THE EXTENDED TOPOLOGICAL QUANTUM FIELD THEORY OF THE FUKAYA CATEGORY IN YANG-MILLS THEORY

Maxim Jeffs

October 2016

A thesis submitted for the degree of Bachelor of Philosophy (Science)
of the Australian National University

ABSTRACT

The Atiyah-Floer conjecture links symplectic topology and low-dimensional geometry. It claims that the Donaldson-Fukaya categories of Atiyah-Bott moduli spaces describe the behaviour of gauge-theoretic invariants of 3- and 4-manifolds under gluing operations. This claim can be formulated as the existence of an extended topological quantum field theory arising from Yang-Mills theory in dimensions 2, 3 and 4. More precisely, the conjecture claims the existence of a ('natural') isomorphism between the instanton Floer homology $\mathrm{HF}_I(Y)$ of a homology 3-sphere Y, and the Lagrangian intersection Floer homology $\mathrm{HF}_L(L_1, L_2)$ of the two (generally immersed) Lagrangian submanifolds L_1, L_2 of the (symplectic) moduli space $\mathcal{M}_\Sigma$ of flat connections (over a Riemann surface Σ) arising by restriction from a Heegaard splitting $Y = Y_1 \cup_\Sigma Y_2$ of Y along Σ. Although $\mathcal{M}_\Sigma$ is a monotone symplectic manifold (whenever it is smooth) because L_1 and L_2 are immersed, the Lagrangian intersection Floer homology may fail to exist due to the appearance of anomalies. Using the obstruction theory of Fukaya-Oh-Ohta-Ono, and its extension to the immersed case by Akaho-Joyce [AJ10], whenever L_1 and L_2 are unobstructed, suitable bounding cochains b_{L_1}, b_{L_2} can be used to deform the boundary map for Lagrangian intersection Floer homology and hence define $\mathrm{HF}_L((L_1, b_{L_1}), (L_2, b_{L_2}))$. In his 2015 paper [Fuk15], Fukaya shows that this is indeed the case: the Lagrangians L_1, L_2 are unobstructed and moreover we have a canonical choice of bounding cochain. In this case, Fukaya claims

THEOREM 0.1. *(Fukaya, 2015) Whenever $\mathcal{M}_\Sigma$ is smooth, we have*

$$\mathrm{HF}_I(Y_1 \cup_\Sigma Y_2) \cong \mathrm{HF}_L((L_1, b_{L_1}), (L_2, b_{L_2}))$$

This thesis explains the above statement, defining the groups on both sides of this isomorphism in the case where all the relevant moduli spaces are transversal. In this case, one obtains a considerable simplification of the obstruction theory of Fukaya-Oh-Ohta-Ono when one uses instead a de Rham model of cohomology. Finally, we discuss how Fukaya claims to prove the above statement, and the various directions in which this result might be taken.

DECLARATION:

This thesis is my own work except where otherwise stated.

Maxim Jeffs

CONTENTS

ACKNOWLEDGEMENTS

The author wishes to thank: A/Prof. Bryan (Bai-Ling) Wang for supervising this thesis and numerous advanced courses over the past two and a half years; Dr. Brett Parker, Dr. David Smyth, Dr. Vigleik Angeltveit, and Prof. John Urbas for helpful conversations on subjects related to this thesis; the Australian Mathematical Sciences Institute for the award of a Vacation Research Scholarship that supported some of this work; the Mathematical Sciences Institute at the Australian National University for providing an office and funding for the author to attend the conference *Gromov-Witten theory, Gauge theory and Dualities* in January 2016, as well as for the award of a scholarship that partially supported the writing of this thesis.

... was mich nicht umbringt, macht mich stärker.

F. Nietzsche, Maxim 8, Götzen-Dämmerung (oder, Wie man mit dem Hammer philosophirt).

Chapter 1

INTRODUCTION: TOPOLOGICAL QUANTUM FIELD THEORY

The Atiyah-Floer conjecture, first posed by Sir Michael Atiyah in 1988, represents the final stage in the construction of the extended topological quantum field theory arising from Yang-Mills theory in dimensions 4,3 and 2. In full generality, the conjecture claims the existence of an isomorphism between the instanton Floer homology of a homology 3-sphere M, and the Lagrangian intersection Floer homology of the two (generally immersed) Lagrangian submanifolds of the (symplectic) moduli space of flat connections $\mathcal{M}_\Sigma$ arising from a Heegaard splitting $M = M_1 \cup_\Sigma M_2$ of M (see Figure 1.1). Despite Atiyah's famous heuristic argument ('stretch the neck') for the veracity of the conjecture, progress was slow in coming; of the several programmes announced in the early 1990s, only one, that of Kenji Fukaya, has approached a proof of the conjecture in full generality. After announcing his programme in a series of papers starting in 1992, Fukaya spent the next twenty years establishing the foundations of Lagrangian intersection Floer homology, leading to the 'monumental' work [FOOO09] with Yong-Geun Oh, Hiroshi Ohta and Kaoru Ono. In June 2015, Fukaya uploaded the first of a series of papers to the `arXiv` [Fuk15], claiming to give a proof of the SO(3) Atiyah-Floer conjecture under the hypothesis that $\mathcal{M}_\Sigma$ is a smooth manifold. 'Now it's the time to complete the project we started 20 years ago', writes Fukaya in his introduction; let us see, then, where this story begins.

The story begins with the gluing formulas for Donaldson invariants of 4-manifolds, conveniently formalised as a *topological quantum field theory* (henceforth a TQFT) [Don02, p.2]:

DEFINITION 1.1. *A **topological quantum field theory** in dimension n consists of two functions Z and H:*

- *H assigns a complex Hilbert space to every $(n-1)$-dimensional compact oriented manifold (without boundary);*
- *$Z(X)$ assigns a vector in $H(M)$ for every n-dimensional compact manifold X with boundary M. If $\partial X = \emptyset$, then we take $Z(X) \in \mathbb{C}$.*

satisfying the following axioms:

1. *$H(M_1 \sqcup M_2) = H(M_1) \otimes H(M_2)$;*
2. *$H(\overline{M}) = H(M)^*$, where $\overline{M}$ denotes the manifold with the opposite orientation;*
3. *If $\partial X = M \sqcup \overline{M} \sqcup N$ and X' is obtained by gluing M and $\overline{M}$ together, then $Z(X') = Z(X)^c$, the contraction of $Z(X) \in H(M) \otimes H(M)^* \otimes H(N)$.*

In the case of Yang-Mills theory, we take $n = 4$ and $Z(X)$ to be $\mathscr{D}(X)$, the (degree 0) Donaldson invariant of the 4-manifold X. The vector space $H(M)$ associated to a 3-manifold will be the Instanton Floer homology $\mathrm{HF}_{\text{inst}}(M)$; calculating these groups would, in theory, allow us to compute the Donaldson invariants of 4-manifolds via gluing operations. In general, however, the Instanton Floer homology groups are no easier to compute, but we might hope for a similar gluing formula where we glue 3-manifolds together along Riemann surfaces (see Figure 1.1). To motivate the form this should take, observe that the previous definition could be rewritten to say that a topological quantum field theory is a *tensor functor* from the category of $(n-1)$-manifolds with respect to cobordism to the category of complex vector spaces; by axioms 1 and 2 a cobordism X between $(n-1)$-manifolds M_1 and M_2 yields a linear map $\xi_X : H(M_1) \to H(M_2)$ and by axiom 3, composition of cobordisms corresponds to composition of linear maps. This suggests that one might define a *1-extended topological quantum field theory* to be a 2-tensor functor from the '2-category of $(n-2)$-manifolds with respect to cobordisms' to the 2-category of (cocomplete) linear categories. Even without formulating the definition of the above cobordism category precisely (cf. [Lur09]), one can see heuristically that to every $(n-2)$-dimensional

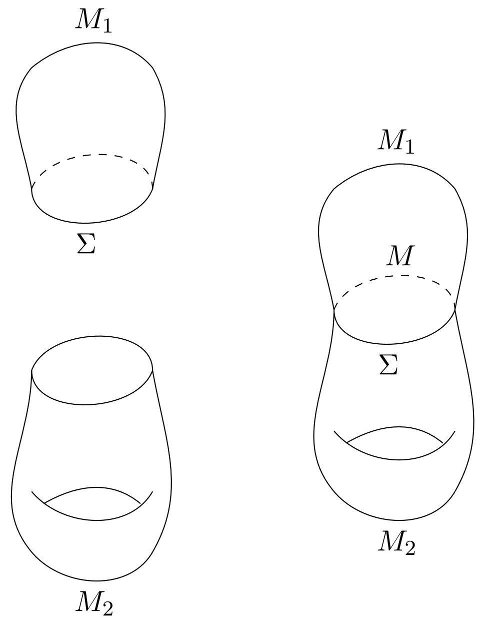

Figure 1.1: A Heegaard splitting of a 3-manifold M along a Riemann surface Σ. Here M_1, M_2 are handlebodies.

manifold Σ one should assign a linear category $\mathscr{C}(\Sigma)$ such that

1. If M is a compact oriented $(n-1)$-manifold with boundary Σ, then there is a corresponding object $\mathcal{R}(M)$ of the category $\mathscr{C}(\Sigma)$;
2. If $\partial M_1 = \Sigma = \overline{\partial M_2}$ for $(n-1)$-manifolds M_1, M_2, and M is obtained by gluing M_1, M_2 together along Σ, then $H(M)$ should be isomorphic to $\mathrm{Hom}_{\mathscr{C}(\Sigma)}(\mathcal{R}(M_1), \mathcal{R}(M_2))$ in some coherent manner.

In the case of Yang-Mills theory, this means associating a linear category $\mathscr{C}(\Sigma)$ to any Riemann surface Σ. But, to any Riemann surface Σ with a specified SU(2) vector bundle, one (ideally) has a naturally associated symplectic manifold, the Atiyah-Bott moduli space $\mathcal{M}_\Sigma$. Moreover, to any symplectic manifold, one can naturally associate many kinds of linear 'Lagrangian' categories, called under the general name of *Fukaya categories*. With these choices, as we shall see, the second axiom above becomes the Atiyah-Floer conjecture, suggesting that this is indeed the correct definition to use to construct an extended topological quantum field theory from Yang-Mills theory.

Making the above discussion rigorous, however, is where the difficulty lies; many of these difficulties are yet to be surmounted and some are possibly insurmountable. The above discussion, therefore, should largely be considered heuristic. We shall return to these technical difficulties in Chapter 6; now we proceed to give a chapter-wise outline of the contents of this thesis, indicating those parts that are new.

1.1 OUTLINE OF THE THESIS

The author himself makes no claims to originality except in exposition. Nowhere have the symplectic and gauge-theoretic Floer homology theories been treated in parallel, with an emphasis on the techniques and ideas they have in common. Nor has an exposition of the recent results in [Fuk15] yet appeared. Some minor points of exposition, however, are entirely original, and are described below.

Chapter 2 provides the analytic foundations for the rest of the thesis, proving the crucial compactness results for the pseudoholomorphic curve equation and the anti-self-dual equation. These results are treated in great detail in many standard textbooks ([MS12], [DK90], [Don02]); rather than proving these results in complete generality, we sketch the main idea in the simplest possible cases and then state the general result that will be required later. In most cases, these results can be developed entirely in parallel between these two theories, an idea which seems to be implicit throughout the literature. The idea of using the theory of harmonic maps to simplify the proofs of the main compactness results for pseudoholomorphic curves is the author's own, as is the proof of the Gromov-Uhlenbeck compactness theorem in the special case of harmonic maps from the Riemann sphere. In presenting the analysis in this thesis, we have attempted to imitate the acclaimed style of S. K. Donaldson (see for instance [DK90] [Don02]) by giving the central ideas rather than the often tedious details of the proofs.

Chapter 3 discusses the general formalism for the construction of moduli spaces in differential geometry, *Fredholm systems*, and proceeds to give the standard examples coming from gauge theory and symplectic geometry. We also discuss compactification and orientation problems for these moduli spaces. After this chapter, all moduli spaces will be assumed to be transversal. Finally, we motivate the notion of virtual integration on moduli spaces following [CLW14]; we hope that one day this may be used to remove the transversality assumptions used in this thesis. The author believes that no general introduction to moduli theory in differential geometry can be found elsewhere. Nor is a rigorous and complete construction of the Atiyah-Bott moduli space of flat connections to be found anywhere in the literature. Hence much of the material in 3.3 and 6.2 is original, though it is largely a straightforward modification of the instanton moduli theory.

Chapter 4 is similar, developing instanton Floer homology and Lagrangian intersection Floer homology in parallel, emphasising the analogies with the finite-dimensional case of Morse homology. All of this material is entirely standard; the author believes that the presentation here clarifies many points that are unclear in [Don02] and [FOOO09]. Note, however, that we do not claim (nor aim) to give a completely rigourous and systematic account of Floer theory in this thesis. Such a development would be enormously lengthy and can be found elsewhere; in Chapter 4 we aim primarily to develop the tools required to explain and motivate the Atiyah-Floer conjecture in the final chapter.

Chapter 5 follows [FOOO09] in using an A_∞ algebra associated to a Lagrangian submanifold to study the obstructions to the existence of Lagrangian intersection Floer homology for general symplectic manifolds. This discussion of the obstruction theory, assuming transversality and using a 'virtual' de Rham model of cohomology, is based upon original work of Bryan Wang (unpublished), as presented at the ANU Floer theory seminar in 2016. As this work is ongoing, this chapter should be considered more as a preliminary sketch, the full details of which will appear at a later date. In the second section, we follow [AJ10] to define the Lagrangian intersection Floer homology for immersed Lagrangians, now using the de Rham model for the relevant A_∞ algebras. Throughout, we ignore a number of technical subtleties related to convergence of formal power series.

Chapter 6 makes the introductory discussion above rigorous; we discuss how instanton Floer homology forms part of a topological quantum field theory in dimensions 3 and 4, and explain the motivation behind the Atiyah-Floer conjecture. We conclude by explaining Fukaya's progress towards a proof of this conjecture in [Fuk15] and the various points that remain to be addressed. The promised second and third parts of [Fuk15] are still yet to appear (after a year and a half of waiting), so we cannot do much more than simply describe the approach that Fukaya claims will prove the conjecture, as many crucial details are still lacking.

For convenience, a review of gauge theory and symplectic geometry is contained in Appendix A. Readers unsure of notation and terminology should consult this appendix, the index, or [KN69, Nic07]. A summary of results concerning Sobolev spaces on manifolds is to be found in Appendix B, along with a discussion of how they may be extended to non-compact manifolds having the form of tubes. In this thesis, the word **manifold** without further qualification refers to a smooth, finite-dimensional, second-countable, Hausdorff manifold without boundary.

Chapter 2

CONFORMALLY INVARIANT VARIATIONAL PROBLEMS

We devote this chapter to presenting a quick review of the two basic conformally-invariant geometric variational problems from theoretical physics that are of great importance in the construction of moduli spaces in differential geometry. Since the analysis is similar in both cases, we treat them in parallel.

2.1 EXAMPLE: HARMONIC MAPS

Suppose Σ is a compact Riemann surface and X is a compact Riemannian manifold with metric g. In various models of string theory one defines the **energy density** of a map $u : \Sigma \to X$ to be $e(u) = |du|^2$, and the corresponding **energy** by

$$E(u) = \frac{1}{2}\int_\Sigma |du|^2 \mathrm{d}\mu_\Sigma$$

One can observe that this energy functional is invariant under conformal automorphisms of the Riemann surface Σ [Jos11, p.496]. This means that conformal transformations of the domain should leave the maxima and minima unchanged. A function $u : \Sigma \to X$ that is a stationary point of the energy functional is called a **harmonic map**. Since the energy of u corresponds to the L^2 norm of its differential, u is an absolute minimum of E if and only if $E(u) = 0$, that is, if and only if u is constant. We therefore instead wish to consider functions that minimise energy in their *homotopy class*. Now suppose that X is a symplectic manifold with symplectic form ω and a compatible almost-complex structure J that together give the Riemannian metric g (see Appendix A for a summary of symplectic geometry). In this case we may write $du = \partial_J u + \bar{\partial}_J u$ with respect to the complex structure on Σ given by i. One can then see that [Don02, p.31]

$$E(u) = \frac{1}{2}\int_\Sigma |\partial_J u|^2 + \frac{1}{2}\int_\Sigma |\bar{\partial}_J u|^2$$

One can now place lower bounds on the energy among maps in a given homotopy class by making use of topological invariants, such as the quantity defined by

$$h(u) = \int_\Sigma u^*\omega$$

which depends only on the homotopy class of u for fixed ω [MS12, p.21]. Also, if Σ has boundary that is mapped by u into some Lagrangian submanifold L of X, then $h(u)$ is also independent of the homotopy class of u (among maps having boundary in L). To see this, suppose we have a smooth family of maps $u_t : \Sigma \times I \to X$ such that $u_t|_{\partial\Sigma}$ is in L for all $t \in I$. Because ω is closed, we have $\mathrm{d}(u_t^*\omega) = 0$, and hence Stokes' Theorem gives

$$\int_\Sigma u_0^*\omega - \int_\Sigma u_1^*\omega = \int_{\partial\Sigma\times I} u_t^*\omega$$

Because L is Lagrangian, $(u_t|_{\partial\Sigma})^*\omega = 0$ and we see that the right hand side is zero. In the case considered above where X is a symplectic manifold, one can show that [Don02, p.31]

$$h(u) = \int_\Sigma |\partial_J u|^2 - \int_\Sigma |\bar{\partial}_J u|^2$$

Hence we must have $E(u) \geq h(u')/2$ for all maps u in the same homotopy class as u', with equality if and only if $\bar{\partial}_J u = 0$. Maps $u : \Sigma \to X$ satisfying $\bar{\partial}_J u = 0$ are called **pseudoholomorphic** or **J-holomorphic**; the above discussion shows that they are precisely the maps that minimise the energy in their homotopy class and have $E(u) = h(u)$. This fact that the energy is a topological invariant for pseudoholomorphic maps will be of importance later.

2.2 EXAMPLE: YANG-MILLS THEORY

Let G be a compact Lie group, X a smooth, compact, oriented Riemannian manifold (without boundary), and $\pi : P \to X$ a principal G-bundle over X (see Appendix A for a summary of bundle theory). In the following we shall take $G = \mathrm{SU}(2)$. In gauge theory, one seeks to minimise the **Yang-Mills functional** $\mathrm{YM} : \mathcal{A}_P \to \mathbb{R}$ defined on connections $A \in \mathcal{A}_P$ by

$$\mathrm{YM}(A) = \frac{1}{8\pi^2}\int_X |F_A|^2 = -\frac{1}{8\pi^2}\int_X \mathrm{Tr}(F_A \wedge *F_A)$$

It is important to observe that this functional is **conformally invariant** *in dimension 4* [DK90, p.41]; multiplying the metric on a 4-manifold X by a positive scalar leaves the Yang-Mills functional unchanged, and hence also leaves its maxima and minima unchanged. Taking the Euler-Lagrange equation yields the **Yang-Mills equation** $d_A^* F_A = 0$. When X is an oriented Riemannian 4-manifold, the Hodge star gives an involution $* : \Omega^2(\mathrm{ad}\ P) \to \Omega^2(\mathrm{ad}\ P)$ such that $*^2 = 1$. We write its ± 1 eigenspaces as $\Omega^{2,\pm}(\mathrm{ad}\ P)$ and call these forms **self-dual** or **anti-self-dual** respectively. Observing that we may write $d_A^* = *d_A*$, it is clear from the Bianchi identity $d_A F_A = 0$ that if $F_A = \pm * F_A$, then A will satisfy the Yang-Mills equations. We call these particular solutions **instantons** and the equations $F_A = + * F_A$ and $F_A = - * F_A$ the **self-dual** (SD) and **anti-self-dual** (ASD) equations respectively. For an arbitrary connection A, we may write $F_A = F_A^- + F_A^+$, where $F_A^\pm$ are the SD and ASD components. Observe that it is immaterial which one we choose to work with as reversing the orientation of the manifold swaps the two; hence we can (and will) switch between the two with impunity.

It it a basic result in Chern-Weil Theory [DK90, p.40] that the second Chern number $c_2(P)[X] \in \mathbb{Z}$ may be written as

$$c_2(P)[X] = \frac{1}{8\pi^2}\int_X \mathrm{Tr}(F_A \wedge F_A)$$

for any connection A on P. We define $k = -c_2(P)[X]$ to be the **topological quantum number** or **instanton number**. Writing $F_A = F_A^+ + F_A^-$ and using the fact that

$$\mathrm{Tr}(F_A^+ \wedge F_A^+) = \mathrm{Tr}(F_A^+ \wedge *F_A^+) = -|F_A^+|^2$$

and similarly, that $\mathrm{Tr}(F_A^- \wedge F_A^-) = |F_A^-|^2$, we find

$$c_2(P)[X] = \frac{1}{8\pi^2}\int_X |F_A^-|^2 - \frac{1}{8\pi^2}\int_X |F_A^+|^2$$

Furthermore, using the orthogonality of the two spaces $\Omega^{2,\pm}(\mathrm{ad}\ P)$, we may also write

$$\mathrm{YM}(A) = \frac{1}{8\pi^2}\int_X |F_A^+|^2 + \frac{1}{8\pi^2}\int_X |F_A^-|^2$$

Therefore $\mathrm{YM}(A) \geq k$, with equality if and only if $F_A^- = 0$. That is, the absolute minima of the Yang-Mills functional occur exactly when A is a SD instanton. This observation that YM is a topological invariant for instantons is of principal importance in the analysis.

It will sometimes be important later to discuss the case where $G = \mathrm{SO}(3)$. In this case the topological quantum number is instead given by the Pontrjagin class

$$k = p_1(P)[X] = -\frac{1}{4\pi^2}\int_X \mathrm{Tr}(F_A \wedge F_A)$$

and the same discussion as the above applies [DK90, p.42].

2.3 BUBBLING PHENOMENA

Sequences of instantons or pseudoholomorphic curves with small energy satisfy certain compactness properties. When we consider the case of large energy we find a qualitative difference in behaviour; the energy density of a sequence of solutions can concentrate at a single point. The conformal invariance of the equation makes it possible for us to rescale the equation near this singularity and hence regard this concentration as a 'bubble' solution on a different space. Since the general method of proof, due to Sacks and Uhlenbeck, is widely applicable, we spend some time sketching the main ideas. We shall find that very similar 'Gromov-Uhlenbeck' compactness results holds for both the instanton equation and the pseudoholomorphic curve equation. This theorem shall form the basis for the compactification of moduli spaces in the subsequent chapter.

In this Chapter we shall consider the simplest possible case, of pseudoholomorphic maps with domain $\hat{\mathbb{C}}$ (the one-point compactification of $\mathbb{C}$) which we shall identify with S^2 via the conformal equivalence given by stereographic projection. Considerable simplification of the compactness theory (as presented for instance in [MS12]) can be obtained if one instead works in the larger class of *harmonic maps*. Of course, by the lower-semicontinuity of the energy functional, the subset of pseudoholomorphic maps will be preserved under even the weakest types of convergence. Suppose we take a sequence of harmonic maps ϕ_n with uniformly bounded energy. Then we have a subsequence converging weakly in $W^{1,2}$ (ϕ_n is certainly uniformly bounded in L^2 whenever the target space X is a compact manifold). Passing to a subsequence, we therefore have strong convergence of ϕ_n in L^2, and passing to a further subsequence therefore gives convergence pointwise almost everywhere. We might hope that this convergence is in fact uniform. But when the dimension n of the domain Σ is 2, we have $1 - 2/2 = 0$ and hence we are in the borderline case for the compact embedding of $W^{1,2}$ into C^0 (see Appendix B for a summary of Sobolev theory). What actually happens is much more subtle; we instead have *Uhlenbeck convergence*, a kind of blowup phenomenon where the limiting energy $\liminf_{n\to\infty} E(\phi_n)$ is strictly greater than the energy $E(\phi_\infty)$ of the limit. We know from §1 of this chapter that the limiting map must therefore a different homotopy class; here the energy escapes by concentrating at points where it blows up to form a *bubble*.

The standard example is as follows. Consider the sequence of holomorphic maps $\phi_n : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \times \hat{\mathbb{C}}$ given by $\phi_n(z) = (z, nz)$ (appropriately extended to $\infty \in \hat{\mathbb{C}}$). It is easy to show using conformal invariance that $E(\phi_n) = 2E(z : \hat{\mathbb{C}} \to \hat{\mathbb{C}}) = 8\pi$. We may observe that ϕ_n will converge uniformly away from $0 \in \hat{\mathbb{C}}$ to the map $\phi_\infty : z \mapsto (z, \infty)$, which has energy 4π. The remaining energy has concentrated at 0 in a bubble. To see this, we shall rescale ϕ_n at zero. Define $\hat{\phi}_n : D_n(0) \to D_{1/n}(0) \to \hat{\mathbb{C}} \times \hat{\mathbb{C}}$ by $z \mapsto z/n \mapsto (z/n, z)$. Then $|d\hat{\phi}_n|$ is uniformly bounded on any compact subset of $\mathbb{C}$ and so the energy density cannot concentrate at a single point. We see that now $\hat{\phi}_n$ converges in C^∞ on every compact subset of $\mathbb{C}$ to $\hat{\phi}_\infty : \mathbb{C} \to \hat{\mathbb{C}} \times \hat{\mathbb{C}}$ given by $\hat{\phi}_\infty : z \mapsto (0, z)$. This map can be smoothly extended over the point $\infty \in \hat{\mathbb{C}}$ by setting $\hat{\phi}_\infty(\infty) = (0, \infty)$. We can also use the conformal invariance of the energy to see that the energy for the sequence ϕ_n in a small ball $B_\varepsilon(0)$ about 0 is given by

$$E(\phi_n; B_\varepsilon(0)) = E(\hat{\phi}_n; B_{n\varepsilon}(0)) \to E(\hat{\phi}_\infty; \hat{\mathbb{C}})$$

as $n \to \infty$. Hence the map $\hat{\phi}_\infty$ carries away the missing energy of 4π in a bubble at the point 0. The sense in which this bubble is 'attached' to $\mathbb{C}$ will be considered in our discussion of gluing theory.

The first main result in the theory of harmonic maps or pseudoholomorphic curves is a regularity estimate for small energy.

THEOREM 2.1. *($\hbar$-Regularity) [Ohm08, p.21] Suppose $1 < p < \infty$. Then there is a universal constant $\hbar > 0$, depending only on X, such that whenever $E(\phi; D_{r_0}) \leq \hbar$ for $\phi : \hat{\mathbb{C}} \to X$ a smooth harmonic map, then for any $0 < r < r_0$, we have the estimate*

$$||d\phi||_{W^{1,p}(D_r)} \leq C||d\phi||_{L^2(D_{r_0})} = CE(\phi; D_{r_0})$$

where C is a constant independent of ϕ, r.

The most important consequence is compactness for small energy.

COROLLARY 2.1. *Suppose $\phi_n : D_{2r} \to X$ have uniformly small energy $E(\phi_n; D_{2r}) \leq \hbar$, then there exists a subsequence converging in $C^\infty(D_r; X)$ to ϕ_∞ a smooth harmonic map.*

Proof. Using Theorem 2.1, we have an estimate $||d\phi_n||_{w^{1,4}(D_r)} \leq C\hbar$. The compact embedding of $W^{1,4}$ into C^0 therefore gives a subsequence of ϕ_n with uniformly convergent derivatives. We also have a uniform estimate for ϕ_n in C^0 because X is compact; applying Arzelá-Ascoli therefore gives a further subsequence of ϕ_n that converges in $C^1(D_r, X)$. Elliptic estimates may then be used to show that the limiting map ϕ_∞ is indeed a smooth solution and improve this convergence to C^∞ convergence (see Lemma 4.6.6 in [MS12] for the standard argument). ∎

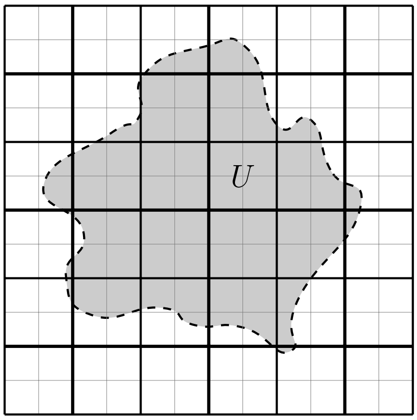

Figure 2.1: The dyadic subdivision of the domain U.

The second important consequence of this theorem is the so-called *quantisation of energy*, which implies that there are only finitely many points where the energy of a sequence of harmonic maps can 'concentrate'.

THEOREM 2.2. *(Quantisation of Energy)[Ohm08, p.24] There exists some universal constant $\hbar > 0$, depending only on X such that $E(\phi) > \hbar$ for any harmonic map $\phi : \hat{\mathbb{C}} \to X$ that is non-constant.*

This may be proved by studying the dependence of the constant C in Theorem 2.1 on the radius r_0, so as to derive the *mean-value inequality* [MS12, p.77].

The third important consequence of 2.1 follows from conformal invariance. Because the harmonic map problem is conformally invariant, if U is a domain in $\mathbb{C}$ that is conformally equivalent to D_r, then the $\hbar$-regularity estimate must also hold for harmonic maps on U, with the *same value of* $\hbar$ (but possibly with a different constant C). Using this observation, we may now prove the prototypical convergence result.

THEOREM 2.3. *(Sacks-Uhlenbeck) Suppose $\phi_n : U \to X$ is a sequence of harmonic maps with energy $E(\phi_n; U)$ on $U \subseteq \mathbb{C}$ uniformly bounded by some constant E_0. Then there exists a subsequence of ϕ_n that converges in C^∞_{loc} away from finitely many points $\{p_1, \dots, p_\ell\} \in \mathbb{C}$ to some $\phi_\infty : U \to X$ that is smooth and defined on all of U. Moreover, there exist weights $m_i \geq \hbar$ such that*

$$e(\phi_n) \to e(\phi_\infty) + \sum_{i=1}^{\ell} m_i \delta_{p_i}$$

as Radon measures (where ϕ_n denotes the subsequence in the first part of the theorem).

DEFINITION 2.1. *The convergence of ϕ_n to ϕ_∞ described above is called* ***Uhlenbeck convergence*** *or* ***convergence modulo bubbling****.*

Proof. Since the same proof applies more generally, we sketch the main ideas here. The idea is to cut U into open dyadic cubes of side length 2^{-m} for $m \in \mathbb{N}$ and use these successive subdivisions to 'hunt down' the points of concentration: see Figure 2.3 for the two-dimensional case. In order for this to work we must overcome several technical problems. Firstly we shall round off the corners of each cube to make a smooth domain that is conformally equivalent to the disk D. We then replace each rounded cube by a slightly enlarged (i.e. scaled by a factor of 2^{-m+1}) version of itself in order to obtain an open cover of U. The important fact about this cover is that each such enlarged, rounded cube (called a *box*) will intersect at most 8 of its neighbours, regardless of the value of m. We call $M = 8$ the intersection number. Part of the difficulty of applying this proof to general manifolds is controlling this intersection number. When we are not simply working in the complex plane, it is instead necessary to use a cover consisting of geodesic balls (so that they are equivalent under rescaling of the metric). It is then possible to control the intersection number of such balls given a bound on the curvature of the metric. This argument is somewhat subtle but underpins the generalisations of this proof (see [Ohm08] for the argument). The final technical problem is the boundary of the closure of U; if we only allow boxes that lie entirely inside U then for every value of m there will always exist points that are not contained in our 'open cover'. So we consider instead the sets $U_n = \{z \in U : d(z, \partial U) \geq 1/n\}$ and apply the same argument below to each U_n, starting with $m \geq n + 1$. A diagonal argument then allows us to deduce the theorem for U from the theorem for all U_n [Ohm08].

Call a box on which the upper bound of the energy of the sequence is greater than $\hbar$ a *bad* box. Otherwise, the box is called *good*, and by the conformal invariance of the $\hbar$-regularity estimate noted above, we can apply Corollary 2.1 to extract a subsequence of ϕ_n that converges smoothly on a slightly smaller box. Define an integer L by

$$L = \left\lfloor \frac{ME_0}{\hbar} \right\rfloor + 1$$

and suppose that K boxes have energy greater than $\hbar$. As sum over the energies contained in each box cannot exceed the total energy bound, we must have $E_0 \geq K\hbar/M$ and hence $K \leq L$. Hence we see that the number of bad boxes is always bounded above by our constant L, independently from the value of m.

Now we run the main argument. For a fixed value of m, we can extract subsequences of ϕ_n that converge smoothly on any of the good boxes. Because of the condition in Corollary 2.1, this convergence only holds on slightly smaller versions of the good boxes; this subtlety is easily remedied by henceforth replacing the boxes under discussion with another collection of boxes, of size intermediate between the original boxes and the dyadic cubes. If we extract these subsequences sequentially for each box, that is, take the subsequence for each box to be a subsequence of the subsequence for the previous box, we may assume that we have one subsequence that converges smoothly on all of the good boxes. When we move to $m+1$, this subsequence will still converge on all of the boxes obtained by subdivisions of the good boxes, and we can perform exactly the same procedure again with $m+1$, but now starting with the subsequence we found for m. At each stage there can be at most L bad boxes, and the number of bad boxes can only *decrease*. Finally, we can take the diagonal subsequence of the collection of subsequences for each m; this will be our desired subsequence of ϕ_n.

Because the number of bad boxes is a decreasing sequence of natural numbers, it must attain its limiting value after some finite number of steps. Call this value ℓ. After this point, consider the sequences consisting of the midpoints of the bad boxes. Because the subdivision is dyadic, they must form Cauchy sequences and hence will have limiting points, call them $p_1, \ldots, p_\ell$. It is immediate from the construction of the subsequence that that it must converge in C^∞_{loc} away from these points. By Corollary 2.1, the limiting map ϕ_∞ is smooth, harmonic, and defined on all of $U - \{p_1, \ldots, p_\ell\}$. Next we shall appeal to:

THEOREM 2.4. *(Removable Singularities) [Ohm08, p.25] Suppose $\phi : D \setminus \{0\} \to X$ is a harmonic map with finite energy. Then ϕ extends to a smooth harmonic map $\hat{\phi} : D \to X$.*

The fact that the limiting map ϕ_∞ must also have finite energy is immediate from the lower semicontinuity of the energy functional. Hence ϕ_∞ extends uniquely to a smooth map defined on all of U. This proves the first statement of the theorem.

Now we shall define the **bubble energies** by

$$m_i = \lim_{\delta \to 0^+} \lim_{n \to \infty} E(\phi_n; D_\delta(p_i))$$

where ϕ_n is now used to denote our subsequence. It is clear that we must have $m_i \geq \hbar$, for otherwise we would have smooth convergence at p_i by Corollary 2.1. Now suppose f is a continuous function on U. For all $\delta > 0$, we may use a partition of unity to break f into a sum of continuous functions $f_i(\delta)$ supported in $D_\delta(p_i)$ with a continuous function supported away from the p_i. Convergence for the part of f supported away from the p_i is a trivial consequence of smooth convergence, and we are left with a sum over estimates of the form

$$\begin{aligned}
&\left| \int_{D_\delta(p_i)} f_i(\delta) e(\phi_n) - \int_{D_\delta(p_i)} f_i(\delta) e(\phi_\infty) - f_i(\delta)(p_i) m_i \right| \\
&\leq \left| \int_{D_\delta(p_i)} f_i(\delta) e(\phi_n) - f_i(\delta)(p_i) E(\phi_n; D_\delta(p_i)) \right| + |f_i(\delta)(p_i)| \left| E(\phi_n; D_\delta(p_i)) - \lim_{n \to \infty} E(\phi_n; D_\delta(p_i)) \right| \\
&+ |f_i(\delta)(p_i)| \left| \lim_{n \to \infty} E(\phi_n; D_\delta(p_i)) - m_i \right| + \left| \int_{D_\delta(p_i)} f_i(\delta) e(\phi_\infty) \right|
\end{aligned}$$

which converge to 0 as $n \to \infty$ and $\delta \to 0$. This completes the proof of the theorem. ■

Remark 2.1. Observe that this proof, modulo choosing an appropriate cover with bounded intersection numbers, applies in exactly the same way to *any* conformally-invariant variational principle satisfying an analogous result to Corollary 2.1 and a removable singularities theorem.

Remark 2.2. Now we discuss how the concentration points p_i of the previous theorem can be interpreted as *bubbles* carrying away the bubble energies m_i. So let ϕ_n be the subsequence constructed in the proof above. Choose a concentration point p and take $\delta > 0$ sufficiently small so that no other concentration point is contained in the ball of radius δ around

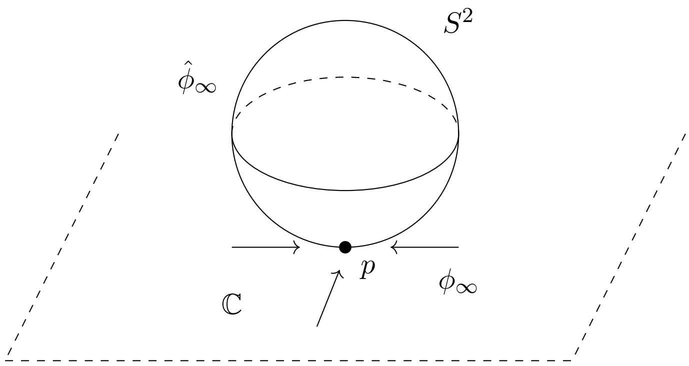

Figure 2.2: A sphere bubble.

p. Let z_n be the point $z \in D_\delta(p)$ where $|d\phi_n(z)|$ attains its maximum value, and let $b_n = |d\phi_n(z_n)|$ be this value. If b_n were bounded, then we would have uniform convergence at p, and therefore we can extract a subsequence, also called b_n, such that $\lim_{n\to\infty} b_n = \infty$. For the corresponding subsequence of points z_n, we must have $\lim_{n\to\infty} z_n = p$ or else we would again have uniform convergence at p. Now we rescale ϕ_n near p as in the example earlier in this section, by defining

$$\hat{\phi}_n(z) = \phi_n\left(\frac{z}{b_n} + z_n\right)$$

for $z \in D_{b_n\delta}(z)$, where we consider $\phi_n : D_\delta(p) \to X$ as a map $T_p\Sigma \xrightarrow{\sim} \mathbb{C} \to X$ via the exponential map. By conformal invariance, the maps $\hat{\phi}_n$ are also harmonic and we see that $|d\hat{\phi}_n| \leq 1$ also. Hence $\hat{\phi}_n$ necessarily converges smoothly on compact subsets of $\mathbb{C}$ to some map $\hat{\phi}_\infty$. By lower semicontinuity and conformal invariance of the energy functional, $\hat{\phi}_\infty$ has less energy than $\hat{\phi}_n$. Hence by the removable singularities theorem we may extend $\hat{\phi}_\infty$ smoothly over $\infty \in \hat{\mathbb{C}}$ to give a harmonic map $\phi_\infty : S^2 \to X$, where we regard $S^2 \setminus \{\infty\}$ as $\mathbb{C}$ via the (conformal) stereographic projection. We refer to the map ϕ_∞ as the **sphere bubble** at p (see Figure 2.2 for the origin of this terminology).

It is easy to see using conformal invariance of the energy that

$$E(\hat{\phi}_\infty; S^2) + E(\phi_\infty; \Sigma) \leq \liminf_{n\to\infty} E(\phi_n; \Sigma)$$

and hence that the map $\hat{\phi}_\infty$ captures some of the missing energy of the subsequence. However, we cannot in general capture *all* of the missing energy in this manner because of the possibility that bubbles can form on these bubbles themselves. However, if one continues this process, accounting for all of the bubbles, then one can show that in fact all of the energy will be captured. This process is called the construction of a *bubble tree*, which we shall take up again in Chapter 3.

INSTANTONS

Much of the above theory applies in the same way to ASD instantons, and we may observe similar behaviour in explicit examples. Let our 4-manifold X be $\mathbb{R}^4$, equipped with the trivial $\mathrm{SU}(2)$-bundle. This is conformally equivalent to $S^4 \setminus \{\infty\}$ and hence any instanton over $\mathbb{R}^4$ that decays appropriately at ∞ can be regarded as an instanton on the 4-sphere. We can identify $\mathbb{R}^4$ with the quaternions $\mathbb{H}$, and the group $G = \mathrm{SU}(2)$ with the space of unit quaternions. We may also identify the Lie algebra of $\mathrm{SU}(2)$ with the imaginary quaternions. Hence we have a collection of $\mathfrak{su}(2)$-valued 1-forms on $\mathbb{R}^4$ given as a function of the quaternionic variable $x \in \mathbb{H}$ by

$$A_\lambda(x) = \frac{\mathrm{Im}(x\mathrm{d}\bar{x})}{|x|^2 + \lambda^2}$$

where λ is a strictly positive real parameter, $\bar{x}$ denotes the quaternionic conjugate, and $dx = dx_1 + idx_2 + jdx_3 + kdx_4$ is the quaternionic differential. This is called the **BPST instanton**. We can see that the curvature is given by

$$F_\lambda(x) = \frac{dx \wedge d\bar{x}}{(|x|^2 + \lambda^2)^2}$$

and it is not difficult to see that A_λ is ASD by expanding out $dx \wedge d\bar{x}$ [DK90, Chapter 3]. However, as $\lambda \to 0$, this curvature 'concentrates' at $x = 0$ and does not converge smoothly to an instanton on a neighbourhood of this point.

Under suitable rescaling, we would observe the formation of a bubble at this point. Thus we might hope to carry out a similar analysis to the above.

However, it is easy to see that instantons cannot satisfy a version of 2.1; the presence of a large gauge symmetry group prevents the equation from possessing elliptic estimates unless some means of 'gauge fixing' is chosen. To see this, observe that there are connections A that are gauge equivalent to the trivial connection (and hence are ASD) such that the L^2 norm of F_A (the energy of A), does not control the higher derivatives of A, let alone the size of the local connection matrices [DK90, p.54]. Hence we need some means of *gauge fixing*, which is typically done using the *Coulomb gauge*. We will see that, with this choice, we have local results very similar to those for the pseudoholomorphic curve equation and hence a similar convergence theorem, but now taking into account gauge transformations.

DEFINITION 2.2. *A connection B is said to be in* ***relative Coulomb gauge*** *with respect to a connection A if $d_A^*(B-A)=0$. A trivialisation of a principal G bundle is said to be in Coulomb gauge with respect to A if $d^*A^\alpha = 0$ holds for the local connection matrices A^α associated to this trivialisation.*

The Coulomb gauge, if it exists, is a solution to the minimisation problem for the map $B \mapsto ||B-A||^2_{L^2}$ and so is heuristically the most likely to yield 'small' matrices and make the above estimates valid [DK90, p.55]. The existence problem for the Coulomb gauge, and the consequent $\hbar$-regularity statement, is settled by the following local result of Uhlenbeck:

THEOREM 2.5. *([DK90], Theorem 2.3.7) There is a universal constant $\hbar > 0$ such that any connection A on the trivial bundle over the 4-ball $\bar{B}^4$ with energy $E(A) \leq \hbar$ is gauge equivalent to a connection $\tilde{A}$ in Coulomb gauge that tends to zero at the boundary ∂B^4 and which satisfies the estimate*

$$||\tilde{A}||_{W^{1,2}} \leq C||F_{\tilde{A}}||_{L^2} = CE(A; B^4)$$

where the constant C is independent of A. Moreover, the connection $\tilde{A}$ is uniquely determined up to a constant gauge transformation.

From this we can immediately deduce the analog of Corollary 2.1:

COROLLARY 2.2. *([DK90], Corollary 2.3.9) Any sequence A_n of ASD connections over B^4 with energy $E(A; B^4) \leq \hbar$ has a subsequence A_n with gauge equivalent connections $\tilde{A}_n$ that converge in C^∞ on B^4.*

By Remark 2.1, we can immediately deduce the Uhlenbeck compactness result below; essentially the same proof as for harmonic maps above can be adapted to the case of Corollary 2.2 by keeping track of the gauge transformations and applying Uhlenbeck's removable singularities theorem for the ASD equation (see [DK90, p.163]). The generalisation of the argument the non-compact case is somewhat more subtle.

THEOREM 2.6. *(Uhlenbeck Compactness) [Don02, p.11] Suppose X is a (not necessarily compact) oriented Riemannian 4-manifold, P a principal G-bundle over X for G compact, and suppose A_n is a sequence of ASD connections with uniformly bounded energy. Then there exists a subsequence of A_n, a finite collection of points $p_1, \ldots, p_\ell$, and a principal G-bundle Q over $X - \{p_1, \ldots, p_\ell\}$ such that*

1. *There exist bundle isomorphisms $\rho_n : Q \to P|_{X-\{p_1,\ldots,p_\ell\}}$ such that $\rho_n(A_n)$ converges smoothly on all compact subsets of $X - \{p_1, \ldots, p_\ell\}$ to some ASD instanton A;*
2. *The connection A extends over $\{p_1, \ldots, p_\ell\}$ to give a smooth ASD connection on X;*
3. *Over X, we have that*

$$e(A_n) \longrightarrow e(A) + \sum_{i=1}^{\ell} \delta_{p_i}$$

in the sense of Radon measures.

With these results in hand, let us proceed to our discussion of moduli spaces.

Chapter 3

MODULI SPACES IN DIFFERENTIAL GEOMETRY

3.1 FREDHOLM SYSTEMS AND TRANSVERSALITY

Moduli spaces in differential geometry are constructed using the formalism of *Fredholm systems*, which give a geometric means of studying the solution sets of geometric non-linear partial differential equations. For basic definitions and theorems concerning Banach manifolds, see [Lan72]. *We shall assume throughout that all our Banach manifolds are separable.*

DEFINITION 3.1. *[CLW14] A **Fredholm system** consists of a Banach bundle $\mathcal{E} \xrightarrow{\pi} \mathcal{B}$ over a Banach manifold $\mathcal{B}$, along with a C^k section $s : \mathcal{B} \to \mathcal{E}$ of this Banach bundle such that the vertical linearisation $D^V_x s : T_x\mathcal{B} \to \mathcal{E}_x$ is a Fredholm operator for all $x \in \mathcal{B}$. We say that the Fredholm system is **regular** or **transversal** if $D^V_x s$ is surjective with bounded right inverse for all $x \in s^{-1}(0)$. We shall say that $\mathcal{M} = s^{-1}(0)$ is the **moduli space** associated to this Fredholm system.*

The central result is the following:

THEOREM 3.1. *If a C^k Fredholm system is regular, then the associated moduli space $\mathcal{M}$ is a finite-dimensional C^k manifold of dimension given by the Fredholm index of $D^V s$*

This follows essentially from the implicit function theorem for Banach manifolds

THEOREM 3.2. *(Implicit Function Theorem) [MS12, p.541] Suppose $f : X \to Y$ is a C^k map between (separable) Banach manifolds such that $D_x f$ has a (bounded) right inverse for all $x \in f^{-1}(y)$. Then $f^{-1}(y)$ is a C^k Banach manifold with tangent space at x given by* $\ker D_x f$.

