corollary 3.25, since the standard height function has gradient flow lines that begin at the critical point $q$ of index 1, and end at the critical point $r$ of the same index. However, as we will see in a second, there exists an arbitrarily small perturbation of the standard height function on $T^2$ which is Morse-Smale. One way to visualize such a perturbation is to use the standard height function on a torus which has been slightly tilted. In a picture:

![Figure 3.5: Stable and unstable manifolds of the tilted torus $T^2$ with standard height function $f$.](image)

In the above example, we constructed a Morse-Smale function by applying a small perturbation to our height function. To prove that this construction is valid we need an important result about the ‘abundance’ of Morse-Smale functions. In 1963 both Kupka [17] and Smale [26] proved that that Morse-Smale gradient vector fields are in some sense generic. We will begin by giving a short definition of what we mean by “generic” after which we will state the Kupka-Smale Theorem without proof.
**Definition 3.27.** A subset $A$ of a topological space $X$ is called *residual* if and only if it is a countable intersection of open dense subsets of $X$, i.e.

$$A = \bigcap_{i=1}^{\infty} G_i,$$

where $G_i$ is open and dense in $X$ for every $i \in \mathbb{N}$. A subset of a topological space $X$ is called *generic* if and only if it contains a residual set. We call $X$ a *Baire space* if and only if every generic subset is dense in $X$.

**Theorem 3.28 (Kupka-Smale Theorem).** Let $(M, g)$ be a finite dimensional compact smooth Riemannian manifold, then the set of Morse-Smale gradient vector fields of class $C^k$ is a generic subset of the set of all gradient vector fields on $M$ of class $C^k$ for all $1 \leq k \leq \infty$.

For a detailed proof of this statement we refer to [4]. Using a homeomorphism induced by the Riemannian metric, we can identify the space of smooth gradient vector fields with $C^\infty_R(M)$ modulo the relation that two functions that differ by a constant are equivalent. Now the Kupka-Smale Theorem implies that the set of Morse-Smale functions is a generic subset of $C^\infty_R(M)$. Baire’s Theorem says that every complete metric space is a Baire space. Since $C^\infty_R(M)$ is a complete metric space [4], the Kupka-Smale Theorem tells us that the set of smooth Morse-Smale functions is dense in $C^\infty_R(M)$. This proves our claim in Example 3.26.

A next step in our treatment of Morse-Smale functions would be to prove the following corollary, since the result is necessary to show that the construction of the Morse-Smale-Witten complex in the next section is well defined.

**Corollary 3.29.** If $p$ and $q$ are critical points of relative index one, i.e. if $\lambda_p - \lambda_q = 1$, then

$$\overline{W(p,q)} = W(p,q) \cup \{p,q\}.$$ 

Moreover, $W(p,q)$ has finitely many components, i.e. the number of gradient flow lines from $p$ to $q$ is finite.

Corollary 3.29 is a corollary of the $\lambda$-lemma, one of the crucial theorems in the theory of smooth dynamical systems. A version of this theorem, as well as several important corollaries that are essential to Morse Homology, were announced in a paper by Smale [25] in 1960. However, the proofs of the theorem and the corollaries did not appear in print until 1969 when Palis, one of Smale’s students published his doctoral thesis [20]. We will not go into the $\lambda$-lemma here, since it would not be very useful given our context. Instead, we will try to give an idea of how the corollaries of the $\lambda$-lemma in Morse Homology lead to Corollary 3.29. A detailed proof of both the $\lambda$-lemma and the corollaries can be found in [4]. The main consequence we can draw from the $\lambda$-lemma in our context is that we have transitivity for the gradient flows. That is, if $p,q,r \in \text{Cr}(f)$ and $W(r,q), W(q,p) \neq \emptyset$ then $W(r,p) \neq \emptyset$. Moreover

$$\overline{W(r,p)} \supset W(r,q) \cup W(q,p) \cup \{p,q,r\}.$$ 

This allows us to define a partial ordering on the critical points of $f$: $q \succ p$ if and only if there exists a gradient flow line from $q$ to $p$. The set of critical points $\text{Cr}(f)$, together with
the partial ordering \( \succ \) is called a phase diagram of \( f \). Now, using only some basic topology and properties of the stable and unstable manifolds we arrive at a result which says that if \( p \succ q \) we have

\[
W(p, q) = W^u(p) \cap W^s(q) = \bigcup_{\hat{p} \succ \hat{q} \succ q} W(\hat{p}, \hat{q}),
\]

where the union runs over all critical points between \( p \) and \( q \) in the phase diagram. Suppose that \( p \) and \( q \) have relative index one. Since there are no intermediate critical points between \( p \) and \( q \) in the phase diagram of \( f \) (by Corollary 3.25), the above expression reduces to the statement that \( W(p, q) \cup \{ p, q \} \) is closed. Thus, \( W(p, q) \cup \{ p, q \} \subset M \) is compact since it is a closed subset of a compact space. The gradient flow lines from \( p \) to \( q \) form an open cover of \( W(p, q) \) and this can be extended to an open cover of \( W(p, q) \cup \{ p, q \} \) by including small open sets around \( p \) and \( q \). Since every open cover of a compact space has a finite subcover, the number of gradient flow lines from \( p \) to \( q \) is finite.

### 3.5 The Morse Homology Theorem

In this section we will construct the Morse-Smale-Witten complex, and present the Morse Homology Theorem, which states that the homology of this complex coincides with the singular homology. As already mentioned before, the Witten complex, as it is often referred to in the literature, has an interesting history. The story started with a Comptes Rendus Note of the French Academy of Sciences by René Thom in 1949 [27] and culminated with Witten’s paper in 1982 [30], where the boundary operator was explicitly written down for the first time. Quite unexpectedly, Witten arrived at this boundary operator through supersymmetric quantum mechanics. Giving a thorough investigation of the why’s and how’s of Witten’s discovery will be the main purpose of the next chapters. In this section, however, we will consider the complex from a purely mathematical point of view. We first discuss orientations on stable and unstable manifolds and define the Morse-Smale-Witten boundary operator. We then state the Morse Homology Theorem and give some computations of homologies using the chain complex. A rigorous proof of the Morse Homology Theorem will not be given since it goes beyond the scope of this thesis. In section 5.2 we have given Witten’s proof using supersymmetric quantum mechanics. However, Morse and Smale, whose work we have explained in the previous sections, had already found the ideas required to make rigorous Witten’s physicist’s proof. A formal proof due to Salamon using the Conley index can be found in [4].

Before we can go into the Morse Homology Theorem, we have to give some orientation conventions. Let \( (M, g) \) be a finite dimensional compact smooth oriented Riemannian manifold and let \( f : M \to \mathbb{R} \) be a Morse-Smale function. Recall that the tangent space at every critical point \( p \) splits as

\[
T_pM = T^s_pM \oplus T^u_pM,
\]

where \( T_pM \) has been given an orientation. We choose a basis of the vector space \( T^u_pM = T_pW^u(p) \) which gives us an orientation of \( T^u_pM \). By Remark 2.12 this determines an orientation of \( T^s_pM = T_pW^s(p) \). That is, the embedded submanifolds \( W^u(p) \) and \( W^s(p) \) have
orientations at $p$ compatible with the orientation of $M$ at $p$. For arbitrary $x \in W^u(p)$ we have an orientation on $T_xW^u(p)$ induced by the embedding $E^u: T_p^u M \to W^u(p)$ defined in the Stable/Unstable Manifold Theorem. Similarly, we have an orientation on $T_xW^s(p)$ for all $x \in W^s(p)$ determined by the embedding $E^s: T_p^s M \to W^s(p)$. These conventions are most easily understood from pictures which will be provided in the examples to come.

One last thing we have to define is the intersection number $n(p,q)$, which can be obtained from counting flow lines with sign. Consider two critical point $p$ and $q$ of relative index one, i.e. $\lambda_p - \lambda_q = 1$. Recall that the unstable manifold $W^u(p)$ and the stable manifold $W^s(q)$ intersect transversally. Assume $W(p,q) = W^u(p) \cap W^s(q) \neq \emptyset$, and let $\gamma: \mathbb{R} \to M$ be a gradient flow line from $p$ to $q$. At any point $x \in \gamma(\mathbb{R}) \subset W(p,q)$ we can complete $-(\nabla f)(x)$ to a positive basis $(-(\nabla f)(x), w_x^u)$ of $T_xW^u(p)$, providing the orientation of $W^u(p)$ at $x$. If we pick a positive basis $v_x^s$ of $T_xW^s(q)$, providing an orientation of $W^s(q)$ at $x$, then $(v_x^s, w_x^u)$ is a basis for $T_xM$. If $(v_x^s, w_x^u)$ is positive for the given orientation on $T_xM$ then we assign $+1$ to the the flow $\gamma$. Otherwise we assign $-1$ to the flow. By definition of the orientations on $W^u(p)$ and $W^s(q)$ in terms of the maps $E^u$ and $E^s$ it is clear that this assignment does not depend on the point $x \in \gamma(\mathbb{R})$. We write $n_{\gamma}$ for the sign associated to the flow $\gamma$ which can be either $+1$ or $-1$.

From Corollary 3.29 we know that $W(p,q) \cup \{p,q\}$ is a compact 1-dimensional manifold. It has action of $\mathbb{R}$ given by the flow. Therefore, $\mathcal{M}(p,q) = W(p,q)/\mathbb{R}$ is a compact zero-dimensional manifold. That is, it consists of a finite number of elements, one for each flow $\gamma$ from $p$ to $q$. We thus have $\#\mathcal{M}(p,q) = \text{“number of flow lines from } p \text{ to } q\text{”}$. To every flow $\gamma$ we have assigned a number $+1$ or $-1$ based on the orientation. The intersection number $n(p,q) \in \mathbb{Z}$ is defined to be the sum of these numbers. In formulas,

$$n(p,q) := \sum_{\gamma \in \mathcal{M}(p,q)} n_{\gamma}.$$  

**Definition 3.30** (Morse-Smale-Witten Complex). Let $f: M \to \mathbb{R}$ be a Morse-Smale functions on a smooth compact oriented Riemannian manifold $M$ of dimension $m < \infty$, and assume that orientations for the unstable manifolds of $f$ have been chosen. Let $C_k(f)$ be the free abelian group generated by the critical points of index $k$, i.e. $C_k(f) = \mathbb{Z} [\text{Crit}_k(f)]$, and let

$$C_* (f) = \bigoplus_{k=0}^m C_k(f).$$

Define the homomorphism $\partial_k : C_k(f) \to C_{k-1}(f)$ by

$$\partial_k(p) = \sum_{q \in \text{Crit}_{k-1}(f)} n(p,q).$$

The pair $(C_*(f), \partial_*)$ is called the Morse-Smale-Witten chain complex of $f$.

**Theorem 3.31** (Morse Homology Theorem). The pair $(C_*(f), \partial_*)$ is a chain complex and its homology is isomorphic to the singular homology $H_*(M)$.  

34
A nice feature of the Morse-Smale-Witten complex is that the boundary operator is described geometrically. If we can draw the space in question it is often straightforward to compute the homology using the Morse-Smale-Witten complex.

**Example 3.32 (The circle).** Consider $M = S^1$ oriented clockwise and the height function $f : S^1 \to \mathbb{R}$ in the following picture, where we have chosen to orient $T^u_p M$ from left to right as indicated. The gradient flow is downward, so $-\nabla f(x)$ agrees with the orientation of $T^u_x M$ when $x$ is on the right side and disagrees with the orientation of $T^u_x M$ when $x$ is on the left side. Thus, to make $(-\nabla f(x), w^u_x)$ a positive basis for $T^u_x M$ we need to take $w^u_x = +1$ when $x$ is on the right, and $w^u_x = -1$ when $x$ is on the left. If we give $T^u_q M = \{0\}$ the orientation +1, then the induced orientation on $T^s_q M$, and thus on $T^s_x W^s(q)$, for which we write $v^s_x$, agrees with the orientation of $S^1$. Therefore, $(v^s_x, w^u_x)$ is a positive orientation of $T^s_x M$ when $x$ is on the right and a negative orientation of $T^s_x M$ when $x$ is on the left. The flow on the right is assigned +1 while the flow on the left is assigned −1. Thus, $n(p,q) = 0$ and the Morse-Smale-Witten complex reads

$$0 \longrightarrow \mathbb{Z} \xrightarrow{0} \mathbb{Z} \longrightarrow 0,$$

which yields

$$H_k((C_\bullet, \partial_\bullet)) = \begin{cases} \mathbb{Z} & \text{if } k = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

as expected.

**Example 3.33 (The $n$-sphere).** Let

$$S^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} | x_1^2 + \ldots + x_{n+1}^2 = 1\}$$

be the $n$-sphere and define the height function $f : S^n \to \mathbb{R}$ by $f(x_1, \ldots, x_{n+1}) = x_{n+1}$. There are two critical points, the north pole $N = (0, \ldots, 0, 1)$ and the south pole $S = (0, \ldots, 0, -1)$ with index $\lambda_N = n$ and $\lambda_S = 0$ respectively. The Morse-Smale-Witten complex is given by

$$0 \longrightarrow C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \ldots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \longrightarrow 0.$$
When \( n = 0 \) it is clear that the boundary operators are all zero, and in Example 3.32 we have shown that \( \partial = 0 \) for \( n = 1 \). When \( n > 1 \) there are no points of relative index 1. Hence, for \( n \geq 0 \) the boundary operators in the Morse-Smale-Witten complex of the height function on \( S^n \) are all zero. The complex is given by

\[
\begin{align*}
0 & \to \mathbb{Z} \\
0 & \to 0 \\
0 & \to 0 \\
\cdots \\
0 & \to 0 \\
0 & \to \mathbb{Z} \\
0 & \to 0 
\end{align*}
\]

As a consequence

\[
H_k((C_\bullet, \partial_\bullet)) = \begin{cases} 
\mathbb{Z} \oplus \mathbb{Z} & \text{if } k = 0 \\
0 & \text{otherwise} 
\end{cases}
\]

when \( n = 0 \) and

\[
H_k((C_\bullet, \partial_\bullet)) = \begin{cases} 
\mathbb{Z} & \text{if } k = 0, n \\
0 & \text{otherwise} 
\end{cases}
\]

for all \( n > 0 \) as expected.