Note that surjective Fredholm operators always have bounded right inverses [MS12, Appendix A]. The finite-dimensionality follows from the fact that $D^V_x s$ is Fredholm for all $x \in s^{-1}(0)$, and that the tangent space $T_x\mathcal{M}$ isomorphic to $\ker(D^V_x s)$. Even if the Fredholm system is not regular, we may still consider the moduli space to be $s^{-1}(0)$ as a topological space. In fact, this set will typically have more structure than just a topology, a problem we return to at the end of this chapter in our discussion of virtual techniques.

The main problems with the construction of moduli spaces in differential geometry are firstly that the Fredholm systems under consideration will certainly *not* be regular in general, and secondly, that the resulting moduli space $\mathcal{M}$ will not in general be compact. These problems can be studied fairly independently. The first may be solved in some cases using *positivity*, *perturbation* and *genericity* arguments, where extra parameters are introduced in order to make vertical linearisation maps surjective. A more systematic means of doing this is treated in our discussion of virtual neighbourhoods. The second problem may be solved by introducing an appropriate *compactification* and this will typically involve some of the detailed analysis from Chapter 2.

In this Chapter we shall introduce some examples of Fredholm systems and how their associated moduli spaces can be shown to exist in special cases. We shall begin with the simplest example, the Atiyah-Bott moduli space of flat connections, which illustrates many points of the general theory without the complications of transversality. The construction of the instanton moduli space is then very similar, but will require some further arguments to show regularity. Moduli spaces of pseudoholomorphic curves are more subtle; we shall only prove their existence in some simple special cases. The difficulties in the general case will be discussed in later sections and the appropriate results stated.

3.2 FINITE-DIMENSIONAL MODELS

Suppose G is a compact Lie group with Lie algebra $\mathfrak{g}$, and with a smooth action on manifold X. We then have an **infinitesimal action**, that is, a Lie algebra anti-homomorphism $\phi : \mathfrak{g} \to \Gamma(TX)$ given by

$$\phi(v)_x = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{e}^{tv} \cdot x$$

for all $x \in X$, where e denotes the exponential map $\mathfrak{g} \to G$. Note that if one uses the reverse convention for the Lie bracket, this will be a Lie algebra homomorphism rather than an anti-homomorphism.

Now suppose that X is a symplectic manifold with a symplectic form ω. We then say that the action of G on X is **symplectic** if the diffeomorphisms induced by G on X all preserve the symplectic form. A special case of this is a *Hamiltonian* action.

DEFINITION 3.2. *An action of G on X is called a* ***Hamiltonian action*** *if there exists a* ***co-moment map*** $J : X \to \mathfrak{g}^*$ *that is* ***co-adjoint equivariant*** *(that is, $R_g^* J = \mathrm{Ad}_g^*(J)$) such that for all $v \in \mathfrak{g}$, the smooth function $\langle J, v\rangle \in C^\infty(X)$ obtained by pairing J pointwise with v is the Hamiltonian function generating the vector field $\phi(v)$.*

Note that such an action is always symplectic and necessarily preserves the level sets of J. Given such an action, we may then perform *symplectic reduction*:

THEOREM 3.3. *(Marsden-Weinstein) [CdS01, p.141] Suppose G is a compact Lie group with a Hamiltonian action on X that is free on $J^{-1}(0)$. Then the quotient $J^{-1}(0)/G$ (called the* ***Marsden-Weinstein quotient*** *or the* ***symplectic reduction****) is a smooth symplectic manifold with symplectic form ω_r satisfying $\iota^*\omega = \pi^*\omega_r$ (for $\pi : J^{-1}(0) \to J^{-1}(0)/G$ the (smooth) quotient map and $\iota : J^{-1}(0) \to X$ the inclusion).*

We introduce now the formalism of *deformation complexes* for these types of problems. Firstly, given $x \in X$, the orbit $G \cdot x$ will be a compact embedded submanifold diffeomorphic to G whenever G is compact and acts freely. When the action is only locally free (that is, the stabilisers are all discrete), then the orbit will be diffeomorphic to G/G_x where G_x is the stabiliser of x, a closed Lie subgroup of G. Note that G/G_x may not in general be a Lie group, but instead a Lie orbifold. The tangent space $T_x(G \cdot x)$ is given by the image of the pointwise infinitesimal action $\phi_x : \mathfrak{g} \to T_xX$. Furthermore, the kernel of the map ϕ_x is the Lie (sub)algebra of $\mathfrak{g}$ corresponding to stabiliser subgroup G_x of G. In particular, when the G-action is locally free at x, the kernel of ϕ will be zero. When 0 is a regular value of J, then $J^{-1}(0)$ will be a smooth embedded submanifold of X by the implicit function theorem. The tangent space $T_xJ^{-1}(0)$ is given by the kernel of the map $D_xJ : T_xX \to \mathfrak{g}^*$. As the G-action preserves the level sets of J, we must have $D_xJ \circ \phi_x = 0$. Hence we have a chain complex

$$0 \to \mathfrak{g} \xrightarrow{\phi_x} T_xX \xrightarrow{D_xJ} \mathfrak{g}^* \to 0$$

called the **deformation complex**. The tangent space to the quotient manifold $J^{-1}(0)/G$ at x should then be given by the first cohomology group $H^1 = \ker(D_xJ)/\mathrm{im}\ (\phi)$. Heuristically, taking the space of 'infinitesimal deformations' of $J^{-1}(0)$ and quotienting out by those arising from the group action should yield the space of 'infinitesimal deformations' of the quotient space. If X has a Riemannian metric, then the **normal space** to the the G-orbit through x may be defined as $\ker \phi_x^*$ for $\phi_x^* : T_xX \to \mathfrak{g}$ the adjoint mapping. Then H^1 is given by the kernel of the **deformation operator** $\phi_x^* \oplus D_xJ : T_xX \to \mathfrak{g} \oplus \mathfrak{g}^*$. In the special case where the G-action is locally free at x, and x is a regular point for J, this operator will be surjective; this is the 'ideal' case. Then the dimension of the quotient manifold (that is the dimension of H^1) will be given by the dimension of the kernel of the deformation operator.

Remark 3.1. It is possible to give a more geometric meaning to the surjectivity of the deformation operator. If G acts freely on X, then, as we shall see below, the quotient map $X \to X/G$ gives a principal G-bundle. Since J is G-equivariant, it descends to give a section of the associated bundle $X \times_{\mathrm{ad}^*} \mathfrak{g}^*$ over X/G. The vertical derivative of J on the quotient space is then equivalent to the deformation operator and so we see that $J^{-1}(0)/G$ is a smooth manifold if the deformation operator is everywhere surjective.

Remark 3.2. In the special case where the G-action is *Hamiltonian*, the fact that the action of G on $J^{-1}(0)$ is locally free actually *implies* that 0 is a regular value of J. This is because we have (by definition) the relation

$$\omega_x(\phi_x(w), v) = \langle D_xJ(v), w\rangle$$

for all $v \in T_xX$, where $\langle\cdot,\cdot\rangle$ denotes the dual pairing on $\mathfrak{g}$. This shows that the image of D_xJ is the adjoint of the Lie algebra of the stabiliser (that is, the set of $w \in \mathfrak{g}^*$ such that $\langle v, w\rangle = 0$ for all $v \in \ker(\phi_x)$). When $\ker(\phi_x) = 0$, the derivative D_xJ must hence be surjective.

The main part of the proof of the theorem above is the following standard lemma. We provide a proof here as this forms a prototype for many of the infinite-dimensional arguments later in this thesis.

LEMMA 3.1. *[CdS01, p.142] Suppose G is a compact Lie group acting freely on a manifold X. Then X/G is a smooth manifold and the projection map $\pi : X \to X/G$ is smooth.*

Proof. The fact that the quotient space is Hausdorff follows from the compactness of G purely as a matter of point-set topology. Now we begin by selecting a point $x \in X$ and considering the orbit $G \cdot x$. Since $G \cdot x$ is an embedded submanifold of X, we may choose local coordinates $y_1, \ldots, y_n$ in an open neighbourhood B_x of x in which $G \cdot x$ is given by the vanishing locus $y_1 = \cdots = y_k = 0$ for some $k \leq n$. We then define the **slice at** x as the submanifold S_x of the neighbourhood B_x given by the vanishing locus $y_{k+1} = \cdots = y_n = 0$. Define a map $\psi : G \times S_x \to X$ given by $\psi(g, s) = g \cdot s$. We claim that by taking B_x sufficiently small, this map will give a diffeomorphism with a G-invariant open neighbourhood U_x of $G \cdot x$. Firstly, we show that $D_{(e,x)}\psi$ is necessarily bijective; it is the direct sum of the differentials of the two maps $\psi_e : x \mapsto e \cdot x$ and $\psi_x : g \mapsto g \cdot x$. The differential of the map ψ_e gives the inclusion $T_x S_x \subseteq T_x X$ and so is injective. The differential of the map ψ_x can be identified with the infinitesimal action $\phi_x : \mathfrak{g} \to T_x X$. As we have seen above, when the G action is locally free, this will be an injection whose image is the tangent space $T_x(G \cdot x)$ to the orbit through x. From the choice of local coordinates above, we see that $T_x X = T_x(G \cdot x) \oplus T_x S_x$ and hence that $D_{(e,x)}\psi$ must be surjective also. By G-equivariance it then follows that $D_{(g,x)}\psi$ is bijective for all $g \in G$.

Now the implicit function theorem implies that there must exist an open neighbourhood of $G \times \{x\} \subseteq G \times S_x$ such that ψ maps this neighbourhood diffeomorphically onto some neighbourhood U_x of $G \cdot x$. Note that U_x is necessarily G-invariant. From this discussion we hence see that U_x/G is homeomorphic to the smooth manifold S_x. We may use these homeomorphisms as coordinate charts for X/G; we simply need to check that the overlaps are smooth. Suppose two such neighbourhoods $U, V \subseteq X$ are given with corresponding slices S, T. We can then see that $S \cap V$ and $T \cap U$ will also be slices corresponding to the intersection $U \cap V$. Hence $G \times (S \cap (U \cap V))$ and $G \times (T \cap (U \cap V))$ must be diffeomorphic and hence the transition $S \cap (U \cap V) \to T \cap (U \cap V)$ must be a smooth map. Finally, it is easy to see that $\pi : X \to X/G$ is smooth since it is smooth in all of the above coordinate charts (by construction). ∎

Remark 3.3. We can see from the above proof that the set of points at which the group G acts freely must be open.

With this, the proof of the Marsden-Weinstein Theorem is straightforward. To see why the symplectic form must descend to the quotient manifold, see [CdS01, p.142]; it is because the tangent spaces $T_x(G \cdot x)$ to orbits in $J^{-1}(0)$ are necessarily *isotropic* by equation 3.2, so that the symplectic form is well-defined on the quotient space H^1.

3.3 EXAMPLE: THE ATIYAH-BOTT MODULI SPACE

In this section we shall use an infinite-dimensional analog of symplectic reduction to put the structure of a smooth finite-dimensional symplectic manifold on the set of gauge equivalence classes of flat connections on a principal bundle over a Riemann surface Σ. This approach was first sketched in [AB83] and so we shall refer to this moduli space as the **Atiyah-Bott moduli space**. So let Σ be a Riemann surface, and $\pi : P \to \Sigma$ a principal G-bundle for a compact Lie group G (we shall mainly be interested in the cases $G = \mathrm{SU}(2)$ and $G = \mathrm{SO}(3)$). The space $\mathcal{A}_P$ of all connections on P can be (for the moment, heuristically) regarded as an infinite-dimensional manifold with a natural 2-form ω given by

$$\omega(\alpha, \beta) = \int_\Sigma \mathrm{Tr}(\alpha \wedge \beta)$$

where α, β are Lie-algebra-valued 1-forms in $\Omega^1(\Sigma, \mathrm{ad}\ P)$, which may be identified with the tangent space $T_A\mathcal{A}_P$ to the affine space $\mathcal{A}_P$ at any point $A \in \mathcal{A}_P$. Here Tr denotes the trace operator on the Lie algebra of G. This form is evidently closed because it does not depend on the choice of point $A \in \mathcal{A}_P$. On a Riemann surface, the natural metric gives a Hodge star $* : \Omega^1(\Sigma, \mathrm{ad}\ P) \to \Omega^1(\Sigma, \mathrm{ad}\ P)$ that satisfies $*^2 = -1$ and hence gives a complex structure on the infinite-dimensional manifold $\mathcal{A}_P$. In fact, because

$$\omega(\alpha, *\alpha) = \int_\Sigma \mathrm{Tr}(\alpha \wedge *\alpha) = -||\alpha||_{L^2(\Sigma, \mathrm{ad}\ P)}$$

it follows that ω is a non-degenerate 2-form, and hence that $(\mathcal{A}_P, \omega)$ is a symplectic manifold with compatible almost-complex structure given by the Hodge star.

Now we wish to consider the 'symplectic quotient' of the infinite-dimensional symplectic manifold $\mathcal{A}_P$ by the action of the infinite-dimensional group $\mathscr{G}_P$ of gauge transformations of P. By considering $\mathscr{G}_P$ as the space $\Gamma(\Sigma, \mathrm{Ad}\ P)$ of sections of the adjoint bundle Ad P over Σ, we can regard the (infinite-dimensional) Lie algebra of $\mathscr{G}_P$ as the space $\Omega^0(\Sigma, \mathrm{ad}\ P)$ of sections of the adjoint bundle ad P of Lie algebras. We then have an exponential map $\mathrm{e} : \Omega^0(\Sigma, \mathrm{ad}\ P) \to \Gamma(\Sigma, \mathrm{Ad}\ P)$ given by applying the exponential map $\mathrm{e} : \mathfrak{g} \to G$ pointwise. Now suppose that $\alpha \in \Omega^0(\Sigma, \mathrm{ad}\ P)$ and that $A \in \mathcal{A}_P$. Then we can compute that

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{e}^{t\alpha} \cdot A = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left(\mathrm{e}^{-t\alpha}\mathrm{d}(\mathrm{e}^{t\alpha}) + \mathrm{e}^{-t\alpha} A \mathrm{e}^{t\alpha}\right) = \mathrm{e}^{-t\alpha}(\mathrm{d}\alpha + [A, \alpha])\mathrm{e}^{t\alpha}\bigg|_{t=0} = \mathrm{d}_A\alpha$$

Hence the infinitesimal action $\phi : \Omega^0(\Sigma, \mathrm{ad}\ P) \to \Gamma(\mathcal{A}_P, T\mathcal{A}_P)$ sends a section $\alpha \in \Omega^0(\Sigma, \mathrm{ad}\ P)$ to the vector field $X_\alpha : \mathcal{A}_P \to \Omega^1(\Sigma, \mathrm{ad}\ P)$ given by $X_\alpha(A) = d_A\alpha$ for all $A \in \mathcal{A}_P$. The infinitesimal action $\phi_A : \Omega^0(\Sigma, \mathrm{ad}\ P) \to \Omega^1(\Sigma, \mathrm{ad}\ P)$ at a point $A \in \mathcal{A}_P$ is thus given simply by the associated covariant derivative $\mathrm{d}_A : \Omega^0(\Sigma, \mathrm{ad}\ P) \to \Omega^1(\Sigma, \mathrm{ad}\ P)$ on the adjoint bundle.

Now, we may identify $\Omega^2(\Sigma, \mathrm{ad}\ P)$ as the dual of the Lie algebra $\Omega^0(\Sigma, \mathrm{ad}\ P)$ via the non-degenerate pairing

$$\langle \alpha, \beta \rangle = \int_\Sigma \mathrm{Tr}(\alpha \wedge \beta)$$

for $\alpha \in \Omega^2(\Sigma, \mathrm{ad}\ P)$ and $\beta \in \Omega^0(\Sigma, \mathrm{ad}\ P)$. Then we claim:

THEOREM 3.4. *Under this identification, the action of $\mathscr{G}_P$ on $\mathcal{A}_P$ is Hamiltonian, that is, satisfies equation 3.2 with co-moment map J given by the curvature map $F : \mathcal{A}_P \to \Omega^2(\Sigma, \mathrm{ad}\ P)$.*

Proof. To prove this, we must compute the differential of the curvature function. So suppose $A \in \mathcal{A}_P$ and take $\alpha \in \Omega^1(\Sigma, \mathrm{ad}\ P) \cong T_A\mathcal{A}_P$. Then we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} F_{A+t\alpha} = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left(F_A + t d_A\alpha + t^2 \alpha \wedge \alpha\right) = d_A\alpha$$

and hence $D_AF : \Omega^1(\Sigma, \mathrm{ad}\ P) \to \Omega^2(\Sigma, \mathrm{ad}\ P)$ is given by the exterior covariant derivative d_A. Now suppose an element $\alpha \in \Omega^0(\Sigma, \mathrm{ad}\ P)$ of the Lie algebra of $\mathscr{G}_P$ is given. Then the function $\mathcal{A}_P \to \mathbb{R}$ given by the pointwise pairing $A \mapsto \langle F_A, \alpha \rangle$ has differential that acts on tangent vectors $\beta \in \Omega^1(\Sigma, \mathrm{ad}\ P) \cong T_A\mathcal{A}_P$ by

$$\mathrm{d}\langle F, \alpha \rangle(\beta) = \int_\Sigma \mathrm{Tr}(d_A\beta \wedge \alpha)$$

At the same time, we have

$$\omega_A(\phi(\alpha), \beta) = \int_\Sigma \mathrm{Tr}(d_A\alpha \wedge \beta)$$

But by Stokes' theorem, since Σ is a closed surface,

$$0 = \int_\Sigma \mathrm{d}\,\mathrm{Tr}(\alpha \wedge \beta) = \int_\Sigma \mathrm{Tr}(d_A\alpha \wedge \beta) - \int_\Sigma \mathrm{Tr}(\alpha \wedge d_A\beta)$$

and hence it follows that $\omega_A(\phi(\alpha), \beta) = \mathrm{d}\langle F, \alpha \rangle(\beta)$. Thus the action is Hamiltonian. ■

Since the zero locus of the curvature function $F : \mathcal{A}_P \to \Omega^2(\Sigma, \mathrm{ad}\ P)$ is exactly the set of flat connections, taking the symplectic quotient $F^{-1}(0)/\mathscr{G}_P$ will yield exactly the space of gauge equivalence classes of flat connections on $P \to \Sigma$. We shall denote this space as $\mathcal{M}_\Sigma$ when the bundle in question is not significant.

SOBOLEV COMPLETIONS

Of course, the 'Theorem' above does not strictly speaking make sense as we have not clarified the sense in which $\mathcal{A}_P$ and $\mathscr{G}_P$ are infinite-dimensional smooth manifolds. In order to mimic the proof of Lemma 3.1, it will therefore be important first to establish the Banach manifold structures on the various spaces involved so that we may apply a Banach manifold version of the implicit function theorem. This is typically performed by taking appropriate *Sobolev completions* (see Appendix B). We shall describe now how this this done in general; let X denote a smooth compact manifold of dimension n with a principal G-bundle $P \to X$, and take k to be an integer strictly greater than $n/2$.

THEOREM 3.5. *[Mor98, p.86] The $W^{2,k}$ completion of the gauge transformation group $\mathscr{G}_P$ is a smooth Banach Lie group with Banach Lie algebra given by the space of $W^{2,k}$ sections of* ad P. *This Banach Lie group acts smoothly on the $W^{2,k-1}$ completion of the space $\mathcal{A}_P$ of connections on P via gauge transformations.*

Proof. Firstly, we can take the $W^{2,k}$ completion of the vector space of sections $\Omega^0(X, \text{ad } P)$ to get a Banach space (and hence a smooth Banach manifold), denoted $\Omega^0_k(X, \text{ad } P)$. Similarly, we can define the Sobolev completion $\Omega^1_{k-1}(X, \text{ad } P)$ as the space $\Omega^0_{k-1}(X, \bigwedge^2 T^*X \otimes \text{ad } P)$. If we fix a smooth connection $A_0 \in \mathcal{A}_P$ then the space of connections differing from A_0 by an element of $\Omega^0_k(X, \text{ad } P)$ inherits the structure of a smooth Banach manifold and is denoted $\mathcal{A}_k$. By taking the Lie group G to be an embedded submanifold of some $\mathrm{GL}(V)$ for a vector space V, we can regard the bundle Ad P as a sub(fibre)bundle of $\mathrm{End}(E)$ where E is the vector bundle associated to the representation of G in V (see Appendix A). Then we can take the $W^{2,k}$ Sobolev completion of the vector space $\Omega(X, \mathrm{End}(E))$ and show that the completion $\mathscr{G}_k$ of the space of sections $\Gamma(X, \text{Ad } P)$ inherits a Banach manifold structure as a submanifold of $\Omega^0_k(X, \mathrm{End}(E))$. For instance, let $G = \mathrm{SO}(3)$. Recall that one typically shows that $\mathrm{SO}(3)$ is an embedded submanifold of $\mathrm{M}_3(\mathbb{R})$ by showing that the map $\phi : M_3(\mathbb{R}) \to M_3(\mathbb{R})$ given by $A \mapsto A^T A$ has the identity matrix as a regular value. This same map ϕ induces a smooth map on $\mathrm{End}(E)$ and hence induces a smooth map Φ on $\Omega^0_k(\Sigma, \mathrm{End}(E))$ since $k > n/2$ (see Appendix B). The derivative of this induced map Φ on sections is just multiplication by the (finite-dimensional) derivative of ϕ. Hence we see that Φ has a regular value at the identity section in the Banach space $\Omega^0_k(\Sigma, \mathrm{End}(E))$ as the finite-dimensional derivative of ϕ is invertible there. Hence the Banach space inverse function theorem applies to show that $\mathscr{G}_k$ has the structure of a smooth Banach manifold for such values of k. The case $G = \mathrm{SU}(2)$ is analogous.

More importantly, $\mathscr{G}_k$ has the structure of a *Banach Lie group* (whenever $k > n/2$). Multiplication in $\mathscr{G}_k$ is smooth as it is a restriction of the smooth multiplication map on $\Omega^0_k(X, \mathrm{End}(E))$ (see Appendix B) to a Banach submanifold. Inversion can also seen to be smooth; for instance when $G = \mathrm{SO}(3)$, it is the restriction to $\mathscr{G}_k$ of the map $A \mapsto A^T$ that is smooth on $\Omega^0_k(X, \mathrm{End}(E))$. The exponential map described above, from the Lie algebra $\Omega^0_k(X, \text{ad } P)$ to $\mathscr{G}_k$, defined as the pointwise composition with the exponential map $\mathfrak{g} \to G$, will yield a smooth bundle map ad $P \to \mathrm{Ad}P$. Since $k > n/2$, composition with this smooth bundle map yields a smooth map on sections and so this exponential map is smooth with respect to these Sobolev completions.

To see that the group action of $\mathscr{G}_k$ on $\mathcal{A}_{k-1}$ is smooth, it suffices to work in a trivialisation where we have the local formula

$$(g, A) \mapsto g^{-1}dg + g^{-1}Ag$$

where $g \in \mathscr{G}_k$ and $A \in \Omega^1_{k-1}(X, \text{ad } P)$. The first term above is smooth as the product of two smooth maps, $g \mapsto g^{-1}$ from $\mathscr{G}_k \to \mathscr{G}_k \subseteq \Omega^0_k(\mathrm{End}(E))$ and $g \mapsto \mathrm{d}g$ from $\Omega^0_k(X, \mathrm{End}(E))$ to $\Omega^1_{k-1}(X, \mathrm{End}(E))$. The second term is simply multiplication, and hence for k sufficiently large, this action is indeed smooth. We can also check that the curvature gives a smooth map $F : \mathcal{A}_{k-1} \to \Omega^2_{k-2}(X, \text{ad } P)$. Firstly, because we have

$$F_{A+ta} = F_A + t d_A a + t^2 a \wedge a$$

all the derivatives of F of order greater than or equal to 3 will be zero. We can see that the first derivative $a \mapsto d_A a$ is linear and bounded as a map $\Omega^1_{k-1}(X, \text{ad } P) \to \Omega^2_{k-2}(X, \text{ad } P)$ using the fact that $d_A a = da + [A, a]$ and the Sobolev multiplication theorem (when k is sufficiently large). The second term, $a \wedge a$, is simply multiplication and is also covered by the multiplication theorem in Appendix B. This completes the proof. ■

Henceforth, we shall leave the routine details of checking the smoothness of Banach completions to the reader.

LOCAL THEORY

Still following the analogy with finite-dimensional symplectic reduction, we shall now try to understand the infinitesimal structure of the moduli space of flat connections. We have seen above that the infinitesimal action at a point $A \in \mathcal{A}_P$ is given by the covariant derivative $d_A : \Omega^0(\Sigma, \text{ad } P) \to \Omega^1(\Sigma, \text{ad } P)$, and that the differential of the curvature at A is given by the exterior covariant derivative $d_A : \Omega^1(\Sigma, \text{ad } P) \to \Omega^2(\Sigma, \text{ad } P)$. When A is a flat connection, then $d_A \circ d_A = 0$ and these two maps fit together to give a chain complex

$$0 \to \Omega^0(\text{ad } P) \xrightarrow{\mathrm{d}_A} \Omega^1(\text{ad } P) \xrightarrow{\mathrm{d}_A} \Omega^2(\text{ad } P) \to 0$$

which is entirely analogous to the *deformation complex* considered in §2. In particular, the twisted de Rham cohomology group, denoted H^1_A or $H^1_A(\Sigma, \text{ad } P)$, can be identified with the tangent space $T_A\mathcal{M}$ to the moduli space of flat connections at A. From an analytic perspective what is important about the above complex is that it is an *elliptic complex*: the sequence of principal symbols of the partial differential operators is *exact*. It is a purely formal consequence of this fact that the associated deformation operator $T_A = d_A \oplus d_A^* : \Omega^1(\text{ad } P) \to \Omega^0(\text{ad } P) \oplus \Omega^2(\text{ad } P)$ is elliptic [DK90, p.131]. Hodge theory for elliptic partial differential operators on compact manifolds then implies that, whenever T_A

is surjective, the kernel $\ker T_A$ can be identified with H^1_A, just as in the finite-dimensional situation above [DK90]. Moreover, it follows that the operator T_A is Fredholm, so that when T_A is surjective the dimension of the space H^1_A of infinitesimal deformations is given by the *negative* of the Fredholm index of T_A.

We may calculate this index using the **Atiyah-Singer index theorem**. Let E denote the complexification of ad P; after tensoring with $\mathbb{C}$, the elliptic complex becomes

$$0 \to \Omega^0(\Sigma, E) \xrightarrow{d_A} \Omega^1(\Sigma, E) \xrightarrow{d_A} \Omega^2(\Sigma, E) \to 0$$

Let $T = T_A$ and observe that the (complexified) principal symbol [LM89, Chapter 3] of $\sigma(T)$ is equal to $\sigma(D) \otimes 1_E$ where $D = d^* \oplus d$ is the Euler characteristic operator on Σ. The the Atiyah-Singer Index Theorem gives the cohomological formula

$$\operatorname{ind}(T) = \left\{\Phi^{-1}\mathrm{ch}[\sigma(T)] \smile \mathrm{Td}(T\Sigma)\right\}[\Sigma]$$

where Φ is the Thom isomorphism and Td is the Todd class [BB85, p.376]. Then we can write, using the multiplicative property of the Chern character:

$$\operatorname{ind}(T) = \left\{\mathrm{ch}(E) \smile \Phi^{-1}(\mathrm{ch}[\sigma(D)]) \smile \mathrm{Td}(T\Sigma)\right\}[\Sigma]$$

We have $\mathrm{ch}(E) = \dim G$ because $c_2(E)$ is zero for dimensional reasons and $c_1(E)$ is zero because the structure group of ad P always reduces to $\mathrm{SU}(\dim G)$ [AHS78]. Therefore,

$$\operatorname{ind}(T) = \dim G \operatorname{ind}(D)$$

and so if we take $G = \mathrm{SO}(3)$ or $G = \mathrm{SU}(2)$, we have $\dim G = 3$ and $\chi(\Sigma) = 2 - 2g$, yiedling $\operatorname{ind}(T) = -(6g - 6)$.

In fact, because of the analogy with symplectic reduction, we have a very simple criteria for this to be the case; by Remark 3.2 in the previous section, the operator T_A will be surjective whenever $\mathcal{G}_P$ acts locally freely on A, that is, whenever the stabiliser of A is discrete. However, observe that some elements of $\mathcal{G}_P$ will stabilise *every* element of $\mathcal{A}_P$, namely those constant gauge transformations taking values in the centre of G. We shall identify this subgroup with $Z(G)$ and recall that when G is a semisimple Lie group the centre of G is discrete. We shall call A an **irreducible connection** if the stabiliser of A is just $Z(G)$; if it is larger, we call A **reducible**. This brings us to our main result in this section.

THEOREM 3.6. *Suppose $P \to \Sigma$ is a principal* $\mathrm{SU}(2)$ *or* $\mathrm{SO}(3)$ *principal bundle over a genus g surface Σ with no reducible flat connections. Then the Atiyah-Bott moduli space $\mathcal{M}_\Sigma$ is a smooth symplectic manifold of dimension* $6g - 6$.

GLOBAL THEORY

To prove the theorem above we shall return to the perspective of Fredholm systems, in order to show that these local infinitesimal deformations integrate out to give a 'global' smooth structure. If we recall Remark 3.1 from the previous section and use $\mathcal{B}_P$ to denote the quotient space $\mathcal{A}_P/\mathcal{G}_P$, then we may regard the curvature map $F : \mathcal{A}_P \to \Omega^2(\Sigma, \mathrm{ad}\, P)$ as a section of the (infinite-dimensional) associated vector bundle $\mathcal{A}_P \times_{\mathrm{ad}} \Omega^2(\Sigma, \mathrm{ad}\, P)$ over $\mathcal{B}_P$. Then the symplectic reduction $\mathcal{M}_\Sigma$ is given by the zero locus $F^{-1}(0)$ of this section, just as in the case of Fredholm systems. As in Remark 3.1, the vertical linearisation of this section F at a flat connection A is given by the deformation operator T_A, which is Fredholm and surjective whenever A is irreducible. Since we assume above that every flat connection A is irreducible, we would therefore like to apply Theorem 3.1, after taking suitable Banach completions, to conclude that $\mathcal{M}_\Sigma$ is indeed a smooth manifold of dimension $\operatorname{ind}(T) = 6g - 6$. The difficulty here is showing that the base space $\mathcal{B}_P$ is actually a smooth Banach manifold itself. This will not be true in general because there may exist (non-flat) reducible connections in $\mathcal{A}_P$, preventing us from applying an infinite-dimensional analog of Theorem 3.1. Instead, we must work with the space $\mathcal{A}^*_P$ of *irreducible* connections on P. If we let $\mathcal{B}^*_P = \mathcal{A}^*_P/\mathcal{G}_P$, then we may instead regard $\mathcal{M}_\Sigma$ as the zero locus of the Fredholm system given by $F : \mathcal{B}^*_P \to \mathcal{A}^*_P \times_{\mathrm{ad}} \Omega^2(\Sigma, \mathrm{ad}\, P)$ because all of the flat connections in $\mathcal{M}_\Sigma$ are assumed to be irreducible.

To begin, as in Theorem 3.1 above, we wish to prove a local slice theorem for the appropriate Sobolev completions of $\mathcal{G}_P$ and $\mathcal{A}^*_P$. In our specific case where $X = \Sigma$ is a Riemann surface, we can take $k > 1$ in Theorem 3.5 above so that the spaces and group action are all smooth. Observe that ultimately it should not matter which value of $k > 1$ we choose because all *flat* connections must in fact be smooth by elliptic regularity [DK90]. However, it is not clear why the resulting topology on $\mathcal{M}_\Sigma$ is independent of k; the proof is a simple adjustment of that found in [DK90, p.130]. Now, the proof of this local slice theorem follows the model of Theorem 3.1, but is considerably simpler because the space $\mathcal{A}_{k-1}$ is affine. Just as in the finite-dimensional case, we wish to form the gauge-invariant orthogonal complement to the group orbit through $A \in \mathcal{A}^*_{k-1}$. Now we can simply do this by identifying $\mathcal{A}_{k-1}$ with $T_A\mathcal{A}_{k-1}$ by taking A as a reference connection, and then using the decomposition $T_A\mathcal{A}^*_{k-1} \cong \Omega^1_{k-1}(\text{ad } P) = \text{im } d_A \oplus \ker d^*_A$, where im d_A is the tangent space to the gauge group orbit. This is clearly an orthogonal complement with respect to L^2, and elliptic theory for T_A tells us that this decomposition is also compatible with the higher Sobolev structure on $\Omega^1_{k-1}(\Sigma, \text{ad } P)$ [DK90, p.131]. Then we have exactly the same result as before:

LEMMA 3.2. *(Local Slice Theorem) Suppose $k > n/2$. Then for each $A \in \mathcal{A}^*_{k-1}$ there exists some $\mathcal{G}_k$-invariant neighbourhood in the space $\mathcal{A}^*_{k-1}$ that is locally diffeomorphic to $\ker d^*_A \times \mathcal{G}_k$ in an equivariant fashion.*

Proof. We use here the Banach-manifold version of the implicit function theorem. Let A be as above and define the map $\Psi : \ker d^*_A \oplus \tilde{\mathcal{G}_k} \to \mathcal{A}^*_{k-1}$ by $(a, s) \mapsto s^*(A + a)$. We want to show that this has an invertible differential at $(0, I)$. But $D_{(I,0)}\Psi : \ker d^*_A \oplus \Omega^0(\Sigma, \text{ad } P) \to \Omega^1(X, \text{ad } P)$ is given by the direct sum $D_{(I,0)}\Psi = D_0\Psi_I \oplus D_I\Psi_0$. The map $\Psi_I(a)$ is simply $A + a$ and hence $D_0\Psi_I = I$, which is surjective onto $\ker d^*_A$ and injective. The map $D_I\Psi_0$ is exactly the infinitesimal action at A, which is injective because A is irreducible. By the elliptic theory for T we have the decomposition $\Omega^1_{k-1}(\Sigma, \text{ad } P) = \ker d^*_A \oplus \text{im } d_A$ and thus $D_{(I,0)}\Psi$ is surjective also. The exact same equivariance argument as in Theorem 3.1 completes the proof. ■

Note that this result implies that $\mathcal{A}^*_P$ is a open subset of $\mathcal{A}_P$ by Remark 3.3, and hence is a smooth Banach manifold itself. Now (almost) the exact same reasoning as in Theorem 3.1 will complete the proof that $\mathcal{A}^*_P \to \mathcal{B}^*_P$ is a smooth Banach bundle over a Banach manifold. The one difference is that the gauge group $\mathcal{G}_P$ is not compact, so we shall require the following (quite general) lemma to see that the quotient space is Hausdorff.

LEMMA 3.3. *Let G be a compact Lie group and let $P \to X$ be a principal G-bundle over a compact n-manifold X. If $k > n/2$ then quotient space of $\mathcal{A}_k$ under the action of $\mathcal{G}_{k-1}$ is a Hausdorff topological space.*

Proof. The L^2 metric on $\mathcal{A}_{k-1}$ given by $d(A, B) = ||A - B||_{L^2}$ descends to a function on the quotient $\mathcal{B}_{k-1}$ by taking an infimum over gauge equivalence classes:

$$d([A], [B]) = \inf_{\substack{A\in[A]\\ B\in[B]}} d(A, B)$$

Because the L^2 topology is coarser than the $W^{k,2}$ topology, to see that $\mathcal{B}_{k-1}$ is Hausdorff it suffices to show that d is non-degenerate on $\mathcal{B}_{k-1}$. We recall here the standard argument from [DK90, p.64]. Suppose that $d([A_\infty], [B_\infty])$ is zero. Then we have a sequence of connections A_i and a sequence of gauge transformations u_i such that A_i and $B_i := u_iA_i$ converge in L^2 to the connections A_∞ and B_∞ respectively. We claim that some subsequence of u_i also converges in L^2. We may write

$$B_i = u_iA_iu_i^{-1} - du_iu_i^{-1}$$

and hence $du_i = u_iA_i - B_iu_i$. Because G is a compact Lie group, u_i are necessarily uniformly bounded in L^2, and so the equation just derived also implies the uniform boundedness of du_i in L^2. Similarly, we can derive L^2 estimates for higher derivatives of u_i by differentiating this equation and using the usual 'bootstrapping' argument. Because X is compact, the Rellich-Kondrachov theorem then implies that u_i has an L^2-convergent subsequence. Hence we have that A_∞ and B_∞ are gauge equivalent, that is $[A] = [B]$, and hence d is non-degenerate. ■

3.4 EXAMPLE: INSTANTON MODULI SPACES

Throughout this section, X will denote a compact, oriented Riemannian 4-manifold, and $\pi : P \to X$ will denote a $G = \text{SU}(2)$-principal bundle over X with topological quantum number k. In this section we wish to put the structure of a manifold on the set of self-dual (SD) connections on the bundle P over X to produce an **instanton moduli space** $\mathcal{M}_X$. In fact, we have already done most of the work in the previous section. After taking appropriate Sobolev completions, we have a principal $\mathcal{G}_P$ bundle $\mathcal{A}^*_P$ over the Banach manifold $\mathcal{B}^*_P$. This has an associated vector bundle $\mathcal{A}^*_P \times_{\text{ad}} \Omega^{2,-}(X, \text{ad } P)$ and the anti-self-dual part F^-_A of the curvature descends to give a smooth section $F^- : \mathcal{B}^*_P \to \mathcal{A}^*_P \times_{\text{ad}} \Omega^{2,-}(X, \text{ad } P)$. The associated deformation complex at a SD connection A is given by

$$0 \longrightarrow \Omega^0(\text{ad } P) \xrightarrow{\mathrm{d}_A} \Omega^1(\text{ad } P) \xrightarrow{\mathrm{d}^-_A} \Omega^{2,-}(\text{ad } P) \longrightarrow 0$$

where d^-_A is the projection of d_A onto the anti-self-dual subspace $\Omega^{2,-}(X, \text{ad } P)$ of $\Omega^2(X, \text{ad } P)$. We denote the cohomology groups of this complex by H^0_A, H^1_A and H^2_A and use h^0, h^1, h^2 to denote their dimensions. The tangent space $T_A\mathcal{M}$ can again be identified with the middle cohomology group H^1_A, and the ellipticity of the complex implies that we have an elliptic deformation operator $T_A = \mathrm{d}^*_A \oplus \mathrm{d}^-_A : \Omega^1(X, \mathrm{ad}\ (P)) \to \Omega^0(X, \mathrm{ad}\ (P)) \oplus \Omega^{2,-}(X, \mathrm{ad}\ (P))$ with the index of T equal to $h^0 - h^1 + h^2$. Hence we have described the moduli space $\mathcal{M}^*_X$ of *irreducible* SD instantons as the zero locus of a Fredholm system. The existence of reducible SD connections can be prevented by placing certain topological conditions on $b^+(X)$ and the topological quantum number k of P (see [FU84]), in which case $\mathcal{M}_X$ is given by a smooth Fredholm system also. Whenever $h^0 = h^2 = 0$, then the operator T_A will be surjective and hence Theorem 3.1 will imply that $\mathcal{M}^*_X$ is a smooth manifold of dimension $-\mathrm{ind}(T_A)$.

The problem now is that the operator T_A *need not* be surjective, that is, that the SD Fredholm system is not transversal in general. When A is an irreducible SD connection we must have $h^0 = 0$ since $H^0_A = \ker d_A$, but to give criteria under which $h^2 = 0$ is more difficult. The simplest explicit result in this direction is the following:

THEOREM 3.7. *(Atiyah-Hitchin-Singer) [AHS78] Suppose that X is a self-dual 4-manifold and has non-negative scalar curvature that is somewhere non-zero (for instance, $X = S^4$ with the standard Riemannian metric). Then $h^2 = 0$ and $\mathcal{M}_X$ is a smooth manifold of dimension*

$$8c_2(\mathrm{ad}\ P) - \frac{1}{2}\dim(G)(\chi(X) - \sigma(X))$$

where σ is the signature of X.

We ought to say a word about what it means for a Riemannian 4-manifold to be **self-dual**. In the standard way, we may regard the Riemann curvature tensor as an endomorphism of $\bigwedge^2 TM$ and decompose it canonically as $R = R_1 + R_2 + W$ where R_1 is the constant-curvature part, R_2 is the traceless Ricci part and W is the conformal part, often called the Weyl tensor. Under the splitting of $\bigwedge^2 TM$ into positive and negative parts under the Hodge star, we have a corresponding decomposition $W = W^+ + W^-$. We then say the manifold is *self-dual* if $W^- = 0$. Admittedly, this condition is rather opaque; to see where this hypothesis is used in [AHS78] it is necessary to go through some detailed calculations with indices, which we shall not be doing here (but see [BB85, p.377]).

Proof. The first part of Theorem 3.7 follows from an elegant application of *Bochner's method*. To show that $\mathrm{coker}T = 0$ it suffices to show that the adjoint operator T^* has zero kernel. We can find an explicit formula for T by taking $\alpha \in \Omega^0(X, \mathrm{ad}\ P)$ and $\beta \in \Omega^{2,-}(X, \mathrm{ad}\ P)$ and observing that

$$\langle \alpha \oplus \beta, Ts\rangle = \langle \alpha, \mathrm{d}^*_A s\rangle + \langle \beta, \mathrm{d}^-_A s\rangle = \langle \mathrm{d}_A\alpha + \mathrm{d}^*_A\beta, s\rangle$$

so that $T^* = \mathrm{d}_A + \mathrm{d}^*_A$. Now we can observe that $\langle \mathrm{d}_A\alpha, \mathrm{d}^*_A\beta\rangle = \langle F_A\alpha, \beta\rangle = 0$ since A is self-dual and β is anti-self-dual. Hence $||T^*(\alpha \oplus \beta)||^2 = ||\mathrm{d}_A\alpha||^2 + ||\mathrm{d}_A\beta||^2$, and so $\alpha \oplus \beta$ is in the kernel of T if and only if $\mathrm{d}_A\alpha = \mathrm{d}^*_A\beta = 0$. We have already seen that $\mathrm{d}_A\alpha = 0$ implies that $\alpha = 0$ whenever A is irreducible. So now we need to show that $\mathrm{d}^*_A\beta = 0$ implies that $\beta = 0$. We observe that $\mathrm{d}^*_A\beta = 0$ actually implies that $\mathrm{d}_A\beta = 0$ by the fact that β is ASD. Hence β is in the kernel of the generalised Laplace operator $\Delta_A = \mathrm{d}_A\mathrm{d}^*_A + \mathrm{d}^*_A\mathrm{d}_A$ associated to the elliptic deformation complex. Now we can use a **Bochner-Weiztenböck identity** for ASD 2-forms on self-dual 4-manifolds: $\Delta_A = \mathrm{d}^*_A\mathrm{d}_A + \kappa/3$, where κ denotes the scalar curvature of X (see [BB85, p.377]). Now we can use the standard Bochner argument; if $\Delta_A\beta = 0$, then $0 = \langle \Delta_A\beta, \beta\rangle = ||d_A\beta||^2 + \langle \kappa\beta, \beta\rangle/3$. This implies that β is a constant and hence

$$\frac{\beta}{3}\int_X \kappa = 0$$

But the integral above is strictly positive by our assumption on X. Hence β is identically zero.

The second part of the theorem follows by calculating the index of T when $h^2 = 0$. As in our calculation before, we begin by complexifying and observing that T has symbol $\sigma(T) = 1_{\mathrm{ad}\ (P)} \otimes \sigma(T_0)$, where $T_0 : \Omega^1(X) \to \Omega^0(X) \oplus \Omega^{2,-}(X)$ is the standard 'untwisted' operator on forms given by $\mathrm{d}^* \oplus \mathrm{d}^-$. By similar arguments to those in the previous section, the Atiyah-Singer index theorem then implies that

$$\mathrm{ind}(T) = -8\ c_2(\mathrm{ad}\ P) + \dim(G)\mathrm{ind}(T_0)$$

where c_2 is used to denote the second Chern class of the complexification of ad P. Now the index of T_0 may be evaluated by elementary arguments by noting that it is the deformation operator associated to the usual de Rham complex

$$0 \longrightarrow \Omega^0(X) \xrightarrow{\ \mathrm{d}\ } \Omega^1(X) \xrightarrow{\ \mathrm{d}^-\ } \Omega^{2,-}(X) \longrightarrow 0$$

so that $\mathrm{ind}(T_0) = b_0 - b_1 + b_2^-$. Since we have $b_0 = 1$, and $\chi = b_0 - b_1 + b_2 - b_3 + b_4 = 1 - 2b_1 + 2b_2$ (by Poincaré duality) as well as $\sigma = b_2^+ - b_2^-$, with $b_2^+ + b_2^- = b_2$, we therefore have that $\mathrm{ind}(T_0) = 1/2(\chi - \sigma)$, completing the index calculation. ∎

We would like to consider instanton moduli spaces on more general 4-manifolds X, and we can do this by allowing the Riemannian metric on X to vary. The smoothness of $\mathcal{M}_X$ a suitably 'generic' metric follows from the result of Freed and Uhlenbeck:

THEOREM 3.8. *(Generic Metrics) [DK90, 4.3.18] Suppose X is a compact, simply-connected, oriented four-manifold and $G = \mathrm{SO}(3)$ or $\mathrm{SU}(2)$. Then there is a dense subset of the space of conformal classes of metrics on X such that for all representatives g of equivalence classes of metrics g in this set, every equivalence class in the moduli space of irreducible self-dual connections on X has $h^2 = 0$. In this case we say that g is a* ***generic metric****.*

We shall indicate in the next section how this theorem can be proved.