**Example 3.34** (The tilted torus). Consider the torus \( \mathbb{T}^2 \) and the Morse-Smale function \( f : \mathbb{T}^2 \to \mathbb{R} \) from Example 3.26. In Figure 3.8 the orientation chosen for \( \mathbb{T}^2 \) is indicated by a solid vector followed by a dashed vector. This gives the orientations on the tangent spaces of the critical points as seen in the figure. The orientations chosen for \( T^u_q \mathbb{M} \) and \( T^u_r \mathbb{M} \) are indicated by the dashed arrows pointing outwards from point \( q \) and \( r \) respectively. This gives an orientation on \( W^u(q) \) and \( W^u(r) \) which is indicated by the single arrows. Moreover, the induced orientations on \( W^s(q) \) and \( W^s(r) \), coming from the orientation convention \( T^u_q \mathbb{M} = T^u_q \mathbb{M} \oplus T^u_q \mathbb{M} \) and \( T^s_r \mathbb{M} = T^s_r \mathbb{M} \oplus T^s_r \mathbb{M} \), are also indicated by the single arrows. The orientation of \( W^s(s) \) agrees with the orientation of \( \mathbb{T}^2 \) when we choose the orientation +1 for \( T^u \mathbb{M} = \{0\} \).

The reader should now verify that these choices of orientation induce the signs for the gradient flow lines. To give an idea of how this should be done we perform one such a verification explicitly. From the picture we see that when \( x \in \gamma(\mathbb{R}) \subset W(p, r) \), where \( \gamma \) is the solid gradient flow line from \( p \) to \( r \), the vector \( -(\nabla f) (x) \) points downwards in the direction of the solid vector giving the first element of the orientation of \( T_x W^u(p) \). Hence, \( w^u_x \) can be taken as the dashed vector pointing to the left. In the picture we have chosen to orient \( W^u(r) \) in the direction of the dashed vector pointing outwards from point \( r \), and hence the induced
orientation on $W^s(r)$ is from left to right. Thus, if we pick a positive basis $v^s_x$ for $T_x W^s(r)$ we see that $v^s_x$ and $-(\nabla f)(x)$ have opposite direction, and therefore $(v^s_x, w^u_x)$ is a negative basis for $T_x M$. This shows that $n_\gamma = -1$ for this gradient flow line. The other signs are determined by similar arguments.

Since $p, q, r$ and $s$ have indices $\lambda_p = 2, \lambda_q = 1, \lambda_r = 1$ and $\lambda_s = 0$ respectively, the Morse-Smale-Witten complex is given by

$$0 \to \mathbb{Z} \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to 0.$$

Therefore

$$H_k((C_*, \partial_*)) = \begin{cases} 
\mathbb{Z} & \text{if } k = 0, 2 \\
\mathbb{Z} \oplus \mathbb{Z} & \text{if } k = 1 \\
0 & \text{otherwise}
\end{cases}$$

Note that we have treated here only the trivial examples where the boundary operators are zero. See for example section 5.2 for a discussion of the ‘deformed’ torus, where some of the boundary operators are non-zero. Also, one could consult [4] for Morse theory on Grassmann manifolds.

It is rather interesting to note that the construction of Morse homology generalizes to infinite dimensions. Floer homology is the infinite-dimensional analogue of Morse homology: the Morse-Smale functions on finite-dimensional manifolds are replaced by ‘functionals’ on infinite-dimensional manifolds. Its involvements have been crucial in recent achievements in symplectic geometry and especially for the proof of the Arnold conjecture [2].
4 Supersymmetry

In this chapter we will consider supersymmetry (SUSY) in its simplest form: a supersymmetric field theory of 0+1 dimensions, that is, a theory describing the motion of a particle in quantum mechanics. In quantum field theories supersymmetry relates particles with integer spin, the bosons, to particles with half-integer spin, the fermions. It serves as a possible extension of the standard model of elementary particles and it is one of the fundamental ingredients of string theory. The aim of this chapter is to give the reader a suitable background for understanding Witten’s 1982 paper on Supersymmetry and Morse Theory [30], to be discussed in chapter 5. In section 4.1, we will present some of the main qualitative features of supersymmetric quantum mechanics. In principle, only this section is required to understand most of Witten’s analysis. However, in [30] Witten describes a tunneling calculation which he does not perform explicitly. In section 4.2, 4.3 and 4.4 we will develop the machinery, formulated in the path integral formalism, to make this instanton calculation precise.

4.1 SUSY in 0+1 dimensions

A theory which stays invariant under an interchange of fermions with bosons is called supersymmetric. As we know from Noether’s theorem [23] we can associate to every symmetry of the Lagrangian a set of conserved charges. In the case of supersymmetry, these are Hermitian operators, often denoted by $Q$, which map fermions to bosons and vice versa

$$\text{fermion} \xleftarrow{Q} \text{boson}.\$$

More formally, a supersymmetry theory consists of a $\mathbb{Z}_2$-graded Hilbert space $\mathcal{H} = \mathcal{H}^b \oplus \mathcal{H}^f$, decomposed into the spaces of bosonic and fermionic states, together with a set of Hermitian supersymmetry operators $Q_i, i = 1, ..., N$ mapping $\mathcal{H}^b$ to $\mathcal{H}^f$ and vice versa. The operator $(-1)^F$ which distinguishes the fermionic and bosonic subspaces, i.e. $(-1)^F |\psi\rangle = |\psi\rangle$ for $|\psi\rangle \in \mathcal{H}^b$ and $(-1)^F |\psi\rangle = -|\psi\rangle$ for $|\psi\rangle \in \mathcal{H}^f$, must satisfy the supersymmetry condition

$$(-1)^F Q_i = Q_i (-1)^F.\$$

Moreover, we have a Hamiltonian operator $H$ generating the time translation which satisfies

$$HQ_i + Q_i H = 0.\$$

We have further conditions specifying the algebraic structure

$$H = Q_i^2, \quad Q_i Q_j + Q_j Q_i = 0 \quad \text{(for } i \neq j). \quad \text{(4.1)}$$

38
In the simplest case, where we have two supersymmetry operators $Q_1$ and $Q_2$, we can perform a change of basis

$$Q = \frac{1}{2}(Q_1 - iQ_2), \quad Q^* = \frac{1}{2}(Q_1 + iQ_2),$$

such that we are left with only one supersymmetry operator $Q$ and its adjoint $Q^*$ mapping $\mathcal{H}^b$ to $\mathcal{H}^f$ and vice versa. One easily checks that the supersymmetry algebra (4.1) reads

$$Q^2 = (Q^*)^2 = 0, \quad H = QQ + Q^*Q.$$

It is this form of the algebra that we will mostly use. At this point, one might wonder what the physical ramifications are of a theory with the above mathematical structure. One of the major implication of such a theory, as we will see in a bit, is the existence of so-called superpartners: for every elementary particle of the bosonic type, there exists a particle of the fermionic type with equal energy, and thus, since energy and mass are equivalent, with equal mass. Dually, every fermionic particle has a bosonic superpartner of the same energy and mass. This one-to-one correspondence between bosons and fermions, however, has not been observed in nature. So, for supersymmetry to play a role in nature it must be spontaneously broken.

A symmetry is said to be *spontaneously broken* when the vacuum $|0\rangle$ of the theory, the state of empty space, is *not* invariant under the symmetry. Although the equations of motion themselves are symmetric, the symmetry is hidden: there are no states which are left invariant by the symmetry. Consider for example the potential in Figure 4.1. It is clear that the potential itself is mirror symmetric. However, the ground state of the system, centered at one of the minima, is not. In defining the vacuum we have to *choose* between one of the minima. By putting, for example, the vacuum in the right well we force a spontaneous symmetry breaking.

In the case of supersymmetry, we would like to know if the vacuum is invariant under a supersymmetry transformation. In other words, we would like to know if the vacuum $|0\rangle$ is annihilated by the supersymmetry operators, i.e.

$$Q|0\rangle = Q^*|0\rangle = 0. \quad (4.2)$$

This might seem a little bit confusing. Obviously, the operator $Q$ does not leave $|0\rangle$ invariant, but maps it to 0, so why call $|0\rangle$ invariant if equation (4.2) is satisfied? Recall that $Q$ is the *generator* of the symmetry. Applying $Q$ to a state corresponds to applying an infinitesimal
supersymmetry transformation. Therefore, equation (4.2) tells us that $|0\rangle$ does not change under a ‘small’ supersymmetry transformation. To make this statement precise, we consider the unitary operator $e^{i\delta Q}$ generated by $Q$. Here $\delta$ is a small parameter. Whenever $|0\rangle$ satisfies equation (4.2), we have

$$e^{i\delta Q}|0\rangle = (1 + i\delta Q - \delta^2 Q^2 + ...) |0\rangle = |0\rangle,$$

which shows that the state is indeed invariant under the supersymmetry.

The special form of the supersymmetry algebra allows us to rewrite the condition for supersymmetry breaking. The supersymmetry algebra tells us that the Hamiltonian can be written as

$$H = QQ^* + Q^*Q.$$

This relation captures all strange features of supersymmetry. First of all, we note that the Hamiltonian is positive definite, since we have

$$\langle \psi | H | \psi \rangle = \langle \psi | QQ^*| \psi \rangle + \langle \psi | Q^*Q| \psi \rangle = \|Q^*| \psi \rangle\|^2 + \|Q| \psi \rangle\|^2 \geq 0.$$

As an immediate consequence the spectrum of $H$ is non-negative, i.e. the energy eigenvalues $E$ satisfy

$$E \geq 0.$$

This gives us a natural lower bound on the energy. If we have a state of zero energy, it must be the ground state of the system. Moreover, determining whether a state has zero energy is equivalent to determining whether a state is annihilated by $Q$ and $Q^*$, i.e.

$$H |\psi\rangle = 0 \iff Q|\psi\rangle = Q^*|\psi\rangle = 0.$$

This is, in fact, Proposition 2.20 from Hodge theory. From the above discussion we draw the important conclusion that

**the supersymmetry is broken if and only if the vacuum has energy $E > 0$.**

In the case that the vacuum does have zero energy $E = 0$ we say that the supersymmetry is *unbroken*. So what happens to the states with energy $E > 0$? The crucial observation is that the states of non-zero energy are paired by the operator $Q$. For every bosonic state $|\psi\rangle$ with energy $E > 0$ we have a fermionic state $Q|\psi\rangle$ of the same energy

$$H(Q|\psi\rangle) = Q(H|\psi\rangle) = Q(E|\psi\rangle) = EQ|\psi\rangle.$$

Thus, we get the following picture, see Figure 4.2. The only exception is given by a state of zero energy, since this is a state which is annihilated by the symmetry operators

$$Q|\psi\rangle = Q^*|\psi\rangle = 0.$$

This is the only state which does not have a supersymmetric counterpart and thus, this is the only state which is left invariant by the symmetry.
Summarizing the above discussion, one has a very straightforward procedure for determining whether supersymmetry is spontaneously broken or not: one must check if the vacuum energy is zero. In attacking this problem a physicist might be inclined to use perturbation theory: it is often much easier to work with a local approximation of the potential. However, as it turns out, having found that supersymmetry is unbroken in perturbation theory does not allow one to conclude that the supersymmetry is also unbroken in the exact system [28]. In this sense, supersymmetry is quite different from other symmetries. If in some approximation the minimum of the potential is zero, an arbitrarily small quantum effect, shifting the potential by a tiny amount, could lift the minimum to a small but non-zero value. Then, the supersymmetry would be spontaneously broken. See for example Figure 4.3. Thus, it is often a very delicate claim that, in a given theory, supersymmetry is not spontaneously broken. An approximate calculation including many effects showing that the vacuum energy is zero in a certain approximation always leaves open the possibility that even smaller effects that have been neglected could raise the ground state energy slightly above zero.

As an example, consider the double-well potential in Figure 4.4. We have already seen that the mirror symmetry in this system is spontaneously broken, but what about the supersymmetry? At first sight, the supersymmetry seems to be unbroken, since the potential allows states of zero energy. However, as it turns out, the exact spectrum does not contain states
of zero energy, even though two such states exist in all orders of perturbation theory [14].
The effects that break the symmetry here are non-perturbative, i.e. they do not show up in perturbation theory to any finite order. Intuitively, we could have guessed this, since in perturbation one expands around a zero of the potential which shows that the approximation is completely insensitive to the existence or non-existence of a second zero of the potential. Only a calculation which reveals the existence of more than one zero of the potential could possibly tell us something about the supersymmetry being broken or not. These calculations are so-called instanton calculations, in which the tunneling from one zero to another is evaluated. The non-perturbative effect, in which a quantum mechanical particle tunnels through a potential barrier, is often responsible for lifting the vacuum energy slightly above zero [10].
The result is a spontaneous breakdown of supersymmetry.