3.5 EXAMPLE: MODULI SPACES OF PSEUDOHOLOMORPHIC CURVES

The general construction of moduli spaces of pseudoholomorphic curves is rather intricate. We consider in this section simply the closed genus 0 case with no marked points and describe the general construction in subsequent sections. Let Σ be a genus 0 Riemann surface (without boundary) and let (X, ω) be a symplectic manifold with a compatible almost-complex structure J (see Appendix A). We will fix a homology class $\beta \in H_2(X)$ for the maps $u : \Sigma \to X$ by specifying that $u_*[\Sigma] = \beta$. We use $\mathcal{M}(X, \beta, J)$ to denote the set of all (smooth) J-holomorphic curves $u : \Sigma \to X$ representing this homology class β. We wish to give this set the structure of a smooth manifold. We begin by reformulating the J-holomorphic curve equation in terms of a Fredholm system. The base space $\mathcal{B}$ will be some Banach completion of $C^\infty(\Sigma; X, \beta)$, the space of smooth maps $\Sigma \to X$ representing the homology class β. As described in Appendix B, this has a Sobolev completion $\mathcal{B}_{k,p}$ defined as a smooth Banach manifold for $kp > 2$. The tangent space to $\mathcal{B}$ at the map $u : \Sigma \to X$ is given formally by $\Omega^0(\Sigma, u^*TX)$, and under Sobolev completion, this becomes $\Omega^0_{k,p}(\Sigma, u^*TX)$. The derivative du of a map $u \in C^\infty(\Sigma, X)$ may be regarded as a section of the complex vector bundle $T^*\Sigma \otimes_J u^*TX$ over Σ. The complex anti-linear part $\bar\partial_J u$ (as described in Appendix A) may then be considered as an element of $\Omega^0(\Sigma; \wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$. Taking this space to be the fibre $\mathcal{E}_u = \Omega^0_{k-1,p}(\Sigma; \wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$ yields a Banach bundle $\mathcal{E} \to \mathcal{B}$ over the base space $\mathcal{B}$. It is not difficult to show that $\bar\partial_J : \mathcal{B} \to \mathcal{E}$ gives a smooth section [MS12, p.44]. The moduli space $\mathcal{M}(X, \beta, J)$ of J-holomorphic curves is then the zero locus $\bar\partial_J^{-1}(0)$ of this section. The moduli space will consist of smooth maps and so will be independent of the choice of (sufficiently large) k by the regularity theory for harmonic maps [Ohm08]. This moduli space will however generally depend on our choice of J, but only up to cobordism; in Chapter 6 we shall discuss how such results may be proved.

By the general theory discussed in §1 of this chapter, if the vertical linearisation $D_u^V \bar\partial_J$, considered as an operator

$$D_u : W^{k,p}(\Sigma, u^*TX) \to W^{k-1,p}(\Sigma, \wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$$

can be shown to be both Fredholm and surjective for all J-holomorphic maps u, then the moduli space $\bar\partial_J^{-1}(0)$ will be a smooth manifold of dimension given by $\mathrm{ind}(D_u)$. The first of these requirements can be shown by explicit calculation of the operator D_u, but the second will not be true in general. However, away from the *multiply covered curves*, we shall show that for a generic compatible almost-complex structure J, the moduli space will indeed be smooth. In the course of the proof we also describe how a similar argument may used to show the generic metrics result for instanton moduli spaces stated above.

DEFINITION 3.3. *A* ***branched covering map*** *is any holomorphic map of Riemann surfaces. A curve $u : \Sigma \to X$ is called* ***multiply covered*** *if it is of the form*

$$\begin{tikzcd} \Sigma' \arrow[r, "\phi"] \arrow[dr, dashed] & \Sigma \arrow[d, "u"] \\ & X \end{tikzcd}$$

for ϕ some branched covering map of degree 2 or greater. Else, the curve is called ***simple****. The curve u is called* ***somewhere injective*** *if there exists a point $z \in \Sigma$ such that $du(z) \neq 0$ and z is the only point in Σ that maps to $u(z)$. We call such a point an* ***injective point****.*

Recall the following result (see [MS12], Prop 2.3.1)

PROPOSITION 3.1. *Any simple pseudoholomorphic curve is somewhere injective. In fact, the set of injective points is open and dense in Σ.*

We now use $\mathcal{B}^*$ to denote the space of maps $\Sigma \to X$ that are simple. The resulting moduli space will now be denoted $\mathcal{M}^*(\beta, J)$ and will consist of *simple* pseudoholomorphic curves. We aim to prove the following theorem:

THEOREM 3.9. *[MS12, p.45] For a Baire set of smooth ω-compatible almost complex structures J on (X, ω), the operator D_u us surjective for all simple J-holomorphic curves u, so that the moduli space $\mathcal{M}^*(\beta, J)$ of simple curves is a smooth manifold.*

The discussion of the surjectivity of D_u will rest upon having an appropriate explicit formula. In local holomorphic coordinates $z = s + it$ on Σ and local coordinates ψ^α on X, we may consider u as a map $u^\alpha : \mathbb{C} \to \mathbb{R}^{2n}$. The anti-linear differential $\bar{\partial}_J u$ takes the form

$$\bar{\partial}_J(u^\alpha) = \frac{1}{2}(\partial_s u^\alpha + J(u^\alpha)\partial_t u^\alpha)\mathrm{d}s + \frac{1}{2}(\partial_t u^\alpha - J(u^\alpha)\partial_s u^\alpha)\mathrm{d}t$$

and hence we see that the pseudoholomorphic curve equation in local coordinates is equivalent to the single first-order Cauchy-Riemann type partial differential equation [MS12, p.19]

$$\frac{\partial u^\alpha}{\partial s} + J(u^\alpha)\frac{\partial u^\alpha}{\partial t} = 0$$

If we take a family u_λ with $\frac{\mathrm{d}}{\mathrm{d}\lambda}|_{\lambda=0} u_\lambda = X^\alpha$, then we can calculate $D_u X$ in local coordinates by

$$\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0}\left(\frac{\partial u^\alpha_\lambda}{\partial s} + J(u^\alpha_\lambda)\frac{\partial u^\alpha_\lambda}{\partial t}\right)$$

which we easily find to be

$$D_u X^\alpha = \bar{\partial}_J X^\alpha + (\nabla_{X^\alpha} J)(u)\partial_t u^\alpha$$

From this local expression we can see that D_u is a linear Cauchy-Riemann operator, differing from the J-holomorphic differential by a zero-order perturbation term. Hence, by standard theory [MS12, Appendix A] this operator will be elliptic and Fredholm.

To derive a global formula for D_u we need to choose a *connection*. Taking $u_t \in C^\infty(\Sigma, X)$ a family of curves, we have $\bar{\partial}_J(u)$ in the space $\Omega^0(\wedge^{0,1}T^*\Sigma \otimes_J u_t^* TX)$, but we have no way of identifying these different vector spaces. Identifying them is equivalent to defining a connection on $\mathcal{E}$ over $\mathcal{B}$. This connection must preserve J in order to preserve the complex linear and anti-linear parts. If ∇ is the Levi-Civita connection on X associated to the metric $\omega(\cdot, J\cdot)$ induced from the ω-compatible almost-complex structure J, then ∇ preserves J if and only if J is integrable [MS12, p.41]. When this is not the case, we can define a new covariant derivative $\tilde{\nabla}$ by setting $\tilde{\nabla}_X Y = \nabla_X Y - \frac{1}{2}J(\nabla_X J)Y$, which will always preserve both J and the metric $\omega(\cdot, J\cdot)$. Evidently, it cannot still be torsion-free. Now, given some $\alpha \in \Omega^{0,1}(u^*TX)$, we may use parallel transport along geodesics to identify this with some element in $\Omega^{0,1}(u_t^*TX)$. Formally, given $z \in \Sigma$, then for all $V \in T_z\Sigma$ we have $\alpha(V) \in T_{u(z)}X$. We write $u_t = \exp_u(tX)$ for some $X \in \Omega^0(u^*TX)$; this means $\exp_u(tX)(z) = \exp_{u(z)}(tX(z))$, the geodesics being taken with respect to the connection ∇, not $\tilde{\nabla}$. Then we can transport $\alpha(V)$ along $\exp_u(tX)(z)$ using $\tilde{\nabla}$ (not ∇) to give some $\Phi_{u_t}(\alpha(V)) \in T_{u_t(z)}X$. In this way we produce $\Phi_{u_t}(\alpha) \in \Omega^{0,1}(u_t^*TX)$. It is clear that this map is a linear isomorphism.

Now we can take $u_t = \exp_u(tX)$ and calcuate, for $z \in \Sigma$, $V \in T_z\Sigma$,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\left(\Phi^{-1}_{u_t(z)}\bar{\partial}_J(u_t)(z)(V)\right) = \left.\tilde{\nabla}_{X_{u_t}(z)}\bar{\partial}_J(u_t)(z)(V)\right|_{t=0}$$

by the definition of the covariant derivative of a section along a curve $t \mapsto u_t(z)$. The rest of the calculation is straightforward, simply expanding out $\tilde{\nabla}$ in terms of the definition above and applying the torsion-freeness of ∇ [MS12, p.42]. This yields

$$D_u X = \frac{1}{2}(\nabla X + J(u)\nabla X \circ i) - \frac{1}{2}J(u)(\nabla_X J)(u)\partial_J u$$

which we can see also takes the form of a Cauchy-Riemann operator plus a zero-order perturbation. The same discussion carries over to this operator. Furthermore, we have a formal adjoint operator D_u^*, which by general theory will also be a first-order elliptic operator with coefficients of the same regularity as D_u [MS12, p.48]. Alternatively, this can be verified by computing a local expression for D_u^* as above.

The explicit form of the operator D_u also allows us to discuss the orientation and dimension of this moduli space. As before, the tangent space $T_u\mathcal{M}(\beta, J)$ is given by the kernel of the operator D_u and so whenever D_u is surjective then the index indD_u will give the dimension of the moduli space. Recall that the index of a Fredholm operator is homotopy invariant with respect to the operator norm topology. Using the homotopy given by

$$t \mapsto \frac{1}{2}(\nabla X + J(u)\nabla X \circ i) - (1-t)\frac{1}{2}J(u)(\nabla_X J)(u)\partial_J u$$

for $t \in [0,1]$, we can homotop D_u to the Dolbeault operator $\bar{\partial}_J$, whose index is simply given by the Riemann-Roch theorem:

$$\dim \mathcal{M}(\beta, J) = n(2-2g) + 2c_1(u^*TX)$$

see [MS12, Appendix B] for the details of this calculation. In the case where J is integrable, then the operator D_u is actually complex linear and so $T_u\mathcal{M}(\beta, J)$ inherits a natural complex structure and hence a canonical orientation. When this is not the case, one can demonstrate that the orientation can be transported in a compatible way along the homotopy from D_u to $\bar{\partial}_J$ above, hence yielding an orientation for the moduli space (see A.27 in [MS12]).

Now we describe the 'generic metrics argument' for simple pseudoholomorphic curves and irreducible instantons. We wish to introduce extra parameters into our moduli space that will give the operator D_u extra degrees of freedom to be surjective. In the case of pseudoholomorphic curves, this extra parameter will be the almost-complex structure J on X in the equation $\bar{\partial}_J u = 0$. We allow this to vary over the set $\mathcal{J}(X,\omega)$ of ω-compatible almost-complex structures J. In the case of ASD instantons, the extra parameter will be the metric g on the 4-manifold X, appearing in the equation $F_A^+ = 0$ as part of the definition of the Hodge star. In fact, because of the conformal invariance of this equation, it suffices to allow g to vary over $\mathcal{C}$, the space of conformal classes $[g]$ of metrics on X.

Proof. (of Theorems 3.9 and 3.8) To use the Banach-manifold formalism described earlier, we will not be able to work with smooth almost-complex structures (or smooth metrics), but will have to work with almost-complex structures of class C^ℓ for some $\ell \geq 1$ in order for $\mathcal{J}(X,\omega)$ to be a Banach manifold (as opposed to a Fréchet manifold). Showing that $\mathcal{J}^\ell$ is indeed a Banach manifold follows the general pattern outlined earlier, applying an implicit function theorem argument to show that it is a submanifold of the Banach space $\mathrm{End}_{C^\ell}(TX)$. The tangent space will be a subspace of this Banach space, given by those sections satisfying those conditions found by differentiating the conditions on J. Because we need the coefficients of D_u to have sufficient regularity, now we will need to take $1 \leq k \leq \ell$ in order to have D_u an operator from $W^{k,p}(u^*TX) \to W^{k-1,p}(\wedge^{0,1}t^*\Sigma \otimes_J u^*TX)$ whose kernel is independent of ℓ. The regularity theory for J-holomorphic curves will then imply that every J-holomorphic curve with $J \in C^\ell$ is actually itself C^ℓ [MS12, Proposition 3.1.10]; by taking ℓ sufficiently large, we can assume that every pseudoholomorphic curve we have to deal with in the following discussion has sufficiently many derivatives for all of the above maps to be well-defined. Once we have finished the argument, we can recover the $J \in C^\infty$ case of the statement of the Theorem by taking an intersection over all ℓ sufficiently large; since the topology on $\mathcal{J}$ actually changes to the C^∞ topology, one needs a small argument to show that this actually works (see [MS12, p.55]). The discussion of the space $\mathcal{C}^\ell$ of conformal classes of C^ℓ metrics is largely the same; here we will need to take $\ell \geq 3$ in order for the relevant operators to be well-defined.

Now one introduces the *universal moduli space*. Consider the Banach manifold given by $\mathcal{B}^*_{k,p} \times \mathcal{J}^\ell$ and take this to be the base space of our Fredholm system. Over this we define a Banach bundle with fibres

$$\mathcal{E}^{k-1,p}_{(u,J)} = W^{k-1,p}(\wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$$

The catch is now that $\mathcal{E}$ is only a $C^{\ell-k}$ Banach bundle. Earlier we defined trivialisations for our Banach bundle in terms of parallel transport with respect to the connection $\tilde{\nabla}$ that preserved the almost-complex structure J. When J is only of class C^ℓ, this parallel transport, and hence the corresponding trivialisations, will only be of class $C^{\ell-k}$ on the relevant Sobolev spaces [MS12, p.50]. We define a section F of this Banach bundle by $F(u,J) = \bar{\partial}_J(u)$; this is of class $C^{\ell-k}$ and hence we have a Fredholm system. The corresponding moduli space is called the **universal moduli space** and is denoted $\mathcal{M}^\ell(\beta, J)$. In the ASD case, we take the base space to be $\mathcal{B}^* \times \mathcal{C}$, with the section now given by mapping $(A,[g])$ to $F_A^{+,g}$, the anti-self-dual part of the curvature with respect to the conformal class $[g]$, in the space $\Omega^{2,+_g}(X, \mathrm{ad}\, P) \times_{\mathrm{ad}} \mathcal{A}^*$.

Now we prove that the universal moduli space is a Banach manifold of class $C^{\ell-k}$ by showing that the operator $D_{(u,J)}F$ is surjective for all u simple and J-holomorphic. The operator $D_{(u,J)}F$ is defined as a map

$$D_{(u,J)}F : W^{k,p}(u^*TX) \oplus T_J\mathcal{J}^\ell \to W^{k-1,p}(\wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$$

and one can easily show, using a similar calculation to that for D_u above, that it is given by

$$D_{(u,J)}F(X,Y) = D_uX + \frac{1}{2}Y(u) \circ du \circ i = D_uX + C_u(Y)$$

From this we can observe that the coefficients of the first-order part are of class C^ℓ when u is J-holomorphic. Firstly, we shall restrict to the case where $k=1$ and show that $D_{(u,J)}F$ is surjective on the spaces above with $k=1$. The analysis in this case is slightly subtle, since maps of class $W^{1,p}$ are *a priori* only continuous, not differentiable; we shall see that actually this is not a problem because we shall only need to use pointwise arguments for *pseudoholomorphic* curves. More importantly, while the above operator $D_{(u,J)}F$ is *not* a Fredholm operator, it has closed range, as a perturbation of the Fredholm operator D_u by the closed operator $C_u(Y)$. Now, since a closed subspace is equal to its closure, to show that $D_{(u,J)}F$ is surjective, it suffices to show that the range is dense. Here we can use the **orthogonal-to-the-image** trick, applying the Hahn-Banach theorem in the form

LEMMA 3.4. *(Hahn-Banach) If Y is a subspace of a Banach space X and $x \in X$ is a point such that $d(x, Y) > 0$, then there exists a linear functional $\phi \in X^*$ such that $\phi|_Y \equiv 0$ and $||\phi|| = 1$.*

Hence, to show that the range of $D_{(u,J)}F$ is dense, it suffices to show that any element Z of the dual space $L^{p^*}(\wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$ that is zero on the range of $D_{(u,J)}F$ is necessarily identically zero. The application of this trick to the universal linearisation operator the case of ASD instantons is identical. Next we use a **unique continuation** trick. If Z is zero on the range of $D_{(u,J)}F$ then we must have

$$\int_\Sigma \langle Z, D_u X\rangle = \int_\Sigma \langle Z, C_u(Y)\rangle = 0 \tag{3.1}$$

for every $(X, Y) \in W^{1,p}(u^*TX) \oplus T_J\mathcal{J}^\ell$. The first of these identities implies that Z is a weak solution of $D_u^* Z = 0$. By our earlier observations about D_u^*, we may apply elliptic regularity to conclude that $Z \in W^{\ell+1,r}$ for any $r > 0$, since the coefficients of D_u^* are also in C^ℓ. Now we observe that $D_u D_u^* Z = 0$. Since $D_u D_u^*$ is an elliptic operator with highest-order term equal to the Laplacian Δ, we can apply **Aronszajn's Theorem** on unique continuation of solutions to elliptic PDE (see [MS12, p.22]) to conclude that if Z vanishes on some open set, then Z is identically zero. For irreducible ASD instantons, we use a very similar unique continuation lemma (see [DK90, p.154]). We shall now use the extra condition provided by the second equality in 3.1, along with the extra degrees of freedom in $T_J\mathcal{J}^\ell$, to force Z to vanish on some small open set. The important result here is that the endomorphisms in $T_J\mathcal{J}$ act *transitively* on TX at every point (see [MS12, Lemma 3.2.2]).

In order to derive this result, we shall assume u is simple and J-holomorphic and use proposition 3.1 to conclude that there is an open dense set of points z in Σ where $du(z) \neq 0$. If $Z(z) = 0$ at every such point, we are done. Suppose otherwise, that $Z(z) \neq 0$ for some injective point $z \in \Sigma$ where u is injective. Then one can use the fact that $T_J\mathcal{J}$ acts transitively on the tangent space at the point $T_{u(z)}X$ to find some $Y \in T_J\mathcal{J}$ such that

$$\langle Z(z), C_u(Y)\rangle = \langle Z(z), Y \circ du(z) \circ i(z)\rangle > 0$$

because $Z(z), du(z)$ are both non-zero. Extending Y to a small neighbourhood of z and multiplying Z by a smooth cutoff function we see by continuity that the above inequality will hold in some neighbourhood of z. Hence we have

$$\int_\Sigma \langle Z, C_u(Y)\rangle > 0$$

for this choice of Y. This contradiction establishes the result. To extend this surjectivity to general k, we suppose that we have some $Z \in W^{k-1,p}(\wedge^{0,1}T^*\Sigma \otimes_J u^*TX)$ and use surjectivity in the $k = 1$ case to write $D_{(u,J)}F(X, Y) = Z$ for some pair (X, Y) having only one weak derivative in L^p. But then $D_u X = Z - \frac{1}{2}C_u(Y)$ is in $W^{k-1,p}$ since u is J-holomorphic and $Y \in C^\ell$. Hence by elliptic regularity for D_u we have $X \in W^{k,p}(u^*TX)$ also. Thus we have deduced the smoothness of the universal moduli space $\mathcal{M}^\ell(\beta, J)$ of simple curves.

To complete the proof of Theorem 3.9, we use the following elementary Lemma concerning Fredholm operators

LEMMA 3.5. *[MS12, Lemma A.3.6] Suppose $D : X \to Y$ is Fredholm and $L : Z \to Y$ is bounded. Furthermore, assume that $D \oplus L : X \oplus Z \to Y$ is surjective. Then $D \oplus L$ has a bounded right inverse. Moreover, $\Pi : \ker(D \oplus L) \to Z$ is Fredholm also, with* $\ker \Pi \cong \ker D$ *and* $\mathrm{coker}\Pi \cong \mathrm{coker}D$.

In the case of pseudoholomorphic curves, we consider this lemma with $X = T_u B$, $Y = \mathcal{E}_{(u,J)}$ and $Z = T_J\mathcal{J}$, with $D = D_u : T_u\mathcal{B} \to \mathcal{E}_{(u,J)}$ and $L = C_u : T_J\mathcal{J} \to \mathcal{E}_{(u,J)}$. The sum $D \oplus L$ is then the surjective operator $D_{(u,J)}F$. The operator Π arises as follows. Consider the projection map $\pi : \mathcal{M}^\ell(\beta, J) \to \mathcal{J}^\ell$. This is a map of class $C^{\ell-1}$ between Banach manifolds of class $C^{\ell-1}$ and the derivative $D_{(u,J)}\pi$ of π is simply the operator Π when restricted to $T_{(u,J)}\mathcal{M}^\ell(\beta, J)$. Moreover, the fibres of π are exactly the moduli spaces $\mathcal{M}^*(\beta, J)$ for different values of J. Now we may apply the **Sard-Smale Theorem**:

THEOREM 3.10. *(Sard-Smale) [FU84, p.70] Suppose $f : X \to Y$ is a C^ℓ map (for $\ell \geq 1$) between (separable) Banach manifolds such that $D_x f$ is Fredholm for all $x \in X$, with index less than or equal to $\ell - 1$. Then the set of $y \in Y$ such that $D_x f$ has a (bounded) right inverse for all $x \in f^{-1}(y)$ is a Baire set in Y.*

Whenever ℓ is large enough so that $\ell - 2 \geq \text{ind}D_{(u,J)}\pi = \text{ind}D_u$, the set of regular values of π is a Baire set in $\mathcal{J}^\ell$. But by Lemma 3.5 we know that $D_{(u,J)}\pi$ is surjective precisely when D_u is surjective for the almost-complex structure J, and hence we know that there is a Baire set in $\mathcal{J}^\ell$ such that the moduli space $\mathcal{M}^*(\beta, J)$ is transversal. This completes the proof. The argument for ASD instantons is completely identical at this stage. ■

We can actually deduce further results from this setup when we consider *paths* of almost-complex structures, or Riemannian metrics. This shall feature in our discussion of topological quantum field theory in Chapter 6.

3.6 COMPACTIFICATION OF MODULI SPACES

The moduli spaces $\mathcal{M}(\beta, J)$ described in the previous section will not in general be compact. To see this, recall that the J-holomorphic curve equation is invariant under conformal reparametrisations of the domain. In the case $\Sigma = \mathbb{CP}^1$ we therefore have an action of the non-compact automorphism group PSL$(2, \mathbb{C})$ on the moduli space $\mathcal{M}(\beta, J)$. Hence the moduli space cannot possibly be compact; we must consider the quotient by the action of the reparametrisation group. However, because this action is *not* smooth, we cannot use the infinitesimal theory developed in the previous section in the context of gauge theory.

This is an example of the phenomenon that we have observed in Chapter 2, namely that sequences pseudoholomorphic maps with uniformly bounded energy could have energy concentrate at a single point and fail to have a smoothly convergent subsequence. Instead, they converged to a new map that was no longer in the moduli space as it had a different homotopy class. However, in the case where all the maps in a sequence represent a homology class with minimal (non-zero) energy, then it is clear from the Gromov-Uhlenbeck compactness theorem for pseudoholomorphic maps in Chapter 2 that no bubbles can occur. In this case the moduli space $\mathcal{M}(\beta, J)/\text{PSL}(2, \mathbb{C})$ will be compact in the C^∞ topology.

In general, however, the issues involved in the compactification are quite subtle. First, some useful terminology.

DEFINITION 3.4. *Suppose $\phi_n : S^2 \to S^2$ is a sequence of conformal automorphisms of S^2 that converges smoothly on compact subsets of $S^2 \setminus \{z_\infty\}$ to the constant map at z_0. Then we say that ϕ_n is a sequence of rescalings with **source** z_∞ and **target** z_0.*

These rescalings can be regarded as 'magnifying' the area around the target 'bubbling point' where the energy is concentrating. Now, recalling the rescaling argument from Remark 2.2, and combining this with the Gromov-Uhlenbeck compactness theorem for the resulting rescaled map, it is not hard to see that in general we have:

THEOREM 3.11. *(Bubble Formation; [MS12], Prop 4.7.1) Suppose u_n is a sequence of J-holomorphic maps $\hat{\mathbb{C}} \to X$ that converges in C^∞ away from some concentration point $z_0 \in S^2$ to a J-holomorphic curve $u : \hat{\mathbb{C}} \setminus \{z_0\} \to X$. Assume furthermore that the bubble energy m_0 at z_0 is strictly positive. Then there exists a J-holomorphic map $v : \hat{\mathbb{C}} \to X$; a subsequence of u_n; a sequence of rescalings $\phi_n : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ with source $z_\infty \in \hat{\mathbb{C}}$ and target z_0; and a finite collection of points $z_1, \ldots, z_n \in \hat{\mathbb{C}}$ such that:*

- *the sequence $u_n \circ \phi_n$ converges in C^∞ to v away from the points $z_1, \ldots, z_n$ and z_∞;*
- *the bubbling energies $m_1, \ldots, m_n$ at $z_1, \ldots, z_n$ exist and are strictly positive;*
- *(Bubbles Connect) we have $u(z_0) = v(z_\infty)$;*
- *(Conservation of Energy) the energy m_0 is equal to $E(v) + \sum_i m_i$;*
- *(Stability) if v is constant then $n \geq 2$.*

The problem with defining the compactification in general is precisely the points $z_1, \ldots, z_n$ on the sphere bubble v. They represent bubbles that have formed on the bubble v itself. One could imagine applying this theorem repeatedly, producing a *bubble tree* of holomorphic spheres. This behaviour does indeed occur; the model example is as follows. Define $u_n : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \times \hat{\mathbb{C}} \times \hat{\mathbb{C}}$ by $z \mapsto (z, nz, n^2 z)$. This will then converge smoothly to $u : z \mapsto (z, \infty, \infty)$ on compact subsets of $S^2 \setminus \{0\}$. All of the maps in the sequence have fixed energy, equal to 12π by conformal invariance, but the limiting map u will only have energy of 4π. Hence we have non-zero energy concentrating at 0. So take $\phi_n(z) = z/n$ and extend to a holomorphic map $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$. Then ϕ_n converges smoothly to 0 on compact subsets of $\hat{\mathbb{C}} \setminus \{\infty\}$ and so has source ∞ and target 0. The rescaled sequence $u_n \circ \phi_n(z) = (z/n, z, nz)$ then converges smoothly on compact subsets of $\hat{\mathbb{C}} \setminus \{0, \infty\}$ to $v(z) = (0, z, \infty)$. Here we see that $v(\infty) = u(0)$, and so $\infty \in S^2$ is the point connecting the sphere bubble v to u. But the energy of the map v is only 4π, whereas the energy of the maps v_n is always 4π. Hence the point $0 \in S^2$ is another positive-energy bubbling point for the sequence $v_n = u_n \circ \phi_n$. Now we can apply the same procedure again. If we define a new sequence of rescalings, $\psi_n(z) = z/n^2$, then we see that the sequence $w_n = u_n \circ \psi_n(z) = (z/n^2, z/n, z)$ converges smoothly to the map $w(z) = (0, 0, z)$ on compact subsets of $\hat{\mathbb{C}} \setminus \{\infty\}$.

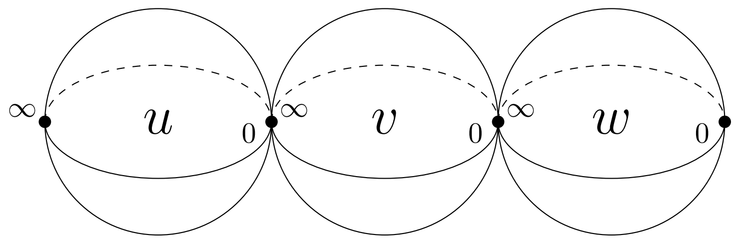

Figure 3.1: A simple example of a bubble tree.

Again, $w(\infty) = v(0)$ and hence we can regard this as another bubbling point. The energy of w is then given by 4π and so now we have recovered all the energy of the original sequence u_n. An illustration of this situation is given in Figure 3.1.

3.6.1 STABLE MAPS AND THE GROMOV TOPOLOGY

Providing a compactification therefore requires a new formalism, where the domains of the maps are allowed to change.

DEFINITION 3.5. *(Stable Maps) A* ***prestable J-holomorphic map*** *consists of a finite collection of J-holomorphic spheres $u_i : S^2 \to X$, joined to each other to make a finite tree via* ***nodal points*** *$x_{u_i,u_j} \in S^2$ with $u_i(x_{u_i,u_j}) = u_j(x_{u_j,u_i})$, along with a finite collection of* ***marked points*** *$z_k \in S^2$ each of which is on some particular sphere u_i. We say that such a collection is a* ***stable J-holomorphic map*** *if there are no constant maps u_i (called* ***ghost bubbles****) where the sum of the number of marked points and stable points on u_i is 2 or less. Finally, we say that two prestable J-holomorphic maps $\{u_i\}$ and $\{v_i\}$ are* ***equivalent*** *if there exists a collection of automorphisms $\phi_i \in \mathrm{PSL}(2,\mathbb{C})$ and a bijection $f : \{v_i\} \to \{u_i\}$ such that $v_i \circ \phi_i = f(v_i)$, the nodal points have $\phi_i(x^v_{v_i,v_j}) = x^u_{f(v_i),f(v_j)}$, and if z^v_k is a marked point on some v_i, then $\phi_i(z^v_k) = z^u_k$.*

A few remark are in order. With this notion of equivalence, it follows that stable maps will have a finite automorphism group [MS12, p.116]. We can define the energy of a prestable map in the obvious way, as the sum over the energies of each of the spheres. We also have an obvious notion of the homology class of a stable map. We now define $\mathcal{M}_k(A, J)$ now to be the set of equivalence classes of stable maps with a total of k marked points representing the homology class $A \in H_2(X; \mathbb{Z})$. Since the marked points are fixed under equivalence, we can define an evaluation map $\mathcal{M}_k(A, J) \to X^k$. We can also fix the number of holomorphic spheres in the stable map, and the tree formed by their connecting nodes; these are called the **strata** of the moduli space. We define the **domain of a stable map** to be the collection of spheres joined together as a tree along the nodal points, considered as a nodal Riemann surface. Finally, we must be careful that these moduli spaces will not always be the same as those discussed in the previous section when $A = 0$ due to the stability requirement.

Now we need to describe what it means for a sequence of maps $u_n : S^2 \to X$ to converge to a stable map in the above sense; we shall use our example discussed above to motivate the conditions we impose. Firstly, we had two sequences of rescaling automorphisms, $\phi_n(z) = z/n$ and $\psi_n(z) = z/n^2$ such that $u_n \circ \phi_n$ converged to the first sphere bubble v away from nodal points, and $u_n \circ \psi_n$ converged to the second sphere bubble w away from nodal points. Secondly, we observe that the bubbling energy at each nodal point is equal to the energy of all the sphere bubbles originating from that point. Finally, rather than regarding the two bubbles v, w as rescaled versions of the sequence u_n, *we can also regard them simply as originating from adjacent bubbles by rescaling*. For instance, we can view w as the C^∞ limit of the compositions $v_n \circ \phi_n = u_n \circ \psi_n \circ (\phi_n)^{-1}$. Here we regard $\psi_n \circ (\phi_n)^{-1}$ as being the 'transition' scaling function between the adjacent sphere bubbles v and w. In particular it is a rescaling with source ∞ and target 0. Similarly, we can view v as the C^∞ limit of the compositions $w_n \circ (\phi_n)^{-1} = u_n \circ \psi_n \circ (\phi_n)^{-1}$. This idea shall serve as the motivation for the following:

DEFINITION 3.6. *(Gromov Convergence) We say that a sequence $u_n : S^2 \to X$ of J-holomorphic spheres* ***Gromov-convergent*** *to a stable map $\{u_i\}$ if there exists a collection of sequences $\phi_{n,i} \in \mathrm{PSL}(2,\mathbb{C})$ such that $u_n \circ \phi_{n,i}$ converges smoothly to u_i away from the nodal points, and the following conditions are satisfied:*

- *(Conservation of Energy) the bubbling energy at every nodal point is equal to the sum over the energies of all bubbles originating from that point;*
- *(Transitions) for every nodal point x_{u_i,u_j}, the sequence $\phi_{n,i} \circ (\phi_{n,j})^{-1}$ has source x_{u_i,u_j} and target x_{u_j,u_i}.*

Note that by the stability condition in 3.11, we do indeed expect the sequence to converge to a *stable* map. The energy conservation requirement expresses the fact that no energy is lost in the rescaling process; we always perform the 'minimal' rescaling at each point. There is also the obvious corresponding notion for sequences of J-holomorphic spheres with marked points, which requires the marked points to converge also. Given such a notion of convergence (and extending it to allow for convergence of sequences of stable maps), we can define a topology on $\mathcal{M}_k(A, J)$ by declaring a set to be closed if and only if it contains all of its sequential limits. This is the **Gromov topology**. The crucial (and difficult) concerning the Gromov topology is the compactness it provides:

THEOREM 3.12. *(Gromov Compactness)[MS12, p.150] In the Gromov topology, the space $\mathcal{M}_k(A, J)$ is* ***compact*** *and* ***metrisable*** *(and in particular, Hausdorff). Moreover, the evaluation map* $\mathrm{ev} : \mathcal{M}_k(A, J) \to X^k$ *is continuous with respect to this topology.*

We have not, however, shown that $\mathcal{M}_k(A, J)$ is a smooth manifold. In fact, this is not true in general, or even generically, as the different strata will have differing dimensions. We will discuss in the next section to what extent can can still make sense of these spaces as *moduli spaces*. Also, we considered the above compactification only in the case of holomorphic spheres; we shall require the case of disks in Chapter 4, where the appropriate modifications that must be made will be discussed.

3.7 VIRTUAL INTEGRATION

Returning to the problem discussed at the start of this chapter, we can ask ourselves what kind of structure the zero locus of a Fredholm system possesses when its section fails to achieve transversality, beyond simply that of a topological space. In many of the applications of the above moduli spaces to defining invariants and homology theories, the only operation it will be necessary for us to perform with these spaces is *integration*. Thus we wish to find some theory that allows us to extend integration to topological moduli spaces defined by Fredholm systems.

First we shall need to recall some basic facts about the de Rham cohomology of vector bundles. Let $\pi : E \to X$ be a rank n *oriented* vector bundle over a compact manifold X. We use $\Omega^*_{cv}(E)$ to denote the set of differential forms on E that have compact support in the *vertical direction*; this is a covariant functor under inclusions of open subsets, and hence we can define a de Rham cohomology $H^*_{cv}(E)$ with compact vertical support, just as in the case with compactly supported cohomology [BT82]. We define a map $\pi_* : \Omega^*_{cv}(E) \to \Omega^{*-n}(X)$ called **integration along fibres** as follows. Firstly, suppose E we have a trivialisation $E \cong \mathbb{R}^n \times X$ of E. Then any differential form β on E in $\Omega^*_{cv}(E)$ can be written as a sum over

$$\phi(x^1, \dots, x^n, y) \cdot \pi^*\alpha \wedge dx^{i_1} \wedge \cdots dx^{i_k}$$

where $x^1, \dots, x^n$ are the coordinates on $\mathbb{R}^n$, α is a form on $X \times \mathbb{R}^n$, and $\phi(x^1, \dots, x^n, y)$ is a function on $\mathbb{R}^n \times X$ compactly supported in the x variables. Then we define $\pi_*(\beta)$ to be 0 if β has no summands containing $dx^1 \wedge \cdots \wedge dx^n$, and otherwise

$$\pi_* \left(\phi(x^1, \dots, x^n, y) \cdot \pi^*\alpha \wedge dx^1 \wedge \cdots dx^n\right) = \pi^*\alpha \cdot \int_{\mathbb{R}^n} \phi(x^1, \dots, x^n, y) \mathrm{d}x^1 \cdots \mathrm{d}x^n$$

This clearly yields a well-defined map $\pi_* : \Omega^*_{cv}(E) \to \Omega^{*-n}(X)$. In the case where E is non-trivial, we use a partition of unity on X with local trivialisations of E to glue these π_* maps together. It is easy to check that this is independent of trivialisations and commutes with the de Rham differential d. Hence the integration along fibres map induces a **pushforward** $\pi_* : H^*_{cv}(E) \to H^{*-n}(X)$ on cohomology. The *Thom isomorphism theorem* states that this map is an isomorphism. Hence there is some $\Theta \in \Omega^n_{cv}(E)$ such that $\pi_*\Theta = 1$; this is called the **Thom form**.

Now suppose $s : X \to E$ is a section of the vector bundle E, and suppose furthermore that s is transversal to the zero section. Then the level set $S = s^{-1}(0)$ is an oriented embedded submanifold of X, and we have

THEOREM 3.13. *The Poincaré dual of $S = s^{-1}(0)$ is exactly $s^*\Theta$. In other words, for all $\omega \in \Omega^*_c(X)$, we have*

$$\int_S \omega = \int_X \omega \wedge s^*\Theta$$

This follows from combining Theorems 12.4 and 12.8 in [BT82]. What will be so important about this theorem is that the right hand side of the equation still makes sense even when s is *not* transversal to the zero section, and hence not necessarily a smooth manifold. Hence what we can do is use the left hand side as a *definition* of **virtual integration**, written

$$\int_S^{\mathrm{vir}} \omega := \int_X \omega \wedge s^*\Theta$$

It is worth noting that the definition of this integral on S depends on the *whole system* $X \xrightarrow{s} E$; the section and the vector bundle are part of the data, although they are not included in the notation. Many useful theorems still hold for virtual integration, such as a version of Stokes' theorem:

$$\int_S^{\mathrm{vir}} d\omega = \int_X d\omega \wedge s^*\Theta = \int_X d(\omega \wedge s^*\Theta) = \int_{\partial X} \omega \wedge s^*\Theta = \int_{\partial S}^{\mathrm{vir}} \omega$$

where we have used the fact that Θ is closed. We can also define **virtual integration along fibres**. Firstly, suppose the base space X has a map $\pi : X \to L$ where L is some compact, oriented smooth manifold, and suppose that the restriction of π to S is a smooth submersion. Then for any $\alpha \in \Omega^*(S)$ and $\beta \in \Omega^*(L)$, we have the transfer equation

$$\int_L \pi_*(\alpha) \wedge \beta = \int_S \alpha \wedge \pi^*\beta$$

Again, the right hand side can be defined 'virtually' even when S is not a smooth manifold. By Poincaré duality applied to L, giving the values of the linear functional $\int_L \pi_*(\alpha) \wedge \beta$ for all $\beta \in \Omega^*(L)$ uniquely determines a form on L, which we can denote $\pi_*^{\mathrm{vir}}(\alpha)$.

Since the failure of transversality is the main difficulty in constructing moduli spaces, the advantage of the virtual integration is that we still have some notion of integration even when this fails. However, the above is only defined for finite-dimensional manifolds. We shall first need to show how the various infinite-dimensional Fredholm systems considered above can be locally reduced to finite-dimensional transversality problems. This technique is called **cokernel perturbation**, and proceeds as follows. Suppose we have a Fredholm system $\mathcal{B} \xrightarrow{s} \mathcal{E}$, with $\mathcal{M} = s^{-1}(0)$ the moduli space. If the cokernel of the vertical linearisation D_xS of S is non-zero, then the moduli space $\mathcal{M}$ need not be a smooth manifold at $x \in \mathcal{M}$. Moreover, the dimension of the kernel of D_xS must be less than the *virtual dimension* given by $\mathrm{ind}(D_xS)$. By the upper semicontinuity of the the dimension of the cokernel of DS, there exists an open neighbourhood U_x of x over which $\mathcal{E}$ is trivial, along with a lifting $\tau_x : \mathrm{coker}(D_xS) \to \mathcal{E}_x$ of the cokernel to $\mathcal{E}_x$, such that for all $y \in U_x$, the natural map $\mathrm{im}\ (\tau_x) \to \mathcal{E}_y \to \mathrm{coker} D_yS$ is surjective. Let $V_x = \mathrm{im}\ (\tau_x)$. Because $\mathcal{E}$ is trivial over U_x, we can identify the fibres $\mathcal{E}_x \to \mathcal{E}_y$ via a linear isomorphism $\phi_{x,y}$. Now we can define a new section $\tilde{S} : U_x \times V_x \to \mathcal{E}|_{U_x}$ by $\tilde{S}(y,v) \mapsto s(y) + \phi_{x,y}(v)$. Because the differential of $\tilde{S}$ is now clearly surjective, it follows that $\mathbf{V}_x = \tilde{S}^{-1}(0)$ is genuinely a smooth manifold, called the **extended moduli space** (in U_x). It also has finite dimension, equal to the virtual dimension of $\mathcal{M}$, plus the dimension of $\mathrm{coker}(D_xS)$. There is a finite-dimensional smooth vector bundle $\mathbf{E}_x \to \mathbf{V}_x$ over $\mathbf{V}_x$ defined by $\mathbf{E}_x = \mathbf{V}_x \times \mathrm{coker}(D_xS)$, called the **obstruction bundle**. This has a canonical section $\sigma_x(y,v) = ((y,v),v)$. Then the original moduli space $\mathcal{M} \cap U_x$ is given exactly by the zero locus $\sigma_x^{-1}(0)$. In fact, there is a homeomorphism of $\sigma_x^{-1}(0)$ onto an open subset of $\mathcal{M}$. This collection, $\mathbf{V_x} \xrightarrow{\sigma_x} \mathbf{E_x}$ is called a **virtual neighbourhood** of x. Hence we have succeeded in locally reducing the structure of $\mathcal{M}$ to a zero locus of a canonical *finite-dimensional* section over a smooth manifold. When $\mathcal{M}$ is compact, we can take a finite open covering of $\mathcal{M}$ by such virtual neighbourhoods of points x_i. Over each of these virtual neighbourhoods, we can now apply the virtual integration formulation from above. However, each of these virtual neighbourhoods can have $\mathbf{V}_{x_i}$ of different dimension, and $\mathbf{E}_{x_i}$ of different rank (but always having the same difference in dimension). Thus it is not clear how the local information fits together.

The answer to this question is provided by the notion of a *virtual manifold*:

DEFINITION 3.7. *(Virtual Manifold) A* ***virtual manifold*** *is a collection* $\{X_I\}$ *indexed by subsets* I *of the powerset of* $\{1,\ldots,n\}$*, satisfying the following three conditions:*

1. *(Vector Bundle Condition) if* $I \subseteq J$*, then we have open sets* $X_{JI} \subseteq X_J$ *and* $X_{IJ} \subseteq X_I$ *and a smooth map* $\Phi_{IJ} : X_{JI} \to X_{IJ}$ *that forms a vector bundle;*
2. *(Cocycle Condition) for all* $I \subseteq J \subseteq K$*, we have* $\Phi_{IK} = \Phi_{IJ} \circ \Phi_{JK}$ *on the intersection of their domains, that is, we have a commutative diagram:*

$$\begin{tikzcd}
X_{K,J} \supset X_{K,I} \arrow[dd] \arrow[dr] & \\
& X_{K,J} \subseteq X_{J,I} \arrow[dl] \\
X_{IJ} \supset X_{I,K} &
\end{tikzcd}$$

3. *(Fibre Product Condition) for all* I, J*, the diagram*

$$\begin{tikzcd}
& X_{I\cup J, I\cap J} \arrow[dl] \arrow[dr] & \\
X_{J, I\cup J} \arrow[dr] & & X_{J, I\cup J} \arrow[dl] \\
& X_{I\cap J, I\cup J} &
\end{tikzcd}$$

is a pullback diagram in the category of vector bundles.

We then have a notion of a **virtual vector bundle** over a virtual manifold; it consists of a virtual manifold $\{E_I\}$ and a collection $\pi_I : E_I \to X_I$ of vector bundles (with the same indexing), satisfying a number of compatibility conditions, such as that $E_{JI} = E_J|_{X_{JI}}$ and that $E_{JI} = (\Phi_{IJ})^*(E_{IJ} \oplus X_{JI})$. The latter condition in particular implies that the dimensions of E_I and X_I change in the same way. Similarly, we can define a **virtual section**, consisting of a collection of maps $\sigma_I : X_I \to E_I$ that make all the obvious diagrams commute. Taking the zero locus $\left\{\sigma_I^{-1}(0)\right\}$ of a virtual section will yield another virtual manifold. We say that a topological space M has the structure of a **virtual system** if there exists a virtual section of a virtual vector bundle over a virtual manifold such that the zero locus $\left\{\sigma_I^{-1}(0)\right\}$ forms an open cover of M. If these virtual sections σ_I are all transversal to the zero section, then in fact this topological space M will in fact be a *smooth manifold*. By the cokernel perturbation described above, we can always give the moduli space $\mathcal{M}$ of a Fredholm system the structure of a virtual system. The extra structure of a virtual system is precisely what we need in order to define virtual integration on the topological space M.

In order to do this, we would like to take a partition of unity and a compatible collection of Thom forms and patch the local virtual integrals together. But there are a number of complications to applying this method. Firstly, it is not obvious that such partitions of unity must exist in general. We then also need to show that the resulting integration theory is independent of these choices. There are also more fundamental problems with using Fredholm systems to define all moduli spaces. In our discussion of stable maps above, we allowed the domain of the J-holomorphic map to change, and possibly include nodal curves. We also needed to consider the action of the reparametrisation group on the moduli space. Since this action will not generally be smooth, we cannot quotient out our Fredholm system by the group action to leave a new smooth Fredholm system as we did in the case of instantons; the resulting Fredholm system will only be *weakly smooth*. The work done by [CLW14] allows one to extend virtual integration to this case. However this is still a work in progress.

For the rest of this thesis we shall assume that all moduli spaces are smooth manifolds.

Chapter 4

INFINITE-DIMENSIONAL MORSE THEORY

4.1 FINITE-DIMENSIONAL MODELS

To motivate the following constructions, it will be useful to spend some time describing an alternative formulation of classical Morse homology inspired by a famous paper of Witten [Wit82]. Let M be a compact Riemannian manifold and let $f : M \to \mathbb{R}$ be a generic smooth function, with all critical points non-degenerate. We define the Morse chain group C to be given by the free $\mathbb{Z}$-module generated by the critical points of f, with grading given by the index of the critical points. The idea of Witten is to define a boundary operator between critical points by 'counting' the number of gradient flow lines of f between them. We carry this out in an excessively formal setting in order to show the analogies with the constructions we will perform later.

For critical points x, y of f, we define a Banach manifold, the **path space** $\mathcal{B}(x, y)$ between the two critical points, to be the $W^{1,2}$ completion of the space of maps $\mathbb{R} \to M$ that tend to x, y at $\pm\infty$ respectively. Over this space we have a Banach bundle $\mathcal{E} \to \mathcal{B}(x, y)$ with fibres $\mathcal{E}_u$ given by $L^2(\mathbb{R}, u^*TM)$, along with a smooth section $P : \mathcal{B} \to \mathcal{E}$ given by $P(u) = \frac{\mathrm{d}u}{\mathrm{d}t} - (\nabla f)(u)$. This yields a Fredholm system and it can be shown, using arguments similar to those in Chapter 3, that for a Baire set of metrics on M this Fredholm system is regular. Hence we have a moduli space, denoted $\mathcal{M}(x, y)$ and called the (unreduced) **moduli space of gradient flow lines**. It has dimension given by $\mathrm{ind}(P) = \mathrm{ind}(y) - \mathrm{ind}(x)$. See the book [Sch93] for all the details of the above.

The manifold $\mathcal{M}(x, y)$ will not in general be compact; a sequence of gradient flow lines between a pair of critical points can tend to a *broken* flow line, one that travels via a third critical point z. The concept is formulated rigorously as follows:

DEFINITION 4.1. *A* ***translation vector*** $\bar{T}$ *is a collection of real numbers* $T(1) < T(2) < \cdots < T(n)$*. We say a sequence* u_k *of gradient flow lines between critical points* x *and* y *is* ***chain convergent*** *if there exists a sequence* $\bar{T}_k$ *of translation vectors with* $T(i)_k - T(i+1)_k \to \infty$ *as* $k \to \infty$*, and a collection* $(v^1, \cdots, v^n)$ *of gradient flow lines connecting* x *to* y*, such that for any compact interval in* $\mathbb{R}$*, there exists a subsequence of* u_k *such that the sequence* $u_k(t + T(i)_k)$ *converges smoothly to* $v^i(t)$*. We call* $(v^1, \cdots, v^n)$ *a* ***broken gradient flow line****. We say* n *is the* ***length of the chain****.*

The following theorems are central:

THEOREM 4.1. *[Mer14, p.7] Suppose* x, y *are critical points with index difference* $m + 1$*.*

- *(Compactness) Any sequence in* $\mathcal{M}(x, y)$ *with no convergent subsequence possesses a subsequence that is chain convergent to a broken gradient flow line of length at most* m*;*
- *(Gluing) Conversely, for any broken gradient flow line of length at most* m*, there exists a sequence of gradient flow lines from* x *to* y *that is chain convergent to it.*

Hence we can define a *compactification*, denoted $\overline{\mathcal{M}}(x, y)$, of $\mathcal{M}(x, y)$ by adding in all of the limits of the subsequences to the moduli space. The above two theorems provide the two directions of the bijection in the following corollary:

COROLLARY 4.1. *Suppose the index difference of* x *and* y *is* 2*. Then the boundary of the moduli space* $\overline{\mathcal{M}}(x, y)$ *is given by*

$$\partial\overline{\mathcal{M}}(x, y) = \coprod_{z:\, \mathrm{ind}(x) - \mathrm{ind}(z) = 1} \mathcal{M}(x, z) \times \mathcal{M}(z, y) \tag{4.1}$$

With further work, one can even show that $\overline{\mathcal{M}}(x, y)$ will be orientable [Mer14, p.8]. Note however that all of the moduli spaces considered above have a free $\mathbb{R}$ action by translating solutions in the 'time' variable. The quotient is therefore a smooth manifold also and will have dimension $\mathrm{ind}(y) - \mathrm{ind}(x) - 1$. *From now on, we shall use $\mathcal{M}(x, y)$ to denote the translation-reduced moduli space.* Observe now that if the *difference* in the index of x, y is 1, then $\overline{\mathcal{M}}(x, y)$ will simply be a finite set of (signed) points, as we have $m = 0$ in the theorems above. Thus we can define a boundary map by

$$d(x) = \sum_{\mathrm{ind}(y)-\mathrm{ind}(x)=1} \#\left\{\overline{\mathcal{M}}(x, y)\right\} y$$

that is, by counting (with signs) the number of gradient flow lines between x and y. Of course, if we compute d^2, we should find that it is zero:

$$d^2(x) = \sum_{\mathrm{ind}(y)-\mathrm{ind}(x)=1} \sum_{\mathrm{ind}(y)-\mathrm{ind}(z)=1} \#\left\{\overline{\mathcal{M}}(x, y)\right\} \#\left\{\overline{\mathcal{M}}(y, z)\right\} z = \#\left\{\partial\overline{\mathcal{M}}(x, z)\right\} z$$

But counting all the gradient flow lines x to y to z is the same as counting the broken flow lines between x and z, that is, counting the number of boundary points of the compactified moduli space $\overline{\mathcal{M}}(x, z)$. Of course, the signed count of the number of boundary points of a compact one-dimensional oriented manifold with boundary must always be zero, demonstrating that, indeed, $d^2 = 0$. We call the resulting homology groups the **Morse homology groups** associated to the function f on M.

Importantly, as we can see from the formulas above, at no point did we need the value of the index of a single critical point, only the *difference* in the index between two critical points. This opens up an avenue for studying Morse homology in infinite-dimensional situations where the number of positive and negative eigenvalues is not necessarily finite. Although proving that the moduli spaces of paths are are actually smooth orientable manifolds may require substantial analysis, so long as there exists a compactification with boundary given by equation 4.1, the fact that $d^2 = 0$ then reduces to a statement that is almost trivial.

In fact, the requirement that the moduli spaces $\mathcal{M}(x, y)$ actually be smooth manifolds is overly restrictive. When $\mathcal{M}(x, y)$ is genuinely a smooth compact oriented 0-dimensional manifold then the signed count $\#\mathcal{M}(x, y)$ of points is equal to the integral

$$\int_{\mathcal{M}(x,y)} 1$$

As we saw in Chapter 3, this integral could instead be defined using *virtual integration*; we call this **virtual counting**. Now suppose that we wish to evaluate $\#\left\{\partial\overline{\mathcal{M}}(x, z)\right\}$. Then using Stokes' theorem for virtual integration, we have

$$\#\left\{\partial\overline{\mathcal{M}}(x, z)\right\} = \int_{\partial\overline{\mathcal{M}}(x,z)} 1 = \int_{\overline{\mathcal{M}}(x,z)} d(1) = 0$$

and hence we still have $d^2 = 0$ if we instead use virtual counting to define the boundary map in Morse homology. This illustrates two important points; firstly, that the discussion of transversality for moduli spaces of gradient flow lines is largely irrelevant; secondly, that applying Stokes' theorem to the boundary structure of moduli space leads to interesting algebraic structures. This idea will be central in Chapters 5 and 6.

4.2 EXAMPLE: THE CHERN-SIMONS FUNCTIONAL

The idea of instanton Floer homology is to perform an infinite-dimensional version of the Morse homology described in Chapter 2, now using the **Chern-Simons functional** defined on the space of gauge equivalence classes of flat SU(2) connections on a (compact, oriented, Riemannian) 3-manifold Y. As with the discussion of the action functional above, great care must be taken to choose the appropriate covering space.

Let $P \to Y$ be a principal SU(2) bundle over a 3-manifold Y. Because $\mathbb{HP}^\infty = \mathrm{BSU}(2)$ has 4-skeleton given by $\mathbb{HP}^1$, cellular approximation implies that P is necessarily trivial [FU84, Appendix A]. *For the rest of this section we fix a trivialisation of P.* We shall see at the end that this choice does not affect the resulting theory. Given this fixed trivialisation, we have a specified identification of the space of SU(2) connections $\mathcal{A}_P$ on P with $\Omega^1(Y, \mathrm{ad}\ P)$, which are simply Lie algebra valued 1-forms. We denote such sections using the symbol A. Now, given a 4-manifold X of the form $Y \times Z$ for a 1-manifold Z, we use $\mathbf{P}$ to denote the pullback of P to X. It is also trivial, and there is a canonical trivialisation of $\mathbf{P}$ coming from our chosen trivialisation of P. Hence we may identify $\mathcal{A}_\mathbf{P}$ with the space of sections $\Omega^1(X, \mathrm{ad}\ \mathbf{P})$ in the same way. Then because $X = Y \times Z$, we have an identification $\Gamma(Z, \Omega^1(Y, \mathrm{ad}\ P) \oplus \Omega^0(Z, \mathfrak{g})) \cong \Omega^1(X, \mathrm{ad}\ \mathbf{P}))$ given by $(A, A_0) \mapsto A(t) + A_0(t)dt$, where t is the coordinate on Z. We denote such connections using the symbol

$\mathbf{A}$. We also need to carefully distinguish the gauge groups $\mathcal{G}_P$ and $\mathcal{G}_{\mathbf{P}}$. Using the trivialisation of P, we can identify $\mathcal{G}_P$ with $C^\infty(Y, G)$ and $\mathcal{G}_{\mathbf{P}}$ with $C^\infty(X, G)$; we write $\mathcal{B}_P$ for $\mathcal{A}_P/\mathcal{G}_P$ and $\mathcal{B}_{\mathbf{P}}$ for $\mathcal{A}_{\mathbf{P}}/\mathcal{G}_{\mathbf{P}}$. Then we have as in Chapter 2 that $\mathcal{A}_P$ is a principal $\mathcal{G}_P$-bundle over $\mathcal{B}_P$ when we restrict to irreducible connections. We shall also identify $\mathcal{G}_P$ as the subgroup of $\mathcal{G}_{\mathbf{P}}$ consisting of functions $X \to G$ that are constant along Z.

DEFINITION 4.2. *Given a trivialisation as above, we define the* ***Chern-Simons functional*** $\theta : \mathcal{A}_P \to \mathbb{R}$ *by*

$$\theta(A) = \frac{1}{8\pi^2} \int_Y \mathrm{Tr}(A \wedge dA + A \wedge A \wedge A)$$

Now we wish to consider the behaviour of the Chern-Simons functional under the action of the gauge group $\mathcal{G}_P$. We will need a preliminary lemma.

LEMMA 4.1. *Suppose* $\mathbf{A} \in \Omega^1(X, \mathrm{ad}\, P)$ *as above. Then we have the identity:*

$$d(\mathrm{Tr}(\mathbf{A} \wedge d\mathbf{A} + \frac{2}{3}\mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A})) = \mathrm{Tr}(F_{\mathbf{A}} \wedge F_{\mathbf{A}})$$

(for notation, see Appendix A).

Proof. Since this calculation is somewhat slippery, we provide details here. We will first need to observe that $\mathrm{Tr}([\alpha \wedge \beta]) = 0$ for any two $\alpha, \beta \in \Omega^*(X, \mathrm{ad}\, P)$. In particular, since we have

$$(A \wedge A \wedge A) \wedge A - (-1)^3 A \wedge (A \wedge A \wedge A) = [A \wedge A \wedge A, A]$$

we see that $\mathrm{Tr}(A \wedge A \wedge A \wedge A) = 0$. Now we begin by writing

$$A \wedge dA + \frac{2}{3} A \wedge A \wedge A = A \wedge F_A - \frac{1}{3} A \wedge A \wedge A$$

using the definition $F_A = dA + A \wedge A$. Since d commutes with trace, we then have

$$d(\mathrm{Tr}(A \wedge dA + \frac{2}{3} A \wedge A \wedge A)) = \mathrm{Tr}(dA \wedge F_A - A \wedge dF_A - \frac{1}{3}(dA \wedge A \wedge A - A \wedge dA \wedge A + A \wedge A \wedge dA))$$

Now we claim that $\mathrm{Tr}(dA \wedge A \wedge A) = \mathrm{Tr}(A \wedge A \wedge dA)$ and $\mathrm{Tr}(A \wedge dA \wedge A) = -\mathrm{Tr}(A \wedge A \wedge dA)$. This follows from the observations

$$\begin{aligned} dA \wedge A \wedge A - (-1)^2 dA \wedge A \wedge A &= [dA, A \wedge A] \\ A \wedge dA \wedge A - (-1)^3 A \wedge A \wedge dA &= [A \wedge dA, A] \end{aligned}$$

and the vanishing of the trace on commutators. Hence we get

$$\mathrm{Tr}(dA \wedge F_A - A \wedge F_A - A \wedge A \wedge dA)$$

Now we use $F_A = dA + A \wedge A$ again to yield

$$\mathrm{Tr}(F_A \wedge F_A - A \wedge A \wedge F_A - A \wedge dF_A - A \wedge A \wedge dA)$$

and the Bianchi identity $d_A F_A = dF_A + [A, F_A] = 0$ to give

$$\mathrm{Tr}(F_A \wedge F_A - A \wedge A \wedge F_A + A \wedge [A, F_A] - A \wedge A \wedge dA) = \mathrm{Tr}(F_A \wedge F_A - A \wedge (A \wedge F_A + A \wedge dA - [A, F_A]))$$

Finally, using $dA = F_A - A \wedge A$ and the fact that $\mathrm{Tr}(A \wedge A \wedge A \wedge A) = 0$, we derive

$$\mathrm{Tr}(F_A \wedge F_A - A \wedge (A \wedge F_A + A \wedge F_A - [A, F_A]))$$

Because we have

$$\mathrm{Tr}(A \wedge A \wedge F_A) = -\mathrm{Tr}(A \wedge F_A \wedge A)$$

by an argument similar to the above, this yields

$$\mathrm{Tr}(F_A \wedge F_A - A \wedge (A \wedge F_A - F_A \wedge A - [A, F_A])) = \mathrm{Tr}(F_A \wedge F_A)$$

since $[A, F_A] = A \wedge F_A + F_A \wedge A$. ∎

Now suppose we have $A \in \mathcal{A}_P$ and a gauge transformation $g \in \mathscr{G}_P$. Take a path $A(t)$ from A to A^g inside $\mathcal{A}_P$. This yields a connection on $\mathbf{P}$ over $Y \times [0,1]$ using the identification given above. Then Stokes' Theorem gives

$$\theta(A) - \theta(A^g) = \frac{1}{4\pi^2} \int_{Y \times [0,1]} \mathrm{Tr}(F_{\mathbf{A}}^2) \tag{4.2}$$

We claim that this is an integer, given by the degree of g, considered as a map $Y \to \mathrm{SU}(2)$. First we need to describe some simple topological constructions. An element of $\pi_1(\mathcal{B}_P)$ can be lifted to a map $[0,1] \to \mathcal{A}_P$ whose endpoints differ by a gauge transformation; the set of these maps is denoted $C^\infty_{\mathscr{G}}([0,1], \mathcal{A}_P)$, and by specifying particular lifts we may produce a map $\ell : \pi_1(\mathcal{B}_P) \to C^\infty_{\mathscr{G}}(I, \mathcal{A}_P)$. Given a path in $C^\infty_{\mathscr{G}}([0,1], \mathcal{A}_P)$, we can associate an element of $\pi_0(\mathscr{G}_P)$. Applying these sequentially is equivalent to taking the connected component of the holonomy corresponding to the loop in $\pi_1(\mathcal{B}_P)$, giving an element of $\pi_0(\mathscr{G}_P)$. The connected components of $\mathscr{G}_P$ are given exactly by the homotopy classes of maps $Y \to \mathrm{SU}(2)$. Using the clutching construction, these are in bijection with $\mathrm{SU}(2)$ vector bundles over $Y \times S^1$. In turn, these are classified by the second Chern class c_2. The Chern class of $\mathbf{P}$ over $Y \times S^1$ may be calculated by taking any connection $\mathbf{A}$ on $\mathbf{P}$ and performing the Chern-Weil integral $\int_X F_{\mathbf{A}}^2$. But the set of all connections on bundles $\mathbf{P}$ over $Y \times S^1$ is in bijection with $C^\infty_{\mathscr{G}}([0,1], \mathcal{A}_P)$; given a path in $C^\infty_{\mathscr{G}}([0,1], \mathcal{A}_P)$ whose endpoints differ by a gauge transformation g, this will yield a connection on the bundle $\mathbf{P}$ over $Y \times S^1$ obtained by the clutching construction from g. Hence we may rewrite equation 4.2 above as

$$\theta(A) - \theta(A^g) = \frac{1}{4\pi^2} \int_{Y \times S^1} \mathrm{Tr}(F_{\mathbf{A}}^2) = c_2(\mathbf{P})$$

which is necessarily an integer. That this integer may be identified with the degree of g follows from the standard clutching theory. We may summarise the above discussion as

THEOREM 4.2. *Then diagram below commutes:*

$$\begin{tikzcd}
& \pi_1(\mathcal{B}_P) \arrow[dl, "\ell"'] \arrow[dr, "\mathrm{hol}", "\sim"'] & \\
C^\infty_{\mathscr{G}}(I, \mathcal{A}_P) \arrow[rr] \arrow[dd, leftrightarrow] \arrow[dddr, bend left] & & \pi_0(\mathscr{G}_P) \arrow[d, "\sim"'] \\
& & \langle Y, \mathrm{SU}(2) \rangle \arrow[d, "\sim"'] \arrow[ddl, bend right, "\deg"'] \\
\coprod_{\mathbf{P}} \mathcal{A}_{\mathbf{P}} \arrow[rr] \arrow[dr, "\int_X F_\cdot"'] & & \mathrm{Vect}_{\mathrm{SU}(2)}(Y \times S^1) \arrow[dl, "c_2", "\sim"'] \\
& \mathbb{Z} &
\end{tikzcd}$$

An alternative way to derive this result is as follows. Recall that for a compact (semisimple) Lie group G, the third cohomology $H^3(G; \mathbb{Z})$ is always isomorphic to $\mathbb{Z}$. Moreover, this $\mathbb{Z}$ is generated by the Maurer-Cartan form

$$\frac{1}{24\pi^2} \mathrm{Tr}((g^{-1} \mathrm{d}g)^3)$$

One can then calculate explicitly that

$$\theta(A) - \theta(A^g) = \frac{1}{24\pi^2} \int_Y \mathrm{Tr}((g^{-1} \mathrm{d}g)^3) = \langle [Y], g^*[\mathrm{SU}(2)] \rangle = \deg(g)$$

The above theorem has a number of other important consequences. We now use $\mathscr{G}_P^0$ to denote the subgroup of $\mathscr{G}_P$ consisting of those gauge transformations of P that have degree 0 as maps $Y \to \mathrm{SU}(2)$. By the above theorem, $\mathscr{G}_P^0$ forms one connected component of $\mathscr{G}_P$ and the quotient $\mathscr{G}_P / \mathscr{G}_P^0$ is isomorphic to $\mathbb{Z}$. Hence the projection $\mathcal{A}_P / \mathscr{G}_P^0 \to \mathcal{A}_P / \mathscr{G}_P = \mathcal{B}_P$ forms the universal $\mathbb{Z}$-cover of $\mathcal{B}_P$. We also have a $\mathscr{G}_P^0$ principal bundle $\mathcal{A}_P \to \mathcal{A}_P / \mathscr{G}_P^0$ as before.

COROLLARY 4.2. *The Chern-Simons functional is well-defined on the quotient $\mathcal{A}_P / \mathscr{G}_P^0$, and well-defined as a map to $\mathbb{R}/\mathbb{Z}$ on the quotient $\mathcal{A}_P / \mathscr{G}_P$. In particular, θ is independent of our original choice of trivialisation for P up to some integer. Moreover, we have a commutative diagram*

$$
\begin{tikzcd}
\mathcal{A}_P \arrow[r, "\theta"] \arrow[d] & \mathbb{R} \arrow[dd] \\
\mathcal{A}_P/\mathscr{G}_P^0 \arrow[ur, "\theta"] \arrow[d] & \\
\mathcal{B}_P \arrow[r, "\tilde{\theta}"] & \mathbb{R}/\mathbb{Z}
\end{tikzcd}
$$

We can compute the derivative of the Chern-Simons functional θ by writing

$$\theta(A+a) = \frac{1}{8\pi^2}\int_Y \mathrm{Tr}\left((A+a)\wedge d(A+a) + \frac{2}{3}(A+a)\wedge(A+a)\wedge(A+a)\right)$$

for some $a \in \Omega^1(Y, \mathrm{ad}\ P)$. A similar calculation to the above then yields

$$\theta(A+a) = \theta(A) + \frac{1}{8\pi^2}\int_Y \mathrm{Tr}\left(2F_A\wedge a + a\wedge d_A a + \frac{2}{3}a\wedge a\wedge a\right)$$

and hence we must have

$$D_A\theta(a) = \frac{1}{8\pi^2}\int_Y \mathrm{Tr}(F_A\wedge a)$$

which is a 1-form on $\mathcal{A}_P$. Beware that equation (2.18) in [Don02] is incorrect and will not yield $D_A\theta$; the correct formula may be found in [Mor98, p.93]. Since the tangent space $T_{[A]}\mathcal{B}_P$ is isomorphic to $\Omega^1(Y, \mathrm{ad}\ P)/\mathrm{im}\ d_A$ and $D_A\theta$ is zero on the image of D_A (by Stokes' Theorem), the above expression descends to give a 1-form on $\mathcal{B}_P$. On $T_A\mathcal{A}_P = \Omega^1(Y, \mathrm{ad}\ P)$ we have an L^2 inner product given by

$$\langle\alpha,\beta\rangle = -\frac{1}{8\pi^2}\int_Y \mathrm{Tr}(\alpha\wedge *_3\beta)$$

This will induce a metric on the quotient space $\mathcal{B}_P$ also. The non-degeneracy of this L^2 norm implies in particular that $D_A\theta = 0$ if and only if $F_A = 0$, that is, A is flat. Therefore the critical points of the Chern-Simons functional on $\mathcal{B}_P$ are identified with the gauge equivalence classes of flat connections on P. We can also observe that

$$D_A(\theta)(a) = \frac{1}{8\pi^2}\int_Y \mathrm{Tr}(F_A\wedge a) = \langle *_3F_A, a\rangle$$

and hence that the gradient of the Chern-Simons functional on $\mathcal{A}_P$ is given by the map $A \mapsto *_3F_A$. Describing the gradient flow lines is somewhat more subtle and will involve our discussion of covering spaces above.

We begin by giving a simple description of the gradient flow lines on $\mathcal{A}_P$. As above, we may write any connection $\mathbf{A}$ on $\mathcal{A}_\mathbf{P}$ as a sum $A_0(t)\mathrm{d}t + A(t)$ for $A_0(t) \in \Gamma(\mathbb{R}, \Omega^0(Y, \mathrm{ad}\ P))$ and $A(t) \in \Gamma(\mathbb{R}, \Omega^1(Y, \mathrm{ad}\ P))$. We claim that there exists a gauge transformation $u \in \mathscr{G}_\mathbf{P}$ such that $u\mathbf{A}$ has $A_0 = 0$; this is called the **temporal gauge**. Firstly observe that

$$\nabla^{\mathbf{A}}_{\partial_t} s = \partial_t s + A_0(t)s$$

for any section $s \in \Gamma(X, E)$, where ∇^A is the covariant derivative induced by A on the vector bundle E. We hence want to find a gauge transformation u such that $u^{-1}\nabla^{\mathbf{A}}_{\partial_t} u = \partial_t$. Taking a section s, this is saying that

$$u(\partial_t(u^{-1}s) + A_0u^{-1}s) = \partial_t s$$

which becomes

$$u^{-1}(\partial_t u)s + u^{-1}A_0(us) = 0$$

hence giving an ordinary differential equation for u as a function of t:

$$\partial_t u + A_0(t)u = 0$$

See [DK90, p.48] for a similar argument. Because this ODE is linear and has smooth coefficients, given *any* smooth initial condition defined on a single time slice, we produce a smooth gauge transformation. Also, observe that if $A_0 = 0$ already, this equation simply becomes $\partial_t u = 0$, that is, u is constant in the t direction. Therefore we conclude that writing a connection $\mathbf{A}$ in temporal gauge is always possible, and is unique up to an action of $\mathscr{G}_P$. Now we shall write the ASD equation on $\mathbf{P} \to Y \times \mathbb{R}$ in this temporal gauge. By equations 4.3 below, it is clear that any ASD 2-form on $Y \times \mathbb{R}$ is given by $\phi \wedge \mathrm{d}t + *_3\phi$, where $\phi \in \Gamma(\mathbb{R}, \Omega^2(Y, \mathrm{ad}\ P))$. Writing out the curvature in temporal gauge gives

$$F_{\mathbf{A}} = d\mathbf{A} + \frac{1}{2}[\mathbf{A}, \mathbf{A}] = dA(t) + \frac{1}{2}[A(t), A(t)] = \frac{\partial A(t)}{\partial t} \wedge \mathrm{d}t + F_{A(t)}$$

Hence in order for the connection to be ASD we must have

$$\frac{\partial A(t)}{\partial t} = *_3 F_{A(t)}$$

which is precisely the gradient flow equation for the Chern-Simons functional. Hence: *the gradient flow lines of the Chern-Simons functional on $\mathcal{A}_P$ are exactly the instantons over $Y \times \mathbb{R}$ in temporal gauge.*

The statement above is extremely awkward for several reasons. We want to study the gradient flow lines of the Chern-Simons functional on $\mathcal{B}_P$, rather than $\mathcal{A}_P$. Also, we want to consider the gradient flow lines of the Chern-Simons functional as *gauge equivalence classes* of ASD connections, a subset of $\mathcal{B}_{\mathbf{P}}$, rather than fixing a gauge. The difficulty here is that $\mathcal{B}_{\mathbf{P}}$ has a much larger gauge group than $\mathcal{B}_P$. As described above, there is clearly a canonical correspondence between $C^\infty(\mathbb{R}, \mathcal{A}_P)$ and $\mathcal{A}_{\mathbf{P}}$. However, there is *no* canonical way to produce the desired map $C^\infty(\mathbb{R}, \mathcal{B}_P)$ to $\mathcal{B}_{\mathbf{P}}$. To produce a connection on $\mathcal{A}_{\mathbf{P}}$ from a path of connections in $\mathcal{B}_P$, one must first lift this path (canonically) to one in $\mathcal{A}_P$. This is the same as specifying a connection in the principal $\mathscr{G}_P$-bundle $\mathcal{A}_P \to \mathcal{B}_P$, that is, choosing a splitting for the exact sequence

$$0 \to \Omega^0(Y, \mathrm{ad}\ P) \xrightarrow{d_A} T_A\mathcal{A}_P \xrightarrow{\mathrm{d}\pi} T_{[A]}\mathcal{B}_P \to 0$$

using the identification of $\Omega^0(Y, \mathrm{ad}\ P)$ with the Lie algebra of $\mathscr{G}_P$. However, using the metric on $\mathcal{A}_P$ given above, there is a canonical choice of splitting given by the adjoint operator $d_A^* : \Omega^1(X, \mathrm{ad}\ P) \to \Omega^0(X, \mathrm{ad}\ P)$. Then we can define a **horizontal lift** for a map $[A](t)\mathbb{R} \to \mathcal{B}_P$ to be one satisfying the equation:

$$d^*_{A(t)}\left(\frac{\partial A(t)}{\partial t}\right) = 0$$

This lift is well-defined up to choice of starting point, that is, of gauge equivalence class of $A(0)$; the resulting connections on $\mathbf{P}$ will differ by a gauge transformation in $\mathscr{G}_{\mathbf{P}}$ that is constant in t and hence we will have a well-defined map to $\mathcal{B}_{\mathbf{P}}$, as desired. Moreover, one can see that all ASD instantons will automatically satisfy this horizontal condition when reformulated in terms of connections in $\mathcal{A}_{\mathbf{P}}$ as the requirement that $\iota_{\partial_t}(d^*_{\mathbf{A}}F_{\mathbf{A}}) = 0$. Furthermore, because the vector field $A \mapsto *_3F_A$ on $\mathcal{A}_P$ is horizontal, (it satisfies $d_A^*(*_3F_A) = 0$), and is $\mathscr{G}_P$-equivariant, it is hence pseudotensorial and so by the general theory in Appendix A it descends to give a well-defined vector field on the base space $\mathcal{B}_P$. Then one can see that paths in $\mathcal{B}_P$ that are the gradient flow lines of this vector field will exactly correspond to elements of the moduli space of instantons on $Y \times \mathbb{R}$. This will be the key to the construction of Floer homology for the Chern-Simons functional.

4.2.1 FREDHOLM THEORY

As with the compact case, it will be important to understand the linearisation operator $D_{\mathbf{A}}$ for the instanton equation over $Y \times \mathbb{R}$. Firstly, we will need to understand the case where $\mathbf{A}$ is a connection constant along $\mathbb{R}$, equal to the pullback of some $A \in \mathcal{A}_P$. Hence we shall spend some time demonstrating how to calculate it. Throughout this section, we shall use $\Lambda^k(Y)$ to denote $\Gamma(\mathbb{R}, \Omega^k(Y, \mathrm{ad}\ P))$.

We first recall some identifications. As above, we have $\Omega^0_{\mathfrak{g}}(X) \cong \Lambda^0(Y)$ and $\Omega^1(X, \mathrm{ad}\ P) \cong \Lambda^0(Y) \oplus \Lambda^1(Y)$. When Y is an oriented Riemannian manifold, it has a Hodge star $*_3$ and a volume form ω_Y. The orientation of Y induces a canonical orientation on $X = Y \times \mathbb{R}$ given by taking the volume form of X to be $\omega_X = \mathrm{d}t \wedge \omega_Y$. This yields a Hodge star $*$ on X with the property that

$$*(\phi \wedge \mathrm{d}t) = - *_3 \phi \tag{4.3}$$

$$*\psi = - *_3 \psi \wedge \mathrm{d}t \tag{4.4}$$

for $\phi \in \Lambda^1(Y)$ and $\psi \in \Lambda^2(Y)$, as can easily be seen. Then the self-dual 2-forms on X (with values in $\mathfrak{g}$) are exactly those having the form $\alpha \wedge \mathrm{d}t - *_3\alpha$ for some $\alpha \in \Omega^1_{\mathfrak{g}}(Y)$. Hence we have an identification of $\Omega^{2,+}_{\mathfrak{g}}(X)$ with $\Lambda^1(Y)$. Therefore the linearisation operator $D_{\mathbf{A}} : \Omega^1(X, \mathrm{ad}\ \mathbf{P}) \to \Omega^0(X, \mathrm{ad}\ \mathbf{P}) \oplus \Omega^{2,+}(X, \mathrm{ad}\ \mathbf{P})$ can be instead regarded as an operator $D_A : \Lambda^0(Y) \oplus \Lambda^1(Y) \to \Lambda^0(Y) \oplus \Lambda^1(Y)$.

THEOREM 4.3. *Under the above identifications, the linearisation operator $D_A : \Lambda^0(Y) \oplus \Lambda^1(Y) \to \Lambda^0(Y) \oplus \Lambda^1(Y)$ is given by*

$$D_A(\alpha, \beta) = \left(-\frac{\mathrm{d}}{\mathrm{d}t} + L_A\right)(\alpha, \beta) = -\frac{\mathrm{d}}{\mathrm{d}t}(\alpha, \beta) + (-d_A^*\beta, -d_A\alpha + *_3 d_A\beta)$$

Proof. Firstly, it is clear that we will have

$$d_{\mathbf{A}}\alpha(t) = \frac{\mathrm{d}\alpha}{\mathrm{d}t} \wedge \mathrm{d}t + d_A\alpha$$

and hence $d_{\mathbf{A}}(\alpha) = (\alpha'(t), d_A\alpha) \in \Lambda^0(Y) \oplus \Lambda^1(Y)$. Now we can write an inner product in $\Lambda^0(Y) \oplus \Lambda^1(Y)$ in the form

$$\langle(\eta, \xi), d_{\mathbf{A}}\omega\rangle = \langle(\eta, \xi), (\omega', d_A\omega)\rangle = \langle\eta, \omega'\rangle + \langle\xi, d_A\omega\rangle$$

Observe that we can write the inner product on $\Lambda^k(Y)$ as

$$\langle\alpha, \beta\rangle = \int_{-\infty}^{\infty} (\alpha, \beta)$$

for (α, β) the inner product on $\Omega^k(Y, \mathrm{ad}\ P)$. Hence using the (3-dimensional) adjoint of d_A and integration by parts will yield

$$\langle(\eta, \xi), d_{\mathbf{A}}\omega\rangle = \langle-\eta' + d_A^*\xi, \omega\rangle$$

and therefore the (4-dimensional) adjoint $d_{\mathbf{A}}^* : \Lambda^0(Y) \oplus \Lambda^1(Y) \to \Lambda^0(Y)$ is given by $d_{\mathbf{A}}^*(\alpha, \beta) = -\alpha' + d_A^*\beta$. Next we can compute $d_{\mathbf{A}}^+ : \Lambda^0(Y) \oplus \Lambda^1(Y) \to \Lambda^1(Y)$ by taking

$$d_{\mathbf{A}}^+(\alpha, \beta) = (1 + *)d_{\mathbf{A}}(\alpha\mathrm{d}t + \beta) = (1 + *)(\beta' \wedge \mathrm{d}t + d_A\beta + d_A\alpha \wedge \mathrm{d}t)$$

essentially using the definition of the exterior covariant derivative. Expanding out the Hodge star $(1 + *)$ and using the formulas in equation 4.3 will give

$$\big((\beta' + d_A\alpha) \wedge \mathrm{d}t + d_A\beta\big) - \big(- *_3 \beta' - *_3 d_A\beta \wedge \mathrm{d}t - *_3 d_A\alpha\big)$$

which we can rewrite as

$$\left(\left(\frac{\mathrm{d}}{\mathrm{d}t} - *_3 d_A\right)\beta + d_A\alpha\right) \wedge \mathrm{d}t - *_3\left(\left(\frac{\mathrm{d}}{\mathrm{d}t} - *_3 d_A\right)\beta + d_A\alpha\right)$$

Hence under the isomorphism given above, $d_{\mathbf{A}}^+ : \Lambda^0(Y) \oplus \Lambda^1(Y)\Lambda^1(Y)$ is given by $d_{\mathbf{A}}^+(\alpha, \beta) = d_A\alpha + \beta' - *_3 d_A\beta$. Finally, we see that the linearisation operator $D_{\mathbf{A}}$ is

$$D_{\mathbf{A}}(\alpha, \beta) = -(d_{\mathbf{A}}^* \oplus d_{\mathbf{A}}^+)(\alpha, \beta) = (-\alpha' - d_A^*\beta, -\beta' - d_A\alpha + *_3 d_A\beta)$$

exactly as claimed. ■

Remark 4.1. The above formula seems to be incorrect in [Don02], equation (2.24). The final results are entirely equivalent if we reverse the direction of $t \in \mathbb{R}$.

It will be important to observe that L_A is a self-adjoint first-order elliptic operator. Furthermore, the following proposition is not difficult to show using the techniques from Chapter 3, section 2:

PROPOSITION 4.1. *[Don02, p.24] The kernel of the operator $L_A : \Omega^0_{\mathfrak{g}}(Y) \oplus \Omega^1_{\mathfrak{g}}(Y) \to \Omega^0_{\mathfrak{g}}(Y) \oplus \Omega^1_{\mathfrak{g}}(Y)$ is the direct sum of H^0_A and H^1_A, the twisted cohomology groups associated to A.*

We shall say that the connection A is **non-degenerate** if H^1_A is zero, and call A **reducible** if $H^0_A = 0$ (we shall see in Chapter 6 that this is equivalent to the definition of reducibility given in Chapter 3). When both are the case, the operator L_A will be invertible and we shall call A **acyclic**.
Remark 4.2. Usually with Morse homology we would simply have the kernel of L_A equal to H^1_A, which is the kernel of the Hessian. Here the extra H^0_A term arises from the group action and measures the size of the stabiliser of A.

Now we can state the main theorem we aim to prove in this section:

THEOREM 4.4. *Under the definitions of Sobolev spaces on the tube $Y \times \mathbb{R}$ given in Appendix B,*

1. *When A is an acyclic connection, then $D_A = \frac{\mathrm{d}}{\mathrm{d}t} + L_A$ is a bounded, invertible operator $W^{1,2}(X) \to L^2(X)$;*
2. *When $\mathbf{A}$ is a connection over X with flat acyclic limits at $\pm\infty$, then the operator $D_{\mathbf{A}}$ is bounded and Fredholm on $W^{1,2}(X) \to L^2(X)$;*

Proof. The proof of (1) involves **separation of variables**. Firstly, using the theory of Sobolev spaces on tubular manifolds considered in Appendix B, we see immediately that D_A defines a bounded linear operator $W^{1,2}(X) \to L^2(X)$. Now, since L_A is an invertible self-adjoint elliptic operator on the *compact* manifold Y, the space $L^2(Y)$ has an orthogonal decomposition into smooth eigenfunctions ϕ_λ of L_A, with eigenvalues λ that are real, non-zero, of finite multiplicity and unbounded (see for instance [Nic07, Theorem 10.4.19]). To see that D_A is surjective, suppose ρ is a given smooth compactly supported section on $Y \times \mathbb{R}$. We can decompose ρ as an infinite sum over eigenfunctions

$$\rho(t) = \sum_\lambda \rho_\lambda(t)\phi_\lambda$$

for $\rho_\lambda : \mathbb{R} \to \mathbb{R}$ compactly supported smooth functions. With respect to this eigenfunction decomposition, the equation $D_A f = \rho$ becomes

$$\frac{\mathrm{d}f_\lambda}{\mathrm{d}t} + \lambda f_\lambda = \rho_\lambda$$

if we imagine taking a similar eigenfunction decomposition of f into $\{f_\lambda(t)\}_\lambda$. This is simply an ordinary differential equation, with solution given explicitly by

$$f_\lambda(t) = \mathrm{e}^{-\lambda t}\int_{-\infty}^{t} \mathrm{e}^{\lambda s}\rho_\lambda(s)\mathrm{d}s$$

when $\lambda < 0$, and by

$$f_\lambda(t) = -\mathrm{e}^{-\lambda t}\int_{t}^{\infty} \mathrm{e}^{-\lambda s}\rho_\lambda(s)\mathrm{d}s$$

Since all these solutions decay exponentially as $t \to \pm\infty$, it is easy to argue that the series defined by

$$f(t) = \sum_\lambda f_\lambda(t)\phi_\lambda$$

converges uniformly over $Y \times \mathbb{R}$ to a weak solution of $D_A f = \rho$. Elliptic regularity then implies that f is actually smooth. By the density result proved in Appendix B, it follows that D_A is surjective. It is also immediate from separation of variables that $f = 0$ is the only L^2 solution to $Df = 0$. Hence D_A is injective also.

The proof of (2) involves two steps. Firstly, we consider the case where $\mathbf{A}$ actually reaches the limiting flat acyclic connections at finite values of t, that is, there exists some $T > 0$ such that for all $|t| > T$, $\mathbf{A}(t)$ is constant at some acyclic flat connection. It is clear that $D_{\mathbf{A}}$ is still a bounded operator; to show that $D_{\mathbf{A}}$ is Fredholm, we shall show that it is invertible modulo compact operators, it has a **parametrix**. The proof of this step involves a 'patching' argument to glue together different inverses for $D_{\mathbf{A}}$ over different domains, as is done, for instance, to construct a parametrix for an elliptic pseudodifferential operator on a compact manifold (see [LM89] for instance). See [Don02, p.50] for the full argument.

To finish the proof of (2), we now let $\mathbf{A}$ be an arbitrary connection with flat acyclic limits. Then some small perturbation $\mathbf{A}'$ of $\mathbf{A}$ will actually satisfy the condition of the previous step, that is, be exactly flat over the ends. Since invertibility is an open condition for operators, the operator $D_{\mathbf{A}(t)}$ will hence still be invertible for all sufficiently large $|t|$. The proof for the previous step then applies exactly as before. ∎

A very similar argument to the third step above, using the invariance of the Fredholm index under perturbations, shows that the Fredholm index of $D_{\mathbf{A}}$ is independent of the choice of $\mathbf{A}$ connecting the flat acyclic limits at either end of $Y \times \mathbb{R}$ [Don02, p.51, Proposition 4.5]. Hence we can associate, to any pair of flat acyclic connections on Y, the invariant $\mathrm{ind}(D_{\mathbf{A}}) \in \mathbb{Z}$ given by the Fredholm index. This will provide the relative grading in Floer's instanton theory. However, we still need to consider the action of the gauge transformations on these flat connections. Consider an instanton $\mathbf{A}$ on $Y \times \mathbb{R}$ connecting $A \in \mathcal{A}_P$ to $A^g \in \mathcal{A}_P$ where $g \in \mathscr{G}_P$ is a gauge transformation. Topologically, we may regard this instead as a connection $\mathbf{A}'$ on the principal SU(2) bundle over $Y \times S^1$ obtained by gluing the two ends of $\mathbf{P}$ together via the gauge transformation g. The Atiyah-Singer index theorem applied to the new operator $D_{\mathbf{A}'}$ over $Y \times S^1$ yields, as in Chapter 3:

$$\mathrm{ind}D_{\mathbf{A}'} = 8c_2\left(\mathbf{P}_{Y\times S^1}\right) - 3(1 - b_1(Y \times S^1) + b_2^+(Y \times S^1)) = 8\deg(g)$$

where we have used Theorem 4.2 to identify $\mathbf{P}_{Y\times S^1}$ with $\deg(g)$, and the observation that $b_1(Y \times S^1) = 1$ and $b_2(Y \times S^1) = 0$ whenever Y is a homology 3-sphere. The fact that $\mathrm{ind}D_{\mathbf{A}} = \mathrm{ind}D_{\mathbf{A}'}$ is justified by appealing to the additivity of the index:

THEOREM 4.5. *(Additivity of the Index) [Don02, Proposition 3.8] Suppose X is a 4-manifold with two ends, of the form $Y \times \mathbb{R}$ and $\bar{Y} \times \mathbb{R}$, along with a principal $\mathrm{SU}(2)$ bundle $\mathbf{P} \to X$ and a connection $\mathbf{A}$ on $\mathbf{P}$ with flat* ***acyclic*** *limits over the two ends differing by a gauge transformation. Then the 4-manifold X' obtained by gluing together the two infinite ends with reverse orientation has a principal $\mathrm{SU}(2)$ bundle $\mathbf{P}' \to X'$ with a connection $\mathbf{A}'$, and the Fredholm index of $D_{\mathbf{A}}$ is equal to that of $D_{\mathbf{A}'}$.*

Now suppose there is a connection $\mathbf{B}$ over $Y \times \mathbb{R}$ connecting a flat acyclic connection B at one end to A at the other. Taking $X = Y \times \mathbb{R} \coprod Y \times \mathbb{R}$ and following the procedure described in the theorem above to glue the connection $\mathbf{B} \coprod \mathbf{A}$ over X along the flat connection A, we obtain a new connection $\mathbf{B}'$ over $Y \times \mathbb{R}$ connecting B at one end to A^g at the other. From the additivity of the index, we find that $\mathrm{ind} D_{\mathbf{B}} + \mathrm{ind} D_{\mathbf{A}} = \mathrm{ind}_{\mathbf{B}'}$ and hence the index of $D_{\mathbf{B}'}$ and $D_{\mathbf{B}}$ differ only by a multiple of 8.

4.2.2 INSTANTON FLOER HOMOLOGY

Having described the basic ideas behind using the Chern-Simons functional for Morse theory, we now wish to define the *instanton Floer homology groups* in analogy with the Morse homology groups defined at the start of this chapter. So let Y be a compact oriented Riemannian homology 3-sphere (i.e. $H^1(Y;\mathbb{Z}) = 0$) with metric g_Y and $P \to Y$ be a principal $\mathrm{SU}(2)$ bundle over Y. As we have seen above, the Chern-Simons functional defines a smooth function $\theta : \mathcal{B}_P \to \mathbb{R}/\mathbb{Z}$, with critical points given by the gauge equivalence classes of flat connections on P. Since P is necessarily trivial, the set of critical points will always include the trivial connection on P, which will always be reducible and degenerate, hence preventing us from applying the Fredholm analysis discussed in the previous section. However, all of the non-trivial (equivalence classes of) flat connections will actually be irreducible. To see this, suppose $A \in \mathcal{B}_P$ is a reducible gauge equivalence class of flat connections and let $\rho : \pi_1(Y) \to \mathrm{SU}(2)$ be the corresponding representation of the fundamental group of Y (see Theorem A.1). Since A is reducible, we can apply Theorem 6.1 to conclude that the representation ρ must split as the direct sum of two 1-dimensional unitary representations. But any homomorphism $\pi_1(Y) \to \mathrm{U}(1)$ is necessarily trivial, since $\mathrm{U}(1)$ is abelian and the abelianisation $H_1(Y;\mathbb{Z})$ of $\pi_1(Y)$ is zero by assumption. Hence the representation ρ is itself trivial and by Theorem A.1, the equivalence class A must be trivial. Hence if we let $\mathcal{R}_Y$ denote the set of gauge equivalence classes of irreducible flat connections in $\mathcal{A}_P$, then the critical points of the Chern-Simons functional consist of the union of $\mathcal{R}_Y$ with the trivial connection. The connections in $\mathcal{R}_Y$ will *not* however be non-degenerate in general, but an appropriate 'generic' perturbation of the Chern-Simons functional can be used in order to guarantee that this is the case, called *Floer's holonomy perturbation*. However, this is only known to work for homology 3-spheres Y. For more on this subject, see [Don02, §5.5]. In keeping with our transversality assumptions, we shall assume that the appropriate perturbations have already been made.