4.2 The Path Integral Formalism

In this section the path integral will be constructed along the lines of [9]. Let us begin with the time-independent Schrödinger equation

$$i\hbar \frac{\partial \psi(t)}{\partial t} = \hat{H}\psi(t),$$

which describes the dynamics of a quantum mechanical system. By formally integrating the Schrödinger equation we obtain an operator

$$U(t', t) = e^{-\frac{i}{\hbar}\hat{H}(t'-t)}$$

which generates the time translation. In some sense, this unitary operator tells us everything there is to know about our quantum system, i.e. if we know the state $|\psi(t)\rangle$ at a certain time $t$ we know it at all future times $t'$ by applying the time evolution to it

$$|\psi(t')\rangle = U(t', t)|\psi(t)\rangle.$$  

Determining the operator is equivalent to finding its matrix elements in a given basis. For example, we can use the coordinate basis defined by

$$\hat{x}|x\rangle = x|x\rangle.$$  

We write

$$U(t', x'; t, x) := \langle x'|U(t', t)|x\rangle = \langle x'|e^{-\frac{i}{\hbar}\hat{H}(t'-t)}|x\rangle$$

42
for the matrix element in this basis. Since it represents the probability amplitude for a particle to propagate between points \(x\) and \(x'\) in a time \(t' - t\), this is known as the \textit{propagator} of the theory. The purpose of the path integral is to give a workable expression for the propagator. Recall that the states \(|x\rangle\) constitute an orthonormal basis, i.e.
\[
\langle x|x' \rangle = \delta(x - x'), \quad \int dx |x\rangle \langle x| = 1.
\]
Similarly, the eigenstates of the momentum operator satisfying
\[
\hat{p}|p\rangle = p|p\rangle
\]
also define an orthonormal basis, i.e.
\[
\langle p|p' \rangle = \delta(p - p'), \quad \int dp |p\rangle \langle p| = 1
\]
The inner product of the coordinate and momentum basis states gives the matrix elements of the transformation between the two bases. In fact, they are the Fourier coefficients
\[
\langle x|p \rangle = \frac{1}{\sqrt{2\pi \hbar}} e^{\frac{i}px}.\]
We are now ready to derive a path integral representation for the transition amplitude \(U(t', x'; t, x)\). The basic idea due to Dirac was the following: divide the time interval \([0, t]\) into \(N\) equal chunks of length \(\Delta t\), such that \(t = N\Delta t\). Since we are interested in the continuum limit \(N \to \infty\), we may assume \(N\) to be sufficiently large. Then,
\[
e^{-\frac{i}{\hbar} \hat{H} \Delta t} = \left[ e^{-\frac{i}{\hbar} \hat{T} \Delta t} \right]^N.
\]
From now on we will restrict to a specific class of Hamiltonians having the form \(\hat{H} = \hat{T}(\hat{p}) + \hat{V}(\hat{x})\), where the kinetic term \(\hat{T} = \frac{\hat{p}^2}{2m}\) is quadratic. Note that \(\hat{T}\) only depends on \(\hat{p}\) while \(\hat{V}\) only depends on \(\hat{x}\). In expanding the above expression one has to be careful that the kinetic operator \(\hat{T}\) and potential operator \(\hat{V}\) do not necessarily commute. In fact
\[
e^{-\frac{i}{\hbar} \hat{H} \Delta t} = 1 - \frac{i}{\hbar} (\hat{T} + \hat{V}) \Delta t + \frac{(-i)^2}{2\hbar^2} (\hat{T} + \hat{V})^2 \Delta t^2 + \ldots = e^{-\frac{i}{\hbar} \hat{T} \Delta t} e^{-\frac{i}{\hbar} \hat{V} \Delta t} + \mathcal{O}(\Delta t^2).
\]
The next step in our calculation is the most well-known trick in all of mathematical physics: rewriting the unit operator \(\mathbb{1}\) in a more useful fashion. Using the completeness relation of the bases \(|x\rangle\) and \(|p\rangle\) we have for every \(n\) a resolution of the identity given by
\[
\mathbb{1}_n = \mathbb{1}_{x_n} \mathbb{1}_{p_n} = \int dx_n |x_n\rangle \langle x_n| \int dp_n |p_n\rangle \langle p_n| = \int dx_n dp_n |x_n\rangle \langle x_n| |p_n\rangle \langle p_n| = \int dx_n dp_n |x_n\rangle \langle p_n| \frac{e^{\frac{i}{\hbar} x_n p_n}}{\sqrt{2\pi \hbar}}.
\]
Combining the above and using the fact that $\hat{T}$ is diagonalized by the states $|p\rangle$ and $\hat{V}$ by the states $|x\rangle$, we have

$$\langle x|e^{-\frac{i}{\hbar}Ht}|x_i\rangle = \langle x|\left[ e^{-\frac{i}{\hbar}T\Delta t} e^{-\frac{i}{\hbar}V\Delta t} \right] \cdots e^{-\frac{i}{\hbar}T\Delta t} e^{-\frac{i}{\hbar}V\Delta t}|x_i\rangle$$

$$= \langle x|\mathbb{I}\left[ e^{-\frac{i}{\hbar}\hat{T}_{N-1}} \cdots e^{-\frac{i}{\hbar}\hat{V}} \right]|x_{N-1}\rangle$$

$$= \int \prod_{n=1}^{N} dx_n dp_n \frac{e^{\frac{i}{\hbar}x_n p_n}}{\sqrt{2\pi\hbar}} \langle x_f|x_N\rangle \langle p_N|e^{-\frac{i}{\hbar}\hat{T}(p_{N-1})\Delta t} e^{-\frac{i}{\hbar}V(x_{N-2})\Delta t}|x_{N-2}\rangle \cdots \langle p_1|e^{-\frac{i}{\hbar}\hat{V}(x_1)\Delta t}|x_1\rangle.$$  \hspace{1cm} \text{(4.3)}

At the second equality we have omitted the $\mathcal{O}(\Delta t^2)$ terms since they will disappear in the limit $N \to \infty$. Note that something remarkable has happened! The quantum mechanical operators $\hat{T}$ and $\hat{V}$ have been replaced by simple numbers $T(p_n)$ and $V(x_n)$, which are the eigenvalues of $\hat{T}$ and $\hat{V}$ respectively, i.e. $\hat{T}(p)|p_n\rangle = T(p_n)|p_n\rangle$ and $\hat{V}(x)|x_n\rangle = V(x_n)|x_n\rangle$. Except for the constant $\hbar$ all quantum mechanics has disappeared from the above expression. This is one of the great advantages of the path integral formalism. We can freely move the numbers $e^{-\frac{i}{\hbar}\hat{T}(p_n)\Delta t}$ and $e^{-\frac{i}{\hbar}V(x_n)\Delta t}$ to the front, which leaves us with some inner products of the form $\langle p_n|x_{n-1}\rangle$. Using the identities

$$\langle p_n|x_{n-1}\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}p_n x_{n-1}}, \quad \langle x_f|x_N\rangle = \delta(x_f-x_N)$$

we can rewrite (4.3) as

$$\langle x_f|e^{-\frac{i}{\hbar}\hat{H}t}|x_i\rangle = \int_{x_0=x_i}^{x_f=x_f} \prod_{n=1}^{N} \left[ \int dx_n \prod_{n=1}^{N} \frac{dp_n}{2\pi\hbar} e^{-\frac{i}{\hbar}\Delta t \sum_{n=0}^{N-1} \left(T(p_{n+1})+V(x_{n+2})-p_{n+1}x_{n+2}x_{n+1}\Delta t\right)} \right]. \hspace{1cm} \text{(4.4)}$$

Because $T(p) = \frac{p^2}{2m}$ we can readily evaluate the momentum integrals in (4.4). Completing the square leaves us with a simple Gaussian integral

$$\int \frac{dp_n}{2\pi\hbar} e^{-\frac{i}{\hbar}\Delta t \left( \frac{p_n^2}{2m} - p_n x_{n-1} \frac{x_n - x_{n-2}}{\Delta t} \right)} = \int \frac{dp_n}{2\pi\hbar} e^{-\frac{i}{\hbar}\Delta t \left( \frac{p_n^2}{2m} - 2m p_n x_{n-1} \frac{x_n - x_{n-2}}{\Delta t} \right)}$$

$$= \int \frac{dp_n}{2\pi\hbar} e^{-\frac{i}{\hbar}\Delta t \left( \frac{p_n - m \frac{x_n - x_{n-2}}{\Delta t}}{\frac{2m}{\Delta t}} \right)^2}$$

$$= \frac{1}{2\pi\hbar} \left( \frac{2\pi m\hbar}{\Delta t} \right)^{\frac{1}{2}} e^{-\frac{i m \Delta t}{\Delta t} \left( \frac{x_n - x_{n-2}}{\Delta t} \right)^2}$$

$$= \left( \frac{m}{2\pi i \hbar \Delta t} \right)^{\frac{1}{2}} e^{-\frac{i m \Delta t}{\Delta t} \left( \frac{x_n - x_{n-2}}{\Delta t} \right)^2}.$$

Substituting this back into the transition amplitude we obtain

$$\langle x_f|e^{-\frac{i}{\hbar}\hat{H}t}|x_i\rangle = \left( \frac{m}{2\pi i \hbar \Delta t} \right)^{\frac{N}{2}} \int \prod_{n=1}^{N-1} dx_n e^{-\frac{i}{\hbar}\Delta t \sum_{n=0}^{N-1} \left( \frac{m}{2} \left( \frac{x_{n+1} - x_n}{\Delta t} \right)^2 - V(x_n) \right)}.$$
a curve \( x(t) \) and in the exponent we recognize the definitions of the derivative of a curve and the Riemann integral

\[
\frac{x_{n+1} - x_n}{\Delta t} \sim \dot{x}(t_n), \quad \Delta t \sum_{n=0}^{N-1} \sim \int_0^t dt'.
\]

Finally, we have found the Lagrangian formulation of the path integral

\[
\langle x_f | e^{-i\frac{\hbar}{\tau} \hat{H}t} | x_i \rangle = \int D[x] e^{i S[x]},
\]

where \( S[x] = \int_0^t dt' \left( \frac{1}{2} m \dot{x}^2 - V(x) \right) \) is the classical action and

\[
\mathcal{D}[x] := \lim_{N \to \infty} \left( \frac{m}{2\pi i \hbar \Delta t} \right)^{\frac{N}{2}} \prod_{n=1}^{N-1} dx_n
\]

the path integral measure. Here we integrate over all paths \( x(t) \) satisfying \( x(0) = x_i \) and \( x(t) = x_f \). A quantum mechanical transition amplitude has been expressed in terms of a path integral over all possible paths through coordinate space, weighed by the classical action! This is Feynman’s “sum over histories” idea [12].

### 4.3 The Instanton Method

We have two approximation methods to analytically treat systems which cannot be solved exactly: perturbation theory and semiclassical analysis. However, as we have seen in the example of the double-well potential perturbation theory does not always suffice. By applying perturbation theory \textit{locally} around a minimum we miss important \textit{global} information about the connection of one minimum to another: the possibility of tunneling. This non-perturbative effect can be beautifully calculated via semiclassical trajectories in imaginary time called \textit{instantons}. The instanton method, which was developed by Coleman in [10], consists of two steps: a semiclassical approximation involving Gaussian integrals and a rotation to imaginary time involving Euclidean integrals.

#### 4.3.1 Semiclassical Limit

The goal of this subsection is to compute the leading contribution to the path integral in the formal limit \( \hbar \to 0 \). This boils down to solving a Gaussian path integral. The following elementary Gaussian integrals form an inspiration. Everyone knows that for \( \lambda \in \mathbb{R}_{\geq 0} \)

\[
\int dx \ e^{-\lambda x^2} = \sqrt{\frac{\pi}{\lambda}}.
\]

A slightly harder example is given by the \( n \)-dimensional Gaussian integral

\[
\int dx \ e^{-x^T A x} = (\pi)^{N/2} \det A^{-1/2}, \quad (4.5)
\]

where \( A \) is a positive definite real symmetric \( n \)-dimensional matrix and \( x \in \mathbb{R}^n \). To evaluate the above integral, we can rotate to an eigenbasis \( x \to O x \) which diagonalizes the matrix.
A = O^TDO, where O is orthogonal and D is diagonal. The Jacobian of this change of variable is detO = 1. The integral thus factorizes into a product of one-dimensional Gaussian integrals which gives a product over the eigenvalues \( a_i \) of \( A \), \[ \prod_{i=1}^{n} \frac{1}{\pi a_i} \]. Replacing the product by a determinant we find (4.5).

We would like to generalize the above computations to arbitrary path integrals. Let us consider a transition amplitude

\[ \langle x_f | e^{\frac{i}{\hbar}Ht} | x_i \rangle = \int \mathcal{D}[x] e^{\frac{i}{\hbar}S[x]} \]

When \( \hbar \to 0 \), the weight factor in the path integral, namely \( e^{\frac{i}{\hbar}S[x]} \), is a phase multiplied by a large quantity. Mathematically, it is therefore not surprising that the dominant contribution to the path integral arises from the paths near the ones which extremize \( S[x] \). From the principle of least action we know that these are precisely the classical trajectories. When \( \hbar \) is a small parameter, we can try to apply some sort of saddle point method by expanding the action around the classical trajectory [11]. Recall that the classical trajectory \( \bar{x}(t) \) satisfies the Euler-Lagrange equation

\[ \frac{\delta S[x]}{\delta x(t)} \bigg|_{x=\bar{x}} = 0. \]

Therefore, it corresponds to an extremum of the exponent in the path integral. We can expand the action around the classical trajectory

\[ x(t) = \bar{x}(t) + y(t). \]

Note that the Jacobian of the change of variable is trivial, i.e. \( \mathcal{D}[x] = \mathcal{D}[y] \). Also, \( y(t_i) = y(t_f) \), since the classical solution already satisfies the boundary conditions \( \bar{x}(t_i) = x_i \) and \( \bar{x}(t_f) = x_f \). We can Taylor expand \( S[\bar{x}(t) + y(t)] \) in \( y(t) \) using the usual definitions for functional derivatives [11]

\[ S[\bar{x}(t) + y(t)] = S[\bar{x}(t)] + \int dt \frac{\delta S[x]}{\delta \bar{x}(t)} y(t) + \frac{1}{2} \int dt dt_1 \int dt_2 y(t_1) \frac{\delta S[x]}{\delta \bar{x}(t_1)} \delta \bar{x}(t_2) \bigg|_{x=\bar{x}} y(t_2) + O(y^3) \]

\[ = S[\bar{x}(t)] + \frac{1}{2} \int dt dt_1 \int dt_2 y(t_1) \frac{\delta S[x]}{\delta \bar{x}(t_1) \delta \bar{x}(t_2)} \bigg|_{x=\bar{x}} y(t_2) + O(y^3), \]

where we have used that the classical path \( \bar{x}(t) \) minimizes the action. Plugging this back into the transition amplitude we obtain

\[ \langle x_f | e^{\frac{i}{\hbar}Ht} | x_i \rangle = \int \mathcal{D}[y] e^{\frac{i}{\hbar} \left( S[\bar{x}] + \frac{1}{2} \int dt dt_1 \int dt_2 y(t_1) \frac{\delta S[x]}{\delta \bar{x}(t_1) \delta \bar{x}(t_2)} \bigg|_{x=\bar{x}} y(t_2) + O(y^3) \right)} \]

\[ = N e^{\frac{i}{\hbar} S[\bar{x}]} \int \mathcal{D}[y] e^{\frac{i}{\hbar} \sum dt dt_1 \int dt_2 y(t_1) \frac{\delta S[x]}{\delta \bar{x}(t_1) \delta \bar{x}(t_2)} \bigg|_{x=\bar{x}} y(t_2) [1 + O(h)] } \]

\[ = \frac{N}{\sqrt{\det \left( \frac{\delta S[x]}{\delta \bar{x}(t_1) \delta \bar{x}(t_2)} \bigg|_{x=\bar{x}} \right)}} e^{\frac{i}{\hbar} S[\bar{x}] [1 + O(h)]}, \]
where $N$ is just the Jacobian of rescaling with $\hbar$. Its only role is to fix the right overall normalization of the propagator. Some comments are in order. First, note that one should be careful at the last equality of (4.6), since the second variation operator is not a matrix as was the case in (4.5). However, a straightforward generalization is possible when we diagonalize the operator using orthonormal eigenfunctions [11]. Secondly, we have to consider the approximation made by throwing away the $O(y^3)$ terms. This is precisely where we need that $\hbar$ is small. One can show that the $O(y^3)$ terms in the exponent lead to a $1 + O(\hbar)$ contribution to the propagator. See for example [21]. In the semiclassical limit the path integral thus reduces to a Gaussian. Thirdly, we have to make sense of the determinant of an operator as appears in the square root. Clearly, the last step in our evaluation breaks down if

$$\det \left( \frac{\delta S[x]}{\delta x(t_1)\delta x(t_2)} \bigg|_{x=\tilde{x}} \right) = 0.$$

Normally, this occurs when there is some symmetry present and in that case we have to be more careful in the evaluation of the path integral. The conclusion of the semiclassical analysis is that the quantum propagator of a particle can be decomposed into two parts:

- an exponent $e^{i\hbar S[x]}$ which is given by the classical trajectory $\tilde{x}(t),$
- a tedious determinant which is given by the quantum fluctuations $y(t)$ around the classical trajectory over which must integrated.

### 4.3.2 Euclidean Path Integral and Instantons

The fundamental tool in the instanton method is the imaginary time version of the path integral, the so-called Euclidean path integral. For simplicity, we will consider a theory of a particle moving in a potential in one dimension:

$$H = \frac{1}{2} \left( \frac{dx}{dt} \right)^2 + V(x).$$

The idea is that instead of computing the transition amplitude $\langle x_f | e^{-i\hbar H t} | x_i \rangle$ for real times $t \in \mathbb{R}$ we could well have computed the transition amplitude for imaginary times $t = -iT$ giving us

$$\langle x_f | e^{-\frac{1}{\hbar}HT} | x_i \rangle, \quad T \in \mathbb{R}_{\geq 0}. $$

One advantage of this procedure is that the wild oscillatory exponent becomes a well-behaved damped exponent, which is often easier to evaluate. It is straightforward to repeat the same splicing procedure as before and write the imaginary time amplitude as a path integral

$$\langle x_f | e^{-\frac{1}{\hbar}HT} | x_i \rangle = \int D[x] E e^{-\frac{1}{\hbar}S_E[x]},$$

(4.7)

where

$$S_E[x] = \int_{-T/2}^{T/2} dt \left[ \frac{1}{2} \left( \frac{dx}{dt} \right)^2 + V(x) \right],$$
is the Euclidean action and \( \mathcal{D}[x]_E \) denotes integration over all functions \( x(t) \), obeying the boundary condition \( x(-T/2) = x_i \) and \( x(T/2) = x_f \). This is done in [21]. We could interpret (4.7) as the ‘analytic continuation’ of the path integral to imaginary time. Notice that the proper time interval \( ds^2 = -dt^2 + dx^2 \) corresponding to the Minkowski metric gets mapped to \( ds^2_E = d\tau^2 + dx^2 \) which has Euclidean signature explaining the name Euclidean path integral.

The path integral in (4.7) is interesting because it can be readily evaluated at the semiclassical limit \( \hbar \to 0 \). The functional integral is then dominated by the stationary point of \( S_E \). For simplicity we will assume that there is only one stationary point \( \bar{x} \), i.e.

\[
\frac{\delta S_E[x]}{\delta x(t)} \bigg|_{x=\bar{x}} = -\frac{d^2 \bar{x}}{dt^2} + V'(\bar{x}) = 0,
\]

where the prime means differentiation with respect to \( x \). If there are more stationary points one has to sum over all of them. The next step in our calculation is to look for eigenfunctions of the second variational derivative of \( S_E \) evaluated at \( \bar{x} \):

\[
\frac{\delta S_E[x]}{\delta x(t_1) \delta x(t_2)} \bigg|_{x=\bar{x}} = -\frac{d^2}{dt^2} + V''(\bar{x}).
\]

That is, we take the functions \( x_n \) to satisfy

\[-\frac{d^2 x_n}{dt^2} + V''(\bar{x}) x_n = \lambda_n x_n,
\]

where the \( \lambda_n \) are the eigenvalues. From (4.6) we know that in the small-\( \hbar \) limit the integral becomes a product of Gaussians

\[
\langle x_f | e^{-\frac{1}{\hbar}HT} | x_i \rangle = Ne^{-\frac{1}{2} S_E[\bar{x}]} \prod_n \lambda_n^{-\frac{1}{2}} [1 + \mathcal{O}(\hbar)]
\]

\[= Ne^{-\frac{1}{2} S_E[\bar{x}]} \left[ \det \left(-\frac{d^2}{dt^2} + V''(\bar{x})\right) \right]^{-\frac{1}{2}} [1 + \mathcal{O}(\hbar)].
\]

Note that (4.8) is the equation of motion for a particle moving in a potential minus \( V \). That is,

\[
E = \frac{1}{2} \left( \frac{d\bar{x}}{dt} \right)^2 - V(\bar{x})
\]

is a constant of motion. We can use (4.9) to determine the solution \( \bar{x} \) to (4.8) by inspection.

As a simple example consider the double-well potential in Figure 4.5. The potential is even, \( V(x) = V(-x) \), and its minima are denoted by \( \pm a \). We attempt to compute the amplitudes

\[
\langle a | e^{-\frac{1}{\hbar}HT} | -a \rangle = \langle -a | e^{-\frac{1}{\hbar}HT} | a \rangle
\]

by approximating the path integral using its semiclassical limit. From Figure 4.5, which shows the inverted potential \(-V\), it is clear that there are at least two solutions to the classical Euclidean equations of motions: those in which the particle stays fixed on top of one or the other of the two hills. There is another solution, satisfying the boundary conditions, where the a particle begins at the top of one hill at time \(-T/2\) and moves to the top of the
Figure 4.5: Left: the double-well potential $V$ with minima at $a$ and $-a$. Right: the inverted double-well potential $-V$.

other at time $T/2$. Since we plan to take $T \to \infty$, we will focus on solutions in this limit, where the particle arrives at the tops at times plus and minus infinity. In this case, we are dealing with solutions to (4.9) with vanishing $E$, giving

$$\frac{dx}{dt} = \sqrt{2V(x)}.$$  \hfill (4.10)

Equivalently,

$$t = t' + \int_0^x dx' \sqrt{2V}.$$

The classical solution $\bar{x}(t)$ is sketched in Figure 4.6; it is called an instanton, according to Bott [8], since just as in the great scheme of things our lives take only an instant to live, these instantons stay put at $-a$ and $a$ most of the time and then whip from $-a$ to $a$ in an instant.

From (4.10) it is easy to derive an expression for the action of an instanton from $-a$ to $a$:

$$S_E[\bar{x}] = \int dt \left[ \frac{1}{2} \left( \frac{dx}{dt} \right)^2 + V(x) \right] = \int dt \left( \frac{dx}{dt} \right)^2 = \int_{-a}^a dx \sqrt{2V(x)}.$$

Thus, the tunneling amplitude is given by

$$\langle a | e^{-\frac{i}{\hbar}HT} | -a \rangle = N e^{-\frac{i}{\hbar}S_E[\bar{x}]} \left[ \det \left( -\frac{d^2}{dt^2} + V''(\bar{x}) \right) \right]^{-\frac{1}{2}} \left[ 1 + O(\hbar) \right] \propto e^{-\int_{-a}^a dx \sqrt{2V(x)}/\hbar},$$

which is the usual semiclassical tunneling coefficient. To get the full answer we should evaluate the determinant prefactor. This turns out to be a rather painful calculation due to a zero mode arising from a translation symmetry of the instantons. The answer can be found in [10].
4.4 Grassmann Variables and the Fermionic Path Integral

To use the instanton method in supersymmetry, one requires the notion of a fermionic path integral. The fermionic path integral is different from the usual one, developed in section 4.2, because it uses coordinate functions which depend not only on the usual coordinate variables but also on a set of anti-commuting variables, called Grassmann variables. These ‘odd’ variables would upon quantization produce fermions satisfying the Pauli exclusion principle, which gave them the name fermionic variables.

Definition 4.1. An n-dimensional Grassmann algebra $\mathcal{G} = \{\psi^i\}$ is formed by generators $\psi^i, i = 1, ..., n$, satisfying

$$\{\psi^i, \psi^j\} := \psi^i \psi^j + \psi^j \psi^i = 0.$$

The coefficients are taken to be in $\mathbb{C}$. In particular, a Grassmann variable $\psi$ squares to zero

$$\psi^2 = 0$$

suggesting already at a classical level the Pauli exclusion principle, which says that one cannot put two identical fermions in the same quantum state. Functions of Grassmann variables have a finite Taylor expansion. For example, an arbitrary function of two variables is given by

$$f(\psi_1, \psi_2) = f_0 + f_1 \psi_1 + f_2 \psi_2^2 + f_3 \psi_1 \psi_2.$$

Terms with an even number of Grassmann variables are called bosonic and terms with an odd number of variables are called fermionic. The rules of integration over Grassmann variables are different from ordinary variables and are defined by

$$\int d\psi = 0, \quad \int d\psi \psi = 1.$$

In the case of many Grassmann variables we have

$$\int d\psi^1 \cdots d\psi^n \psi^n \cdots \psi^1 = 1.$$

The integrals involving permutation of the $n$ variables are given by $\pm 1$ depending on the parity of the permutation. Let us consider some elementary anticommuting integrals. Since $e^{-a\psi^2} = 1$ one needs at least two Grassmann variables to have a nontrivial exponential function. For example

$$e^{-a\psi^1 \psi^2} = 1 - a \psi^1 \psi^2,$$

which gives the corresponding Gaussian integral

$$\int d\psi^1 d\psi^2 e^{-a\psi^1 \psi^2} = a.$$  

For complex Grassmann variables $\psi^i$ and $\bar{\psi}^i [5]$, $\lambda_i \in \mathbb{C}$ and toy action $S[\psi] = \sum_i \bar{\psi}^i \lambda_i \psi^i$ we have

$$\int \prod_i d\psi^i d\bar{\psi}^i e^{-S[\psi]} = \prod_i \int d\bar{\psi}^i d\psi^i e^{-\bar{\psi}^i \lambda_i \psi^i} = \prod_i \lambda_i.$$
If \( S[\psi] = \sum_{i,j} \bar{\psi}^{i} A_{ij} \psi^{j}, \) where \( A \) is a Hermitian matrix, we can reduce the integral to the calculation in the previous example by diagonalizing \( A \). Then
\[
\int \prod_{i} d\bar{\psi}^{i} d\psi^{i} e^{-S[\psi]} = \det A.
\]