We may now define the **Floer chain complex** $\mathrm{CF}(Y, g_Y)$ to be the free abelian group generated by $\mathcal{R}_Y$. This group has a natural $\mathbb{Z}_8$ grading given by the Fredholm index. If we fix some $\alpha \in \mathcal{R}_Y$ and a representative $A \in \mathcal{A}_P$, then for any given $\beta \in \mathcal{R}_Y$ we can choose an equivalence class $B \in \mathcal{A}_P$ and an instanton $\mathbf{A}$ over $Y \times \mathbb{R}$ connecting A to B. The Fredholm index of the operator $D_{\mathbf{A}}$ is then well-defined, and, as we have seen above, independent of the choice of $\mathbf{A}$. Moreover, by the index calculation above, it is independent of the choices of representatives A, B up to multiples of 8. We denote the resulting grading by $\mu(\alpha, \beta) \in \mathbb{Z}_8$. Note that if we had instead chosen to define the Floer chain complex to be generated by non-trivial equivalence classes of flat connections in the universal cover $\mathcal{A}_P/\mathscr{G}_P^0$, then we would have a well-defined $\mathbb{Z}$-grading. However, in this case we would have infinitely many generators of Floer chain complex and we would face issues about formal convergence of sums. It is possible to cope with these issues by using a version of the Novikov Morse theory described in the next sections.

Now we wish to define a boundary map on the Floer chain complex in analogy with Morse homology. Given a pair of equivalence classes $\alpha, \beta \in \mathcal{R}_Y$ and a representative $A \in \mathcal{A}_P$ of α, we can consider the index of the operator $D_{\mathbf{A}}$ for any connection $\mathbf{A}$ over $Y \times \mathbb{R}$ joining A to a representative of β. Since, as we have seen above, acting by a gauge transformation of non-zero degree shifts the index by 8, if we require $\mathrm{ind} D_{\mathbf{A}}$ to have a certain integer value, then there is a uniquely determined representative $B \in \mathcal{A}_P/\mathscr{G}_P^0$ of β such that this is the case. Moreover, we can see that, even if we identify connections A and A^g on P differing by a gauge transformation of non-zero degree, it is not possible for us to have gradient flow lines from a single critical point to itself of relative index 1. The (1-dimensional) moduli space of gradient flow lines between α and β, denoted $\mathcal{M}(\alpha, \beta)$, is then defined to be the set of gauge equivalence classes of ASD instantons $\mathbf{A}$ on $\mathbf{P} \to Y \times \mathbb{R}$ with *finite energy* such that $\lim_{t \to -\infty} \mathbf{A}(t) = A$ and $\lim_{t \to \infty} \mathbf{A}(t)$ is equivalent to B via a degree 0 gauge transformation. This space is necessarily independent of the choice of representative A, since all of the resulting instantons can simply by acted on by the corresponding constant gauge transformation. As in Chapter 3, this can be written as Fredholm system by taking appropriate Sobolev completions. By the *exponential decay theorem* [Don02, Proposition 4.4], every instanton over the tube $Y \times \mathbb{R}$ with flat acyclic limits must decay exponentially with all derivatives to these limiting connections. Hence, regardless of our choice of Sobolev space on $Y \times \mathbb{R}$, it will still contain all such instantons. Indeed, much of the theory is identical, with the complications arising from the fact that $Y \times \mathbb{R}$ is not a compact 4-manifold. We have done most of the difficult work in the previous subsection where we demonstrated that the linearisation operator is actually Fredholm (which need not be the case for a general elliptic operator over a tube). However, now we *cannot* make a generic metrics argument now because the identification of instantons on $Y \times \mathbb{R}$ with gradient flow lines of the Chern-Simons functional relies on the fact that we chose to use the product metric on $Y \times \mathbb{R}$. To achieve transversality, one therefore must appeal to appropriate perturbation arguments for the instanton equation. Following our philosophy from Chapter 3, we shall disregard these difficulties; the details may be found in [Don02, §5.5]. In particular, Proposition 5.17 in [Don02] states that in this case the dimension of $\mathcal{M}(\alpha, \beta)$ is indeed given by the index $\mu(\alpha, \beta)$. As as the case with Morse theory, we have a free $\mathbb{R}$-action on $\mathcal{M}(\alpha, \beta)$ by translation, and we shall use the same notation $\mathcal{M}(\alpha, \beta)$ to denote the translation-reduced moduli space, which has dimension $\mu(\alpha, \beta) - 1$.

Next we shall need to consider the compactness properties of these moduli spaces. The source of non-compactness in the instanton moduli spaces over $Y \times \mathbb{R}$ is 'sliding' behaviour along the tube; the energy of an instanton can run off to infinity. Under the conformal equivalence of $S^4 \setminus \{0\}$ and $S^3 \times \mathbb{R}$, this sliding corresponds exactly to the conformal rescaling that is the source of non-compactness in the closed case. We can use *exactly* the same definitions of chain-convergence from our discussion of Morse homology in §1 and apply them to solutions of the instanton equation over $Y \times \mathbb{R}$. Simple arguments from the Uhlenbeck compactness theorem in Chapter 2 then yield the following result, entirely analogous to that from Morse homology:

THEOREM 4.6. *(Compactness Principle) ([Don02], p.121) If $\mathcal{M}(\alpha, \beta)$ has dimension less than or equal to 8, then any sequence has a chain-convergent subsequence. Moreover, if the dimension of $\mathcal{M}(\alpha, \beta)$ is less than or equal to 4, then any subsequence converges to a chain of length less than or equal to the dimension of $\mathcal{M}(\alpha, \beta)$ that does not pass through the trivial connection.*

Here the dimension restrictions prevent the occurrence of *interior bubbling*, in which energy concentrates at a single point in the interior of $Y \times \mathbb{R}$ rather than escaping to infinity. By the index calculation in Chapter 3, we can see that this phenomenon causes the dimension of the space of infinitesimal deformations of the instanton moduli space to decrease by a multiple of 8; if the energy of the limiting instanton strictly decreases, then so must the second Chern number, which is multiplied by a factor of 8 in the index formula. Hence this behaviour cannot possibly occur when the dimension of the moduli space $\mathcal{M}(\alpha, \beta)$ is sufficiently small. The condition that the chain does not pass through the trivial connection is crucial, since we do not allow this connection into our chain group. As in the case of Morse homology, the converse to the above theorem is given by appropriate gluing results (see [Don02, §4.4]), as sketched in the following section. This allows us to establish that the codimension 1 boundary of the moduli space $\mathcal{M}(\alpha, \beta)$ with $\mu(\alpha, \beta) = 2$ is given by equation 4.1.

This lets us define a boundary map $d : \mathrm{CF}(Y, g_Y) \to \mathrm{CF}(Y, g_Y)$ via

$$d(\alpha) = \sum_{\substack{\beta \in R_Y \\ \mu(\alpha,\beta)=1}} \# \{\mathcal{M}(\alpha, \beta)\} \cdot \beta$$

where the count is guaranteed to be finite by the compactness result above. The exact same argument as in §1 then shows that $d^2 = 0$ and so we have a well-defined homology group, denoted $\mathrm{HF}(Y, g_Y)$, and called the **instanton Floer homology**. It is clear that the map d has degree 1 by the additivity of the index and hence $\mathrm{HF}(Y, g_Y)$ inherits the $\mathbb{Z}_8$ grading of $\mathrm{CF}(Y, g_Y)$. The question of the independence of $\mathrm{HF}(Y, g_Y)$ of the metric g_Y will be taken up in Chapter 6.

4.2.3 GLUING THEORY

In this subsection study how 'approximate' solutions to the instanton and pseudoholomorphic curve equations can be perturbed in some neighbourhood to give genuine solutions. The general method, due to Taubes, is widely applicable and involves a simple argument where a right inverse is constructed for the desired PDE using the contraction mapping principle.

Suppose $\alpha, \beta, \gamma \in R_Y^*$ are non-degenerate and irreducible, and suppose we have $[A_1(t)] \in \mathcal{M}(\alpha, \beta)$ and $[A_2(t)] \in \mathcal{M}(\beta, \gamma)$. Then by the exponential decay theorem, there exists some representative $A_1(\infty)$ of β such that $|A_1(t) - A_1(\infty)| \to 0$ exponentially in C^k as $t \to \infty$, and similarly for the other flat limits of A_1, A_2. What we wish to do is glue A_1 and A_2 together along β to give an ASD instanton over $Y \times \mathbb{R}$ in $\mathcal{M}(\alpha, \gamma)$. A similar situation arises when gluing ASD instantons on closed 4-manifolds together to give an instanton on the connected sum. Now fix some $T \in \mathbb{R}$. The naïve way to go about gluing A_1, A_2 together would be to take a cutoff function ρ supported on $[-1, \infty)$ and define a new connection on $Y \times \mathbb{R}$ by

$$A^* = A_1 \#_T A_2 = \begin{cases} A_1(t+T) & t \leq -1 \\ \beta + \rho(-t)(A_1(t) - \beta) + \rho(t)(A_2(t) - \beta) & t \in [-1,1] \\ A_2(t-T) & t \geq 1 \end{cases}$$

called the **pregluing solution**. However, because the ASD equation is non-linear, this connection will not in general be ASD also, but only *approximately* ASD. Hence we can try to find a nearby solution by appropriately perturbing A^* by some $a(t)$ to give $A^* + a(t)$. For this to be an ASD instanton, $a(t)$ must then satisfy the equation

$$0 = F^+_{A^*+a} = d^+_{A^*} a + (a \wedge a)^+ + F^+_{A^*}$$

which is another non-linear partial differential equation for a. To rewrite this equation in a more approachable form, we shall first appeal to the following lemma:

LEMMA 4.2. *If A_1, A_2 are regular, that is, the operators $\mathrm{d}^+_{A_1}$ and $\mathrm{d}^+_{A_2}$ are surjective, then so is $\mathrm{d}^+_{A^*}$.*

The proof of this lemma is very similar to the proof of Fredholmness earlier in this section; we glue together right inverses coming from the two parts. Hence we know that $d^+_{A^*}$ has a bounded right inverse operator, which we shall call P. Hence we may let $a = Pz$ for some smooth section $z \in L^2_\delta(\mathbb{R}, \Omega^{2,+}_Y(\text{ad } P))$ which we imagine as being a very small perturbation. The equation in terms of z can then be written as

$$z = -F^+_{A^*} - (Pz \wedge Pz)^+$$

By taking T sufficiently large, we can always ensure that the pregluing solution has $||F^+_{A^*}||_{L^2} \leq \varepsilon_T$, so that the right hand side of this equation can be viewed as a small perturbation. Now we claim that there exists a unique solution z to this equation with $||z||_{L^2_\delta} \leq \varepsilon$. Write $N(z) = (Pz \wedge Pz)^+$ for the non-linear term; we wish to solve the equation $z = -F^+_{A^*} - N(z)$ via the contraction mapping principle. In fact, we have an estimate of the form

$$||N(z_1) - N(z_2)||_{L^2} \leq C(||z_1||_{L^2} + ||z_2||_{L^2})||z_1 - z_2||_{L^2}$$

and hence we do indeed get a contraction mapping on a sufficiently small ball around 0. This gluing procedure can be used to prove:

THEOREM 4.7. *(Gluing) There exists a smooth embedding map*

$$\mathcal{M}(\alpha,\beta) \times \mathcal{M}(\beta,\gamma) \times [T,\infty) \to \mathcal{M}(\alpha,\gamma)$$

which gives a smooth neighbourhood of each broken trajectory.

In general, we might wish to glue at a connection β which is reducible. In this case, we have freedom to glue via a gauge transformation. These are called *gluing parameters* and in this case the embedding will take the form of a map $\mathcal{M}(\alpha,\beta) \times \mathcal{M}(\beta,\gamma) \times [T,\infty) \times G \to \mathcal{M}(\alpha,\gamma)$. We may also want to glue in the case where β is degenerate. In this case, we no longer have right inverses. In this case, we might try to mimic the cokernel perturbation argument from Chapter 3, enlarging the moduli space in order to make the linearisation operators surjective. Carrying through the arguments in this case gives rise instead to a **virtual neighbourhood** of the broken trajectory, given by a *virtual gluing map*. We will not present the details here. Similar gluing results also hold for J-holomorphic curves, and can be found in [MS12].

4.3 EXAMPLE: THE ACTION FUNCTIONAL ON PATHS

Let X be a closed symplectic manifold with symplectic form ω, and let L_0, L_1 be two Lagrangian submanifolds (see Appendix A for definitions). The idea of Lagrangian intersection Floer homology is to perform and infinite-dimensional version of the Morse homology described earlier, using the action functional on the space of paths between the two Lagrangian submanifolds. We shall find that the critical points of this functional correspond to (lifts of) intersection points $L_0 \cap L_1$ of the two Lagrangians, and that the gradient flow lines correspond to pseudoholomorphic strips. We shall find, however, that the relation $d^2 = 0$ does *not* hold in general due to the presence of *disk bubbles*, so that we are unable to define the homology groups. Assumptions on $\pi_2(X, L_0)$ can remove these bubbles, but the difficult task of defining Lagrangian intersection Floer homology in general is postponed until Chapter 5.

Let $\Omega(L_0, L_1)$ denote the space of all smooth paths between L_0 and L_1. Formally, the tangent space is given by $T_\ell\Omega = \{X \in \Omega^0(\ell^*TX) : X(0) \in TL_0,\ \ X(1) \in TL_1\}$, and so we define the **action 1-form** by

$$\alpha_\ell(X) = \int_0^1 \omega(\dot{\ell}(t), X(t))\mathrm{d}t$$

In fact, this form will be closed, as we can see from the following calculation. Let $X, Y \in T_\ell\Omega$ and let $\ell_X(s,t)$ be a 1-parameter family of curves in Ω such that $\ell(0,t) = \ell$ and $\frac{\partial}{\partial s}|_{s=0}\ell(s,t) = X(t)$. Similarly for $\ell_Y(s,t)$. Then we have

$$\begin{aligned}(d\alpha)_\ell(X,Y) = X(\alpha(Y)) - Y(\alpha(X)) &= \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0}\left(\int_0^1 \omega(\dot{\ell}_X(s,t), Y(t)) - \omega(\dot{\ell}_Y(s,t), X(t))\mathrm{d}t\right)\\ &= \int_0^1 \omega(\dot{X}(t), Y(t)) - \omega(\dot{Y}(t), X(t))\mathrm{d}t = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\omega(X,Y)\mathrm{d}t\end{aligned}$$

because ω is closed. Hence

$$(d\alpha)_\ell(X,Y) = -\omega(X(0), Y(0)) + \omega(X(1), Y(1)) = 0$$

because L_0, L_1 are Lagrangian. Therefore, by the Poincaré Lemma, an appropriate *covering space* of $\Omega(L_0, L_1)$, the 1-form α will be the differential of some functional, which we will refer to as the **action functional**. We shall now give an explicit construction of this functional.

Because $\Omega(L_0, L_1)$ has infinitely many connected components, we shall choose one by specifying some basepoint ℓ_0; we denote this component by $\Omega(L_0, L_1, \ell_0)$. Any path $\ell \in \Omega(L_0, L_1, \ell_0)$ is connected to ℓ_0 by a path, which we may regard as a map $u : I \times I \to X$ such that $u(0,t) = \ell_0$, $u(0,t) = \ell$, $u(s,0) \in L_0$ and $u(s,1) \in L_1$. Following the usual construction of the universal covering space (see [Hat02]), we define $\tilde{\Omega}(L_0, L_1, \ell_0)$ to be the space of homotopy classes of such maps $u : I \times I \to X$. On this covering space we define the *action functional* by

$$A(\ell, u) = \int_{I\times I} u^*\omega$$

Is is a simple exercise to check that dA is the lift $\pi^*\alpha$ of α to $\tilde{\Omega}(L_0, L_1, \ell_0)$. Hence the critical points of A are given by the lifts of the zeroes of α. By non-degeneracy of the symplectic form ω, if $\alpha_\ell \equiv 0$ we must have ℓ the constant path. This is only possible if ℓ is constant at a point in $L_0 \cap L_1$ and hence the critical points of A correspond to lifts of the constant path ℓ_p at intersection points p. Now, because π is a covering map, it gives a local identification of $T\Omega$ and $T\tilde{\Omega}$. Given an ω-compatible almost-complex structure J on X, we can define a metric on $\tilde{\Omega}(L_0, L_1, \ell_0)$ via this identification, using

$$\langle X_1, X_2\rangle = \int_0^1 \omega(X_1(t), JX_2(t))\mathrm{d}t$$

on $T_\ell\Omega(L_0, L_1)$. From this we see that

$$\alpha_\ell(X) = \langle J\dot{\ell}, X\rangle$$

and hence the gradient of the action functional is given by $\ell \mapsto J\dot{\ell} \in \Gamma(\ell^*TX)$. If $\ell(s)$ is an s-parameterised family of paths in Ω, we may consider it instead as a map $u : \mathbb{R} \times [0,1] \to X$. Then this path is a gradient flow line if and only if

$$\frac{\partial u}{\partial s} = -\nabla A(u) = -J(u)\frac{\partial u}{\partial t}$$

Hence we see that *gradient flow lines of the action functional on Ω correspond exactly to J-holomorphic curves from the strip $\mathbb{R} \times [0,1]$ having boundary in L_0, L_1* (see the Figure above). This is the crucial insight of Lagrangian Floer homology.

In order to study the subtle relations between the gradient flow lines and the covering space theory, it will be necessary to make a digression to define the Maslov index.

4.3.1 THE MASLOV INDEX

There are several different topological constructions in the symplectic geometry literature that are referred to as the *Maslov index*. Firstly, we associate a *universal Maslov index* to any loop in the Lagrangian Grassmannian $\Lambda(n)$. Here we consider $\mathbb{R}^{2n}$ as $\mathbb{C}^n$ and give it the standard symplectic form coming from the complex structure and the Euclidean metric. Then $\Lambda(n)$ denotes the set of all Lagrangian subspaces of $\mathbb{R}^{2n}$ with respect to this symplectic form. Now we

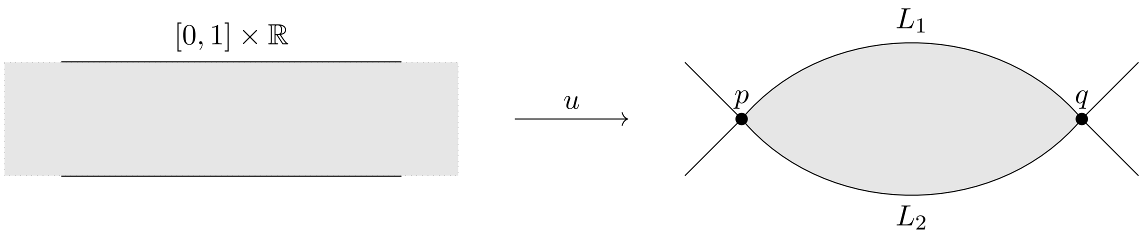

Figure 4.1: A pseudoholomorphic strip connecting the two critical points p, q.

wish to give this space a topology. Firstly, we claim that unitary group $\mathrm{U}(n)$ acts transitively on $\Lambda(n)$. To see this, note that L a subspace of $\mathbb{R}^{2n}$ is Lagrangian if and only if it is orthogonal to iL. If L' is another Lagrangian subspace of $\mathbb{R}^{2n}$, then there is clearly an element of $\mathrm{O}(2n)$ that carries L to L' and iL to iL', as we can see by choosing orthonormal bases for L and L'. This element of $\mathrm{O}(2n)$ then also preserves the symplectic structure so is necessarily corresponds to an element of $\mathrm{U}(n)$. Secondly, we claim that the stabiliser of any Lagrangian subspace is given by $\mathrm{O}(n) \subseteq \mathrm{U}(n)$. This follows by observing that $\mathbb{R}^n \subseteq \mathbb{C}^n$ is a Lagrangian subspace and is preserved exactly by elements of $\mathrm{O}(n)$. Hence we have a fibration $\mathrm{O}(n) \to \mathrm{U}(n) \to \Lambda(n)$ and can regard $\Lambda(n)$ as the quotient $\mathrm{U}(n)/\mathrm{O}(n)$.

From this perspective, we have a well-defined map $\Lambda(n) \to S^1$ given by $\det^2$, since the determinant of any matrix in $\mathrm{O}(n)$ is always ± 1. Letting $\mathrm{S}\Lambda(n) = (\det^2)^{-1}(1)$ be the Grassmannian of oriented Lagrangian subspaces, similar arguments to the above show that we have a fibration $\mathrm{SO}(n) \to \mathrm{SU}(n) \to \mathrm{S}\Lambda(n)$. Altogether, we have a commutative diagram of fibrations:

$$\begin{tikzcd}
\mathrm{SO}(n) \arrow[r] \arrow[d] & \mathrm{O}(n) \arrow[r, "\det"] \arrow[d] & S^0 \arrow[d] \\
\mathrm{SU}(n) \arrow[r] \arrow[d] & \mathrm{U}(n) \arrow[r, "\det"] \arrow[d] & S^1 \arrow[d, "z^2"] \\
\mathrm{S}\Lambda(n) \arrow[r] & \Lambda(n) \arrow[r, "\det^2"] & S^1
\end{tikzcd}$$

Taking the long exact sequence in homotopy groups gives

$$\begin{tikzcd}
 & \pi_1(\mathrm{SU}(n)) \arrow[d] & & & \pi_0(\mathrm{SU}(n)) \arrow[d] & \\
\cdots \arrow[r] & \pi_1(S\Lambda(n)) \arrow[r] \arrow[d] & \pi_1(\Lambda(n)) \arrow[r, "\det^2"] & \pi_1(S^1) \arrow[r] & \pi_0(\mathrm{S}\Lambda(n)) \arrow[r] \arrow[d] & \cdots \\
 & \pi_0(\mathrm{SO}(n)) & & & 0 &
\end{tikzcd}$$

and hence we see that $\det^2$ induces an isomorphism of $\pi_1(\Lambda(n))$ with $\mathbb{Z}$. This integer is called the **Maslov index** of the loop.

From this, we can see that $H^1(\Lambda(n);\mathbb{Z}) \cong \mathbb{Z}$, with generator called the **universal Maslov class**, assigning to each loop γ in $\Lambda(n)$ the integer $\deg(\det^2(\gamma))$. This may be regarded as a characteristic class as follows. The Lagrangian Grassmannian $\Lambda(n)$ is the classifying space for tangent bundles of Lagrangian submanifolds; that is, if $L \to X$ is an immersion, then we have a map $L \to \Lambda(n)$ and we can pull back the universal Maslov class to L. This allows us to assign an index to any loop on a Lagrangian submanifold L.

Next, we have a notion of Maslov index for maps $u : (\Sigma, \partial\Sigma) \to (X, L)$ from surfaces with (possibly immersed) Lagrangian boundary conditions. Since $\partial\Sigma$ is a disjoint union of circles, we have a collection of loops in the Lagrangian submanifold L. Summing over the Maslov index of each of these loops gives yields the **Maslov index** of the map u, denoted $\mu_L(u)$. It is clear that this is homotopy invariant among maps of the same type. An alternative description is often useful. When Σ has connected boundary, then it is contractible, and so any symplectic vector bundle over Σ admits a symplectic trivialisation. Hence $(u|_{\partial\Sigma})^*TL$, as a subbundle of $u^*TX|_{\partial\Sigma}$ will yield a loop in $\Lambda(n)$. The Maslov index of this loop is then $\mu_L(u)$ and is independent of choice of trivialisation. For example, the map $u : D^2 \to \mathbb{C}$ given by $u(z) = z^n$ sends the boundary $\partial D^2 \cong S^1$ to the Lagrangian submanifold $\mathrm{U}(1) \subseteq \mathbb{C}$ and has Maslov index $2n$. This example shall be of importance later in this thesis.

Now we want to assign various notions of Maslov index for paths and loops in the space $\Omega(L_0, L_1)$. Since a loop u in

$\Omega(L_0, L_1)$ is simply a map from the cylinder $S^1 \times I$ with boundary conditions in L_0, L_1 we may hence assign a Maslov index to u using the definition in the previous paragraph. In fact, this depends only on the homotopy class of the loop and so yields a group homomorphism $\mu : \pi_1\Omega(L_0, L_1) \to \mathbb{Z}$. Now we consider u a path in $\Omega(L_0, L_1)$ between paths ℓ_0 and ℓ_1; here u is a map $I \times I \to (X, \omega)$ having top and bottom boundaries in L_0, L_1 and 'side' boundaries in ℓ_0 and ℓ_1. In order to produce a Maslov index, we fix a symplectic trivialisation of u^*TX, and *fix* a path of Lagrangian subspaces from $T_{\ell_0(0)}L_0$ to $T_{\ell_0(1)}L_1$ along ℓ_0, and similarly along ℓ_1. In the case where ℓ_0 is the constant path ℓ_p at an intersection point, then we have the *canonical short path* γ from T_pL_0 to T_pL_1 inside $\Lambda(n)$ by requiring that $\gamma(t)$ is transversal to T_pL_0 for all $t \in I$; this path is unique up to homotopy. Now we define the **relative Maslov index** to be the Maslov index of the loop in $\Lambda(n)$ obtained by concatenation when travelling around the boundary of u. This gives a well-defined map from the set $\pi_2(\ell_0, \ell_1)$ of homotopy classes of paths from ℓ_0 to ℓ_1 to $\mathbb{Z}$, upon making such choices. When we have fixed an ℓ_0 as above, along with a path of Lagrangian subspaces along ℓ_0, we will often write $\pi_2(\ell_1)$ to denote the set $\pi_2(\ell_0, \ell_1)$. Also, when ℓ_0, ℓ_1 are the constant paths at intersection points $p, q \in L_0 \cap L_1$ respectively, we will write instead $\pi_2(p, q)$ to denote this set of homotopy classes, and call the relative Maslov index the **Maslov-Viterbo index**. Using this, we can associate an **absolute Maslov index** to lifts $[\ell_p, w]$ of the constant path at p to the universal covering space $\widetilde{\Omega}(L_0, L_1, \ell_0)$; here $w \in \pi_2(p)$ is a path from ℓ_0 to ℓ_p and hence we can associate a Maslov index to this path just as above.

The first observation we make is that the relative Maslov index is additive under concatenation of paths, and flips sign for paths that are reversed. That is, if we take a path $u \in \pi_2(\ell_0, \ell_1)$ and concatenate with $v \in \pi_2(\ell_1, \ell_0)$, then the Maslov index for $u * v$ as a loop at ℓ_0 is equal to $\mu(u) + \mu(v)$. Similarly, if we suppose that some $B \in \pi_2(p, q)$ is given, then we can express the relative Maslov index $\mu(B)$ in terms of the absolute Maslov index by

$$\mu(B) = \mu([\ell_q, w * B]) - \mu([\ell_p, w])$$

for any $w \in \pi_2(p)$. Now we shall return to our discussion of covering spaces.

COVERING SPACES

We want to consider the gradient flow lines as paths connecting critical points of the action functional. If a solution u to the J-holomorphic curve equation $u : \mathbb{R} \times I \to X$ has finite energy, then by a theorem of Floer, u converges exponentially along the $\mathbb{R}$-direction to the critical points p, q and hence may be regarded from a topological perspective as mapping $I \times I \to X$ after compactification. Thus we may associate to every gradient flow line an element of $\pi_2(p, q)$. The problem we must consider now is the grading. The dimension of the moduli space of gradient flow lines between critical points p, q should be given in Morse homology by the difference in the grading between p and q. However, the moduli space of gradient flow lines representing the class $B \in \pi_2(p, q)$ actually has dimension $\mu_L(B)$. This indicates that we need to lift critical points to an appropriate covering space in order to recover a grading. But when we do this, we do not wish to produce spurious extra gradient flow lines. This problem is solved by introducing the *Novikov covering*.

Returning to our discussion of covering spaces, to a loop $u \in \Omega(L_0, L_1, \ell_0)$ we can associate two numbers: the symplectic area, given by

$$I_\omega(u) = \int_{I \times S^1} u^*\omega$$

and the Maslov index $I_\mu(u) = \mu(u)$. These give group homomorphisms $I_\omega : \pi_1\Omega(L_0, L_1, \ell_0) \to \mathbb{R}$ and $I_\mu : \pi_1\Omega(L_0, L_1, \ell_0) \to \mathbb{Z}$. The covering space corresponding to the subgroup $\ker I_\omega \cap \ker I_\mu$ is called the **Novikov covering** and denoted $\tilde{\Omega}^N(L_0, L_1, \ell_0)$. We may regard it as explicitly constructed from the universal covering space $\tilde{\Omega}(L_0, L_1, \ell_0)$ by identifying two paths w and w' exactly when the concatenation $\bar{w} * w'$ is the the kernel of both I_ω and I_μ. The deck transformation group of this covering space is $\pi_1\Omega(L_0, L_1, \ell_0)/\ker I_\omega \cap \ker I_\mu$ and is necessarily abelian. It acts on the Novikov covering by concatenation; if $[\ell_1, w]$ is an equivalence class consisting of a path $\ell_1 \in \Omega(L_0, L_1, \ell_0)$ and a homotopy class $w \in \pi_2(\ell_1)$, then $g \in \pi_1\Omega(L_0, L_1, \ell_0)$ acts by $g \cdot [\ell_1, w] = [\ell_1, w * g]$. In particular, we have for the action functional

$$\mathcal{A}[\ell_p, w * g] = \int (w * g)^*\omega = \int w^*\omega + \int g^*\omega = \mathcal{A}[\ell_p, w] + I_\omega(g) = \mathcal{A}[\ell_p, w]$$

Hence we see that $\mathcal{A}$ is well-defined on the Novikov covering as a map to $\mathbb{R}$.

The problem now is that the J-holomorphic curves $u : I \times I \to X$ are the gradient flow lines on $\Omega(L_0, L_1, \ell_0)$, not on the covering space. Let $\pi : \tilde{\Omega} \to \Omega(L_0, L_1, \ell_0) = \Omega$ be any covering space projection, so that $d\mathcal{A} = -\pi^*\alpha$. As noted before, since π is a covering space map, π_* gives an isomorphism of tangent spaces near every point and hence we can

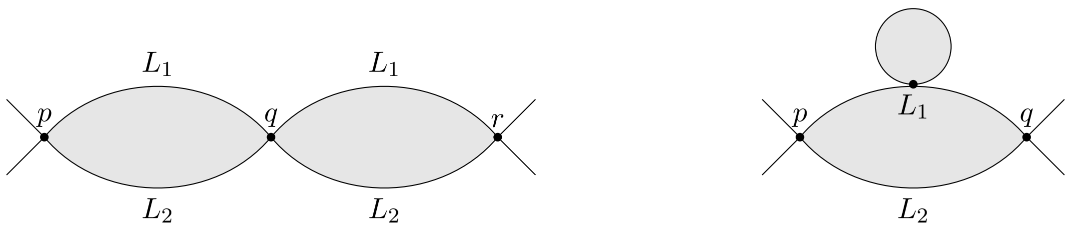

Figure 4.2: Disk breaking (left) and disk bubbling (right).

use π_* to define a metric on $\tilde{\Omega}^N(L_0, L_1, \ell_0)$. Now suppose $X \in T_\ell\tilde{\Omega}$ and V is the vector field associated to α on Ω. Then we observe

$$(\pi^*\alpha)_\ell(X) = \alpha_{\pi(\ell)}(\pi_* X) = \langle \pi_* X, V(\pi(\ell))\rangle = \langle X, (\pi_*)^{-1}V(\pi(\ell))\rangle$$

and hence the vector field associated to $\pi^*\alpha$ is exactly $\tilde{V}(\ell) = (\pi_*)^{-1}V(\pi(\ell))$. Thus the gradient flow equation for a lift $\tilde{u}$ of u to the covering space $\tilde{\Omega}$ must be given by

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}s} = \tilde{V}(\tilde{u}(s))$$

or, equivalently,

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}s} = (\pi_*)^{-1}V(u)$$

which, after applying π_* to both sides, is simply equivalent to the usual equation

$$\frac{\mathrm{d}u}{\mathrm{d}s} = V(u)$$

Conversely, it is clear that every gradient flow line on $\tilde{\Omega}$ can be pushed down to a gradient flow line on Ω. Hence the gradient flow lines on the covering space $\tilde{\Omega}$ are exactly the lifts of the gradient flow lines on Ω. However, as noted before, these different lifts can have different dimensional moduli spaces; now we wish to specialise our discussion to the case of $\tilde{\Omega} = \tilde{\Omega}^N$. To lift a gradient flow line u from p to q, we need to specify a lift $[\ell_p, w]$ of p. Then u flows from $[\ell_p, w]$ to $[\ell_q, w * B]$ where $B \in \pi_2(p, q)$ is the homotopy class of u. But there are now many other gradient flow lines arriving at a given $[\ell_q, w * B]$ coming from different lifts of ℓ_p. We claim that the set of such possible points is in bijection with $\pi_2(p, q)$; but this follows from the transitivity of the covering space action by concatenation. Supposing that $[\ell_q, w * B] = [\ell_q, w * B']$ for $B, B' \in \pi_2(p, q)$, the equivalence relation defined by the *Novikov covering* then requires that $(w * B) * \overline{w * B'} \in \ker I_\omega \cap \ker I_\mu$. By commutativity, we hence have the requirement that $B * \overline{B'} \in \ker I_\mu$, that is, B and B' have the same Maslov index. We also have a condition on the energy. Hence they have the same dimensional moduli spaces and all is well. Therefore we deduce that the moduli space of (finite-energy) gradient flow lines on the covering space is given by

$$\mathcal{M}([\ell_p, w], [\ell_q, w']) = \coprod_{\substack{B \in [w * \overline{w'}] \\ \omega(B) = \mathcal{A}[\ell_q, w'] - \mathcal{A}[\ell_p, w]}} \mathcal{M}(p, q, B)$$

where $\mathcal{M}(p, q, B)$ is simply the moduli space of (finite-energy) gradient flow lines on Ω representing the class B. It will be important to know that for any finite energy $C \in \mathbb{R}$, the number of $B \in \pi_2(p, q)$ with $\mathcal{M}(p, q, B)$ non-empty and $\omega(B) \leq C$ is always finite.

4.3.2 MODULI SPACES OF DISKS

As before, it will be necessary to discuss the moduli spaces of gradient flow lines in order to define Lagrangian intersection Floer homology. However, we will find that they fail to have to desired compactification, due to the presence of *anomalies*. Because of these anomalies, the full definition of Lagrangian intersection Floer homology will only be given in the next Chapter.

Instead of regarding the domains of the maps as strips, it will instead be convenient to consider them as *disks* with two marked points on the boundary mapping to the intersection points $L_0 \cap L_1$. The type of behaviour corresponding to breaking of gradient flow lines is then called **disk breaking** (see Figure 4.2, left). However, sequences of disks with boundary conditions in a pair of Lagrangians can also have other types of limiting behaviour, called **disk bubbling** (see

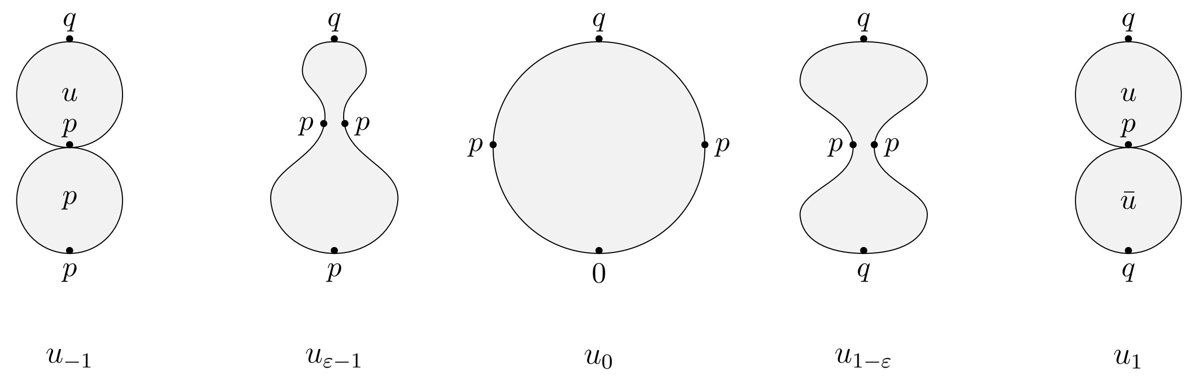

Figure 4.3: Disk bubbling and disk breaking for a simple example.

Figure 4.2, right) where an extra disk forms on the boundary of one Lagrangian away from an intersection point. This can be analysed using a version of the compactness theory for J-holomorphic curves from Chapter 3. These bubbles occur when the derivative du tends to infinity at a certain point. If this occurs in the interior of the disk, then we have a sphere bubble as described in Chapter 2. If instead $|du_i(x_i)| \to \infty$ for a sequence x_i tending to the boundary of the disk, then we have a disk bubble instead.

The standard example of the latter is as follows. Let $X = \mathbb{C}$ be the target manifold with its usual symplectic structure, let $L_1 = S^1 \subseteq \mathbb{C}$ be the unit circle, and let L_2 be the real axis in $\mathbb{C}$. Then $L_1 \cap L_2$ consist of two points, which we shall denote by $p = (1, 0)$ and $q = (0, 1)$. By the Riemann mapping theorem, $\mathcal{M}(p, q)$ has a unique element, given by the obvious holomorphic disk, and similarly for $\mathcal{M}(q, p)$. Hence we must have $d(p) = q$ and $d(q) = p$. Clearly, therefore, we have $d^2(p) \neq 0$. This has occurred because of the failure of the moduli space $\mathcal{M}(p, p)$ to have the correct compactness properties. We can actually obtain an explicit description of the holomorphic maps p to p; they are given by:

$$u_\alpha(z) = \frac{z^2 - \alpha}{1 - \alpha z^2}$$

where $\alpha \in (-1, 1)$. All of these maps represent regular points in the moduli space, so to compactify we simply need to add in $\alpha = 1, -1$. When $\alpha \to 1$, we get precisely the desirable broken trajectory; u_1 goes from p to q and then back to p (see Figure 4.3). However, when $\alpha \to -1$, we see that u_α on the lower half of the disk tends to the constant map at p, but u_α on the upper half of the disk must still once around the entire boundary of the disk D^2 from p to q and back again; hence the derivative is concentrating at q. After appropriate rescaling, what results is a **disk bubble** at the point q, given by the identity map of the disk to itself, attached to the constant map at p.

Evidently is is necessary to give closer consideration to the compactification of moduli spaces of disks. We can discuss moduli spaces of disks using the formalism of Chapter 3, using the following trick.

DEFINITION 4.3. *A **bordered prestable map** consists of a prestable map $\{u_i\}$ commuting with an anti-holomorphic involution τ of the domain fixing all of the marked points and nodal points, along with an orientation of the set $\partial\Sigma$ of fixed points of τ. Given such an orientation, we can define the associated **disk** Σ as the set of points in the domain that form the positively-oriented boundary of $\partial\Sigma$. We say that the bordered prestable map if J-**holomorphic** if the u_i restricted to Σ are J-holomorphic. We say that the bordered map is **stable** if and only if the corresponding map is stable. We say that u_i has **boundary in a Lagrangian** L if $u_i(\partial\Sigma) \subseteq L$.*

We can now reformulate all of the notions in Chapter 3.6.1 incorporating the involution τ; we believe the reader is capable of doing this for themselves. Hence we have in the same way a moduli space $\mathcal{M}_{0,k}(\beta, J, L)$ of disks with boundary in the Lagrangian L, by fixing a homotopy class $\beta \in \pi_2(X, L)$. This moduli space will have many connected components; we shall use $\mathcal{M}_k(\beta)$ to denote the component containing the disk $D^2 \subseteq \mathbb{C}$ with k marked points around the boundary with clockwise ordering. As before, these moduli spaces will not in general be smooth manifolds and will contain a number of strata depending on the different tree structures obtained by bubbling. We do have the following results:

THEOREM 4.8. *(Fukaya-Oh-Ohta-Ono) The moduli space $\mathcal{M}_k(\beta)$ has a topology with respect to which it is compact and Hausdorff [FOOO09, 2.1.29], having virtual dimension*

$$\dim \mathcal{M}_k(\beta) = n + \mu_L(\beta) + k - 3$$

[FOOO09, 2.1.32] and codimension 1 boundary consisting of those disks having exactly one nodal point [FOOO09, 2.1.33].

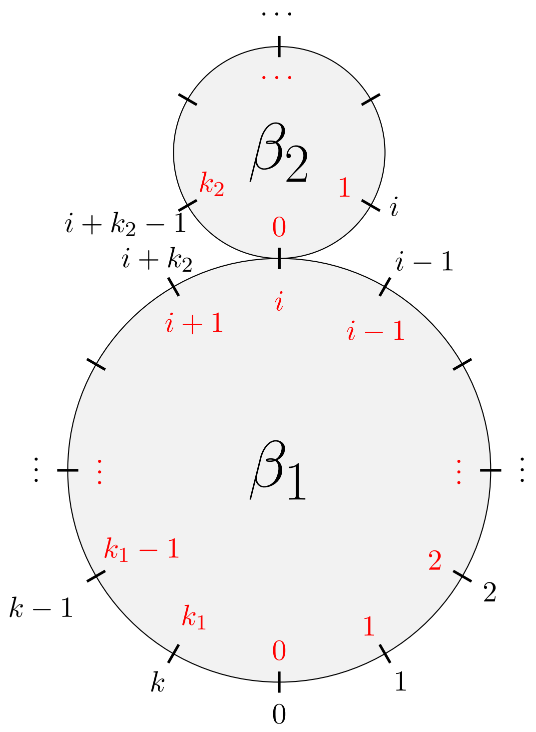

Figure 4.4: Labelling of marked points for a disk breaking into two disks.

We say that $L \subseteq X$ is **relatively spin** if there exists a class in $H^2(X; \mathbb{Z}_2)$ that restricts to the second Stiefel-Whitney class of L. In this case, the moduli spaces $\mathcal{M}_k(\beta)$ will in fact be *orientable* also [FOOO09, 2.1.30].

We can actually obtain a more explicit description of this boundary. Firstly, we can note that sphere bubbles always occur with codimension 2, and hence will not contribute to the codimension 1 boundary. This arises from the gluing theory (cf. Chapter 2, Section 4) by noting that deformation of the complex structure near such a nodal curve has one complex parameter (and hence two real parameters). On the other hand, boundary disk bubbles have only a 1-parameter degeneration, this being where we choose to cut the two strips when we glue them together. In other words, there is no rotational component. Hence the only form of non-compactness we need consider is the breaking of disks along the boundary. This can occur between any pair of marked points, and carry away any number of marked points. It must also carry away part of the energy in the form of a non-zero homotopy class in $\pi_2(X, L)$. From this we may deduce

THEOREM 4.9. *(Compactification, [FOOO09], 7.1.64) The compactification of $\mathcal{M}_{k+1}(\beta)$ in the Gromov topology has codimension 1 boundary given by*

$$\partial^{(1)}\overline{\mathcal{M}}_{k+1}(\beta) = \coprod_{\substack{\beta_1+\beta_2=\beta \\ k_1+k_2=k+1 \\ k_2\geq 0,\ i=1,\ldots,k_1}} \mathcal{M}_{k_1+1}(\beta_1, L, J, L) \times_{\mathrm{ev}_i, \mathrm{ev}_0} \mathcal{M}_{k_2+1}(\beta_2, X, J, L)$$

where we of course must require that $\beta_1, \beta_2 \neq 0$ if the resulting maps would be unstable (that is, $\beta_1 \neq 0$ when $k_1 = 1$ and $\beta_2 \neq 0$ when $k_2 = 1$).

The sense in which $\mathcal{M}_{k+1}(\beta)$ is a moduli space is somewhat controversial, so the word 'boundary' above must be taken as largely heuristic. The labelling may be clarified by referring to Figure 4.4. The black labels around the outer edge refer to the labels on the original disk in $\mathcal{M}_{k+1}(\beta)$. The red labels around the inner edge refer to the internal labels of the marked points for each of the individual disks in the moduli spaces $\mathcal{M}_{k_1+1}(\beta_1, L, J, L)$ and $\mathcal{M}_{k_2+1}(\beta_2, X, J, L)$ that have been glued together.

Chapter 5

LAGRANGIAN FLOER THEORY

5.1 THE A_∞ ALGEBRA ON DIFFERENTIAL FORMS

In order to define the Lagrangian intersection Floer homology in a more general context than in §1 of this Chapter, namely for general symplectic manifolds and immersed Lagrangians, we will follow [FOOO09] and define the (weak) A_∞ algebra associated to a Lagrangian submanifold. In this section we fix a symplectic manifold (X, ω) and an embedded Lagrangian submanifold $L \subseteq X$. For convenience, we include in this Chapter the algebraic definitions and constructions relating to A_∞ algebras.

HIC SUNT DRACONES 5.1. *Readers are warned that there are a number of different sign conventions for A_∞ algebras used in the literature; the sign conventions used in symplectic geometry (such as in [Sei08]) and those used in abstract algebra are mutually incompatible. See [Pol09] for a detailed summary of how to translate between the various conventions. Throughout we follow the conventions in [FOOO09].*

DEFINITION 5.1. *A pair $(A, \{m_k\}_{k=0}^{\infty})$ consisting of a $\mathbb{Z}$-graded $\mathbb{C}$-vector space A, and a set of linear maps $m_k : A^{\otimes k} \to A$ for $k \geq 0$ of degree $2-k$, is a **weak A_∞ algebra** if it satisfies the A_∞**-relations** for all $k \geq 1$*

$$\sum_{\substack{k_1+k_2=k+1 \\ k_2\geq 0, k_1\geq 1 \\ i=1,\ldots,k_1}} (-1)^{|\alpha_1|+\cdots+|\alpha_i|+1} m_{k_1}(\alpha_1, \ldots, \alpha_{i-1}, m_{k_2}(\alpha_i, \ldots, \alpha_{i+k_2}), \alpha_{i+k_2+1}, \ldots, \alpha_k) = 0 \tag{5.1}$$

*for all elements $\alpha_1, \cdots, \alpha_k \in A$ of pure degree. Here we regard m_0 as being an element of A, and we say that A is an A_∞ **algebra** if in addition, $m_0 = 0$.*

When $k = 1$, the first A_∞ relation is:

$$m_2(\alpha_1, m_0) \pm m_2(m_0, \alpha_1) = \pm m_1(m_1(\alpha_1))$$

Hence when $m_0 = 0$, the map m_1 makes A into a cochain complex. We call the cohomology group $H^*(A)$. When $k = 2$, and $m_0 = 0$, the A_∞ relation becomes:

$$m_1(m_2(\alpha_1, \alpha_2)) = m_2(\alpha_1, m_1(\alpha_2)) \pm m_2(m_1(\alpha_1), \alpha_2)$$

If we regard m_2 as a product operation on A, then this equation shows that m_1 is a derivation with respect to this product. This implies that the product structure descends to the cohomology $H^*(A)$ of A with respect to m_1. If we now use a multiplication dot to denote the product with respect to m_2, and assume that $m_1(\alpha_1) = m_1(\alpha_2) = m_1(\alpha_3) = 0$, then the A_∞-relation for $k = 3$ and $m_0 = 0$ becomes

$$\alpha_1 \cdot (\alpha_2 \cdot \alpha_3) - (\alpha_1 \cdot \alpha_2) \cdot \alpha_3 = \pm m_1(m_3(\alpha_1, \alpha_2, \alpha_3))$$

This implies that the product m_2 is associative on $H^*(A)$. The operation m_3 in some sense measures the failure of m_2 to be associative on A.