Fermionic path integrals can be defined analogously to the usual path integral, but with integration over Grassmann fields \( \psi(t) \) and \( \bar{\psi}(t) \) \([5]\), i.e.
\[
\int \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] e^{iS[\psi]}.
\]
These are functions of time that take values in an infinite dimensional Grassmann algebra, \( \psi^{i} \sim \psi(t) \). Consider for example an action of the form \( S[\psi] = \int dt \bar{\psi}(t) A \psi(t) \). We ‘diagonalize’ the differential operator \( A \) working on Grassmann fields in terms of left and right eigenfunctions
\[
A \xi_{n} = \lambda_{n} \eta_{n}, \quad \bar{\eta}_{n} A = \lambda_{n} \bar{\xi}_{n},
\]
where \( \lambda_{n} \in \mathbb{R}_{\geq 0} \) and \( \xi_{n} \) and \( \eta_{n} \) are regular function, which are chosen to be orthonormal
\[
\int dt \bar{\xi}_{n}(t) \xi_{m}(t) = \int dt \bar{\eta}_{n}(t) \eta_{m}(t) = \delta_{nm}.
\]
The fermionic fields can be expanded in the ‘normal modes’
\[
\psi(t) = \sum_{n} \xi_{n}(t) a_{n}, \quad \bar{\psi}(t) = \sum_{m} \bar{\eta}_{m}(t) \bar{a}_{m},
\]
where \( a_{n} \) and \( \bar{a}_{m} \) are Grassmann variables. The integral becomes
\[
\int \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] e^{-S[\psi]} \int d\bar{a}_{0} d\bar{a}_{0} e^{-\sum_{i,j} \int dt \bar{\eta}_{i}(t) \bar{a}_{i} A \xi_{j}(t) a_{j}} = \prod_{n} \lambda_{n} = \det A.
\]
This is true only if \( A \) has no zero modes, where zero modes are defined as functions \( \xi_{0} \) and \( \bar{\eta}_{0} \) such that
\[
A \xi_{0} = 0, \quad \bar{\eta}_{0} A = 0.
\]
Here, \( \xi_{0} \) is a \( \psi \) zero mode and \( \bar{\eta}_{0} \) is a \( \bar{\eta} \) zero mode. If we have for example one \( \psi \) zero mode, then
\[
\psi(t) = \xi_{0}(t) a_{0} + \sum_{n \neq 0} \xi_{n}(t) a_{n}, \quad \bar{\psi}(t) = \sum_{n \neq 0} \bar{\xi}_{n}(t) \bar{a}_{n},
\]
which gives
\[
\int \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] e^{-S[\psi]} = \det A \int d\bar{a}_{0} \cdot 1 = 0,
\]
\footnote{Note that we are actually diagonalizing the positive Hermitian operator \( A \hat{A} \) using the basis of orthonormal eigenfunctions \( \eta_{n} \), i.e. \( A \hat{A} \eta_{n} = \lambda_{n}^{2} \eta_{n} \). We define \( \xi_{n} = (A \eta_{n})/\lambda_{n} \), implying \( A \xi_{n} = \lambda_{n} \eta_{n} \). See [1].}
where $\text{det}'$ denotes the adjusted determinant where the zero mode is discarded. However, if we plug in a linear fermi field, we get

$$
\int \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] e^{-S[\psi]} \psi(t) = \int \prod_{n \neq 0} da_n d\bar{a}_n da_0 \prod_{n \neq 0} (1 - \bar{a}_n a_n \lambda_n) \cdot (a_0 \xi_0(t) + \sum_{m \neq 0} a_m \xi_m(t))
$$

$$
= \xi_0(t) \cdot \prod_n \lambda_n = \xi_0(t) \text{det}' A. \quad (4.12)
$$

Thus, a fermionic insertion 'kills' a zero mode [1].

52
5 Witten on Morse Theory and Supersymmetry

This chapter will explore the first part of Witten’s 1982 paper on *Supersymmetry and Morse Theory*. Witten’s remarkable result will be considered that using supersymmetric quantum mechanics one can derive the classical Morse inequalities. He saw that supersymmetric quantum mechanics shows up in mathematics quite naturally: in the study of differential forms. This identification allows one to reduce the cohomology problem to the search for vacua of the supersymmetric Hamiltonian. The number of supersymmetric vacua is in fact a topological invariant! By looking at a deformed version of the de Rham complex, Witten found a one-to-one correspondence between approximate vacua in perturbation theory and critical points of a Morse function. However, to say something about the exact spectrum, i.e. the topology of our manifold, we need to take into account tunneling paths from one critical point to another. In doing so we are, in fact, giving a description of the Morse complex. In section 5.1 we will go into Witten’s deformation of the de Rham complex and his proof of the weak Morse inequalities. The main focus of section 5.2 will be Witten’s complex defined in terms of critical points and tunneling paths. A sketch will be given of the instanton calculation used to define the boundary operator.

5.1 Witten’s Proof of the Weak Morse Inequalities

Let $M$ be a Riemannian manifold of dimension $n$. Let $\Omega^p(M)$, $p = 0, ..., n$ be the space of $p$-forms on $M$. Let $d$ and $d^*$ be the usual exterior derivative and its adjoint. We consider the following example of a supersymmetric system. Take as a (pre-)Hilbert space the exterior algebra of differential forms

$$\mathcal{H} = \bigoplus_{p=0}^{n} \Omega^p(M),$$

and define the supersymmetry operators

$$Q_1 = d + d^*, \quad Q_2 = i(d - d^*).$$

Using the fact that $d^2 = d^* d^2 = 0$ we have the following supersymmetry relations

$$H = Q_1^2 = Q_2^2, \quad \{Q_1, Q_2\} = Q_1 Q_2 + Q_2 Q_1 = 0$$

where $H$ is the Hamiltonian of the system given by

$$H = dd^* + d^* d.$$
This Hamiltonian corresponds to the usual Laplacian $\Delta$ acting on forms. Including the operator $(-1)^p$, which counts the number of fermions modulo two, we have a set of operators satisfying the supersymmetry algebra. So how does the Hilbert space split under these symmetries? There is a natural way of interpreting the $p$-forms in $\mathcal{H}$ as bosonic or fermionic depending on whether $p$ is even or odd. Now the $Q_i$ map bosonic states into fermionic states and vice versa.

Instead of considering the system above, Witten decided to look at a slightly deformed version. Let $f : M \to \mathbb{R}$ be a Morse function and $t \in \mathbb{R}$. We define the ‘deformed’ exterior derivatives by

$$d_t = e^{-ft}d e^{ft}, \quad d_t^* = e^{ft}d^* e^{-ft}.$$  

It is clear that we still have $d_t^2 = d_t^* = 0$ so we can define

$$Q_{1t} = d_t + d_t^* \quad \text{and} \quad Q_{2t} = i(d_t - d_t^*).$$  

This gives us a deformed Hamiltonian

$$H_t = \Delta_t = d_t d_t^* + d_t^* d_t.$$  

As we will see in a second, Morse theory will make its appearance quite naturally. Since we have the following cochain complex

$$... \xrightarrow{d_t} \Omega^{k-1}(M) \xrightarrow{d_t} \Omega^k(M) \xrightarrow{d_t} \Omega^{k+1}(M) \xrightarrow{d_t} ...$$  

we are in a position to look at the $p$-forms $\psi$ which obey $d_t \psi = 0$ but cannot be written as $\psi = d_t \chi$ for any $\chi$. These are the $p$-th cohomology groups $H^p_t$ which a priori depend on the parameter $t$. However, as a short calculation shows, the mapping $\psi \mapsto e^{-ft} \psi$ is an invertible mapping from $p$-forms which are closed but not exact in the usual sense to $p$-forms which are closed but not exact in the sense of $d_t$. Therefore, it gives us a cochain map, which is an isomorphism of cochain complexes

$$... \xrightarrow{d} \Omega^{k-1}(M) \xrightarrow{d} \Omega^k(M) \xrightarrow{d} \Omega^{k+1}(M) \xrightarrow{d} ...$$  

This ensures that the $t$-dependence of the cohomology falls out. Now Theorem 2.21 from Hodge theory gives us a way of calculating the number of linearly independent forms in the cohomology by considering the number of zero eigenvalues of the deformed Laplacian $\Delta_t$ acting on forms. In formulas,

$$H^p(M) \cong H^p_t(M) \cong \ker(\Delta_t|_{\Omega^p}). \quad (5.1)$$  

The space of supersymmetric vacua is a topological invariant! Recall the definition of the Betti numbers $\beta_p$ as the dimension of the cohomology group

$$\beta_p = \dim H^p(M).$$  

54
The fundamental observation made by Witten was that for $t \to \infty$ we get the semiclassical limit for the operator $\Delta_t$. More concretely, the asymptotic estimates for the eigenvalues of $\Delta_t$, whose spectrum simplifies dramatically for large $t$, can be used to estimate the dimension of $\ker((\Delta_t)_|_{\Omega_p})$. As a result, we are able to place upper bounds on the Betti numbers $\beta_p$ in terms of the critical points of $f$.

It will be useful to introduce some notation. In local coordinates $(x^i)$ we write

$$a^{*i} \alpha = dx^i \wedge \alpha$$

for the exterior multiplication with $dx^i$ on $p$-forms. The adjoint operator $a^i$

$$a^i dx^{j_1} \wedge ... \wedge dx^{j_p} = \sum_{k=1}^p (-1)^k g^{i j_k} dx^{j_1} \wedge ... \wedge dx^{j_{k-1}} \wedge dx^{j_{k+1}} \wedge ... \wedge dx^{j_p}$$

corresponds to interior multiplication. It is straightforward to check that the operators satisfy

$$\{a^i, a^{*j}\} = g^{ij}.$$ 

The objects $a^i$ and $a^{*j}$ should remind you of the fermion creation and annihilation operators which arise in physics, since they map bosons into fermions and vice versa. We are now ready to give an expression for $\Delta_t$ in local coordinates.

**Proposition 5.1.** Choosing an orthonormal coordinate system $(x^i)$ in a flat neighborhood on $M$ we have

$$\Delta_t = \Delta + \frac{t^2}{2} \|df\|^2 + tA,$$

where $A$ is the $C^\infty(M)$-linear operator given by

$$A = \sum_{i, j} \left( \frac{\partial^2 f}{\partial x^i \partial x^j} \right) [a^{*i}, a^j].$$

**Proof.** We prove the proposition by a direct calculation. For $\alpha \in \Omega^p(M)$ we have

$$d_t \alpha = e^{-ft} d(e^{ft} \alpha) = e^{-ft} (e^{ft} \, d\alpha + e^{ft} \, df \wedge \alpha) = d\alpha + tdf \wedge \alpha,$$

where

$$df = \sum_i \frac{\partial f}{\partial x^i} \, dx^i.$$

Therefore,

$$d_t = d + t \sum_i \frac{\partial f}{\partial x^i} a^{*i}.$$

Hence,

$$d^*_t = d^* + t \sum_i \frac{\partial f}{\partial x^i} a^i.$$
This gives us the following expression for $\Delta_t$

$$\Delta_t = \{ d, d^* \} = \{ d + t \sum_i \frac{\partial f}{\partial x^i} a^i, d^* + t \sum_j \frac{\partial f}{\partial x^j} a^j \} = \{ d, d^* \} + l^2 \{ \sum_i \frac{\partial f}{\partial x^i} a^i, \sum_j \frac{\partial f}{\partial x^j} a^j \} + t \{ \sum_i \frac{\partial f}{\partial x^i} a^i, d^* \}$$

$$= \Delta + l^2 \sum_{i,j} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} \{ a^i, a^j \} + tA,$$

where

$$A = \sum_j \{ d, \frac{\partial f}{\partial x^j} a^j \} + \sum_i \{ \frac{\partial f}{\partial x^i} a^i, d^* \}.$$

What remains is to show that $A$ can be written as the operator in the proposition. Define $\partial_i$ by

$$\partial_i (\sum_{j_1,..j_p} u_{j_1..j_p} dx^{j_1} \wedge \ldots \wedge dx^{j_p}) = \sum_i \frac{\partial u_{j_1..j_p}}{\partial x^i} dx^{j_1} \wedge \ldots \wedge dx^{j_p}$$

so that we have the representation

$$d = \sum_i a^i \partial_i.$$

Similarly,

$$d^* = -\sum_i a^i \partial_i$$

where we have used that $\partial_i^* = -\partial_i$. We now calculate for every $j$

$$\{ d, \frac{\partial f}{\partial x^j} a^j \} = d \left( \frac{\partial f}{\partial x^j} a^j \right) + \frac{\partial f}{\partial x^j} a^j d$$

$$= \sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j + \frac{\partial f}{\partial x^j} d a^j + \frac{\partial f}{\partial x^j} a^j d$$

$$= \sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j + \sum_i \frac{\partial f}{\partial x^j} \{ a^i, a^j \} \partial_i$$

$$= \sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j + \sum_i g_{ij} \frac{\partial f}{\partial x^j} \partial_i$$

and

$$\{ \frac{\partial f}{\partial x^j} a^i, d^* \} = \frac{\partial f}{\partial x^j} a^i d^* + d^* \left( \frac{\partial f}{\partial x^j} a^i \right)$$

$$= \frac{\partial f}{\partial x^j} a^i d^* - \sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j + \frac{\partial f}{\partial x^j} d^* a^j$$

$$= -\sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j - \sum_i \frac{\partial f}{\partial x^j} \{ a^i, a^j \} \partial_i$$

$$= -\sum_i \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^j - \sum_i g_{ij} \frac{\partial f}{\partial x^j} \partial_i.$$
Adding both expressions and summing over \( j \) we find
\[
A = \sum_{i,j} \frac{\partial^2 f}{\partial x^i \partial x^j} a^{*i} a^j - \sum_{i,j} \frac{\partial^2 f}{\partial x^i \partial x^j} a^i a^{*j} = \sum_{i,j} \left( \frac{\partial^2 f}{\partial x^i \partial x^j} \right) [a^{*i}, a^j].
\]
This proves the statement. \( \square \)

When we look at the Hamiltonian \( \Delta_t \) in this form, the importance of the critical points becomes evident. For very large \( t \) the ‘potential energy’ \( V(x) = t^2 \| df \|^2 \) blows up, except in the vicinity of the critical points where \( df = 0 \). For this reason, when we take \( t \) to be large, the eigenstates of \( H_t \) are concentrated near the critical points of \( f \). In the process of giving a rough estimation for the eigenvalues it turns out, quite unexpectedly, that we are actually proving the weak Morse inequalities. A mathematician might complain that Witten’s proof in [30] skips over some of the details, so we have tried to give suitable references in those places.