Remark 5.1. The above algebraic formalism is rather naïve. We really wish to consider our A_∞ algebras as modules over the **universal Novikov ring** (with complex coefficients), that is, the ring of formal power series of the form

$$\Lambda_0^{\mathbb{C}} = \left\{ \sum_{i=0}^{\infty} a_i T^{\lambda_i} : a_i \in \mathbb{C}, \text{ for powers } \lambda_i \geq 0 \text{ such that } \lambda_i \to \infty \text{ as } i \to \infty \right\}$$

Here the powers of T are used to keep track of the 'energy'; since there are only finitely many homotopy classes $\beta \in \pi_2(X, L)$ with energy less than some given λ, the convergence of these power series should not usually cause problems. But to make sense of this formal convergence in general, one will need to work with the algebraic formalism of *gapped, filtered A_∞ algebras* as in [FOOO09] and [AJ10]. For simplicity, we shall not introduce these notions here.

Recall the moduli spaces $\mathcal{M}_{k+1}(\beta)$ of marked disks from the previous Chapter. By evaluating these disks at the various marked points (labelled $0, 1, \ldots, k$ as in the previous Chapter), we obtain $\mathrm{ev}_0^\beta : \mathcal{M}_{k+1}(\beta) \to L$ and $\mathrm{ev}_{(k)}^\beta : \mathcal{M}_{k+1}(\beta) \to L^k$. Now, on the chain level, we will take our A_∞-algebra associated to the Lagrangian L to be given by $\Omega^*(L; \Lambda_0^{\mathbb{C}})$. We now define operations $m_k : \Omega^*(L; \Lambda_0^{\mathbb{C}})^{\otimes k} \to \Omega^*(L; \Lambda_0^{\mathbb{C}})$ by

$$m_k(\alpha_1, \ldots, \alpha_k) = \sum_{\beta \in \pi_2(X,L)} m_k^\beta(\alpha_1, \ldots, \alpha_k) = \sum_{\beta \in \pi_2(X,L)} \int^{\mathrm{vir}}_{\mathcal{M}_{k+1}(\beta,L) \xrightarrow{\mathrm{ev}_0^\beta} L} (\mathrm{ev}_{(k)}^\beta)^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k) T^{\int_\beta \omega}$$

where the integral denotes integration over fibres, and we use the notation $\alpha_1 \vartriangle \cdots \vartriangle \alpha_k$ to denote $\pi_1^*\alpha_1 \wedge \cdots \wedge \pi_k^*\alpha_k$, the form on L^k obtained by pulling back the forms from each individual copy of L and wedging them together. This definition requires a number of comments. Firstly, when $\mathcal{M}_{k+1}(\beta)$ is a smooth manifold, we shall have to assume that the map ev_0 is a smooth submersion. In this case, ev_0 may locally be written as the projection from a product and we can apply the definition of the integration along fibres map given in Chapter 3 Section 8. However, as noted in the previous chapter, since $\mathcal{M}_{k+1}(\beta)$ is not in general a smooth manifold, we need some appropriate notion of submersion and a corresponding definition of *virtual integration along fibres*. Given such a definition, we would expect some form of Stokes' theorem to hold. Secondly, when $\beta = 0$, the moduli spaces $\mathcal{M}_{k+1}(\beta)$ are empty by definition unless $k \geq 2$ since otherwise the disks are not stable. Because the only disk with $\beta = 0$ is the constant map, the space $\mathcal{M}_{k+1}(0)$ ought to be the moduli space of complex structures on D^2 with $k+1$ marked points. Rather than being a smooth manifold, this is actually an orbifold when $k \geq 2$. Hence in the special case where $\beta = 0$, we shall instead define $m_1^{\beta=0}$ to be the de Rham differential d on forms, and $m_k^{\beta=0}$ to be zero for all $k \geq 2$.

THEOREM 5.1. *The pair $(\Omega^*(L; \Lambda_0^{\mathbb{C}}), \{m_k\}_{k=0}^\infty)$ is a weak A_∞ algebra.*

Proof. Fix some $\beta \in \pi_2(X, L)$. We shall calculate the quantity

$$\int^{\mathrm{vir}}_{\mathcal{M}_{k+1}(\beta,L) \xrightarrow{\mathrm{ev}_0^\beta} L} \mathrm{d}(\mathrm{ev}_{(k)}^\beta)^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k) \tag{5.2}$$

in two different ways. Since d and the pullback by $\mathrm{ev}_{(k)}^\beta$ commute, we obtain

$$\sum_{i=1}^k (-1)^* (\mathrm{ev}_{(k)}^\beta)^*(\alpha_1 \vartriangle \cdots \vartriangle d\alpha_i \vartriangle \cdots \vartriangle \alpha_k)$$

where $*$ represents some sign that we shall not concern ourselves with. Using the definitions above, we may rewrite this as

$$\sum_{i=1}^k (-1)^* m_k^\beta(\alpha_1, \ldots, m_1^0(\alpha_i), \ldots, \alpha_k) \tag{5.3}$$

The second way to calculate the above quantity is to apply our putative Stokes' theorem for integration along fibres to yield

$$\int^{\mathrm{vir}}_{\partial^{(1)}\mathcal{M}_{k+1}(\beta) \xrightarrow{\mathrm{ev}_0^\beta} L} (\mathrm{ev}_{(k)}^\beta)^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k)$$

Of course, we know the structure of the codimension-1 boundary $\partial^{(1)}\mathcal{M}_{k+1}(\beta)$ from our discussion in the previous chapter. Hence we may rewrite the above as:

$$\sum_{\substack{\beta_1+\beta_2=\beta \\ k_1+k_2=k+1 \\ k_2 \geq 0, i=1,\ldots,k_1}} (-1)^* \int^{\mathrm{vir}}_{\mathcal{M}_{k_1+1}(\beta_1) \times_{\mathrm{ev}_i} \mathcal{M}_{k_2+1}(\beta_2) \xrightarrow{\mathrm{ev}_0^\beta} L} (\mathrm{ev}_{(k)}^\beta)^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k)$$

where we exclude the cases $k_1 = 1, \beta_1 = 0$ and $k_2 = 1, \beta_2 = 0$ from this sum as they are excluded from the boundary of $\mathcal{M}_{k+1}(\beta)$. Since we are integrating over the fibres of the fibre product $\mathcal{M}_{k_1+1}(\beta_1) \times_{\mathrm{ev}_i} \mathcal{M}_{k_2+1}(\beta_2)$, it is clear that we have

$$\int^{\mathrm{vir}}_{\mathcal{M}_{k_1+1}(\beta_1)\times_{\mathrm{ev}_i}\mathcal{M}_{k_2+1}(\beta_2)\xrightarrow{\mathrm{ev}_0^\beta} L} (\mathrm{ev}^\beta_{(k)})^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k) =$$

$$\int^{\mathrm{vir}}_{\mathcal{M}_{k_1+1}(\beta_1,L)\xrightarrow{\mathrm{ev}_0^{\beta_1}} L} (\mathrm{ev}^\beta_{(k_1+1)})^* \left(\alpha_1 \vartriangle \cdots \alpha_{i-1} \vartriangle \left(\int_{\mathcal{M}_{k_2+1}(\beta_2)\xrightarrow{\mathrm{ev}_0^{\beta_2}} L} (\mathrm{ev}^\beta_{(k_2+1)})^*(\alpha_i \vartriangle \cdots \alpha_{i+k_2-1})\right) \vartriangle \alpha_{i+k_2} \vartriangle \cdots \alpha_k\right)$$

by chasing $\alpha_1 \vartriangle \cdots \vartriangle \alpha_k$ around the commutative diagram

$$\begin{tikzcd}
 & \mathcal{M}_{k_1+1}(\beta_1) \times_{\mathrm{ev}_i,\mathrm{ev}_0} \mathcal{M}_{k_2+1}(\beta_2) \arrow[r] \arrow[d] \arrow[dl] & \mathcal{M}_{k_2+1}(\beta_2) \arrow[r, "\mathrm{ev}^{\beta_2}_{(k_2)}"] \arrow[d, twoheadrightarrow, "\mathrm{ev}_0^{\beta_2}"] & L^{k_2} \\
L & \mathcal{M}_{k_1+1}(\beta_1) \arrow[l, "\mathrm{ev}_0^{\beta}", "\mathrm{ev}_0^{\beta_1}"'] \arrow[r, "\mathrm{ev}_i^{\beta_1}"] \arrow[d, "\mathrm{ev}^{\beta_1}_{(k_1-1)}"] & L & \\
 & L^{k_1-1} & L^{k_1} \arrow[l] \arrow[u] & L^k \arrow[l] \arrow[uu]
\end{tikzcd}$$

where ev^{β_1}, ev^{β_2} have been used to denote the various evaluation maps for the moduli spaces $\mathcal{M}_{k_1+1}(\beta_1)$ and $\mathcal{M}_{k_2+1}(\beta_2)$ respectively, and $\mathrm{ev}^{\beta_1}_{(k_1-1)}$ is the map $\mathcal{M}_{k_1+1}(\beta_1) \to L^{k_1-1}$ given by evaluating at all the marked points *except* i. But this is simply

$$\sum_{\substack{k_1+k_2=k+1\\ \beta_1+\beta_2=\beta\\ k_2\geq 0, i=1,\ldots,k_1}} (-1)^* m^{\beta_1}_{k_1}(\alpha_1,\ldots,\alpha_{i-1}, m^{\beta_2}_{k_2}(\alpha_i,\ldots,\alpha_{i+k_2}), \alpha_{i+k_2+1},\ldots,\alpha_k) = 0$$

and we again exclude the cases $k_1 = 1, \beta_1 = 0$ and $k_2 = 1, \beta_2 = 0$. The latter missing terms are precisely those computed in equation 5.3; the former terms must take the form $m^0_1 m^{\beta_2}_{k_2}(\alpha_1,\ldots,\alpha_k) = dm^{\beta_2}_{k_2}(\alpha_1,\ldots,\alpha_k)$ and hence must be zero for dimensional reasons. Combining the two expressions for 5.2 therefore yields the equation

$$\sum_{\substack{k_1+k_2=k+1\\ \beta_1+\beta_2=\beta\\ k_2\geq 0, i=1,\ldots,k_1}} (-1)^* m^{\beta_1}_{k_1}(\alpha_1,\ldots,\alpha_{i-1}, m^{\beta_2}_{k_2}(\alpha_i,\ldots,\alpha_{i+k_2}), \alpha_{i+k_2+1},\ldots,\alpha_k) = 0$$

which is precisely the A_∞ relation for an A_∞ algebra over the Novikov ring $\Lambda_0^{\mathbb{C}}$ (up to some signs). With further effort, it is possible (but unpleasant) to determine the signs also. ■

Remark 5.2. It is not clear from the above that the A_∞ algebra structure on $\Omega^*(L;\Lambda_0^{\mathbb{C}})$ is really independent of the almost-complex structure J. In fact, in [FOOO09]it is only independent of J up to a *homotopy equivalence* of (gapped, filtered) A_∞ algebras. We shall not pursue this here.

5.1.1 REMARKS ON OTHER DEFINITIONS

The work of Fukaya-Oh-Ohta-Ono uses a *homological* version of the above theory. We shall briefly describe here how these pictures are *dual* in the case where all moduli spaces in question are assumed to be smooth manifolds. A **smooth k-cycle** P in a (compact, oriented) smooth manifold L is simply a smooth map $i : P \to L$ from a (compact, oriented) k-manifold P. Denote the set of k-cycles by $\mathcal{C}_k(L)$. These cycles can be used to define a homology theory, with equivalence of cycles given by cobordism and addition given by disjoint union (see [Nic07], p.263). A cycle defines a linear functional $H^*(L) \to \mathbb{R}$ via

$$\alpha \mapsto \int_P i^*\alpha$$

and hence by Poincaré duality, P corresponds to a cohomology class in $H^*(L)$, denoted δ_i; we call this the **Poincaré dual** of the cycle P. Hence we must have

$$\int_P i^*\alpha = \int_L \delta_i \wedge \alpha$$

and therefore from the projection formula for integration along fibres, we see that $\delta_i = i_*(1)$ whenever i is a submersion.

Suppose now that $P_1, \ldots, P_k$ are smooth cycles dual to the forms $\alpha_1, \ldots, \alpha_k$. Now we can define an operation $m_k(P_1, \ldots, P_k) \in \mathcal{C}_*(L)$ to be given by the map ev_0 from the fibre product

$$\mathcal{M}_{k+1}(\beta) \times_{(\mathrm{ev}_1,\ldots,\mathrm{ev}_k),(i_1,\ldots,i_k)} (P_1 \times \cdots \times P_k)$$

and we assume that this map is a submersion. Recall that the fibre product of manifolds exists as a manifold whenever the relevant maps are transversal. We shall assume that this is the case, so that ev_0 does indeed give a smooth cycle. In the case when $P_1, \ldots, P_k$ are embeddings of submanifolds, we may regard the fibre product space $\mathcal{M}_{k+1}(\beta) \times_{(ev_1,\ldots,\mathrm{ev}_k),(i_1,\ldots,i_k)} (P_1 \times \cdots \times P_k)$ as the *moduli space of constrained disks*; those that are forced to intersect the submanifolds $P_1, \ldots, P_k$, while the 0 marked point is left free to vary. Now we claim that the Poincaré dual of the cycle $m_k(P_1, \ldots, P_k)$ is given exactly by $m_k(\alpha_1, \ldots, \alpha_k)$. Firstly, we observe that the Poincaré dual of the cycle $i_1 \times \cdots \times i_k : P_1 \times \cdots \times P_k \to L^k$ is given by $\alpha_1 \vartriangle \cdots \vartriangle \alpha_k$. This is simply an application of Fubini's Theorem (see [Nic07], p.267 for a proof). Now we shall use P to denote $P_1 \times \cdots \times P_k$ and write $i = i_1 \times \cdots i_k : P \to L^k$. We then have a pullback diagram

$$\begin{tikzcd}
\mathcal{M}_{k+1}(\beta) \times_{\mathrm{ev}_k, i} P \arrow[r] \arrow[d, "\pi"] & P \arrow[d, "i"] \\
\mathcal{M}_{k+1}(\beta) \arrow[r, "\mathrm{ev}_k"] \arrow[d, "\mathrm{ev}_0"] & L^k \\
L &
\end{tikzcd}$$

Since the cycle π in $\mathcal{M}_{k+1}(\beta)$ is the pullback of $\delta_i = \alpha_1 \vartriangle \cdots \vartriangle \alpha_k$ in L^k, we have $\mathrm{ev}_k^* \delta_i = \delta_\pi$. Hence the Poincaré dual of the evaluation map $\mathrm{ev}_0 \circ \pi$ from the fibre product is given by

$$\delta_{\mathrm{ev}_0 \circ \pi} = (\mathrm{ev}_0 \circ \pi)_*(1) = (\mathrm{ev}_0)_*(\delta_\pi) = (\mathrm{ev}_0)_*(\mathrm{ev}_k)^*(\alpha_1 \vartriangle \cdots \vartriangle \alpha_k)$$

exactly as claimed.

We may observe from the above that Lagrangian intersection Floer (co)homology can be defined with reference to any homology or cohomology theory that admits a certain set of purely formal constructions, such as pullbacks, pushforwards and fibre products; almost everything in the arguments above was purely formal. The model of the Fukaya category in the previous section, on *virtual de Rham cohomology*, is just one of many possible definitions. The homology model used by Fukaya-Oh-Ohta-Ono is called *smooth singular homology* and is similar to the idea of smooth cycles above except that fibre products exist in general; a suitably generic diffeomorphism is incorporated into the definition in order to achieve transversality. In the sequel we shall often find it convenient to switch between the two pictures.

5.2 ANOMALY AND OBSTRUCTION

Note that the algebra defined above will in general only be a *weak* A_∞ algebra. According to the above definitions, the element $m_0 \in \Omega^*(L; \Lambda_0^{\mathbb{C}})$ is given by

$$m_0 = \sum_{\beta \in \pi_2(X,L)} \left(\int_{\mathcal{M}_1(\beta)} 1 \right) T^{\int_\beta \omega}$$

and hence is exactly a count of the disk bubbles that occur on L. The fact that $m_0 \neq 0$ is precisely what stops us from having $m_1^2 = 0$ for an A_∞ algebra (see Appendix C). It is thus called the **anomaly**. However, if we can 'deform' the A_∞ operations so that $m_0 = 0$, then we will also have $m_1^2 = 0$ and we will be able to define the cohomology of the complex. This is precisely the role of a *bounding cochain*.

Given a *weak* A_∞ algebra, we often wish to produce a (strong) A_∞ algebra. This may be done using the formalism of *bounding cochains*. Let A be a weak A_∞ algebra. An element $b \in A$ is called a **bounding cochain** if it satisfies the **Maurer-Cartan equation**

$$\sum_{k=0}^{\infty} m_k(b, \cdots, b) = 0$$

where some filtration is required to make sense of this infinite sum; we shall pretend that this is not a problem. Note that if $m_0 = 0$, then $b = 0$ is (clearly) a bounding cochain. We use $\mathcal{MC}(A)$ to denote the set of bounding cochains for A. If $\mathcal{MC}(A)$ is non-empty, then we say A is **unobstructed**. Suppose now that we are given $b \in \mathcal{MC}(A)$ for a weak

A_∞ algebra A. Then we can use b to define a strong A_∞ algebra A_b as follows. On the level of modules, we let $A = A_b$ and define **deformed** A_∞ **operations** $m_k^b : A_b^{\otimes k} \to A_b$ by

$$m_k^b(\alpha_1, \ldots, \alpha_k) = \sum_{n_1,\ldots,n_k \geq 0} m_{k+n_0+\cdots+n_k}(\underbrace{b, \cdots, b}_{n_0}, \alpha_1, \underbrace{b, \cdots, b}_{n_1}, \ldots, \alpha_k, \underbrace{b, \cdots, b}_{n_k})$$

In other words, to define m_k^b we 'insert the bounding cochain b into m_k' in every possible manner and then sum. It is clear that b is a bounding cochain exactly when $m_0^b = 0$. What is not so clear is that this actually defines an A_∞ algebra for an arbitrary choice of b (of strictly positive degree). To prove this, it is more convenient to use the *tensor coalgebra* associated to an A_∞ algebra: see [FOOO09, §3.6]

Extending the terminology above, we say that the Lagrangian L is **unobstructed** if there exists a bounding cochain for the weak A_∞ algebra on $\Omega^*(L; \Lambda_0^{\mathbb{C}})$, and we use $\mathcal{MC}(L)$ to denote the set of bounding cochains for this weak A_∞ algebra. Hence when L is unobstructed we can, as described above, produce a (strong) A_∞ algebra associated to L by deforming the A_∞ operations. This may seem excessively formal, but there is also a more geometric formulation of the above procedure. *In the discussion that follows, we shall disregard problems with signs.* Firstly, let G be the set of homotopy classes $\beta \in \pi_2(X, L)$ such that $\int_\beta \omega > 0$. This is an abelian semi-group under addition, and if we write $\beta_1 > \beta_2$ if $\beta_1 = \beta_2 + \beta_3$ for some other $\beta_3 \in G$, then this gives us a partial order on G. We shall call those $\beta \in G$ with lowest non-zero order the **primitive elements**, and will denote such classes by $\beta_{(1)}$; they are the elements of G that cannot be decomposed as sums, and have minimal energy. Hence the corresponding moduli spaces $\mathcal{M}_{k+1}(\beta_{(1)})$ do not require compactification. In particular, the evaluation map $\mathrm{ev}_0 : \mathcal{M}_1(\beta_{(1)}) \to L$ is a smooth singular cycle because the boundary is empty. We denote this by $\mathfrak{o}_{\beta_{(1)}}(L)$, and its Poincaré dual by $\tilde{\mathfrak{o}}_{\beta_{(1)}}(L)$. This is the **first obstruction class**. To proceed to the next step, we must *assume* that this obstruction vanishes for *every* $\beta_{(1)}$, that is, the cycle $\mathrm{ev}_0 : \mathcal{M}_1(\beta_{(1)}) \to L$ is always the boundary of some smooth cycle $\mathrm{ev}_0 : B_{\beta(1)}(L) \to L$, denoted $\mathfrak{b}_{\beta_{(1)}}(L)$, with $\partial B_{\beta(1)}(L) = \mathcal{M}_1(\beta_{(1)})$. Therefore $\mathfrak{o}_{\beta_{(1)}}(L) = \partial \mathfrak{b}_{\beta_{(1)}}(L)$ and hence $\tilde{\mathfrak{o}}_{\beta_{(1)}}(L) = d\tilde{\mathfrak{b}}_{\beta_{(1)}}(L)$ for the Poincaré dual $\tilde{\mathfrak{b}}_{\beta_{(1)}}(L)$ of $\mathfrak{b}_{\beta_{(1)}}(L)$.

Now we consider those elements in G of second-lowest non-zero order, denoted using $\beta_{(2)}$. The boundary of $\mathcal{M}_1(\beta_{(2)})$ is hence given by

$$\coprod_{\beta_{(2)} = \beta_{(1)} + \beta'_{(1)}} \mathcal{M}_2(\beta_{(1)}) \times_{(\mathrm{ev}_1, \mathrm{ev}_0)} \mathcal{M}_1(\beta'_{(1)})$$

where the union is taken over all decompositions of $\beta_{(2)}$ into sums of the form $\beta_{(1)} + \beta'_{(1)}$. We know, however from our assumption above, that $\mathcal{M}_1(\beta'_{(1)}) = \partial B_{\beta'_{(1)}}(L)$. Let us define

$$\mathcal{M}_1(\beta_{(2)}, \beta'_{(1)}) = \mathcal{M}_2(\beta_{(2)} - \beta'_{(1)}) \times_{(\mathrm{ev}_1, \mathrm{ev}_0)} B_{\beta'_{(1)}}(L)$$

Then since $\beta_{(2)} - \beta'_{(1)} = \beta_{(1)}$, the space $\mathcal{M}_2(\beta_{(2)} - \beta'_{(1)})$ requires no compactification and so

$$\partial \mathcal{M}_1(\beta_{(2)}, \beta'_{(1)}) = \mathcal{M}_2(\beta_{(1)}) \times_{(\mathrm{ev}_1, \mathrm{ev}_0)} \mathcal{M}_1(\beta'_{(1)})$$

Hence if we take a smooth singular chain defined by

$$\mathfrak{o}_{\beta(2)}(L) = \mathrm{ev}_0 : \mathcal{M}_1(\beta_2) - \sum_{\beta'_{(1)} < \beta_{(2)}} \mathrm{ev}_0 : \mathcal{M}_1(\beta_{(2)}, \beta'_{(1)})$$

then we see that $\mathfrak{o}_{\beta_{(2)}}(L)$ is indeed a cycle because of the cancellations that occur. The Poincaré dual $\tilde{\mathfrak{o}}_{\beta_{(2)}}(L)$ is called the **second obstruction class**. Again, to proceed further we must assume that $\mathfrak{o}_{\beta_{(2)}}(L)$ is the boundary of some $\mathrm{ev}_0 : B_{\beta_{(2)}}(L) \to L$, for every $\beta_{(2)}$.

We may inductively construct a sequence of obstruction classes $\mathfrak{o}_k(L)$ in $\Omega^*(L)$; if the k-th obstruction $\mathfrak{o}_{\beta_{(k)}}$ is equal to the boundary of some $B_{\beta_{(k)}}(L)$ for every class $\beta_{(k)}$, then we can construct the $(k+1)$-st obstruction $\mathfrak{o}_{\beta_{(k+1)}}(L)$ as follows. Let us define

$$\mathcal{M}_1(\beta_{(k+1)}, \beta_{(i)}, \ldots, \beta_{(i_m)}) = \mathcal{M}_{m+1}(\beta_{(k+1)} - \beta_{(i)} - \cdots - \beta_{(i_m)}) \times_{\mathrm{ev}_m, \mathrm{ev}_0} \prod_{j=1}^m B_{\beta_{(i_j)}}(L)$$

using our assumption that the obstruction of order less than or equal to k all vanish. Then the $(k+1)$-st obstruction is given by

$$\mathfrak{o}_{\beta_{(k+1)}}(L) = \sum_{\substack{m=0,1,\ldots,k \\ i_1,\ldots,i_m \leq k \\ \beta_{(i_1)},\ldots,\beta_{(i_m)} \in G}} \frac{(-1)^*}{m!} \mathcal{M}_1(\beta_{(k+1)}, \beta_{(i)}, \ldots, \beta_{(i_m)})$$

which is indeed a cycle under the given assumptions; the factor of $1/m!$ and the undetermined sign $(-1)^*$ are included to make the combinatorics come out correctly.

We claim now that if all obstruction classes vanish, then in fact L is unobstructed, with bounding cochain $b \in \Omega^*(L; \Lambda_0^{\mathbb{C}})$ given by

$$b = \sum_{\substack{i \in \mathbb{N} \\ \beta_{(i)} \in G}} \tilde{\mathfrak{b}}_{\beta_{(i)}}(L) T^{\int_{\beta_{(i)}} \omega}$$

To prove this, it is helpful to instead work in the Poincaré dual picture as described above. In this model we can see that

$$b = \sum_{\substack{i \in \mathbb{N} \\ \beta_{(i)} \in G}} \mathfrak{b}_{\beta_{(i)}}(L) T^{\int_{\beta_{(i)}} \omega}$$

Now, since we have

$$\sum_{k=0}^{\infty} m_k(b, \cdots, b) = \sum_{\beta \in G} T^{\int_\beta \omega} \sum_{k=0}^{\infty} m_k^\beta(b, \ldots, b)$$

if we expand out the definition of b and simply collect powers of T, we find that they are of the form

$$T^{\int_\beta \omega} \sum_{\substack{m=0,1,\ldots k \\ i_1, \ldots, i_m \leq k \\ \beta_{(i_1)}, \ldots, \beta_{(i_m)} \in G}} \frac{1}{m!} \mathrm{ev}_0 : \mathcal{M}_{m+1}(\beta - \beta_{(i)} - \cdots - \beta_{(i_m)}) \times_{\mathrm{ev}_m, \mathrm{ev}_0} \prod_{j=1}^{m} B_{\beta_{(i_j)}}(L)$$

which are all zero by definition of the obstruction classes and the assumption that they vanish.

5.2.1 IMMERSED LAGRANGIAN FLOER THEORY

We have shown above how to construct a weak A_∞ algebra structure on $\Omega^*(L; \Lambda_0^{\mathbb{C}})$ for a single embedded Lagrangian submanifold, and how to construct a bounding cochain b that we can use to deform the A_∞ operations to yield a (strong) A_∞ algebra. Now we wish to study the corresponding situation for immersed Lagrangian submanifolds L. We briefly describe here the modifications that must be made in this case, due to [AJ10].

Suppose $\iota : L \looparrowright (X, \omega)$ is an immersion of a Lagrangian submanifold $\iota(L)$ into X, having only transversal self-intersections. Let $R = \{(p_+, p_-) \in L \times L : p_- \neq p_+, \ \iota(p_+) = \iota(p_-)\}$ be the set of preimages of these intersection points. Let $\beta \in \pi_2(X, \iota(L))$ be given. To define the moduli space $\mathcal{M}_{k+1,\alpha}(\beta, L, X, J)$ of disks $u : D^2 \to X$ with $k+1$ marked points $z_1, \ldots, z_k$ on the boundary in $\iota(L)$, we require that there exists a subset α of $\{0, \ldots, k\}$, a function $j : \alpha \to R$, and a lift of the map $u : \partial D^2 \setminus \{z_1, \ldots, z_k\} \to \iota(L)$ to a map $\tilde{u} : \partial D^2 \setminus \{z_1, \ldots, z_k\} \to L$, such that for all $i \in \alpha$, the value $\tilde{u}(z_i)$ is equal to $j(i)$ in R. These marked points $z_i, i \in \alpha$ represent points where the disk u can 'jump' between different branches of the Lagrangian L. As before, we require these maps to be stable under the reparametrisation action that maps jumping points $z_i, i \in \alpha$ to other jumping points; hence when $\beta = 0$ we require that $k \geq 2$. Suppose that we have $u(z_i) = \iota(p_\pm)$. In order to obtain a Fredholm system, we shall choose a path of Lagrangian subspaces inside $T_{\iota(p_\pm)} X$ connecting $\iota_* T_{p_-} L$ to $\iota_* T_{p_+} L$, and reparametrise the disk around z_i so as to insert this path. This also allows us to define a Maslov index for u as before, by taking the Maslov index of the loop of Lagrangian submanifolds obtained by going around the boundary ∂D^2 of the disk. Hence we can regard $\mathcal{M}_{k+1,\alpha}(\beta, L, X, J)$ as a Fredholm system as before, with virtual dimension depending on the Maslov index of the disks. The compactification of this moduli space is identical, but with extra labelling for the set α. Broken disks in the moduli space $\mathcal{M}_{k_1+1,\alpha_1}(\beta_1, L, J, L) \times_{\mathrm{ev}_i, \mathrm{ev}_0} \mathcal{M}_{k_2+1,\alpha_2}(\beta_2, X, J, L)$ correspond to disks in $\mathcal{M}_{k+1,\alpha}(\beta, L, X, J)$ with index set of jumping points given by

$$\alpha = \alpha_1 \cup_i \alpha_2 = \{k \in \alpha, k < i\} \cup \{k + k_2 + 1 : \ k \in \alpha_1, \ k > i\} \cup \{k + i : \ k \in \alpha_2\}$$

Now we define the moduli space

$$\mathcal{M}_{k+1}(\beta, L) = \coprod_{\alpha \subseteq \{0, \ldots, k\}, j : \alpha \to R} \mathcal{M}_{k+1,\alpha}(X, L, \beta)$$

which has evaluation maps $\mathrm{ev}_k^\beta : \mathcal{M}_{k+1}(\beta, L) \to L^k \coprod R^k$ and $\mathrm{ev}_0^\beta : \mathcal{M}_{k+1}(\beta, L) \to L \coprod R$. Defining $\tilde{L} = L \coprod R$, we can consider $\Omega^*(\tilde{L}; \Lambda_0^{\mathbb{C}})$ as before. In fact, we still have the exact same A_∞ structure on differential forms when we use the definitions from the previous section (the proofs carry through). Also, the obstruction theory works as before, and so we say that L is unobstructed if we can find a bounding cochain b for $\Omega^*(\tilde{L}; \Lambda_0^{\mathbb{C}})$.

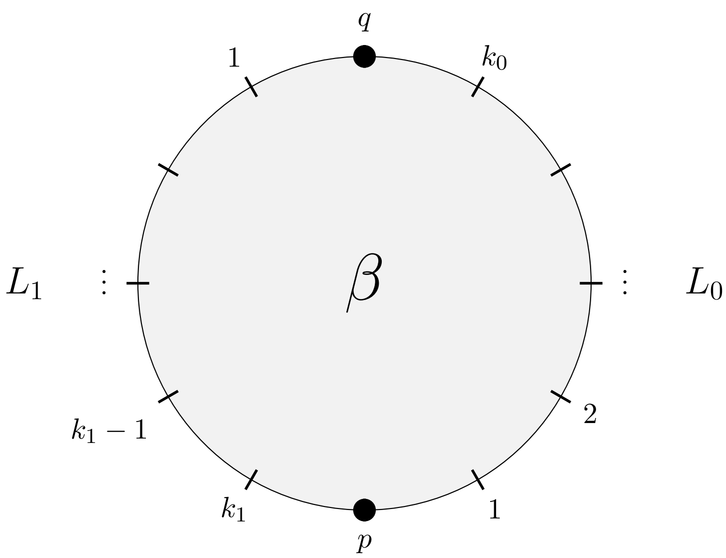

Figure 5.1: Labelling of marked points for moduli spaces of disks with boundary in pair of Lagrangians

5.3 LAGRANGIAN INTERSECTION FLOER HOMOLOGY

Now that we have shown how to construct a (strong) A_∞ algebra associated to an unobstructed and possibly immersed Lagrangian submanifold, we now return to the question of defining Lagrangian intersection Floer homology. Suppose L_0, L_1 are two Lagrangian submanifolds with bounding cochains b_0, b_1 respectively. For $\beta \in \pi_2(X, L_0 \cup L_1)$, we can consider the moduli space $\mathcal{M}_{k_0,k_1}(p, q, \beta)$ of marked J-holomorphic disks having boundary conditions as shown in Figure 5.1.

This comes with an evaluation map $\mathrm{ev}^\beta : \mathcal{M}_{k_0,k_1}(p, q, \beta) \to L_0^{k_0} \times L_1^{k_1}$ as usual. The module $\mathrm{CF}^*(L_0, L_1) = \Lambda_0^{\mathbb{C}} \{L_0 \cap L_1\}$ generated by intersection points of L_0 and L_1 is the chain complex for Lagrangian intersection Floer homology. We wish to endow this space with the structure of a A_∞ *bimodule* over the Lagrangian algebras associated to L_0, L_1. Firstly, given an A_∞ algebra A, we define the **tensor coalgebra** $T(A)$ to be the direct sum

$$T(A) = \sum_{n=0}^{\infty} A^{\otimes n}$$

with the obvious grading coming from the grading of A. Now we define

DEFINITION 5.2. *Given two A_∞ algebras (A_0, m_k^0) and (A_1, m_k^1), a free, graded $\Lambda_0^{\mathbb{C}}$-module $D = \bigoplus_{m\in\mathbb{Z}} D^m$ is called a (A_1, A_0) A_∞ **bimodule** if there are operations $n_{k_0,k_1} : A_1^{\otimes k_1} \otimes D \otimes A_0^{\otimes k_0} \to D$ of degree 1 such that if we define $\hat{d} : T(A_1) \otimes D \otimes T(A_2) \to T(A_1) \otimes D \otimes T(A_2)$ by*

$$\begin{aligned}&\hat{d}(x_1 \otimes \cdots \otimes x_{k_1} \otimes p \otimes y_1 \otimes \cdots \otimes y_{k_0})\\ &= \sum_{k_1' \le k_1,\ k_0' \le k_0} (-1)^* x_1 \otimes \cdots \otimes x_{k_1-k_1'} \otimes n_{k_1',k_0'}(x_{k_1-k_1'+1}, \ldots, x_{k_1}, p, y_1, \ldots, y_{k_0'}) \otimes y_{k_0'+1} \otimes \cdots \otimes y_{k_0})\end{aligned}$$

then we have $\hat{d}^2 = 0$.

We define these operations on $\mathrm{CF}^*(L_0, L_1)$ as follows. Suppose we have $x_i \in \Omega^*(L; \Lambda_0^{\mathbb{C}})$, for $i = 0, 1, \ldots, k_1$ and $y_j \in \Omega^*(L; \Lambda_0^{\mathbb{C}})$ for $j = 1, \ldots, k_0$. Then let

$$n_{k_0,k_1}(x_1, \ldots, x_{k_1}, p, y_1, \ldots, y_{k_0}) = \sum_{\substack{\beta\in\pi_2(X,L_0\cup L_1)\\ q\in L_0\cap L_1}} \int^{\mathrm{vir}}_{\mathcal{M}_{k_0,k_1}(p,q;\beta)} (\mathrm{ev}^\beta)^*(x_1 \vartriangle \cdots x_{k_1} \vartriangle y_1 \vartriangle \cdots y_{k_0}) T^{\int_\beta \omega} \cdot q$$

Now, if we apply n_{k_0,k_1} to the bounding cochains (b_0, b_1) and sum over k_0, k_1, then we get a boundary map $\delta^{b_0,b_1} : \mathrm{CF}^*(L_0, L_1) \to \mathrm{CF}^*(L_0, L_1)$ given by

$$\delta^{b_0,b_1}(p) = \sum_{\substack{q\in L_0\cap L_1\\ k_0,k_1\ge 0\\ \beta\in\pi_2(X,L_0\cup L_1)}} \int^{\mathrm{vir}}_{\mathcal{M}_{k_0,k_1}(p,q;\beta)} (\mathrm{ev}^\beta)^*(\underbrace{b_0 \vartriangle \cdots \vartriangle b_0}_{k_0} \vartriangle \underbrace{b_1 \vartriangle \cdots \vartriangle b_1}_{k_1}) T^{\int_\beta \omega} \cdot q$$

The defining relations for the A_∞ bimodule structure, combined with the Maurer-Cartan equations, show that we have $(\delta^{b_0,b_1})^2 = 0$ [FOOO09, Lemma 3.7.14] and hence we can define the **Lagrangian intersection Floer homology** $\mathrm{HF}((L_0, b_0), (L_1, b_1))$ with respect to the bounding cochains b_0, b_1.

We can now define the **Donaldson-Fukaya category** of a symplectic manifold (X, ω). The objects will be pairs (L, b) consisting of a (possibly immersed) Lagrangian submanifold $L \subseteq X$ with only transversal self-intersections that is unobstructed, along with a choice of bounding cochain b for L. Given objects (L_1, b_1) and (L_0, b_0), we define the set of homomorphisms between them to be given by the Lagrangian intersection Floer homology $\mathrm{HF}((L_0, b_0), (L_1, b_1))$ as defined above. These groups have a natural composition homomorphism

$$\mathrm{HF}((L_0, b_0), (L_1, b_1)) \otimes \mathrm{HF}((L_1, b_1), (L_2, b_2)) \to \mathrm{HF}((L_0, b_0), (L_2, b_2))$$

called the *Donaldson product* that makes this collection of objects into a category; see [FOOO09] for the definition. The product operation will not be important for our purposes in Chapter 6.

Chapter 6

EXTENDED TQFT AND YANG-MILLS THEORY

6.1 INSTANTON FLOER HOMOLOGY AS A TQFT

In Chapter 4 we claimed that the instanton Floer homology was independent of the choice of Riemannian metric on the homology 3-sphere Y and thus was a *topological invariant* (since 3-manifolds have only one possible smooth structure). In order to prove this we shall interpret the instanton Floer homology as part of a *topological quantum field theory*. Our discussion in this section will be largely descriptive, avoiding the various analytic difficulties.

Firstly, we shall introduce some simple *Donaldson invariants*. Let X be a compact oriented Riemannian 4-manifold with boundary Y (such a manifold always exists because the cobordism ring is trivial in dimension 3) and Riemannian metric g_X. Let X' denote the manifold obtained by gluing X to the non-compact tubular manifold $\bar{Y} \times [0, \infty)$ along Y. For every representative ρ of a gauge equivalence class of irreducible flat connections in R_Y^*, we define the set $\mathcal{M}_X(\rho)$ to be the moduli space of instantons on X' with L^2 curvature and flat limit ρ along the tubular end. The moduli theory from the previous Chapters may be adapted easily to this case. Additionally, when the dimensions of the relevant moduli spaces are suitably small, we have a compactification with codimension 1 boundary given by

$$\partial\overline{\mathcal{M}_X(\rho)} = \coprod_{\sigma \in R_Y^*} \mathcal{M}_X(\sigma) \times \mathcal{M}_{Y\times\mathbb{R}}(\sigma, \rho)$$

where $\mathcal{M}_{Y\times\mathbb{R}}(\sigma, \rho)$ is the moduli space of instantons over the tube $Y \times \mathbb{R}$ considered in the previous Chapter. This is saying that the only source of non-compactness in $\mathcal{M}_X(\rho)$ is the breaking of instantons at ∞ along $Y \times [0, \infty)$. Recall that we only identify flat connections on Y in R_Y^* if they differ by a gauge transformation of degree 0. In this case, the gauge transformation may be extended to all of X (purely as a matter of topology) and hence the moduli spaces $\mathcal{M}_X(\rho)$ do not depend on the choice of representative. We now define the *relative Donaldson invariant* $\mathcal{D}_{X,g_X} \in \mathrm{CF}(Y, g_Y)$ by the formula:

$$\mathcal{D}_{X,g_X} = \sum_{[\rho]\in R_Y^*} \#\left\{\mathcal{M}_X^0(\rho)\right\} \cdot \rho$$

where $\mathcal{M}_X^0(\rho)$ denotes the component of the moduli space of virtual dimension 0 (if necessary, virtual counting may be used). We claim that this element is a cycle in $\mathrm{CF}(Y, g_Y)$; we can write

$$d_{Y,g_Y}(\mathcal{D}_{X,g_X}) = \sum_{\rho,\sigma\in R_Y^*} \#\left\{\mathcal{M}_X^0(\rho)\right\} \#\left\{\mathcal{M}_{Y\times\mathbb{R}}^0(\rho, \sigma)\right\} \cdot \sigma = \sum_{\sigma\in R_Y^*} \#\left\{\partial\overline{\mathcal{M}}_X^0(\sigma)\right\} \cdot \sigma$$

which is zero by the usual argument. Hence $\mathcal{D}_{X,g_X}$ descends to give a class in $\mathrm{HF}(Y, g_Y)$, called the (degree 0) **relative Donaldson invariant**. There are two special cases of this we would like to discuss. Firstly, if $Y = \emptyset$, then $\mathcal{D}_{X,g_X}$ is simply an integer instead, given by counting the 0-dimensional moduli space of instantons on X. Second is the case where $Y = \bar{Y}_0 \coprod Y_1$ is disconnected. Then X may be regarded as a cobordism from Y_0 to Y_1, and yields a map $\mathcal{D}_{X,g_X} : \mathrm{CF}^*(Y_0, g_{Y_0}) \to \mathrm{CF}^*(Y_1, g_{Y_1})$ that acts on $\rho \in R_{Y_0}^*$ via

$$\mathcal{D}_{X,g_X}(\rho) = \sum_{\sigma\in R_{Y_1}} \#\left\{\mathcal{M}_X^0(\rho, \sigma)\right\} \cdot \sigma$$

where $\mathcal{M}_X^0(\rho,\sigma)$ is the moduli space of instantons over X' with flat boundary condition ρ over $Y_0 \times [0,\infty)$ and σ over $\bar{Y}_1 \times [0,\infty)$. Note that the same proof above that showed $d_{Y,y_0}(\mathcal{D}_{X,g_X}) = 0$ also shows that $\mathcal{D}_{X,g_X}$ is a chain map.

Remark 6.1. Note that in the case $Y_0 = Y_1$ and $X = Y_1 \times \mathbb{R}$, this map is *different* from the boundary map in instanton Floer homology because we do not quotient by the translation action. Because $\dim \mathcal{M}_X^0(\rho,\sigma) = \mu(\rho,\sigma) = 0$ in this case, the map $\mathcal{D}_{X,g_X}$ has **degree** 0 with respect to the grading on $\mathrm{CF}^*(Y_1, g_{Y_1})$.

The main result we wish to prove is:

THEOREM 6.1. *(Topological Quantum Field Theory) The relative Donaldson invariant $\mathcal{D}_{X,g_X} : \mathrm{CF}^*(Y_0, g_{Y_0}) \to \mathrm{CF}^*(Y_1, g_{Y_1})$ is a chain map. On the level of homology, it is independent of the choice of Riemannian metric g_X on X having fixed restrictions g_{Y_0}, g_{Y_1} on the boundary. Moreover, it satisfies the* ***functorality property*** *for cobordisms; if $\partial X_1 = \bar{Y}_0 \coprod Y_1$ and $\partial X_2 = \bar{Y}_1 \coprod Y_2$, with Riemannian metrics g_{X_1}, g_{X_2} respectively, then*

$$\mathcal{D}_{X_2,g_{X_2}} \circ \mathcal{D}_{X_1,g_{X_1}} = \mathcal{D}_{X_1 \cup_{Y_1} X_2, g_{X_1} \# g_{X_2}}$$

where $X_1 \cup_{Y_1} X_2$ denotes the manifold obtained by gluing X_1 and X_2 together along Y_1.

In fact, this implies our topological invariance immediately:

COROLLARY 6.1. *The group $\mathrm{HF}(Y, g_Y)$ is independent of the metric g_Y up to a canonical isomorphism.*

Proof. Suppose g_Y, g_Y' are two Riemannian metrics on Y and take a path of metrics connecting g_Y to g_Y'. This gives rise to a metric g_X on $X = Y \times [0,1]$. Hence we have a chain map $D_{X,g_X} : \mathrm{CF}^*(Y, g_Y) \to \mathrm{CF}^*(Y, g_Y')$. If we take the reversed path, from g_Y' to g_Y, we get another metric $\bar{g}_X$ on $X = Y \times [0,1]$ and hence another chain map $\mathcal{D}_{X,\bar{g}_X} : \mathrm{CF}^*(Y, g_Y') \to \mathrm{CF}^*(Y, g_Y)$. The functorality property for the relative Donaldson invariant implies that $\mathcal{D}_{X,g_X} \circ \mathcal{D}_{X,\bar{g}_X} = \mathcal{D}_{X,g_X \# \bar{g}_X}$. But the metric $g_X \# \bar{g}_X$ has restriction g_Y at both boundaries of $Y \times [0,1]$. The product metric $g_Y \oplus \mathrm{d}t^2$ also has the same restrictions to the boundary, and hence by the invariance of choice of metric on X, we must have $\mathcal{D}_{X,g_X} \circ \mathcal{D}_{X,\bar{g}_X} = \mathcal{D}_{X,g_Y \oplus \mathrm{d}t^2}$ on homology. But we claim that $\mathcal{D}_{X,g_Y \oplus \mathrm{d}t^2}$ is simply the identity map. This is because the 0-dimensional moduli space $\mathcal{M}^0(\rho,\sigma)$ is empty whenever ρ is not gauge-equivalent to σ (for topological reasons), and consists of a single point (the product connection) otherwise. Hence we see that $\mathcal{D}_{X,g_X} \circ \mathcal{D}_{X,\bar{g}_X} = 1$ on the level of homology. A similar argument shows that $\mathcal{D}_{X,\bar{g}_X} \circ \mathcal{D}_{X,g_X} = 1$ also, and hence we have $\mathrm{HF}(Y, g_Y) \cong \mathrm{HF}(Y, g_Y')$ via these isomorphisms. ■

From this we may deduce another important consequence of Theorem 6.1; recalling the definition of a *topological quantum field theory* from Chapter 1, we have

COROLLARY 6.2. *If we assign the Donaldson invariant $\mathcal{D}_X$ to (compact oriented) 4-manifolds X and the instanton Floer homology $\mathrm{HF}(Y)$ to (compact oriented) homology 3-spheres Y, then these satisfy the axioms (1)-(3) of a* ***topological quantum field theory***.

Of course, not all of the axioms for a topological quantum field theory are actually satisfied. Historically, however, this was the motivating example (see Atiyah's paper).

To prove Theorem 6.1, we will first need to introduce some more moduli spaces. Let X be as above and suppose g_X, g_X' are two metrics on X restricting to g_{Y_0}, g_{Y_1} on the respective boundaries; as before, take a path $g_X(t)$ of Riemannian metrics connecting them. Define

$$\hat{\mathcal{M}}_X = \coprod_{\substack{\rho \in R^*_{Y_0} \\ \sigma \in R^*_{Y_1}}} \{([A], t) : [A] \in \mathcal{M}_{X, g_X(t)}(\rho,\sigma)\} = \coprod_{\substack{\rho \in R^*_{Y_0} \\ \sigma \in R^*_{Y_1}}} \hat{\mathcal{M}}_X(\rho,\sigma)$$

which we call the **parametrised moduli space**; let $\hat{\mathcal{M}}^0(\rho,\sigma)$ denote the 0-dimensional part. This allows us to define a map $H : \mathrm{CF}^{i-1}(Y_0, g_{Y_0}) \to \mathrm{CF}^i(Y_1, g_{Y_1})$ via

$$H(\rho) = \sum_{\sigma \in R^*_{Y_1}} \# \left\{ \hat{\mathcal{M}}^0_{X, g_t(X)}(\rho,\sigma) \right\} \cdot \sigma$$

which should give us our desired chain homotopy between $\mathcal{D}_{X,g_X}$ and $\mathcal{D}_{X,g_X'}$, that is, we want to show that $\mathcal{D}_{X,g_X} - \mathcal{D}_{X,g_X'} = d_{Y_1} \circ H + H \circ d_{Y_0}$. We shall derive this relation by considering the compactification of $\hat{\mathcal{M}}_X^{(1)}$. By perturbing the path $g_X(t)$ of metrics appropriately, we can achieve transversality for $\hat{\mathcal{M}}_X^{(1)}$, so that it is a smooth finite-dimensional manifold. We cannot however guarantee that every metric on the path $g_X(t)$ will be generic, that is, $\mathcal{M}_{X,g_X(t)}$ need not be smooth for any specific t. This is because space of non-generic metrics on X divides the space of all metrics into

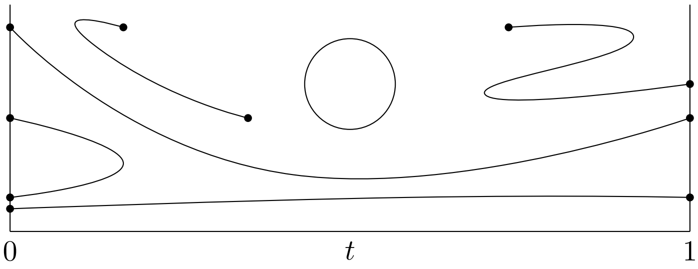

Figure 6.1: A picture of the moduli space $\hat{\mathcal{M}}_X^{(1)}$; the internal endpoints are the result of wall-crossing.

chambers; in order to travel from g_X to g'_X it may be necessary to cross between chambers. This phenomenon is called **wall-crossing** and is illustrated in Figure 6.1. When no wall-crossing occurs then $\hat{\mathcal{M}}_X$ gives the cobordism between $\mathcal{M}_{X,g_X}$ and $\mathcal{M}_{X,g'_X}$ that we claimed existed in Chapter 3. But in general we shall need to include the degeneration points in our compactification of $\hat{\mathcal{M}}_X^{(1)}$.

THEOREM 6.2. *The boundary components of the compactification of $\hat{\mathcal{M}}_X^{(1)}(\rho,\sigma)$ are given by*

- *the endpoints $\mathcal{M}_{X,g_X}(\rho,\sigma)$ and $\mathcal{M}_{X,g'_X}$ (with the reverse orientation);*
- *the internal breaking points of the form*

$$\coprod_{\substack{\rho'\in R^*_{Y_0}\\ \mu(\rho,\rho')=1}} (\mathcal{M}_{Y_0\times\mathbb{R}}(\rho,\rho')/\mathbb{R}) \times \hat{\mathcal{M}}^{(0)}_{X,g_X(t)}(\rho',\sigma)$$

and

$$\coprod_{\substack{\sigma'\in R^*_{Y_1}\\ \mu(\sigma,\sigma')=1}} \hat{\mathcal{M}}^{(0)}_{X,g_X(t)}(\rho,\sigma') \times (\mathcal{M}_{Y_0\times\mathbb{R}}(\sigma,\sigma')/\mathbb{R})$$

For the latter two cases, we need a new kind of gluing theory for moduli spaces with non-generic metrics. In fact, this is a well-known extension of the theory presented in Chapter 4, also due to Taubes; see [Don02, §4.5]

Now the fact that $\#\left\{\partial\overline{\hat{\mathcal{M}}}^{(1)}_{X,g_X}(\rho,\sigma)\right\} = 0$ implies the chain homotopy relation above in the usual way:

$$\begin{aligned}
&\#\left\{\mathcal{M}_{X,g'_X}(\rho,\sigma)\right\} - \#\left\{\mathcal{M}_{X,g_X}(\rho,\sigma)\right\}\\
&= \sum_{\substack{\rho'\in R^*_{Y_0}\\ \mu(\rho,\rho')=1}} \#\left\{(\mathcal{M}_{Y_0\times\mathbb{R}}(\rho,\rho')/\mathbb{R})\right\}\cdot\#\left\{\hat{\mathcal{M}}^{(0)}_{X,g_X(t)}(\rho',\sigma)\right\} + \sum_{\substack{\sigma'\in R^*_{Y_1}\\ \mu(\sigma,\sigma')=1}} \#\left\{\hat{\mathcal{M}}^{(0)}_{X,g_X(t)}(\rho,\sigma')\right\}\cdot\#\left\{(\mathcal{M}_{Y_0\times\mathbb{R}}(\sigma,\sigma')/\mathbb{R})\right\}
\end{aligned}$$

6.2 ATIYAH-BOTT MODULI SPACES AGAIN

We return now to the subject of the Atiyah-Bott moduli space of flat connections over a Riemann surface in order to investigate its relationship with 3-dimensional geometry. We shall introduce some notation. Let Σ be a Riemann surface of genus g and Y be a compact oriented 3-manifold with boundary Σ. Take $P \to Y$ a principal G bundle for some compact semisimple Lie group G, and use $P \to \Sigma$ to denote its restriction Σ. Let $\mathcal{A}_\Sigma$ denote the set of connections on P over Σ, let $\mathcal{A}_Y$ denote the set of connections on P over Y. Let $\mathcal{G}_\Sigma$ denote the gauge group of $P \to \Sigma$ and $\mathcal{G}_Y$ denote the gauge group for P over Y. We shall use $\mathcal{R}_Y$ to denote the set of (all) gauge equivalence classes of flat connections on Y.

Recall that the moduli space $\mathcal{M}_\Sigma$ is a smooth manifold when there are no reducible flat connections on P. As observed in Chapter 4, any principal SU(2) bundle over Σ is necessarily trivial and hence $\mathcal{M}_\Sigma$ will never be a smooth manifold when $G = \mathrm{SU}(2)$. However, in this case $\mathcal{M}_\Sigma$ has instead the structure of a real algebraic variety. By the holonomy theorem (see Appendix A), there is a bijective map between gauge equivalence classes of flat connections on principal G bundles over any manifold X and conjugacy classes of representations of the fundamental group $\pi_1(X)$ in G, given by taking the holonomy around loops. But in general, the moduli space of flat connections $\mathcal{M}_\Sigma$ on a *particular* principal G-bundle P over Σ is only in bijection with the subset of equivalence classes of representations ρ in $\mathrm{Hom}(\pi_1(\Sigma), G)/G$

with the property that, if $\tilde{\Sigma}$ denotes the universal cover of Σ, then the induced principal G-bundle $\tilde{\Sigma} \times_\rho G$ is isomorphic to P [Mro10]. However, because there is a unique SU(2) bundle over any surface Σ (the trivial one), the moduli space $\mathcal{M}_\Sigma$ must be in bijection with the whole set $\mathrm{Hom}(\pi_1(\Sigma), \mathrm{SU}(2))/\mathrm{SU}(2)$. For 3-manifolds Y that are handlebodies, that is, those Y that can be obtained by attaching finitely many 1-handles to the solid 3-ball, then there is only the trivial G-bundle over Y [Weh05b] for any connected Lie group G and hence we have the same bijection between (conjugacy classes of) representations of $\pi_1(Y)$ and (equivalence classes of) flat connections. Hence one can approach the entire problem from the perspective of representation theory and algebraic geometry.

For any affine algebraic group variety G, the set $\mathrm{Hom}(\pi_1(\Sigma), G)$ has the structure of an algebraic variety, called the **representation variety**. For if we are given a presentation of $\pi_1(\Sigma)$ with n generators and r relations, we may regard the representation variety as the closed subset of G^n cut out by the zero locus polynomial equations corresponding to the r relations. To construct the quotient $\mathrm{Hom}(\pi_1(\Sigma), G)/G$ of the representation variety by the conjugation action, one uses geometric invariant theory to identify the ring of invariant functions with the characters of the group representation (see [Sik10]). Hence we can give $\mathrm{Hom}(\pi_1(\Sigma), G)/G$ the structure of an algebraic variety also, called the **character variety**. Note of course that if G is instead an algebraic group over a field that is not algebraically closed (such as SU(2) over $\mathbb{R}$) then G is no longer a variety and it becomes necessary to instead use scheme theory in these constructions to take account of the complex points. At any rate, the techniques of algebraic geometry can be used in this case to show that $\mathcal{M}_\Sigma$ is a smooth manifold away from the singular points, corresponding to reducible representations of $\pi_1(\Sigma)$.

The advantage of this picture is that restriction of flat connections from Y to Σ then has a very simple description in terms of representation theory. If $i : \Sigma \to Y$ denotes the inclusion of the boundary, then there is an induced map on fundamental groups that fits into the long exact sequence

$$\cdots \to \pi_2(Y) \to \pi_2(Y, \Sigma) \xrightarrow{\partial} \pi_1(\Sigma) \to \pi_1(Y) \to \pi_1(Y, \Sigma) \to \cdots$$

associated to the pair (Y, Σ). Whenever Y is a handlebody, we can see from this sequence that $\pi_1(Y)^*\pi_1(\Sigma)/\mathrm{im}\ \partial$ ([Weh05b]). Thus we may identify $\mathcal{R}_Y$ with the subset L_Y of $\mathcal{M}_\Sigma$ consisting of conjugacy classes of representations of the form $\rho : \pi_1(\Sigma) \to G$ such that $\rho(\pi_2(Y, \Sigma)) = \{I\}$. This set L_Y is the image of the map $\mathcal{R}_Y \to \mathcal{M}_\Sigma$ given by pulling back representations from $\pi_1(Y)$ to $\pi_1(\Sigma)$ via the map i_* induced on fundamental groups by i. It is clear that when i_* is not surjective, this map $\mathcal{R}_Y \to \mathcal{M}_\Sigma$ need not be injective. In fact, though the map i_* itself is not in general an immersion ([Sik10]), the image L_Y will in general be an immersed submanifold away from the singular points of $\mathcal{M}_\Sigma$. Moreover, away from these singular points and self-intersection points, the submanifold L_Y will in fact be Lagrangian with respect to the symplectic structure on $\mathcal{M}_\Sigma$, which admits a description purely in terms of representation theory [Sik10]. We shall sketch a more differential-geometric proof of this fact below, inspired by the proof in [Sik10].

If we instead choose to work with $G = \mathrm{SO}(3)$, then one can give a simple topological condition for the non-existence of reducible connections, and hence the smoothness of the moduli space $\mathcal{M}_\Sigma$. We begin by recalling the following standard result:

PROPOSITION 6.1. *[FU84, Theorem 3.1] Suppose P is a principal G-bundle P for G a compact semisimple Lie group and suppose that A is a flat connection on $E = \mathrm{ad}\ P$. Then the following are equivalent:*

1. *the bundle E splits as a sum of vector bundles $E_1 \oplus E_2$ and the connection A can be written as a direct sum $A_1 \oplus A_2$ for A_1, A_2 connections on E_1, E_2 respectively;*
2. *the covariant derivative $d_A : \Omega^0(\mathrm{ad}\ P) \to \Omega^1(\mathrm{ad}\ P)$ has non-trivial kernel;*
3. *the stabiliser Γ_A of A under the action of the gauge group $\mathscr{G}_P$ is strictly larger than the centre $Z(G)$ of G, that is, A is **reducible**.*

Proof. It is easy to see that the kernel of d_A is zero if and only if Γ_A contains no 1-parameter subgroup, since, as we have remarked before, $\ker d_A$ may be regarded as the Lie algebra of the stabiliser of A. Recall [DK90, Lemma 4.2.8] that the stabiliser Γ_A may be identified with the centraliser $Z(H_A)$ of the holonomy group H_A of A (that is, the image of the holonomy representation). Since G is semisimple, $Z(G)$ is discrete, and for any *proper* subgroup $H \subseteq G$, the centraliser must contain a 1-parameter subgroup. This is easy enough to see for $G = \mathrm{SU}(2)$ or $\mathrm{SO}(3)$ where the centraliser of any element not in $Z(G)$ is given by a copy of $U(1)$. Hence we conclude that the kernel of d_A is non-zero if and only if $\Gamma_A = Z(G)$, that is, if and only if A is reducible.

Similarly, we can observe that if A is reducible, then $Z(H_A)$ is strictly larger that $Z(G)$ and so the holonomy group H_A must be a proper subgroup of G. Hence A arises from a reduction of the structure group of P to H_A by [KN69, Theorem 7.1]; this means that we can write P as the induced bundle $P_{H_A} \times_L G$ for some principal H_A-bundle P_{H_A}, and also that A is induced on P from some connection on P_{H_A} (see Appendix A). In this case the induced representation of H_A on $\mathfrak{g}$ has an irreducible sub-representation given by the adjoint representation of H_A on the Lie algebra $\mathfrak{h}_A$. Since

G is semisimple the representation of H_A on $\mathfrak{g}$ therefore splits as a direct sum of $\mathfrak{h}_A$ and another representation ρ of H_A on V. Therefore $\operatorname{ad} P = P_H \times_{\mathrm{ad}} \mathfrak{g}$ splits as a direct sum of the vector bundles $P_H \times_{\mathrm{ad}} \mathfrak{h}_A$ and $P_H \times_\rho V$, with an associated splitting of the connection A. The converse to the above is clear, for if A were to arise from some reduction of the structure group to some H_A, then it must have a large stabiliser under the action of the gauge group (see [KN69, p.105]), given by those automorphisms of P that conjugate H_A inside G. ■

So let $G = \mathrm{SO}(3)$ and suppose that the associated $\mathrm{SO}(3)$ vector bundle $E \to \Sigma$ splits as a direct sum $E = L \oplus E'$. Then by the Whitney sum formula, we have $0 = w_1(E) = w_1(L) + w_1(E')$ and $w_2(E) = w_1(L) \smile w_1(E') = w_1(L) \smile w_1(L)$. However, one can easily see from the description of the $\mathbb{Z}$ cohomology ring of Riemann surface of genus $g \geq 1$ given in [Hat02] that the square of any degree 1 $\mathbb{Z}_2$ cohomology class is zero. Hence $w_2(E) = 0$, so by taking a bundle E with $w_2(E) \neq 0$, we eliminate the possibility of reducible connections. Now suppose $E \to M$ a vector bundle with group $G = \mathrm{SO}(3)$. If M has no boundary then every such bundle with $w_2 = 0$ will be trivial (as observed before). When M has a connected boundary, it will deformation-retract to a 1-manifold and hence every $\mathrm{SO}(3)$ bundle will also be trivial. But in the case where ∂M is disconnected, there will exist a unique $\mathrm{SO}(3)$ vector bundle over M with $w_2(E)$ non-zero on each connected component [Fuk15]. This is the situation considered by Fukaya in his study of instanton Floer homology for 3-manifolds with boundary. If M' is another such 3-manifold with boundary Σ, then this condition also implies that all of the flat connections on $M \sqcup_\Sigma M'$ will be irreducible, allowing us to define $\mathrm{SO}(3)$ instanton Floer homology for this 3-manifold.

Returning to our discussion above:

THEOREM 6.3. *In the case described above, the subset of $\mathcal{M}_\Sigma$ obtained by restricting flat connections from Y to Σ is an (immersed) Lagrangian submanifold.*

Proof. It will be important first to understand the induced map $T_A\mathcal{R}_Y \to T_A\mathcal{M}_\Sigma$ on tangent spaces. So fix some flat connection $A \in \mathcal{A}_Y$ and use this connection to identify $T_A\mathcal{A}_Y$ with $\Omega^1(Y, \operatorname{ad} P)$, and use its restriction to Σ to identify $T_A\mathcal{A}_\Sigma$ with $\Omega^1(\Sigma, \operatorname{ad} P)$. Then the tangent map $T_A\mathcal{A}_Y \to T_A\mathcal{A}_\Sigma$ is simply given by restriction of differential forms from Y to Σ. If we use Hodge theory to identify $T_A\mathcal{R}_Y^* H^1_A(Y)$ with $\ker d_A^* \cap \ker d_A \subseteq \Omega^1(Y, \operatorname{ad} P)$, and the same for $T_A\mathcal{M}_\Sigma$, then the tangent map is again simply given by restriction of differential forms. As this is the same as pulling back by the inclusion $i : \Sigma \to Y$, we can identify the tangent map with the induced map $i^* : H^1_A(Y, \operatorname{ad} P) \to H^1_A(\Sigma, \operatorname{ad} P)$ on twisted cohomology.

So suppose now that $\alpha, \beta \in T_A\mathcal{M}_\Sigma^* \ker d_A^* \cap \ker d_A \subseteq \Omega^1(\Sigma, \operatorname{ad} P)$ are obtained by restriction from $\Omega^1(Y, \operatorname{ad} P)$. Then

$$\int_\Sigma \operatorname{Tr}(\alpha \wedge \beta) = \int_Y \mathrm{d}\operatorname{Tr}(\alpha \wedge \beta) = \int_Y \operatorname{Tr}(d_A\alpha \wedge \beta) - \int_Y \operatorname{Tr}(\alpha \wedge d_A\beta) = 0$$

because $d_A\alpha = d_A\beta = 0$. Hence we see that L_Y is isotropic. To see that the dimension of L_Y is half the dimension of $\mathcal{M}_\Sigma$, take some $A \in \mathcal{R}_Y$ and consider the following commutative diagram of twisted cohomology groups

$$\begin{tikzcd}
H^1_A(Y; \operatorname{ad} P) \arrow[r, "i^*"] \arrow[d, "\sim"] & H^1_A(\Sigma; \operatorname{ad} P) \arrow[r, "\partial^*"] \arrow[d, "\sim"] & H^2_A(Y, \Sigma; \operatorname{ad} P) \arrow[d, "\sim"] \\
H^2_c(Y, \Sigma; \operatorname{ad} P) \arrow[r, "\partial"] & H^1_c(\Sigma; \operatorname{ad} P) \arrow[r, "i_*"] & H^1_c(Y; \operatorname{ad} P)
\end{tikzcd}$$

where H^*_c denotes compactly supported twisted cohomology and the vertical maps come from Poincaré duality for this twisted cohomology (see [BT82, Theorem 7.8]). From this diagram we can observe that i^* and i_* are adjoint maps, and hence that the rank of i^* is equal to the rank of i_*. By the commutativity of the diagram, this is in turn equal to the rank of ∂^*. By the rank-nullity theorem, this rank is equal to $\dim H^1(\Sigma; \operatorname{ad} P) - \dim\ker\partial^*$, and by exactness of the top row we hence conclude that

$$\operatorname{rank} i^* = \dim H^1(\Sigma; \operatorname{ad} P) - \operatorname{rank} i^*$$

from which the conclusion follows. This argument is from [Fre95]; an analogous argument using group co/homology shows that the same result is true in the algebraic case [Sik10]. Next, we must show that the Lagrangian L_Y is actually an immersed submanifold of $\mathcal{M}_\Sigma$. We can do this by considering the commutative diagram

$$\begin{tikzcd}
H^1(Y; \operatorname{ad} P) \arrow[r, "i^*"] \arrow[d, "\sim"] & H^1(\Sigma; \operatorname{ad} P) \arrow[d, "\sim"] \\
T_{[A]}\mathcal{R}_Y \arrow[r, "di^*"] & T_{[A]}\mathcal{M}_\Sigma
\end{tikzcd}$$

Here the top map must have constant rank by our previous argument, and therefore so must di^*. Hence by the constant rank theorem, the image of i must be an immersed submanifold [Sik10].

Finally we must show that the subset $\mathcal{L}_Y$ of $\mathcal{A}_\Sigma$ consisting of restrictions of flat connections is actually gauge-invariant so that it descends to the quotient $\mathcal{M}_\Sigma$. For any Lie group G, we always have $\pi_2(G) = 0$, and when $G = \mathrm{SU}(2)$, then $\pi_1(G) = 0$ also. In this case, the principal $\mathrm{SU}(2)$ bundle P is necessarily trivial and gauge transformations in $\mathscr{G}_\Sigma$ simply correspond to maps $\Sigma \to \mathrm{SU}(2)$, which are all homotopically trivial by cellular approximation. Hence these maps can all be extended to maps $Y \to \mathrm{SU}(2)$, implying that $\mathcal{L}_Y$ is gauge invariant. Note, however, that when $G = \mathrm{SO}(3)$ it need not be the case that $\mathcal{L}_Y$ is gauge-invariant. For instance, if we let Y be the solid torus with boundary $\Sigma = T^2 \cong S^1 \times S^1$ and consider the gauge transformation $T^2 \to \mathrm{SO}(3)$ of the trivial $\mathrm{SO}(3)$ bundle over Σ defined by $(s,t) \mapsto \gamma(s)$ for γ the generator of the group $\pi_1(\mathrm{SO}(3)) \cong \mathbb{Z}_2$, then it is clear that this map cannot extend over Y or else the loop γ would be trivial. We may remedy this difficulty as follows. Let $\mathscr{G}_{\Sigma,Y}$ denote the group of gauge transformations in $\mathscr{G}_\Sigma$ that can be extended to Y. Then we have a commutative diagram:

$$\begin{tikzcd}
\mathcal{A}_Y \arrow[r] \arrow[d] & \mathcal{A}_\Sigma \arrow[d] \\
\mathcal{R}_Y \arrow[r] \arrow[dr, dashed] & \mathcal{A}_\Sigma/\mathscr{G}_{\Sigma,Y} \arrow[d] \\
 & \mathcal{M}_\Sigma
\end{tikzcd}$$

Then we obtain the subset L_Y of $\mathcal{M}_\Sigma$ by composing the map $\mathcal{R}_Y \to \mathcal{A}_\Sigma/\mathscr{G}_{\Sigma,Y}$ with the projection $\mathcal{A}_\Sigma/\mathscr{G}_{\Sigma,Y} \to \mathcal{M}_\Sigma$. Since $\mathscr{G}_\Sigma/\mathscr{G}_{\Sigma,Y}$ is a finite group, this projection map gives a covering space and so the local theory for L_Y does not change; hence we can use the same arguments above to see that L_Y is Lagrangian and immersed. ■

6.3 THE ATIYAH-FLOER CONJECTURE

Returning to the ideas described in the introduction, to construct the extended topological quantum field theory arising from Yang-Mills theory we associate to a Riemann surface Σ the Donaldson-Fukaya category of the moduli space of flat connections $\mathcal{M}_\Sigma$ on the bundle discussed above. If Y_1 is a 3-manifold with boundary Σ, then restriction of flat connections from Y_1 to Σ will yield an immersed Lagrangian submanifold L_{Y_1} of $\mathcal{M}_\Sigma$, which will be an object in the Donaldson-Fukaya category $\mathscr{F}(\mathcal{M}_\Sigma)$. Now, given a homology 3-sphere Y, it is a standard result in the theory of 3-manifolds that there exists a **Heegaard splitting** of Y along an embedded Riemann surface Σ [Ati88]. That is, Y can be obtained by gluing two handlebodies Y_1, Y_2 with $\partial Y_1 = \Sigma = \overline{\partial Y_2}$ along Σ. But in this case, the gluing axiom 2 becomes exactly the Atiyah-Floer conjecture: $\mathrm{HF}_I(Y) \cong \mathrm{Hom}(L_{Y_1}, L_{Y_2}) = \mathrm{HF}_L(L_{Y_1}, L_{Y_2})$. when the Lagrangian intersection Floer homology has no anomalies.

There is a heuristic argument for why we might expect the Atiyah-Floer conjecture to hold. Ideally, the collection of gauge equivalence classes of flat connections on Y will be a finite set, in bijection with the pairs of gauge equivalence classes of flat connections from Y_1 and Y_2 that agree upon restriction to Σ, which in turn should be in bijection with the intersection points of the two Lagrangians L_{Y_1} and L_{Y_2} inside the moduli space of flat connections $\mathcal{M}_\Sigma$. This, at least, indicates that on the level of chain groups, we have an isomorphism between the *instanton* chain group $\mathrm{CF}_I(Y)$ and the *Lagrangian* chain group $\mathrm{CF}_L(L_{Y_1}, L_{Y_2})$. Furthermore, a simple calculation shows that the instantons over the 4-manifold $\Sigma \times [0,1] \times \mathbb{R}$, regarded, as in §4, as maps $[0,1] \times \mathbb{R} \to \mathcal{M}_\Sigma$, satisfy the pseudoholomorphic curve equation when the metric on Σ is shrunk to a point, that is, we use the product metric $\lambda^2 g \oplus \mathrm{d}s^2 \oplus \mathrm{d}t^2$ on $\Sigma \times [0,1] \times \mathbb{R}$ and take $\lambda \to 0$. By conformal invariance, this is equivalent to the case where metric on the 'neck' $[0,1] \times \Sigma$ is stretched along the $[0,1]$-direction, as proposed in [Ati88]. In fact, even more is true. The Euler characteristic of $\mathrm{HF}_I(Y)$ is given by the *Casson invariant* $\lambda(Y)$, which, by a Theorem of Taubes, is equal to the intersection number of the two Lagrangians L_{Y_1} and L_{Y_2} in $\mathcal{M}_\Sigma$. This latter quantity is in turn equal to the Euler characteristic of the Lagrangian intersection Floer homology $\mathrm{HF}_L(L_{Y_1}, L_{Y_2})$ whenever it exists with no anomaly. In fact, the original motivation given by Atiyah for proposing the conjecture in [Ati88] was precisely to explain the coincidence of these Euler characteristics.

In terms of theoretical physics, the Atiyah-Floer conjecture is the observation that $\mathrm{SU}(2)$ gauge theory on a cylindrical space-time of the form $\Sigma \times \mathbb{R} \times [0,1]$ reduces upon Kaluza-Klein reduction of Σ (at least in the semi-classical limit), to the open A-model of topological string theory for maps $\mathbb{R} \times [0,1] \to X$ from the string worldsheet $\mathbb{R} \times [0,1]$ to the background Kähler manifold X given by the moduli space of stable vacua with topological quantum number 0. In the first theory, one places boundary conditions at $\pm\infty$ and these correspond in the second theory to boundary conditions for the open string along D-branes (Lagrangians) inside X. The homotopy between the two chain complexes is then constructed via an *adiabatic limit* argument; if the metric along Σ is squashed sufficiently slowly then the semi-classical ground states in the first system will continuously evolve to those in the second system. From a mathematical perspective, however, this argument is difficult to analyse, largely because differential geometry lacks a formalism for making sense of degenerate Riemannian metrics (though see [DS94] for the special case of mapping cylinders).

Many attempts were made in the early 1990s to prove the Atiyah-Floer conjecture by these direct 'adiabatic limit' arguments, most notably, in the special case of mapping cylinders, by Dostoglou-Salamon [DS94]. These arguments are however largely difficult to make rigorous; focus shifted to the consideration of alternative moduli spaces, culminating in a long series of papers by Salamon-Wehrheim 2002-2009 claiming to prove 'half' of the conjecture [SW08] [Weh05b] [Weh05a]. Of course, only recently has the technology to define the Lagrangian intersection Floer homology in the immersed case become available [AJ10], but it is perhaps surprising that these ideas (such as the bounding cochain) are used in an essential way in the most recent paper by Fukaya [Fuk15]. We shall spend the remaining sections giving a sketch of the proof as it has appeared so far.

Several important technicalities regarding the Atiyah-Floer conjecture were noted in the description of the Atiyah-Bott moduli space, namely that the moduli space of flat connections failed to be a smooth symplectic manifold whenever the vector bundle in question was trivial, or the genus of the surface was small. The latter difficulty is not so problematic in terms of low-dimensional topology: the case $g = 0$ corresponds to taking connected sums of homology 3-spheres and has already been treated in a paper of Fukaya [Fuk96]; the case $g = 1$ is related to knot theory and was treated by Floer himself (see the nice exposition by Donaldson and Braam in [HTWZ95]). However, the restriction that the relevant vector bundles be non-trivial is quite severe and forces Fukaya to work in a context where all 3-manifolds have disconnected boundary.

6.4 RELATIVE FLOER HOMOLOGY

We return now to the situation described in section 2 above with a Riemann surface Σ and a principal SO(3) bundle $P \to \Sigma$ such that the associated moduli space $\mathcal{M}_\Sigma$ is smooth. Given a 3-manifold Y bounding Σ and an immersed Lagrangian $\tilde{L} \to \mathcal{M}_\Sigma$, we wish to associate a *relative Floer homology group* $\mathrm{HF}(Y, L)$. As with the case of Lagrangian intersection Floer homology in the previous chapter, we will have to introduce an A_∞ deformation in order to make this homology group exist. To start, on the level of chain complexes, we define

$$\mathrm{CF}(Y, L) = \Omega^*(L_Y \otimes_{\mathcal{M}_\Sigma} \tilde{L}; \Lambda_0^{\mathbb{C}})$$

Now, recall that we can associate an A_∞ algebra to any immersed Lagrangian submanifold $\tilde{L} \to \mathcal{M}_\Sigma$ of the moduli space of flat connections using the formalism described in Chapter 5. When the self-intersections are transversal, it is given as a module by

$$\mathrm{CF}(L) = \Omega^*(\tilde{L} \times_{\mathcal{M}_\Sigma} \tilde{L}; \Lambda_0^{\mathbb{C}}) = \Omega^*(\tilde{L}; \Lambda_0^{\mathbb{C}}) \oplus \bigoplus_{(p,q) \in L \cap L, p \neq q} \Lambda_0^{\mathbb{C}}[p, q]$$

with a series of operations $m_k : \mathrm{CF}(L)^{\otimes k} \to \mathrm{CF}(L)$ coming from Lagrangian intersection Floer homology. Now we wish to give $\mathrm{CF}(M, L)$ the structure of a **right A_∞ module** over the A_∞ algebra $\mathrm{CF}(L)$; that is, we wish to define a sequence of maps $n_k : \mathrm{CF}(M, L) \otimes \mathrm{CF}(M, L)^{\otimes k} \to \mathrm{CF}(M, L)$ of degree 1 that satisfy the relations

$$\begin{aligned} &\sum_{\ell=0}^{k} n_{k-\ell}(n_\ell(y, x_1, \ldots, x_\ell), x_{\ell+1}, \ldots, x_k) \\ &+ \sum_{0 \leq \ell \leq m \leq k} n_{k-m+l+1}(y, x_1, \ldots, m_{m-\ell}(x_\ell, \ldots, x_{m-1}), \ldots, x_k) = 0 \end{aligned}$$

Note that this is the same as the condition stated in Chapter 5 for $\mathrm{CF}(M, L)$ to be a bimodule over the pair consisting of $\mathrm{CF}(L)$ and the trivial A_∞ algebra $\Lambda_0^{\mathbb{C}}$. As before, if we are given a bounding cochain b for L, then we may define a differential $d^b : \mathrm{CF}(M, L) \to \mathrm{CF}(M, L)$ by setting

$$d^b(y) = \sum_{k=0}^{\infty} n_k(y, b, \ldots, b)$$

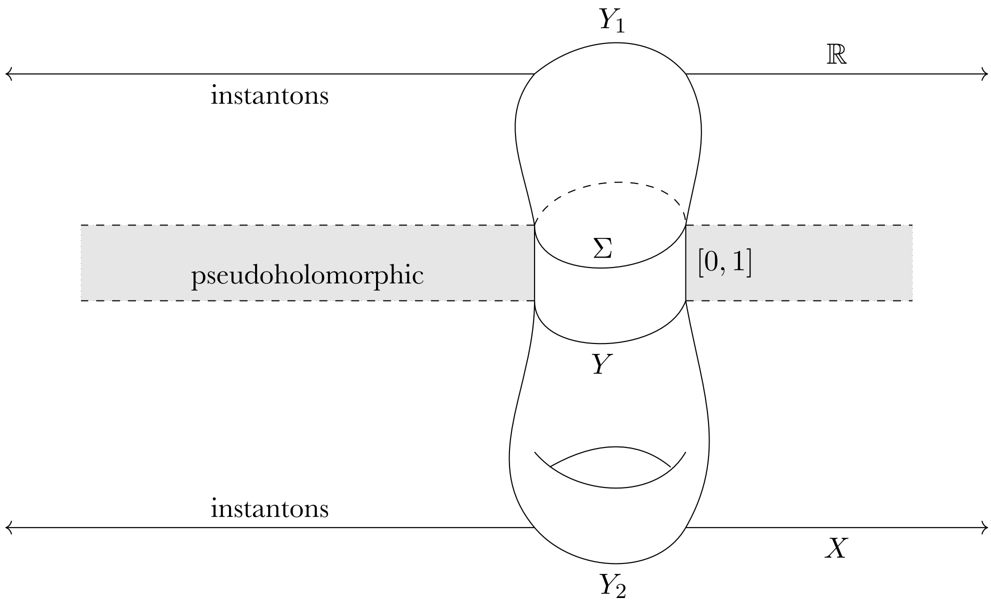

Figure 6.2: Illustration of Fukaya's proof of the Atiyah-Floer conjecture.

The fact that b is a bounding cochain will imply that $d^b \circ d^n = 0$ [FOOO09, Lemma 3.7.14]. Hence if all of the above constructions work we have a homology group $\mathrm{HF}(M, L)$, called the **relative Floer homology**. Now we turn to the construction of the operations n_k.

We begin by defining a hybrid equation on $[0,1] \times \mathbb{R} \times \Sigma$:

$$\begin{cases} \frac{\partial A}{\partial t} - d_A \Psi - *_\Sigma \left(\frac{\partial A}{\partial s} - \mathrm{d}_A \Phi \right) = 0 \\ \chi(s,t)^2 \left(\frac{\partial \Psi}{\partial s} - \frac{\partial \Phi}{\partial t} + [\Phi, \Psi] \right) + *_\Sigma F_A = 0 \end{cases} \tag{6.1}$$

where we write a connection on $[0,1] \times \mathbb{R} \times \Sigma$ as $A + \Phi \mathrm{d}s + \Psi \mathrm{d}t$ and take $\chi(s,t)$ to be a smooth cutoff function on $[0,1] \times \mathbb{R}$ taking values in $[0,1]$, that is zero on $(1/2, 1] \times \mathbb{R}$ and strictly positive on $[0, 1/2) \times \mathbb{R}$. This is the ASD equation that would arise from multiplying the Σ component of the product metric on $[0,1] \times \mathbb{R} \times \Sigma$ by the 'squashing function' $\chi(s,t)$. Instead, one can easily observe that on $(1/2, 1]$, where $\chi \equiv 0$, this equation is, as expected, exactly the pseudoholomorphic curve equation, when we regard the (necessarily flat) connection on $[0,1] \times \mathbb{R} \times \Sigma$ as a map $[0,1] \times \mathbb{R} \to \mathcal{M}_\Sigma$. This is because when $\chi \equiv 0$, the second equation implies that A is flat, and the first equation is then exactly the pseudoholomorphic curve equation $\partial_t \hat{A} - *_\Sigma \partial_s \hat{A} = 0$ when we regard A as a map $\hat{A} : [0,1] \times \mathbb{R} \to \mathcal{M}_\Sigma$. Thus the equation above represents a smooth interpolation between the ASD and pseudoholomorphic curve equations (Figure 6.2 illustrates the idea). As such, its behaviour is easy to analyse away from the line between the two equations: [Fuk98] shows compactness and removable singularity theorems essentially by interpolating between the results of Uhlenbeck and Gromov. A very similar analysis was carried out earlier by [Lip14], to whom Fukaya acknowledges a debt.

Returning to our construction of the relative Floer homology, we take our 3-manifold Y and attach a neck that is isometric to $\Sigma \times [0,1]$ in the product metric. By extending the cutoff function to be identically 1 away from this neck we may hence consider solutions to equation 6.1 over all of Y. By considering moduli spaces of solutions to this equation over Y with boundary condition L, we may construct the operations n_k as in [Fuk15, Definition 2.19] so as to satisfy the module relations above. The description of such moduli spaces is quite involved. Fukaya also acknowledges that he currently has no way of constructing these moduli spaces as they cannot carry 'Kuranishi structures'.

Now suppose we take $\tilde{L} = L_Y$. Then both of the chain complexes for $\mathrm{CF}(L)$ and $\mathrm{CF}(M, L)$ are simply $C(L_Y \times_{\mathcal{M}_\Sigma} L_Y; \Lambda_0^{\mathbb{C}})$ and so we can construct a **unit** for the A_∞ algebra (see [FOOO09, §3.7]) by taking 1_Y to be the fundamental class $[L_Y]$ in $C(L_Y; \Lambda_0^{\mathbb{C}})$ [Fuk15, Definition 3.7]. This proof is entirely elementary. The unit then yields a bounding cochain for the immersed Lagrangian submanifold L_Y, hence allowing us to define the relative Floer homology group $\mathrm{HF}(M, L)$ in the manner described above.

Appendix A

APPENDIX A: GEOMETRY

A.1 BASIC THEORY

We provide here a fairly detailed summary of basic differential geometry: bundles, connections and covariant derivatives. For proofs, readers are referred to the classic [KN69], or the more modern [Tau11] for numerous examples. Throughout, X is a smooth manifold without boundary.

BUNDLES

VECTOR BUNDLES

We begin by recalling the basic notion of a *vector bundle* and some of the relevant theory.

DEFINITION A.1. *A rank n real* ***vector bundle*** $\pi : E \to X$ *over X is a smooth manifold E along with a smooth projection map $\pi : E \to X$ such that the fibres have the structure of vector spaces, satisfying a local triviality condition; there exists an open cover $\{U_\alpha\}_{\alpha \in I}$ of X and smooth maps $\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$ commuting with the projection maps and linear on fibres.*

Given a vector $X \in \pi^{-1}(x)$, we write $\phi_\alpha(X) = (x, X_\alpha)$, where the vector $X_\alpha \in \mathbb{R}^n$ is interpreted as being the coordinates of X with respect to the trivialisation ϕ_α. Over an overlap $U_{\alpha\beta} = U_\alpha \cap U_\beta$, we have a map $\phi_\alpha \circ \phi_\beta^{-1} : U_{\alpha\beta} \times \mathbb{R}^n \to U_{\alpha\beta} \times \mathbb{R}^n$ that must take the form $(x, v) \mapsto (x, g_{\alpha\beta}(x)v)$ for some **transition function** $g_{\alpha\beta}(x) : U_{\alpha\beta} \to \mathrm{GL}_n(\mathbb{R})$. With respect to the vector coordinates, the transitions take the form $X_\alpha = g_{\alpha\beta} X_\beta$. Observe that we must also have $g_{\beta\alpha} = g_{\alpha\beta}^{-1}$, as well as the **cocycle conditon** $g_{\alpha\beta} g_{\beta\gamma} = g_{\alpha\gamma}$; in fact, given a collection of matrix-valued functions satisfying these two conditions, we may reconstruct the vector bundle by defining

$$E = \coprod_\alpha U_\alpha \times \mathbb{R}^n \Big/ (x, v)_\alpha \sim (x, g_{\alpha\beta}(x)v)_\beta$$

along with the obvious projection map. In the important special case where $E = TX$, the tangent bundle of X, and we trivialise TX over coordinate charts $\{U_\alpha\}_{\alpha \in I}$, then the transition functions will be given by the Jacobian matrices of the coordinate changes.

A **section** $s : X \to E$ of a vector bundle must satisfy $\pi \circ s = 1$; we denote the space of sections of E by $\Gamma(X, E)$. From the reconstruction theorem above, we see that specifying a section of E is the same as specifying a collection of functions $v_\alpha : U_\alpha \to \mathbb{R}^n$ satisfying $v_\alpha(x) = g_{\alpha\beta}(x) v_\beta(x)$ for all $x \in U_{\alpha\beta}$. Then a local trivialisation of E over U_α may be equivalently be described as giving a *local frame field*, that is, a collection of n sections s^i_α that form a basis in each fibre of E over U_α; given such a frame field, local trivialisations are given by mapping vectors $X \in \pi^{-1}(x)$ to their coordinates in $\mathbb{R}^n$ with respect to the fibre basis $s^i_\alpha(x)$. Conversely, given a local trivialisation $\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$, we have a local frame field given by $s^i_\alpha(x) = \phi_\alpha^{-1}(x, e^i)$. Given two local frame fields s^i_α, s^i_β associated to ϕ_α, ϕ_β then they are related by $s^i_\beta(x) = (g_{\alpha\beta})^i_j(x) s^j_\alpha(x)$, or, in matrix notation, $s_\beta(x) = g_{\alpha\beta}(x) s_\alpha(x)$; observe that this is the *reverse* of the transition for the coordinates, a fact which will be siginificant in the following discussion.

We want to make a quick note on *vector bundle morphisms* here. A vector bundle morphism $E \to F$ between two vector bundles $E, F \to X$ over the same base space is defined to be a smooth map $E \to F$ commuting with projections to X that is linear on fibres. The set of vector bundle morphisms between two vector bundles may itself be regarded as a vector bundle and we have an important identification of the tensor product bundle $E^* \otimes F$ with this bundle of morphisms $\mathrm{Hom}(E, F)$. Moreover, we may identify vector bundle morphisms (global sections of $\mathrm{Hom}(E, F)$) with maps of sections $\Gamma(X, E) \to \Gamma(X, F)$ that are *tensorial* in the sense that they are linear and only depend on the value at the point. In particular, by taking $F = X \times \mathbb{R}$, we have that sections of E^* are the same as *tensorial* elements of the dual of the space of sections $\Gamma(X, E)^*$. When we want to talk about differential forms with values in a vector bundle, these two perspectives are both useful. We denote by $\Omega^k(X, E)$ the space $\Gamma(X, E \otimes \bigwedge^k T^*X)$; we can regard a section either as a vector of k-forms, or as a vector-valued k-form, that is, a k-form with vector coefficients. Given a vector bundle, we may also construct the **endomorphism bundle** $\mathrm{End}(E)$, whose sections we can think of, under the above correspondence, as being matrix-valued. Applying the same construction with forms, we can consider sections of $\Omega^k(X, \mathrm{End}(E))$ as being either matrices of k-forms or matrix-valued k-forms. Changing between these two perspectives will be useful later in our discussion of connections. See [Tau11, Lemma 11.1, pp.126-127] for proofs and more on this perspective.

PRINCIPAL BUNDLES

There are a number of different possible definitions of a *prinicipal bundle*; the one we give here can be shown to be equivalent to that given in [KN69].

DEFINITION A.2. *A* ***prinicipal*** G***-bundle*** $\pi : P \to X$ *over* X *is a manifold* P *with a* ***right action*** *of* G *on* P*, free and transitive along the fibres of a smooth* G*-invariant map* $\pi : P \to X$ *that has* G*-equivariant local trivialisations; that is, there exists an open cover* $\{U_\alpha\}_{\alpha \in I}$ *of* X *and smooth maps* $\psi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times G$ *commuting with projections to* U_α *and satisfying* $\psi_\alpha(p \cdot g) = \psi_\alpha(p) \cdot g$.

Morphisms of principal G-bundles are then defined in the obvious way; maps $\phi : P \to P'$ commuting with projections and the G-action.

HIC SUNT DRACONES A.1. *Beware that some authors instead use a* ***left action*** *of* G *on* P *by using instead* $p \mapsto p \cdot g^{-1}$*. In this case, all appearances of* g *below must be replaced by* g^{-1}*. Note that it is not possible for* G *to act by left multiplication on* P*.*

For example, we define the **frame bundle** $F(E)$ of a rank n real vector bundle $E \to X$ by setting $P = \coprod_{x \in X} \mathrm{GL}(E_x)$ with the obvious projection and with a right group action of $GL_n(\mathbb{R})$ on a frame $\{v_j\}_{j=1}^n$ of E_x via $\{v_j\} \cdot g = \{w_i\}$ for $w_i = (g^{-1})_i^j v_j$. This is consistent with the action of transitions on frames discussed above. Local trivialisations for P may be constructed from those for E. If a bundle metric on E is specified, then we may define the bundle of orthonormal frames via $P = \coprod_{x \in X} \mathrm{O}(E_x)$; this is a principal $\mathrm{O}(n)$ bundle. If E is oriented then in a similar fashion we have the bundle of oriented orthonormal frames, a $\mathrm{SO}(n)$-bundle. Similar remarks apply in the complex case. The frame bundle of a tangent bundle TX is often referred to as the frame bundle of X itself.