**Proposition 5.2 (Weak Morse Inequalities).** Let \( M_p \) be the number of critical points of a Morse function \( f \) whose Morse index is \( p \). Then
\[
M_p \geq \beta_p,
\]
where \( \beta_p \) is the \( p \)-th Betti number.

**Proof.** First, by means of a Taylor expansion in the vicinity of a critical point with Morse index \( p \) one can introduce local coordinates \( (x_i) \) (chosen so that the critical point corresponds to \( x_i = 0 \)) giving us for the local coordinate expression of \( f \)
\[
f(x) = f(0) + \frac{1}{2} \sum_{i=1}^{n} \lambda_i x_i^2 + O(x^3). \tag{5.2}
\]
Here the \( \lambda_i \) are the eigenvalues of the Hessian matrix. The existence of such an expression is not evident and can be achieved by a suitable choice of metric \( g \) and an orthogonal rotation where the Hessian is diagonalized. See for example [22]. Plugging this expression into the local coordinate representation for the Hamiltonian, omitting the higher order terms, we get the approximation
\[
\tilde{H}_t = \Delta + t^2 \sum_{i=1}^{n} \lambda_i^2 x_i^2 + t \sum_{i=1}^{n} \lambda_i [a^{*i}, a^i]
\]
\[
= \sum_{i=1}^{n} \left( -\frac{\partial^2}{\partial x_i^2} + \lambda_i^2 t^2 x_i^2 + t \lambda_i [a^{*i}, a^i] \right).
\]
It will be quite easy to calculate the spectrum of the operator which appears above. This operator is
\[
\tilde{H}_t = \sum_{i=1}^{n} (H_i + t \lambda_i K_i),
\]
where

\[ H_i = -\frac{\partial^2}{\partial x_i^2} + \lambda_i^2 t^2 x_i^2, \quad K_j = [a^{*j}, a^j]. \]

First we note that \( H_i \) and \( K_j \) mutually commute so they can be simultaneously diagonalized. One easily recognizes \( H_i \) as the Hamiltonian of the one-dimensional harmonic oscillator. The eigenvalues are given by \( t|\lambda_i|(1 + 2n_i) \) for \( n_i = 0, 1, 2, \ldots \). So what can we say about the eigenvalues of the operator \( K_j \)? A quick calculation shows that

\[
[a^{*i}, a^i](f dx_{i_1} \wedge \ldots \wedge dx_{i_p}) = \begin{cases} f dx_{i_1} \wedge \ldots \wedge dx_{i_p} & \text{if } i \in \{i_1, \ldots, i_p\} \\ -f dx_{i_1} \wedge \ldots \wedge dx_{i_p} & \text{if } i \notin \{i_1, \ldots, i_p\} \end{cases}
\]

which gives us \( \pm 1 \) as eigenvalues for \( K_j \). Combining the above we arrive at the following set of eigenvalues for \( \tilde{H}_t \)

\[
t \sum_{i=1}^{n_i} (|\lambda_i|(1 + 2n_i) + \lambda_i m_i), \quad n_i = 0, 1, 2, \ldots \text{ and } m_i = \pm 1. \quad (5.3)
\]

So what can we say about possible zero eigenvalues? First we note that the spectrum above corresponds to \( \tilde{H}_t \) acting on the exterior algebra as a whole. However, we are interested in \( \tilde{H}_t \) working on \( p \)-forms. A moment’s thought about the operator \( K_i \) shows that in this case we must require the number of positive \( m_i \) to be equal to \( p \). Now for (5.3) to vanish we must set all \( n_i \) to zero and choose \( m_i \) to be +1 if and only if \( \lambda_i \) is negative. In other words, around every critical point \( \tilde{H}_t \) has precisely one zero eigenvalue, which is a \( p \)-form if the critical point has Morse index \( p \). All other eigenvalues are proportional to \( t \). Since the number critical points of index \( p \) is \( M_p \) we thus have

\[
\dim(\ker(\tilde{H}_t|\Omega_p))) = M_p.
\]

At this point one should recall that we have used an approximate Hamiltonian since we discarded the \( O(x^3) \) terms in the Morse function. So how good was our approximation? In other words, what can we say about the eigenvalues of \( H_t \) in terms of the eigenvalues of \( \tilde{H}_t \)? One can show that \( \tilde{H}_t \) is the first order approximation of \( H_t \) [30]. To make this statement precise one could write down an asymptotic expansion for the eigenvalues of \( H_t \) in \( t \). The exact verification of this approximation is rather technical and involves a semiclassical analysis of Schrödinger operators. This is done in [24].

For every critical point \( x \), \( \tilde{H}_t \) has just one eigenstate \( |\Psi\rangle \) which does not diverge with \( t \) and it is a \( p \)-form if \( x \) has Morse index \( p \). We will call \( |\Psi\rangle \) a perturbative ground state. However, as we have seen above, it is not necessarily the case that \( H_t \) annihilates all states \( |\Psi\rangle \), since we have only shown that the leading terms in perturbation theory vanish. But what we do know is that \( H_t \) certainly does not annihilate any of the other states, whose energy is proportional to \( t \) for large \( t \). So the number of zero energy \( p \)-forms cannot be larger than the number of critical points of Morse index \( p \). Thus, we have established the weak Morse inequalities

\[
M_p \geq \beta_p.
\]

This proves the statement. \( \square \)
5.2 The Witten Complex

In this section we will explain how Witten came up with the complex appearing in the Morse Homology Theorem. While the reasoning of [30] is followed closely, we have tried to elaborate on some of the details which are left out by Witten. In particular, we have tried to make the instanton calculation, which is crucial in defining the coboundary map, more precise. To avoid confusion, we remark that Witten works in the cohomology picture while we have developed Morse theory in homology framework. However, the de Rham’s Theorem, which we have discussed in the section 2.3, ensures that both outcomes are related.

In section 5.1 we have obtained the weak Morse inequalities from an approximate calculation of the spectrum of $H_t$, so we might hope to strengthen our inequalities by performing a more accurate calculation of the number of zero eigenvalues. One’s first thought might be to try to calculate the higher order terms in perturbation theory. However, as it turns out, all of the other terms in the asymptotic expansion vanish for the states with energies vanishing in lowest order [30]. To learn something new we must perform a calculation which is sensitive to the existence of more than one critical point. Because the ‘potential energy’ $V = t^2 \| df \|^2$ has more than one minimum we must allow for the possibility of tunneling from one critical point to another. From chapter 4 we know that the effect of tunneling can be calculated by means of an instanton calculation.

We begin by considering a particle propagating on a compact Riemannian manifold $(M, g)$. The position of the particle is denoted by $\phi : \mathbb{R} \to M$. Moreover, we have Grassmann fields $\psi$ and $\bar{\psi}$ tangent to $M$ which are complex conjugates of each other. In local coordinates $(x^i)$ we thus have the bosonic variables $\phi^i$ and fermionic variables $\psi^i, \bar{\psi}^i$. The dynamics of the supersymmetric sigma model is described by the action

$$S[\phi, \psi, \bar{\psi}] = \frac{1}{2} \int d\lambda \left[ \sum_{i,j} g_{ij} \left( \frac{d \phi^i}{d \lambda} \frac{d \phi^j}{d \lambda} + \bar{\psi}^i \frac{D \psi^j}{D \lambda} - i \frac{D \bar{\psi}^i}{D \lambda} \psi^j \right) + R_{ijkl} \psi^i \psi^j \psi^k \psi^l \right].$$

Here $\frac{D}{D\lambda}$ denotes a covariant derivative and $R_{ijkl}$ the metric curvature tensor of $M$. Both concepts will not be explained here since they go beyond the scope of this thesis. For simplicity, we will assume to be working in flat space, while remembering that a straightforward generalization to arbitrary manifolds is possible. In flat space matters simplify substantially:

$$S[\phi, \psi, \bar{\psi}] = \int d\lambda \left[ \sum_{i,j} g_{ij} \left( \frac{1}{2} \frac{d \phi^i}{d \lambda} \frac{d \phi^j}{d \lambda} + \bar{\psi}^i \frac{d \psi^j}{d \lambda} \right) \right].$$

It can be shown that the above action is invariant under supersymmetry transformations [13] which by Noether’s procedure gives conserved charges

$$Q = ig_{ij} \frac{d \phi^i}{d \lambda} \bar{\psi}^j, \quad Q^* = -ig_{ij} \frac{d \phi^i}{d \lambda} \psi^j.$$

One’s first question might be how the above action is related to the quantum mechanical system we have studied in section 5.1, which is described by the exterior algebra $\Omega^*(M)$. It turns
out that one can obtain the exterior algebra through canonical quantization of (5.5). This is done in [29]. The operators $\psi_i$ and their Hermitian conjugates $\bar{\psi}_i$ satisfy, after quantization, the algebra

$$\{\psi_i, \psi_j\} = \{\bar{\psi}_i, \bar{\psi}_j\} = 0, \quad \{\psi_i, \bar{\psi}_j\} = g^{ij}, \quad (5.6)$$

which makes them into the fermion annihilation and creation operators respectively. The supersymmetry charges become

$$Q = ip_i \bar{\psi}_i, \quad Q^* = -ip_i \psi_i,$$

where $p_i$ is the momentum conjugate to $\phi^i$. Quantization is not complete unless we specify the representation of the above algebra. So what do the quantum mechanical states of this system look like? We first consider the states for which the fermion Hilbert space is completely empty, i.e. the states that are annihilated by $\psi_i$. The wave function of such a state is an arbitrary function $\varphi(x^k)$ of scalar coordinates $x^k$. When we act on such a state with a fermion creation operator $\psi_i^*$, we get a state containing a single fermion of type $i$. We describe these states with a wave function $\varphi_i(x^k)$, with single index $i$ tangent to the manifold. States with two fermions $i$ and $j$, would be described by a wave function $\varphi_{ij}(x^k)$, which by Fermi statistics, must be antisymmetric with respect to the indices $i$ and $j$. A state of the system is described by a function $\varphi(x^k)$, a vector field $\varphi_i(x^k)$, a second-rank tensor field $\varphi_{ij}(x^k)$, a third-rank tensor field $\varphi_{ijm}(x^k)$ and so on until we reach an $n$-th antisymmetric tensor $\varphi_{i_1 \ldots i_n}(x^k)$, which describes a filled Fermi sea. This is precisely the description of the de Rham complex of $M$! When we identify

$$|0\rangle \leftrightarrow 1, \quad \bar{\psi}_i^* |0\rangle \leftrightarrow dx^i, \quad \bar{\psi}_i^* \bar{\psi}_j^* |0\rangle \leftrightarrow dx^i \wedge dx^j, \quad \ldots, \bar{\psi}_i^* \ldots \bar{\psi}_n^* \leftrightarrow dx^1 \wedge \cdots \wedge dx^n.$$

The operator $Q$ acts on a state with $q$ fermions by adding a new fermion and at the same time differentiating the old wave function, i.e. $Q$ is the exterior derivative $d$. On the other hand, $Q^*$ removes fermions making it the adjoint operator $d^*$. The Hamiltonian

$$H = QQ^* + Q^* Q = dd^* + d^* d$$

is the usual Laplacian acting on forms.

We can modify the action in (5.5) by adding a potential term constructed by the Morse function $f : M \to \mathbb{R}$ and a rescaling by $t$:

$$S[\phi, \psi, \bar{\psi}] = \int d\lambda \left[ \sum_{i,j} g_{ij} \left( \frac{1}{2} \frac{d\phi^i}{d\lambda} \frac{d\phi^j}{d\lambda} + \bar{\psi}_i^* \frac{d\psi^j}{d\lambda} \right) - \frac{1}{2} t^2 g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} - t \frac{\partial^2 f}{\partial x^i \partial x^j} \bar{\psi}_i^* \psi^j \right]. \quad (5.7)$$

This is the action of the supersymmetric non-linear sigma model in flat space. One immediately recognizes the potential $V(x) = \frac{1}{2} t^2 g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} = \frac{1}{2} t^2 \|df\|^2$. A quick calculation shows that the supersymmetry charges are represented as

$$Q_i = \left( ip_i + t \frac{\partial f}{\partial x^i} \right) \bar{\psi}_i, \quad Q_i^* = \left( -ip_i + t \frac{\partial f}{\partial x^i} \right) \psi_i,$$
which are exactly the coordinate expressions for $d_t$ and $d_t^*$ respectively,

$$d_t = \left( \partial_i + t \frac{\partial f}{\partial x^i} \right) a^i, \quad d_t^* = \left( -\partial_i + t \frac{\partial f}{\partial x^i} \right) a^i.$$  

The Hamiltonian of this system is thus given by

$$H_t = \{ Q_t, Q_t^* \} = d_t d_t^* + d_t^* d_t = \Delta_t.$$  

Setting $Q_1 = Q_t + Q_t^*$ and $Q_2 = i(Q_t^* - Q_t)$ we get the system discussed in section 5.1.

Summarizing the above, we find a beautiful correspondence between the physical and mathematical model: the supersymmetric sigma model in (5.4), after quantization, corresponds to the de Rham complex $\Omega^\bullet(M)$. In particular, states of $p$ fermions are $p$-forms and the supersymmetry charges $Q$ and $Q^*$ are the exterior derivative $d$ and its adjoint $d^*$. In [30], Witten had the crucial insight that by applying to the system a potential coming from a Morse function $f : M \to \mathbb{R}$ we are actually deforming the de Rham complex. After rescaling with a parameter $t$ this deformation is $d \mapsto d_t = e^{-f t} d e^{f t}$. As we have seen, for $t \to \infty$ the potential term blows up, which simplifies the spectrum of $H_t$ dramatically.

We are now ready to derive the instanton solutions from the action in (5.7). In doing so we will closely follow [13]. Recall that in section 5.1 we have found a perturbative ground state $\Psi_i$ for every critical point $x_i$ of $f$. Here $\Psi_i$ is a $\mu_i$-form when $x_i$ has index $\mu_i$. However, it is not necessarily the case that each $\Psi_i$ determines a supersymmetric ground state in the full theory. In other words,

$$Q \Psi_i = 0 \quad (5.8)$$

does not necessarily have to be satisfied. Although (5.8) holds to all orders in perturbation theory, in the full theory we should expect an expansion

$$Q \Psi_i = \sum_j \langle \Psi_j, Q \Psi_i \rangle \Psi_j. \quad (5.9)$$

Thus, we want to compute

$$\langle \Psi_j, Q \Psi_i \rangle = \int_M \Psi_j \wedge * (d + df \wedge) \Psi_i. \quad (5.10)$$

These are the non-perturbative corrections to the matrix elements of $Q$. In fact, they are amplitudes associated to tunneling paths between the critical points $x_i$ and $x_j$. If the tunneling amplitudes in (5.9) do not cancel, the sum gives a non-zero contribution, revealing the perturbative ground state $\Psi_i$ to be a state with non-zero energy in the exact system. Notice that, since $\Psi_j$ is a $\mu_j$-form and $Q \Psi_i$ is a $(\mu_i + 1)$-form, the above matrix element (5.10) is non-zero only if

$$\mu_j = \mu_i + 1.$$  

Thus, tunneling corrections are only between states with relative Morse index one.

We would like to rewrite the matrix elements $\langle \Psi_j, Q \Psi_i \rangle$ using the path integral formalism. Recall that in the limit of large $t$ the ground state wave functions are sharply peaked near
the critical points of \( f \), i.e. \( \Psi_i \) is an approximate delta function at \( x_i \) for large \( t \). In this limit the Morse function \( f \) may be viewed as an operator acting on the ground states by

\[
f \Psi_i = f(x_i) \Psi_i.
\]

This implies that

\[
\langle \Psi_j, Q \Psi_i \rangle = \frac{1}{f(x_i) - f(x_j)} \langle \Psi_j, (-f(x_j)Q + Qf(x_i))\Psi_i \rangle = \frac{1}{f(x_i) - f(x_j)} \langle \Psi_j, [Q,f] \Psi_i \rangle
\]

\[
e^{\frac{1}{f(x_i) - f(x_j)} \lim_{T \to \infty} \langle \Psi_j, e^{-TH} [Q,f] e^{-TH} \Psi_i \rangle}
\]

(5.11)

to leading order in \( 1/t \) [13]. Here we have used that in the commutator \([Q,f] = -fQ + Qf\) the operator \(-f\) on the left works on \( \Psi_j \), giving the factor \(-f(x_j)\) while the operator \( f \) on the right works on \( \Psi_i \), giving the factor \( f(x_i) \). For \( \Psi_i \) we can, in fact, take any function which has non-vanishing overlap with the \( i \)-th critical point and vanishes at all others, since the operator \( e^{-TH} \) for \( T \to \infty \) projects the function to the perturbative ground state at the \( i \)-th critical point. Something similar holds for \( \Psi_j \). Note that

\[
[Q,f] \omega = [d + df \wedge, f] \omega = (d + df \wedge) f \omega - f(d + df \wedge) \omega
\]

\[
= df \wedge \omega + f d \omega + df \wedge \omega - f d \omega - df \wedge \omega
\]

\[
= df \wedge \omega.
\]

Thus, \([Q,f] = \frac{\partial f}{\partial x^i} \tilde{\psi}^i\). So why did we bother to rewrite (5.10) in this fashion? If we interpret \( T \) as the Euclidean time, equation (5.11) describes a propagator sandwiching the operator \([Q,f]\). Hence, for \( T \to \infty \) this propagator can be calculated by the Euclidean path integral

\[
\lim_{T \to \infty} \langle \Psi_j, e^{-TH} [Q,f] e^{-TH} \Psi_i \rangle = \int \mathcal{D} \phi \mathcal{D} \psi \mathcal{D} \tilde{\psi} \tilde{\psi}^i \frac{\partial f}{\partial x^i} e^{-S_E[\phi,\psi]}.
\]

(5.12)

Here the integration region is the space of paths satisfying \( \phi(-\infty) = x_i \) and \( \phi(\infty) = x_j \). The Euclidean action \( S_E \) is obtained by replacing in (5.7) the time coordinate \(-i \tau\)

\[
S_E[\phi, \psi] = \int d\tau \left[ \sum_{i,j} g_{ij} \left( \frac{1}{2} \frac{d \phi^i}{d \tau} \frac{d \phi^j}{d \tau} + \tilde{\psi}^i \frac{d \psi^j}{d \tau} \right) + \frac{1}{2} t^2 g_{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} + t \frac{\partial^2 f}{\partial x^i \partial x^j} \tilde{\psi}^i \psi^j \right].
\]

(5.13)

The bosonic part of this action is given by

\[
S_B[\phi] = \frac{1}{2} \int d\tau \left[ \sum_{i,j} g_{ij} \frac{d \phi^i}{d \tau} \frac{d \phi^j}{d \tau} + t^2 g_{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} \right].
\]

Due to the special form of the potential it is fairly easy to show that the minimum action extrema of \( S_B \) with given initial and final conditions are the paths of steepest ascent. In fact, a straightforward manipulation shows that

\[
S_B[\phi] = \frac{1}{2} \int d\tau \left( \frac{d \phi^i}{d \tau} \pm t g^{ij} \frac{\partial f}{\partial x^i} \right)^2 \pm t \int d\tau \frac{df}{d\tau}.
\]

(5.14)
From (5.14) we see that for any trajectory
\[ S_B \geq t|f(\tau = +\infty) - f(\tau = -\infty)|, \]
with equality if and only if
\[ \frac{d\phi^i}{d\tau} \pm t g^{ij} \frac{\partial f}{\partial x^j} = 0. \]  \tag{5.15}
After rescaling this becomes the equation of steepest ascent
\[ \frac{d\phi^i}{d\tau} = g^{ij} \frac{\partial f}{\partial x^j}(\phi). \]  \tag{5.16}
These are the instantons. Here we have chosen the minus sign as to ensure that the path integral is non-vanishing only if \( h(x_j) > h(x_i) \). In the terminology of Morse theory, we see that the minimum action paths between two critical points \( x_i \) and \( x_j \) are the ascending\(^1\) gradient flow lines from \( x_i \) to \( x_j \). The action for such a path is
\[ I = t|f(x_j) - f(x_i)|. \]  \tag{5.17}
We are now interested in the fermionic part of the action. When we define the operator \( D_\pm \) working on Grassmann fields tangent to \( M \) [13]
\[ D_\pm \psi = \frac{d\psi^i}{d\tau} \pm t g^{ij} \frac{\partial f}{\partial x^j} \psi^k \]  \tag{5.18}
we see that the fermionic part of the action (5.13) can be written as
\[ S_F[\psi, \bar{\psi}] = \int d\tau g_{ij} \bar{\psi} \left( \frac{d\psi^i}{d\tau} + t g^{ij} \frac{\partial f}{\partial x^j} \psi^k \right) = \int d\tau g_{ij} \bar{\psi}^j D_+ \psi = - \int d\tau g_{ij} D_- \bar{\psi}^j \psi^j \]
where we used partial integration at the last equality. For the path integral in (5.12) to be non-vanishing, since there is a single insertion of \( \bar{\psi} \), the number of \( \bar{\psi} \) zero modes must be larger than the number of \( \psi \) zero modes by one. Compare this to (4.12). Thus, the path integral is non-vanishing if and only if
\[ \text{Ind} D_- := \dim \ker D_- - \dim \ker D_+ = 1, \]  \tag{5.19}
where \( \text{Ind} D_- \) is the index of the operator \( D_- \). One can show that the index satisfies
\[ \text{Ind} D_- = \Delta \mu, \]
where \( \Delta \mu = \mu_j - \mu_i \) is the relative Morse index. This is done in [13] by means of spectral flows of the Hessian. For any instanton from \( x_i \) to \( x_j \), (5.19) thus tells us that we must have \( \mu_j = \mu_i + 1 \).

We are finally ready to evaluate the path integral in (5.12). Changing the variable \( \phi = \gamma + \xi \), where \( \xi \) is the fluctuation around the instanton \( \gamma \), the action in the quadratic approximation reads
\[ S_E = t|f(x_j) - f(x_i)| + \int d\tau \left[ \frac{1}{2} |D_- \xi|^2 - D_- \bar{\psi} \psi \right]. \]
\(^1\text{Note that these are not equal to the gradient flow lines, but differ by a sign. They go up, instead of down.}\)
Because the kernel of $D_-$ is one-dimensional and there is no kernel of $D_+$ there is one bosonic $\xi$ zero mode, one fermionic $\bar{\psi}$ zero mode and no $\psi$ zero mode. The non-zero modes simply give the ratio of bosonic and fermionic determinants, i.e.

$$\frac{\det' D_-}{\sqrt{\det' D^* D_-}} = \frac{\det' D_-}{|\det' D_-|} = \pm 1,$$

which is the famous cancellation between bosons and fermions due to supersymmetry. Here we have combined the results from (4.6) and (4.11). One can show that the zero mode integrals give a contribution

$$\int_{-\infty}^{\infty} dt \int d\bar{\psi}_0 \bar{\psi}_0 \frac{d\gamma}{d\tau} \partial f \partial x^i (\gamma) = f(j) - f(i),$$

where $\gamma$ is the instanton centered at $\tau$. Intuitively, the bosonic zero mode corresponds to the translation symmetry of the instanton. Collecting the two and recovering the classical factor $e^{-I}$ we obtain the following expression for the contribution of the instanton $\gamma$ to the path integral in (5.12):

$$\pm (f(j) - f(i)) e^{-t(f(j) - f(i))}.$$

Summing over the instantons we obtain

$$\langle \Psi_j, Q\Psi_i \rangle = \sum_{\gamma} n_{\gamma} e^{-t(f(j) - f(i))},$$

where $n_{\gamma}$ is +1 or −1 depending on the instanton $\gamma$. Surprisingly enough, after the dust has settled the final solution to the path integral problem, because of the peculiar nature of supersymmetry, is extremely simple. Determining the sign $n_{\gamma}$ is most easily done as follows. The result in (5.20) shows that the integral $\int_M \Psi_j \wedge *Q\Psi_i$ receives dominant contributions along the paths of steepest ascent [30]. For each path $\gamma$, $n_{\gamma}$ is +1 or −1 depending on whether the orientation determined by $\Psi_j \wedge *Q\Psi_i$ along $\gamma$ matches with the orientation of $M$ or not. This condition can be translated in more familiar terms: the transport of a certain tangent frame along the tunneling path, which allows one to compare the orientation of the frame at $x_i$, to the frame at $x_j$ [3]. Working out the details, as is done in [30], it turns out that we are actually describing the algorithm for the determining the signs in the Morse complex.

Since the factor $e^{-I}$ does not depend on the instanton $\gamma$ we can take it out of the sum making it an overall factor in the amplitude. Only the relative signs $n_{\gamma}$ are important for determining if we have an exact zero energy vacuum. Therefore, we define the operator

$$\partial \Psi_i := \sum_{\mu_j = \mu_i + 1} \Psi_j \sum_{\gamma} n_{\gamma}.$$

This is the rescaled action of the supercharge $Q$ on the perturbative ground states. Because $Q$ is nilpotent, i.e. $Q^2 = 0$, it should also be nilpotent when acting on $\Psi_i$‘s. That is, $\partial^2 = 0$. Thus, if we define the graded space of perturbative ground states

$$C^k := \bigoplus_{\mu_i = k} \mathbb{R}\Psi_i,$$
we have a cochain complex \((C^\bullet, \partial^\bullet)\) with the coboundary operator given by the action of the supercharge
\[
0 \xrightarrow{\partial} C^0 \xrightarrow{\partial} C^1 \xrightarrow{\partial} \ldots \xrightarrow{\partial} C^n \xrightarrow{\partial} 0.
\]
The complex \((C^\bullet, \partial^\bullet)\) is called the Witten complex. The space of supersymmetric ground states is of course the cohomology of this complex. With equation (5.1) from Hodge theory we see that this cohomology is isomorphic to the de Rham cohomology. By reversing the arrows, we see that the Witten cochain complex, is dual to the Morse chain complex. From the de Rham’s theorem, we know that the Witten complex calculates the singular cohomology with real coefficients, which ‘proofs’ that the Morse complex calculates the singular homology. Of course, the above does not amount to a rigorous proof as we require in mathematics, but it gives at least some intuition for why Witten wrote down the construction to begin with.

**Example 5.3** (The tilted torus). Consider the torus \(T^2\) and the Morse-Smale function \(f : T^2 \to \mathbb{R}\) from Example 3.26. The function \(f\) induces a potential \(\frac{1}{2}t^2 \| df \|^2\) which blows up for \(t \to \infty\). For large \(t\), we have four *approximate* vacua, one for each critical point. In Figure 5.1 the instanton paths \(\gamma\) are given with corresponding signs \(n_\gamma\). Note that the paths have opposite direction to the ones in the Morse complex, since we are working in the cohomology framework. From Figure 5.1 it is clear that all tunneling amplitudes are zero. Compare to Example 3.34. This shows that every approximate vacuum is also a vacuum in the exact system. The Witten complex is thus given by
\[
0 \xrightarrow{\partial} \mathbb{R} \xrightarrow{\partial} \mathbb{R} \oplus \mathbb{R} \xrightarrow{\partial} \mathbb{R} \xrightarrow{\partial} 0.
\]

We conclude
\[
H^k((C^\bullet, \partial^\bullet)) = \begin{cases} 
\mathbb{R} & \text{if } k = 0, 2 \\
\mathbb{R} \oplus \mathbb{R} & \text{if } k = 1 \\
0 & \text{otherwise}
\end{cases}
\]

This is precisely the de Rham cohomology of the torus as expected.
Example 5.4 (The ‘deformed’ torus). Consider the ‘deformed’ torus in Figure 5.2 with height function $f$. We have not written down the instanton signs explicitly, since as it turns out most of them cancel. The only exception is given by the tunneling amplitudes $\langle \Psi_p|Q|\Psi_t \rangle$ and $\langle \Psi_u|Q|\Psi_t \rangle$ which turn out have opposite sign. The Witten complex is thus given by

$$0 \to \mathbb{R} \to \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \to \mathbb{R} \oplus \mathbb{R} \to 0,$$

where $\partial$ is

$$\partial(\Psi_r, \Psi_q, \Psi_t) = (\Psi_t, -\Psi_t).$$

Therefore

$$H^k((C^\bullet, \partial^\bullet)) = \begin{cases} \mathbb{R} & \text{if } k = 0, 2 \\ \mathbb{R} \oplus \mathbb{R} & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

as expected.

For a more ‘exotic’ example one could consult [19] where supersymmetric quantum mechanics on $SO(4)$ is examined.

As a final remark, we note that in our treatment of Witten’s paper [30] we have seen that supersymmetric quantum mechanics is just Hodge-de Rham theory. The real aim of [30] is however to prepare the ground for supersymmetric quantum field theory as the Hodge-de Rham theory of infinite-dimensional manifolds. This connection gives a rich collection of possibilities for further exploration.
6 Populaire Samenvatting

Topologie is de wiskunde van vormen. Lijnen, cirkels, driehoeken, sferen, piramides, donuts, gekromde vlakken: dit zijn allemaal objecten die je als wiskundige dagelijks tegenkomt in de topologie. Een topoloog is in het bijzonder geïnteresseerd in eigenschappen die hetzelfde blijven nadat de vorm wordt gekneed, uitgerekt of op een andere ‘mooie’ manier wordt vervormd. 

De vervormingen die zijn toegestaan noemen we continu, wat ongeveer betekent dat we niet mogen scheuren, knippen of plakken. De cirkel kan bijvoorbeeld worden uitgerekt in de vorm van een vierkant of driehoek of iets complexers als de omtrek van een handafdruk. Het is echter niet mogelijk om de cirkel continu vervormen tot een lijn. Daartoe zouden we de cirkel moeten knippen. Twee objecten die in dezelfde vorm kunnen worden gekneed noemen we homeomorf (van het Griekse woord homeo-, wat “gelijk” betekent, en morphe wat “vorm” betekent). Het blijkt in het algemeen erg lastig om te bepalen of twee vormen homeomorf zijn. Eén manier om iets te leren over de vorm is door te kijken naar bepaalde eigenschappen die voor homeomorfe vormen hetzelfde zijn, zogenaamde topologisch invarianten. Een voorbeeld van een topologisch invariant is het aantal losse stukken waaruit de vorm bestaat. Een ingewikkelder voorbeeld van een topologisch invariant, welbekend bij wiskundigen, is de fundamentaal groep. Het volume van een vorm is een voorbeeld van een eigenschap die geen topologisch invariant is: homeomorfe vormen kunnen best verschillende volumes hebben.

Figuur 6.1: Voorbeelden van vormen in het dagelijks leven.

In 1982 heeft de natuurkundige Edward Witten een artikel geschreven, waarin hij de natuurkunde en de wiskunde op een heel onverwachtse manier combineert: hij gebruikt resultaten uit de supersymmetrie (dit is natuurkunde) en Morse-theorie (dit is wiskunde) om een topologisch invariant te bedenken. Het blijkt dat de natuurkunde op deze manier heel natuurlijk opduikt in de topologie. Het doel van deze scriptie is om het artikel van Witten uit te pluizen en zo te laten zien op welke manier de natuurkunde heel erg belangrijk kan zijn voor de wiskunde.
Supersymmetrie is een natuurkundige theorie die gaat over elementaire deeltjes, de kleinste bouwstenen van het universum. Het blijkt dat we alle elementaire deeltjes kunnen opdelen in twee heel verschillende soorten: fermionen en bosonen. Fermionen zijn deeltjes waarmee je materie kunt maken. Bosonen zijn deeltjes die krachten over brengen. Een voorbeeld van een fermion is een elektron, aanwezig in atomen, en een voorbeeld van een boson is een foton, een lichtdeeltje. In het vervolg zullen we ons fermionen en bosonen voorstellen als balletjes van verschillende kleuren, zie Figuur 6.2.

Morse-theorie is een wiskundige theorie, die onderdeel is van de topologie. Het idee van Morse-theorie is dat we een vorm willen begrijpen door een functie te bekijken die aan ieder punt in de vorm een getal koppelt. Deze functie noemen we een Morse-functie. Mooi aan Morse-theorie is dat we bijzonder veel informatie over de vorm kunnen halen uit een aantal losse punten in de vorm: de kritieke punten van de Morse functie. Dit zijn de punten waar de afgeleide van de Morse functie gelijk is aan nul. Als belangrijke voorbeeld bekijken we de donut in Figuur 6.3. De Morse-functie $f$ is hier de hoogte functie, die aan ieder punt op de donut de hoogte van dat punt koppelt. We zien dat de Morse-functie vier kritieke punten heeft, aangegeven met $p, q, r$ en $s$. Voor ieder kritiek punt kunnen we de Morse index definiëren als het aantal richtingen waarin je vanuit het punt naar beneden loopt. Het punt $p$ bovenaan heeft index 2 omdat je in alle twee de richtingen, zowel langs het papier (van links naar rechts) als het papier in (van voor naar achter), naar beneden loopt. De punten $q$ en $r$ zijn zogenaamde zadelpunten, er is één richting waarin je naar beneden loopt (langs het papier) en één richting waarin je omhoog loopt (het papier in), dus die hebben beide index 1. Vanuit het onderste punt $s$ is er geen enkele richting waarin je naar beneden kunt lopen, dus de index van dat punt is 0.

Een manier om in te zien waarom de kritieke punten van de Morse-functie belangrijk zijn voor de topologie van de donut is door het volgende voor jezelf uit te proberen. Neem een kop koffie en een donut en laat de donut langzaam in de koffie glijden. We bekijken nu het
Figuur 6.4: Morse-theorie op de donut.

deel van de donut dat is ondergedompeld in de koffie. In Figuur 6.4 is te zien hoe de vorm van dit deel verandert naarmate een groter stuk van de donut in de koffie zit: de vorm begint als een punt, wordt dan een schijfje, vervolgens een cylinder, dan een donut met een gat in de bovenkant en eindigt als de volledige donut. We zien dus dat er alleen iets topologisch interessants gebeurt met de ondergedompelde vorm wanneer een kritiek punt de koffie raakt. Als we bijvoorbeeld tussen punt $s$ en punt $r$ zitten is de vorm homeomorf aan de schijf. Dit verandert pas wanneer we punt $r$ passeren. Dan wordt de schijf een cylinder. Dit is precies wat Morse-theorie ons vertelt: we kunnen de topologie van de vorm begrijpen door enkel te kijken naar de kritieke punten. Nu is de grote vraag of we een topologisch invariant kunnen vinden die voortkomt uit de Morse-theorie.

Witten gaf in 1982 een natuurkundige interpretatie aan de Morse-functie: hij zag de afgeleide van de Morse-functie als een potentiële, waarin deeltjes kunnen bewegen. De potentiaal kun je zien als een berglandschap op de donut, waar de balletjes over heen kunnen rollen. Merk op dat we voor ieder kritiek punt een potentiaalputje krijgen omdat de afgeleide van de Morse-functie daar gelijk is aan nul. In Figuur 6.5 zijn de putjes rond ieder kritiek punt aangegeven. Wanneer we deeltjes op de vorm laten bewegen dan komen een aantal van die deeltjes in de putjes terecht en wel op zo’n manier dat kritieke punten van even index een boson in hun putje hebben en kritieke punten van oneven index een fermion. Voor de donut hebben we vier deeltjes: twee fermionen en twee bosonen. Zie Figuur 6.5.
Nu komt Witten’s geniale inzicht: het totaal aantal deeltjes in de potentiaalputjes is in het algemeen een topologisch invariant! Om dit te illustreren bekijken we de ‘vervormde’ donut uit Figuur 6.6, die homeomorf is aan de gewone donut. Het deukje zorgt ervoor dat er nu zes kritieke punten zijn en je gaat gemakkelijk na dat ieder kritieke punt inderdaad de index heeft zoals aangegeven in de figuur. Rond ieder kritiek punt krijgen we een potentiaalputje met daarin een bijbehorende deeltje. Er lijkt nu op het eerste gezicht iets vreselijk mis te gaan: het totaal aantal deeltjes is niet vier, zoals bij de donut, maar zes! Zie Figuur 6.7. Het totaal aantal deeltjes lijkt geen topologisch invariant te zijn. Dit probleem is ontstaan omdat we niet helemaal zorgvuldig zijn geweest. De elementaire deeltjes die we bekijken zijn extreem klein en hun gedrag wordt beschreven door quantum mechanica. Binnen de quantum mechanica is het best mogelijk dat er tunneling plaatsvindt, het fenomeen waarbij een quantum deeltje door de ‘muur’ van een potentiaalputje heen tunnelt. Dit is precies het effect waar we rekening mee moeten houden. Het boson linksboven tunnelt naar het fermion in het putje eentje lager en zorgt ervoor dat beide deeltjes uit het potentiaalputje verdwijnen. Zoals te zien is in Figuur 6.7, hebben we opnieuw vier deeltjes. Dit blijkt in het algemeen te werken. Hoe je de donut ook continu vervormt, uiteindelijk blijven er altijd vier deeltjes over.
Kortom, we zien dat Witten met behulp supersymmetrie en Morse-theorie een eigenschap heeft gevonden die voor homeomorfe vormen hetzelfde is: het totaal aantal fermionen en bosonen in de putjes van de potentiaal. Als we deze eigenschap vertalen naar de wiskunde, dan krijgen we een van de belangrijkste topologisch invarianten die er bestaat: de cohomologie van een vorm. De vertaling gaat met behulp van differentiaalmeetkunde. Voor de wiskundigen onder u: de fermionen en bosonen, dat wil zeggen de balletjes in de voorgaande plaatjes, identificeren we met differentiaalvormen. De balletjes die zich in de putjes bevinden zijn nu harmonische differentiaalvormen, het zijn de deeltjes met energie nul. Nu is bekend dat iedere harmonische differentiaalvorm correspondeert met een cohomologieklasse. Op deze manier kunnen we de cohomolgie van een vorm bepalen door fermionen en bosonen te tellen: het aantal voortbrengers van de $k$-de cohomologie groep is gelijk aan het aantal fermionen en bosonen in de putjes behorende bij de kritieke punten van Morse index $k$.

Door met een natuurkundige bril naar een wiskundig probleem te kijken hebben we op een heel onverwachtse manier een oplossing gevonden. De samenwerking tussen natuurkunde en wiskunde blijkt op deze manier vaak enorm vruchtbaar te zijn. Het artikel van Edward Witten uit 1982 was een van de eerste keren dat topologie en theoretische natuurkunde elkaar hielpen, maar zeker niet de laatste. Met de recente ontwikkeling in bijvoorbeeld snaartheorie of topologische quantumveldentheorie is het einde van deze innige samenwerking nog lang niet in zicht.
Bibliography


73