We also have a description of principal bundles in terms of transition functions; over $U_{\alpha\beta}$, we have $\psi_\alpha \circ \psi_\beta^{-1}(x, g) = (x, h_{\alpha\beta}(x)g)$ for some group-valued function $h_{\alpha\beta} : U_{\alpha\beta} \to G$ given by $h_{\alpha\beta}(x) = \pi_2 \circ \psi_\alpha \circ \psi_\beta^{-1}(x, e)$, by G-equivariance of the transition functions. As before, a principal bundle may be recovered from a collection of transition functions satisfying the the cocycle condition. Hence we see that a section of P (defined in the same way as above) must consist of a collection of functions $s_\alpha : U_\alpha \to G$ transforming as $s_\beta = h_{\alpha\beta} s_\beta$, that is, like a *frame*. Continuing the example above, it is easy to see that the transition functions $h_{\alpha\beta}$ for the frame bundle $F(E)$ of a vector bundle E must be exactly the transition functions $g_{\alpha\beta}$ for E. However, since they will in general differ, we shall continue to use different notation to avoid confusion.

ASSOCIATED BUNDLES

Given a representation $\rho : G \to \mathrm{GL}(V)$ of G on a finite-dimensional vector space V, we may construct a vector bundle with fibre V, the **associated bundle** $P \times_\rho V$ (also denoted P_V when the representation in question is clear), as the quotient $P \times_\rho V = P \times V/(p, v) \sim (p \cdot g, \rho(g^{-1})v)$, with projection given by $\pi([p, v]) = \pi(p)$. We interpret $[p, v]$ as an equivalence class of pairs consisting of an internal frame p and a vector v of coordinates with respect to that frame. If we fix some frame $p_0 \in \pi^{-1}(x)$, then because the G action on P is transitive on fibres, we know that any (p, v) is equivalent to $(p_0, \rho(g^{-1})v)$ for some appropriate $g \in G$. We can then define a vector space structure on the fibres of π by $[p_0, v_1] + [p_0, v_2] = [p_0, v_1 + v_2]$; it is easy to check that this is well-defined using the freeness of the G-action on fibres of P. We then have local trivialisations for $P \times_\rho V$ given by $\psi_\alpha[p, v] = (\pi(p), \rho(\pi_2 \circ \psi_\alpha(p))v)$, for $\pi_2 \circ \psi_\alpha : U_\alpha \to G$ the local trivialisations for P. It is not difficult to check that these maps are well-defined and linear on fibres. The transition functions for this vector bundle may then be determined from those of P; they are precisely $\rho(h_{\alpha\beta})$, acting by *left* multiplication. To see this, suppose we have local sections $s_\alpha : U_\alpha \to P$ and a collection of coefficient vectors

$v_\alpha \in V$ for all $\alpha \in I$; this then defines a section of P_V over each U_α via $s(x) = [s_\alpha(x), v_\alpha]$. Then over $U_{\alpha\beta}$, we must have $[s_\alpha(x), v_\alpha] = [s_\beta(x), v_\beta]$ with $v_\alpha = \rho(h_{\alpha\beta})v_\beta$ because

$$(x, v_\beta) = \psi_\beta([s_\alpha, v_\alpha]) = (x, \rho(\pi_2 \circ \psi_\beta(s_\alpha))v_\alpha) = (x, \rho(\pi_2 \circ \phi_\beta \circ \phi_\alpha^{-1}(x, e))v_\alpha) = (x, \rho(h_{\alpha\beta})v_\alpha)$$

From the frame bundle construction above, we can see that every vector bundle arises as the associated bundle of some principal bundle by taking the vector bundle associated to the frame bundle by the standard representation of $\mathrm{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$. With this perspective, we see that the theory of vector bundles reduces to the representation theory of the group G. Moreover, any morphism of G-representations will correspond to a morphism of vector bundles over X. Hence any 'categorical' operation we can perform on group representations we may also perform on vector bundles. For instance, the direct sum of two representations ρ_1, ρ_2 is defined on $V_1 \oplus V_2$ by $\rho(v \oplus w) = \rho_1(v) \oplus \rho_2(w)$; this gives rise to the vector bundle $P_{V_1} \oplus P_{V_2}$. Similarly, the tensor product of representations gives rise to the tensor product of vector bundles. The *dual* of a representation $\rho : G \to \mathrm{GL}(V)$ is defined on a linear functional $w \in V^*$ via $(\rho^*(g)w)(v) = w(\rho(g^{-1})v)$; the inverse is necessary to preserve the homomorphism property. Using the discussion above, we may hence immediately determine the transition functions for the sum, tensor product and dual vector bundles; in particular, we see that the transition functions for the dual bundle are the inverses of those for the original vector bundle.

To any Lie group G, there is a naturally associated representation, namely the adjoint representation $\mathrm{ad} : G \to \mathrm{GL}(\mathfrak{g})$ of the group on its Lie algebra, yielding the **adjoint bundle**, denoted $\mathrm{ad}\ P$. More generally, given any other representation $\rho : G \to \mathrm{GL}(V)$, we have an associated adjoint representation $\mathrm{ad}\ (\rho)$ on $\mathrm{End}(V)$ given by composing with the adjoint representation of $\mathrm{GL}(V)$ (i.e., conjugation). If we have an associated bundle $E = P \times_\rho V$, then $P \times_{\mathrm{ad}\ (\rho)} \mathrm{End}(V)$ is exactly the endomorphism bundle $\mathrm{End}(E)$. Taking the differential of the representation ρ at the identity yields a representation $\rho_* : \mathfrak{g} \to \mathrm{End}(V)$ of the Lie algebra of G, and hence a morphism of representations of G between ad and $\mathrm{ad}\ (\rho)$. By the remarks above, we hence have a morphism of vector bundles $\mathrm{ad}\ P \to \mathrm{End}(E)$. In the special case where $P = F(E)$ is the frame bundle of E and ρ is the defining representation of $\mathrm{GL}_n(\mathbb{R})$, we will have $\mathrm{ad} = \mathrm{ad}\ (\rho)$ and hence an isomorphism of bundles $\mathrm{End}(E) \cong \mathrm{ad}\ P$. This identification will be significant in our discussion of connections on principal bundles below. More generally, if G is a Lie subgroup of $\mathrm{GL}(V)$ (such as $\mathrm{SO}(n) \subseteq \mathrm{GL}_n(\mathbb{R})$) and F_G is the subbundle of G-frames for E (oriented orthonormal frames in the case of $\mathrm{SO}(n)$), then $\mathrm{ad}\ F_G$ is a subbundle of $\mathrm{End}(E)$ consisting of those endomorphisms that pointwise live in the Lie algebra (in the case of $\mathrm{SO}(n)$, are pointwise skew-symmetric).

The adjoint and endomorphism bundles have the extra structure of being bundles of Lie algebras, which allows us to define some further operations. Firstly suppose that α, β are in $\Omega^1(\mathrm{ad}\ P)$ and define

$$[\alpha, \beta](X, Y) = [\alpha(X), \beta(Y)] - [\alpha(Y), \beta(X)]$$

In the case where the bracket in $\mathfrak{g}$ is given by the commutator in a matrix algebra, then define

$$(\alpha \wedge \beta)(X, Y) = \alpha(X)\beta(Y) - \alpha(Y)\beta(X)$$

Note that both of the above expressions are indeed tensorial and antisymmetric in X and Y and hence must define 2-forms. We can see immediately from the above that $[\alpha, \beta] = \alpha \wedge \beta + \beta \wedge \alpha$ and that $\alpha \wedge \alpha = \frac{1}{2}[\alpha, \alpha]$. More generally, for $\alpha, \beta \in \Omega^k(X, \mathrm{ad}\ P)$, choose a basis for $\bigwedge^k T^*X$ and write $\alpha = \sum_i \alpha_i \otimes x_i$ and $\beta = \sum_j \beta_j \otimes y_j$ for k-forms α_i, β_j and sections $x_i, y_j \in \Omega^0(X, \mathrm{ad}\ P)$. Then we define

$$\alpha \wedge \beta = \sum_{i,j} (\alpha_i \wedge \beta_j) \otimes (x_i y_j)$$

and

$$[\alpha, \beta] = \sum_{i,j} (\alpha_i \wedge \beta_j) \otimes [x_i, y_j]$$

It is clear that the above operations are well-defined and agree with the previous definitions in the special case of 1-forms. Moreover one can easily verify that we have $[\alpha, \beta] = \alpha \wedge \beta - (-1)^{|\alpha||\beta|}\beta \wedge \alpha$ and that $[\alpha, \beta] = (-1)^{|\alpha||\beta|}[\beta, \alpha]$. No similar commutation formula holds for $\wedge$ unless the matrices happen to commute pointwise. Since the Lie algebra has a trace operation, we can also define a trace on $\Omega^k(\mathrm{ad}\ P)$ in the obvious way.

We can perform the above construction more generally if G acts smoothly on a manifold F; we form an associated fibre bundle with fibre F using the same procedure. One important example of the above is given by the adjoint action $\mathrm{Ad} : G \to \mathrm{Aut}(G)$ of G on itself by conjugation, giving the **gauge group**, denoted $\mathscr{G}_P$, or simply as $\Gamma(X, \mathrm{Ad}\ P)$. We shall have more to say about this group in later sections. We shall also observe here that if $G = \mathrm{GL}(V)$ and

E is the associated vector bundle, then the bundle Ad P can be regarded as the subbundle of $\mathrm{End}(E)$ consisting of pointwise intertible automorphisms, and gauge transformations as sections of the endomorphism bundle. It will also be of importance later to recall the notion of **reduction of the structure group**. Suppose $H \subseteq G$ is a Lie subgroup, P_G a principal G-bundle and P_H a principal H bundle. Since H has a left action on G by diffeomorphisms given by left multiplication, we may form the associated fibre bundle $P_H \times_{L_h} G$ which is a principal G-bundle. If the principal G-bundle P_G happens to be isomorphic to $P_H \times_{L_h} G$, we say that the structure group of P_G can be **reduced** from G to H. Equivalently, the structure group of P_G reduces to H exactly when all of the transition functions $h_{\alpha\beta}$ may be chosen to lie in the subgroup H. Note also that ad P_H is a subbundle of ad P_G and hence gauge transformations for P_H can be regarded as gauge transformations for P_G. For example, if the structure group $G = \mathrm{GL}_n(\mathbb{R})$, then it may always be reduced to $\mathrm{O}(n)$, but not in general to $\mathrm{SO}(n)$. If it can, we say that P_G is **orientable**. Similar remarks apply in the complex case.

There is a simple description forms with values in associated bundles. Firstly, given a function $s : P \to V$, we say it is **pseudotensorial** with respect to the representation ρ of G if $R_g^* s = \rho g^{-1} s$ where R_g is the right multiplication action on P of $g \in G$. We can then define a section $\tilde{s}$ of $P \times_\rho V$ over X by setting $\tilde{s}(x) = [p, s(p)]$ for *any* $p \in \pi^{-1}(x)$; the pseudotensorality property ensures that this is well-defined. In fact, this correspondence can be seen to be bijective. More generally, we have a correspondence between sections $s \in \Omega^k(X, P \times_\rho V)$ and fibrewise linear maps $\pi^* \wedge^k TM \to V$ that satisfy the same pseudotensorial property with respect to the obvious G-action on the pullback bundle $\pi^* \wedge^k TM$.

CONNECTIONS

COVARIANT DERIVATIVES

We begin with the most concrete manifestation of a *connection*:

DEFINITION A.3. *A* ***covariant derivative*** *on a vector bundle $\pi : E \to X$ is a linear operator $\nabla : \Omega^0(X, E) \to \Omega^1(X, E)$ satisfying the Leibniz rule $\nabla(fs) = df \otimes s + f\nabla s$ for any $f \in C^\infty(X)$. We write $\nabla_X s$ for the pairing $\langle \nabla s, X\rangle \in \Omega^0(X, E)$ when X is a tangent vector field.*

Firstly, if E is the trivial bundle, we have a canonical covariant derivative given by the usual de Rham differential d acting on components with respect to a given trivialisation. We may also observe that, given two connections ∇_1, ∇_2, then their difference has $(\nabla_1 - \nabla_2)(fs) = f(\nabla_1 - \nabla_2)s$ and hence defines a tensorial map on sections $\Gamma(X, E) \to \Gamma(X, T^*X \otimes E)$, or equivalently, a (globally-defined) element of $\Omega^1(\mathrm{End}(E))$. Thus we see that the set of connections $\mathscr{A}$ is an affine space modelled on $\Omega^1(\mathrm{End}(E))$. For a local trivialisation of E over $U_\alpha \subseteq X$, any connection ∇ can hence be written as $d + A^\alpha$ for some $A^\alpha \in \Omega^1(U_\alpha, E)$ acting on E-valued forms by multiplication. These are called the *local connection 1-forms*. Over an overlap $U_{\alpha\beta}$, these connection 1-forms are related by

$$A^\beta = g_{\alpha\beta}^{-1} A^\alpha g_{\alpha\beta} + g_{\alpha\beta}^{-1} dg_{\alpha\beta}$$

for $g_{\alpha\beta}$ the transition functions of E. A collection $\{A^\alpha\}_{\alpha \in I}$ of sections satisfying the above transformation rules will define a unique covariant derivative, denoted d_A. Hence a simple gluing argument shows that every vector bundle admits a connection, and hence infinitely many connections. We can extend the covariant derivative to a *exterior covariant derivative* $d_A : \Omega^k(X, E) \to \Omega^{k+1}(X, E)$ by defining $d_A(s\omega) = (d_A s) \wedge \omega + s d\omega$ and extending by linearity. One can then check that if $s \in \Omega^0(X, E)$, then $d_A(d_A(fs)) = f d_A(d_A s)$ and hence $d_A \circ d_A(\alpha)$ defines tensorial map of sections $\Omega^0(X, E) \to \Omega^2(X, E)$, or equivalently, a (globally-defined) element F_A of $\Omega^2(X, \mathrm{End}(E))$ called the **curvature** of d_A. In a local trivialisation, we will have

$$F_A^\alpha = dA^\alpha + A^\alpha \wedge A^\alpha = dA^\alpha + \frac{1}{2}[A^\alpha, A^\alpha]$$

In particular, the curvature 2-form must be related between trivialisations by $F^\beta = g_{\alpha\beta}^{-1} F^\alpha g_{\alpha\beta}$. It is then easy to derive the *Bianchi identity*, $d_A F_A = 0$, by explicit computation. Alternatively, if in a local trivialisation the covariant derivative on $\Omega^0(U_\alpha, E)$ is given by $d + A^\alpha$, then the exterior covariant derivative then acts on forms in $\Omega^k(U_\alpha, E)$ via $d_A(\omega) = d\omega + [A^\alpha, \omega]$. The Bianchi identity and the above formula for the curvature now become apparent. In this special case, d_A also has derivation properties with respect to the product operations on $\Omega^*(X, \mathrm{ad}\, P)$:

$$\begin{aligned} d_A(\alpha \wedge \beta) &= d_A\alpha \wedge \beta + (-1)^{|\alpha|}\alpha \wedge d_A\beta \\ d\,\mathrm{Tr}(\alpha) &= \mathrm{Tr}(d_A\alpha) \end{aligned}$$

HIC SUNT DRACONES A.2. *There seem to be several different conventions for the above in the gauge theory literature, usually differing by signs or factors of* 2. *In particular, it is important to note that some authors use the notation* $[\alpha \wedge \beta]$ *for* $[\alpha, \beta]$, *a practice which only seems to be a source of confusion. Hence some authors may state the above results differently depending on their different conventions and notation.*

There are several special types of connections that are important in differential geometry. If $F_A \equiv 0$ we say that A is **flat**. If E has a metric, then if d_A satisfies $d(s,t) = (d_A s, t) + (s, d_A t)$, then d_A is **metric compatible**. Given a connection on $E = T^*X$, we can alternatively define the exterior covariant derivative d_A by antisymmetrising the map $\nabla_A : \Gamma(X, E) \to \Gamma(X, E \otimes \bigotimes^k T^*X)$. The **torsion** T_A is then the difference $d - d_A$ between the exterior covariant derivative and the standard exterior derivative; it can be easily shown that $T_A \in \Omega^1(X, \wedge^2 T^*X)$. We say that d_A is **torsion-free** if $T_A \equiv 0$ and in this case, $d_A = d$ on *all* differential forms. The *Fundamental Theorem of Riemannian Geometry* then states that there exists a unique torsion-free metric-compatible connection on T^*X, called the **Levi-Civita Connection**. Henceforth, the notation ∇ refers specifically to this connection unless stated otherwise. The connection 1-forms in this case are referred to as the **Christoffel Symbols**. Given connections d_{A_1}, d_{A_2} on bundles E_1, E_2, they induce connections d_A on $E_1 \oplus E_2$, $E_1 \otimes E_2$ and E_1^* via $d_A(s_1 \oplus s_2) = d_{A_1} s_1 \oplus d_{A_2} s_2$ and $d_A(s_1 \otimes s_2) = d_{A_1} s_1 \otimes s_2 + s_1 \otimes d_{A_2} s_2$ and $d\langle s, t\rangle = \langle d_A s, t\rangle + \langle s, d_{A_1} t\rangle$ respectively. In particular, the Levi-Civita connection induces connections on all tensor bundles on X, which, by a standard abuse of notation, we will continue to denote by ∇. We shall see in our discussion of connections below how the above formulas may be derived in a coherent manner.

It is often useful to have a description of the above in local coordinates. Choosing a local coordinate system over U_α for the manifold X, we have $\langle d_A s, \partial_i\rangle = \partial_i s + A_i^\alpha$ for $A^\alpha = A_i^\alpha dx^i$, where we now regard A^α instead as a 1-form with matrix coefficients A_i^α. Writing $F^\alpha = F_{ij}^\alpha dx^i \wedge dx^j$ we then have

$$F_{ij}^\alpha = \frac{\partial A_j^\alpha}{\partial x^i} - \frac{\partial A_i^\alpha}{\partial x^j} + [A_i^\alpha, A_j^\alpha]$$

which are the *coefficients* with respect to these coordinates.

CONNECTIONS ON PRINCIPAL BUNDLES

We briefly describe the general notion of a connection on a principal bundle. On a principal bundle, there is a natural way to choose a *vertical subspace* of the tangent space at each point, but no canonical choice of a complementary *horizontal subspace*; a connection on a principal bundle is a way of making such a choice:

DEFINITION A.4. *Suppose* $\pi : P \to X$ *is a principal* G *bundle over* X *and* $p \in P$ *is a point. Then the* **vertical subspace** *of* T_pP *is defined to be* $\ker D_p\pi : T_pP \to T_{\pi(p)}X$. *A* **connection** *on* P *is a smooth (that is, has a local basis of smooth sections around every point) choice of* **horizontal subspaces** $H_p \subseteq T_pP$ *such that* $T_pP = H_p \oplus V_p$ *and* $H_{pg} = (R_g)_* H_p$ *(a property called* G*-invariance).*

We may represent the above by a G-equivariant exact sequence of vector bundles $0 \to \ker(D\pi) \to TP \xrightarrow{D\pi} \pi^*TM \to 0$ where $\ker(D\pi) = VP$ is the **vertical tangent space**. A connection H_p is then a (smooth, G-invariant) choice of splitting for this exact sequence. In particular, we see that a connection gives us a coherent way to lift tangent vectors $V \in T_xX$ to $\tilde{V} \in T_pP$ for each $p \in \pi^{-1}(x)$. An alternative way of specifying a splitting is by giving a map $\omega : TP \to \ker(D\pi)$ that is appropriately G-equivariant. Because the fibres of the map π may be identified with the structure group G (non-canonically), we have a (canonical) identification of V_p with $\mathfrak{g}$ for all $p \in P$. It can be shown that the vertical bundle VP is in fact isomorphic to the trivial bundle $P \times \mathfrak{g}$ of Lie algebras. Hence ω may be alternatively be described as a $\mathfrak{g}$-valued 1-form, that fixes vectors in VP under this identification. The G-invariance condition now becomes the adjoint pseudotensorality condition $R_g^*\omega = \mathrm{ad}_{g^{-1}}\omega$. Such a 1-form is called a **connection 1-form** and these are in bijective correspondence with connections on P; one often refers to ω simply as the *connection* also.

We now describe how a connection on P can be used to produce covariant derivatives on an associated bundle $P \times_\rho V$; there are many equivalent ways of doing this, of which we describe only one. Given a section $s \in \Omega^0(X, E)$, we may regard it, as noted above, as a suitably pseudotensorial function $s : P \to V$. Taking the differential d yields a 1-form on P, $ds : TP \to V$, which we may then restrict to π^*TM via the identification coming from the choice of horizontal subspace. This will then satisfy the appropriate pseudotensorality properties by the invariance condition on the horizontal subspaces H_p and hence yield a 1-form ∇s on X, as above. It is easy to check that this procedure will satisfy the properties required of a covariant derivative. To obtain a more explicit description of this covariant derivative, suppose we have two connection 1-forms ω_1 and ω_2; then their difference must vanish identically on VP and hence defines a map $\pi^*TM \to \mathfrak{g}$ that will, by definition, satisfy the pseudotensorality condition required to yield an element of $\Omega^1(X, \mathrm{ad}\, P)$. One can then show that $\nabla_1 - \nabla_2 = \rho_*(\omega_1 - \omega_2)$, where ρ_* is the bundle morphism introduced earlier.

In particular we noted that when P is the frame bundle $F(E)$, then ρ_* is in fact a bundle isomorphism and hence we have a bijection between covariant derivatives on E and connections on P. Moreover, in this case if $\mathrm{O}(E)$ is instead the bundle of orthonormal frames for E we may perform the same construction where ρ is now the defining representation of $\mathrm{O}(n)$. The covariant derivatives that arise in this way are then exactly the **metric-compatible** covariant derivatives on E. Furthermore, given any representation related to ρ, such as via tensor products, direct sums or duals, then the connection on P corresponding to a covariant derivative on E can be used to induce covariant derivatives on all of these associated bundles. The remarks above may be used to find an expression for the local connection 1-forms arising from a connection ω. Suppose we have a local section s^α of P (that is, a local frame field), yielding a local trivialisation for both P and $P \times_\rho V$; the difference $d - \nabla$ in this trivialisation is then the local connection 1-form and can be shown to be exactly $A^\alpha = \rho_*(s^\alpha)^*\omega$.

For later use, it will be important to discuss **gauge transformations**. Automorphisms of the principal bundle P may be identified with elements of the gauge group $\mathscr{G}_P$ introduced earlier. Given a section $s \in \Gamma(X, \mathrm{Ad}\ P)$ we may view it instead as a map $s : P \to G$ with $R_g^* s = \mathrm{Ad}_{g^{-1}} s$ via the induced bundle construction. Then we can define an automorphism $\phi : P \to P$ by taking $p \mapsto p \cdot s(p)$. This is clearly a bijective correspondence and hence there is a natural group structure on $\mathscr{G}_P$ given by composition of automorphisms. Moreover, in the induced bundle perspective on the gauge group we can see that applying the exponential map pointwise to a section in $\Omega^0(X, \mathrm{ad}\ P)$ will yield a section of $\mathrm{Ad}P$ over X; hence $\Omega^0(X, \mathrm{ad}\ P)$ can be regarded as the 'Lie algebra' of the gauge group. We shall make this precise in the main text of the report. In the special case where $P = X \times G$ is the trivial bundle, automorphisms are simply G-valued functions $h : X \to G$ on X with the associated automorphism given by $(x, g) \mapsto h(x) \cdot g$.

Any automorphism of P also induces automorphisms on all of the associated bundles of P. In particular, the action on ad P is simply by pointwise conjugation. The gauge group also acts on connections on P; pulling back a connection 1-form via an automorphism $s : P \to P$ must necessarily yield another connection 1-form because of the G-invariance properties of automorphisms. Over a local trivialisation gauge transformations then act on local connection 1-forms for the associated covariant derivatives; but this is *exactly* the same way in which transition functions for the principal bundle P act on covariant derivatives and hence we must have

$$\phi^* A^\alpha = \rho(h)^{-1} A^\alpha \rho(h) + \rho(h)^{-1} d\rho(h)$$

for $h : X \to G$ the associated G-valued function. When $G = \mathrm{GL}(V)$ or a subgroup, and ρ is the defining representation, we usual drop the ρ in this notation and simply regard gauge transformations as matrix-valued sections of the endomorphism bundle of the assoiciated vector bundle. From this we may observe that the corresponding action on curvature is by conjugation; $F_{\phi^* A} = \rho(h)^{-1} F_A \rho(h)$, as we could have deduced from the above. In particular, flatness of a connection is preserved under gauge transformations.

HOLONOMY

One situation in which the perspective on covariant derivatives as connections on an appropriate principal bundle is helpful is in the study of *holonomy*. Fix a connection A on P. Given a loop $\gamma : [0,1] \to X$ with $\gamma(0) = \gamma(1) = x$ and a point $p \in \pi^{-1}(x)$, we may *lift* γ to a curve $\tilde{\gamma} : [0,1] \to P$ with $\tilde{\gamma}(0) = p$ by requiring $\tilde{\gamma}$ to have tangent vector $\tilde{\gamma}'(t)$ at each $t \in [0,1]$ given by the lift of $\gamma'(t)$ to $\tilde{\gamma}(t)$ using the specified connection A. By standard theorems for ordinary differential equations, we see that such a lift $\tilde{\gamma}$ exists and is unique given these choices. Furthermore, by applying the same uniqueness theorem to $\pi(\tilde{\gamma})$ we can see that $\pi(\tilde{\gamma}(t)) = \gamma(t)$ for all $t \in [0,1]$; in particular, $\tilde{\gamma}(1) \in \pi^{-1}(x)$ also, and hence must be equal to $p \cdot g$ for some uniquely-defined $g \in G$. We call this g the **holonomy** of the connection A around the loop γ and write $g = \mathrm{Hol}_{A,\gamma}(p)$. Then we have the following theorem

THEOREM A.1. *(Holonomy [Tau11, pp.158-159]) The map* $(\gamma, A, p) \to \mathrm{Hol}_{A,\gamma}(p)$ *has the following properties:*

1. *G-Equivariance: we have* $\mathrm{Hol}_{A,\gamma}(pg) = \mathrm{Ad}_{g^{-1}} \mathrm{Hol}_{A,\gamma}(p)$*;*
2. *Homomorphism: the holonomy* $\mathrm{Hol}_{A,\gamma}(p)$ *is independent of the parametrisation of* γ *and if* γ_1, γ_2 *are two loops, then the concatenation* $\gamma_1 * \gamma_2$ *has holonomy* $\mathrm{Hol}_{A,\gamma_1 * \gamma_2}(p) = \mathrm{Hol}_{A,\gamma_1}(p) \cdot \mathrm{Hol}_{A,\gamma_2}(p)$*;*
3. *Homotopy Invariance: if A is a flat connection, then* $\mathrm{Hol}_{A,\gamma}$ *only depends on the homotopy class of* γ*;*
4. *Gauge Invariance: if A, A' are related by an automorphism of P, then* $\mathrm{Hol}_{A,\gamma}(p) \equiv \mathrm{Hol}_{A',\gamma}(p)$ *up to conjugation;*
5. *Classification: conjugacy classes of representations* $\pi_1(X) \to G$ *are in bijection with isomorphism classes of pairs* (P, A) *consisting of a principal bundle P and a flat connection A on P.*

Proofs of the above can be found in [Tau11]; item (5) shall be particularly useful in our discussion of the Atiyah-Bott moduli space.

A.2 SYMPLECTIC GEOMETRY

We recall here the basic definitions in symplectic geometry; see [CdS01] for more.

A smooth manifold X is called **symplectic** if there is a closed 2-form $\omega \in \Omega^2(X)$ which is *non-degenerate* in the sense that the correspondence $X \mapsto \omega(X, \cdot)$ gives an isomorphism of T_pX with T_p^*X for all $p \in X$. Note that X is necessarily even-dimensional. If a **Hamiltonian function** is given, that is, a smooth function $H : X \to \mathbb{R}$, we may use this symplectic structure to define a vector field X_H by the relation $dH(Y) = \omega(X_H, Y)$. The flows $\Phi^t_{X_H} : X \to X$ of this vector field are called **Hamiltonian diffeomorphisms**. It is easy to see that such diffeomorphisms will always preserve the symplectic form. A submanifold L of X is called **Lagrangian** if $\dim L = \frac{1}{2} \dim X$ and the pullback of the symplectic form ω to L is trivial. The most important examples of symplectic manifolds are *cotangent bundles*: if M is any smooth manifold, then $X = T^*M$ has a **tautological 1-form** λ given by $s^*\lambda = s$ for every 1-form $s \in \Gamma(T^*M)$. The differential $\mathrm{d}\lambda$ then gives a canonical symplectic form on X. With this symplectic form, the zero section of T^*M will always be a Lagrangian submanifold.

HIC SUNT DRACONES A.3. *Some authors use the reverse convention, instead taking $dH(Y) = -\omega(X_H, Y)$. This convention has some advantages, but is the source of various inconsistencies in the literature.*

An **almost-complex structure** J on a manifold X is an endomorphism $J : TX \to TX$ of the tangent bundle such that $J^2 = -I$. If X is a complex manifold, it has a canonical almost-complex structure given by $iI : TX \to TX$ where i is the imaginary unit. Such almost-complex structures are called **integrable**. If X is a symplectic manifold with symplectic form ω, an almost-complex structure J on X is said to be **tamed** by ω if $\omega(X, JX) > 0$ for all $X \neq 0$ and **compatible** with ω if $\omega(\cdot, J\cdot)$ gives a Riemannian metric on X. We denote by $\mathcal{J}$ the space of all ω-compatible almost complex structures on X, and by $\mathcal{J}_t$ the space of ω-tamed almost complex structures. Both spaces are non-empty and contractible [CdS01]. Given a map $\phi : X \to Y$ between manifolds with almost-complex structures J_1, J_2 respectively, we may write the derivative as $D\phi = \partial\phi + \bar{\partial}\phi$, the sum of the **holomorphic** and **anti-holomorphic** differentials. The holomorphic differential $\partial\phi$ is the J_1, J_2 linear part of $D\phi$, in the sense that $\partial\phi \circ J_1 = J_2 \circ \partial\phi$. Similarly, the anti-holomorphic differential $\bar{\partial}\phi$ is the J_1, J_2 *anti-linear* part. In the case where J_1, J_2 are both integrable, the map ϕ is holomorphic if and only if the anti-holomorphic differential $\bar{\partial}\phi$ is zero.

Appendix B

APPENDIX B: ANALYSIS ON MANIFOLDS

We summarise here some basic facts concerning Sobolev spaces on manifolds. Throughout we take $1 < p < \infty$ for simplicity.

SOBOLEV SPACES ON COMPACT MANIFOLDS

Let M be a compact, oriented smooth n-manifold without boundary and with Riemannian metric g, and let $E \to M$ a vector bundle with bundle metric h and a metric-compatible connection ∇. Define the Sobolev space $W^{k,p}(M,E)$ to be the completion of $\Gamma(M,E)$ with respect to the Sobolev norm

$$||s||_{k,p} = \sum_{i=1}^{k} \int_M |\nabla^k s|_h^p \mathrm{d}\mu_g$$

where ∇^k is the k-iterated covariant derivative $\Gamma(X,E) \to \Gamma(X, E \otimes \bigotimes^k T^*X)$. The resulting Banach space does not depend on any of the choices made above, up to isomorphism; this can be shown by demonstrating that the identity map between the two spaces is bounded. For $1 < p < \infty$, these spaces are reflexive, and smooth functions form a dense subset. The familiar Sobolev theorems continue to apply; we use $\Sigma(k,p) = k - n/p$ to denote the Sobolev strength.

THEOREM B.1. *If $k_0 \geq k_1$ and $\Sigma(k_0,p_0) \geq \Sigma(k_1,p_1)$, then $W^{k_0,p_0}(M,E)$ embeds continuously into $W^{k_1,p_1}(M,E)$. If these inequalities are strict, then this embedding is compact.*

THEOREM B.2. *If $\Sigma(m,p) \geq k + \alpha$, then $W^{m,p}(M,E)$ embeds continuously into the Hölder space $C^{k,\alpha}(M,E)$. If this inequality is strict, then the embedding is compact.*

The above theorems, with proof, may be found in [Nic07, p.478-479] (beware of typos). In the case where M is not compact, the above results are significantly more subtle, as we shall see below. In particular, the dependence on the metric g becomes important.

Once we have set up the theory of Sobolev spaces, this allows us to prove the central estimates for elliptic partial differential operators on compact manifolds:

THEOREM B.3. *Let $L : \Gamma(M,E) \to \Gamma(M,F)$ be an elliptic partial differential operator of order m acting on sections of vector bundles $E, F \to M$ as above. Then:*

- *If $Lu = v$ weakly, for $u \in L^p(M,E)$ and $v \in W^{k,p}(M,F)$, then we have $u \in W^{k+m,p}(M,E)$ with the estimate*

$$||u||_{k+m,p,E} \lesssim (||v||_{k,p,F} + ||u||_{p,E})$$

- *If $Lu = v$ for $u \in C^{m,\alpha}(M,E)$ and $v \in C^{k,\alpha}(M,E)$, then $u \in C^{k+m,\alpha}(M,E)$ with the estimate*

$$||u||_{k+m,\alpha,E} \lesssim (||u||_{0,\alpha,E} + ||v||_{k,\alpha,F})$$

Versions of the above estimates also hold locally, provided we shrink the domain appropriately [Nic07, p.489].

MULTIPLICATION

First, we note some basic results:

PROPOSITION B.1. *[MS12, p.550] If $u \in W^{k,p}(M)$ and $v \in W^{k,\infty}(M)$, then $||uv||_{k,p} \lesssim ||u||_{k,p}||v||_{k,\infty}$. In particular, if $\phi \in C^\infty(M)$, multiplication by ϕ will induce a bounded linear map on all $W^{k,p}$ spaces.*

More generally, when we have $kp > n$ (so that the functions in $W^{k,p}$ are continuous), pointwise multiplication turns $W^{k,p}$ into an algebra [FU84, p.114]. Moreover, if $W^{j,q} \subseteq W^{k,p}$ (that is, when $\Sigma(j,p) \geq \Sigma(k,p)$ and $j \geq k$), then multiplication gives a continuous bilinear map $W^{k,p} \times W^{j,q} \to W^{k,p}$ [FU84, p.114]. A slightly more general multiplication result is often helpful:

THEOREM B.4. *[BB85, p.381] Suppose $Q : E_1 \times E_2 \to E_3$ is a smooth bilinear map between vector bundles. Then for $q = p$ and the conditions above (that is, $\Sigma(j,q) \geq \Sigma(k,p)$ and $j \geq k$), Q induces a continuous bilinear map $W^{k,p}(E_1) \times W^{j,q}(E_2) \to W^{k,p}(E_3)$.*

We can ask more generally about composing with arbitrary functions:

THEOREM B.5. *[MS12, p.560] Suppose $kp > n$ and $f \in C^\ell(\mathbb{R},\mathbb{R})$ for $\ell \geq k$ with bounded derivatives. Then the map $u \mapsto f \circ u$ is of class $C^{\ell-k}$ on $W^{k,p}(M)$. In particular, composing with a smooth function with bounded derivatives gives a smooth map from $W^{k,p}(M)$ to itself in this situation.*

The same result will hold for vector-valued functions. Finally, we observe that Sobolev spaces in some cases behave well under diffeomorphisms of the domain:

PROPOSITION B.2. *[MS12, p.552] Let $\phi : \bar{U} \to \bar{V}$ be a C^k diffeomorphism with bounded derivatives (actually, weaker hypotheses will suffice) between the closures of two open subsets U, V of M. Then we have the estimate $||u \circ \phi||_{k,p,U} \leq C||u||_{k,p,V}$ for all $1 < p < \infty$. In particular, if ϕ is smooth, then this estimate will hold for all $k \geq 0$.*

Moreover, since this map is linear and bounded, it is actually smooth. Note that C^k spaces do not transform so well under diffeomorphisms of the domain [Wen14]. But smooth diffeomorphisms of the domain will induce smooth maps on Sobolev spaces. Moreover, we can also see that smooth matrix-valued maps will induce smooth maps on vector-valued Sobolev spaces by pointwise multiplication. Hence we can see that, for *any* k and p, the Banach space $W^{k,p}(M,E)$ of sections of a vector bundle can be made into a smooth Banach manifold [MS12, p.561]. However, when we consider Banach spaces of sections of *fibre* bundles (such as the space of maps between two manifolds), we must consider transitions of the form $u \mapsto f \circ u$ for general smooth functions f. As we have seen above, this map will only be smooth in the case where $kp > n$, when the space $W^{k,p}(M,N)$ of maps between manifolds will indeed be a smooth Banach manifold [MS12, p.561]. Moreover, the evaluation map *at a particular point* will then be smooth [Wen14].

HIC SUNT DRACONES B.1. *When working with Sobolev spaces with $kp \leq n$, extreme caution is advised. In particular, it makes no sense to talk about values of such functions at particular points, and hence does not make sense to talk about such functions when the target is not a vector bundle but rather a fibre bundle, unless the fibre is embedded in some Euclidean space. The choice of embedding will then be significant.*

B.1 ANALYSIS ON TUBES

In this thesis, it will also be necessary to consider elliptic operators on certain types of non-compact manifolds. In this case, many of the Sobolev theorems above fail to hold and so we cannot appeal to the familiar results. Let Y be a compact Riemannian manifold, and let $X = Y \times \mathbb{R}$ be the 'tube' with cross-section Y. It has a natural Riemannian metric g induced from that on Y, which we will consider to be fixed. There are two different ways to define the Sobolev space $L^{1,2}(X,E)$ associated to a vector bundle $E \to X$ (with specified bundle metric) on a general non-compact manifold. The first is as the completion of the compactly supported smooth sections $f : X \to E$ under the norm given by

$$||f||^2_{H^{1,2}} = \int_X |Df|^2_g \mathrm{d}V + \int_X |f|^2_g \mathrm{d}V$$

for some first-order elliptic partial differential operator D acting on sections of the vector bundle E over X. This space is typically denoted $H^{1,2}$, or the space of *strongly differentiable* sections. The second definition is as the space of E-valued sections in $L^2(X,g)$ with Df (defined as a distribution) also in $L^2(X,g)$; this space is typically denoted $W^{1,2}(X,E)$, the space of *weakly differentiable* sections. Of course, on a compact manifold, these two definitions are entirely equivalent, and the equivalence of these definitions plays a crucial role in much of the elliptic theory. The key theorem in this section is:

THEOREM B.6. *(Meyers-Serrin Theorem for Tubular Manifolds) $H^{1,2} \equiv W^{1,2}$ when X is a tubular manifold.*

Proof. We attempt to mimic the standard proof for compact manifolds. One inclusion is always obvious: the sections with 'strong' derivatives will always have weak derivatives. Proving the converse requires that we can approximate (in the $H^{1,2}$ norm) any $f \in W^{1,2}(X, E)$ by smooth, compactly supported sections. It is clear that we can approximate and compactly supported function by a sequence of smooth sections (since this is exactly the same as the case of compact manifolds), so it suffices to show that we can approximate a general function $f \in W^{1,2}(X, E)$ by compactly supported functions in $W^{1,2}(X, E)$.

To this end, define a cutoff function β_T supported in $Y \times [-(T+1), T+1]$ that is smooth and has $|\nabla \beta_T| \leq 1$. Then we have

$$D(\beta_T f) = \beta_T Df + (\nabla \beta_T) * f$$

where $*$ denotes some algebraic multiplication operator. It is clear that $||\beta_T f - f||_{L^2(X)} \to 0$ as $T \to \infty$, since $f \in L^2(X, g)$. Now we consider

$$||D(\beta_T f) - Df||^2_{L^2(X)} \leq ||\beta_T Df - Df||^2_{L^2(X)} + ||(\nabla \beta_T) * f||^2_{L^2(X)}$$

The first term goes to zero as $T \to \infty$ for the same reason, since $Df \in L^2(X, g)$. Now, $\nabla \beta_T$ must be supported in the compact strips $[-(T+1), -T]$ and $[T, T+1]$ and hence we have an inequality

$$||\nabla \beta_T * f||^2_{L^2(X)} \leq C \left(\int_{-T-1}^{-T} |f|^2 + \int_T^{T+1} |f|^2 \right)$$

by the multiplication estimate above. Since $f \in L^2(X, g)$, this tends to 0 as $T \to \infty$ and hence we are done. ■

Remark B.1. The Meyers-Serrin theorem is false in general for non-compact manifolds and higher-order Sobolev spaces; see [Heb99] for some conditions under which it does hold.

Remark B.2. In the definitions above we used an arbitrary first-order elliptic operator D. To see that this choice does not affect the resulting space $W^{1,2}(X, E)$ when the principal symbol of D is uniformly bounded, let ∇ be a covariant derivative on sections of E. Then we have a Bochner-Wietzenböck formula of the form:

$$D^* Df = \nabla^* \nabla f + Kf$$

where K is some curvature operator. Since K is necessarily uniformly bounded (by the assumption on D) along the tube $X \times \mathbb{R}$, integrating the above equation against f shows that the two define equivalent $H^{1,2}$ norms.

Now we want to consider Sobolev theorems on tubular manifolds; the proofs of these results all follow a similar pattern. Firstly, some notation; let $B = Y \times (0, 1) \subseteq X$ be 'band' on the tube, and $B^+ = Y \times [-1/2, 3/2]$ be a slightly enlarged band. Let B_n denote the translation of B by the integer $n \in \mathbb{Z}$ and similarly for B_n^+. Each B_n is simply an open subset of the compact manifold B_n^+, where we may apply the familiar Sobolev inequalities, with *the same constant* on each band. Summing over the relevant Sobolev inequality for each band will then yield the inequality for the whole manifold. For instance, the Sobolev inequality $||f||_{L^q} \leq C||f||_{W^{1,p}}$ for $1 - n/p > -n/q$ (where $n = \dim X$) can be obtained by writing

$$||f||^q_{L^q} = \sum_{n=-\infty}^{\infty} \int_{B_n} |f|^q \leq \sum_{n=-\infty}^{\infty} C \left(\int_{B_n} |f|^p + |Df|^p \right)^{q/p} \leq C \left(\sum_{n=-\infty}^{\infty} \int_{B_n} |f|^p + |Df|^p \right)^{p/q} = C||f||_{W^{1,p}}$$

where we have used the fact that $p/q > 1$ in order to apply the triangle inequality. The same approach can be used more generally on non-compact manifolds that possess covers that are suitably uniform [Heb99]. A similar argument shows that we have the usual Sobolev embedding theorem $W^{k,p} \to C^\ell$ for suitable values of k, p, ℓ as above.

Importantly, however, the Rellich lemma (and other similar compactness results) are now obviously *false* for non-compact manifolds such as X. In particular, this means that elliptic operators no longer need to be Fredholm. Proving that the relevant elliptic operators are indeed Fredholm is a crucial part of Floer theory and is discussed in Chapter 4.

BIBLIOGRAPHY

[AB83] M. F. Atiyah and R. Bott, *The Yang-Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A **308** (1983), no. 1505, 523–615.

[ABC+09] Paul S. Aspinwall, Tom Bridgeland, Alastair Craw, Michael R. Douglas, Mark Gross, Anton Kapustin, Gregory W. Moore, Graeme Segal, Balázs Szendrői, and P. M. H. Wilson, *Dirichlet branes and mirror symmetry*, Clay Mathematics Monographs, vol. 4, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2009.

[AHS78] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A **362** (1978), no. 1711, 425–461.

[AJ10] M. Akaho and D. Joyce, *Immersed Lagrangian Floer theory*, J. Differential Geom. **86** (2010), no. 3, 381–500.

[APS75] M. F. Atiyah, V. K. Patodi, and I. M. Singer, *Spectral asymmetry and Riemannian geometry. I*, Math. Proc. Cambridge Philos. Soc. **77** (1975), 43–69.

[Ati88] Michael Atiyah, *New invariants of* 3- *and* 4-*dimensional manifolds*, The mathematical heritage of Hermann Weyl (Durham, NC, 1987), Proc. Sympos. Pure Math., vol. 48, Amer. Math. Soc., Providence, RI, 1988, pp. 285–299.

[Aur14] D. Auroux, *A beginner's introduction to Fukaya categories*, Contact and symplectic topology, Bolyai Soc. Math. Stud., vol. 26, János Bolyai Math. Soc., Budapest, 2014, pp. 85–136.

[BB85] B. Booss and D. D. Bleecker, *Topology and analysis*, Universitext, Springer-Verlag, New York, 1985, The Atiyah-Singer index formula and gauge-theoretic physics, Translated from the German by Bleecker and A. Mader.

[BT82] Raoul Bott and Loring W. Tu, *Differential forms in algebraic topology*, Graduate Texts in Mathematics, vol. 82, Springer-Verlag, New York-Berlin, 1982.

[CdS01] A. Cannas da Silva, *Lectures on symplectic geometry*, Lecture Notes in Mathematics, vol. 1764, Springer-Verlag, Berlin, 2001.

[CLW14] B. Chen, A-M. Li, and B-L. Wang, *Virtual neighborhood technique for pseudo-holomorphic spheres*, preprint: arXiv:1306.3276v2, 2014.

[DK90] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1990, Oxford Science Publications.

[Don02] S. K. Donaldson, *Floer homology groups in Yang-Mills theory*, Cambridge Tracts in Mathematics, vol. 147, Cambridge University Press, Cambridge, 2002, With the assistance of M. Furuta and D. Kotschick.

[DS94] Stamatis Dostoglou and Dietmar A. Salamon, *Self-dual instantons and holomorphic curves*, Ann. of Math. (2) **139** (1994), no. 3, 581–640.

[Flo88a] A. Floer, *Morse theory for Lagrangian intersections*, J. Differential Geom. **28** (1988), no. 3, 513–547.

[Flo88b] Andreas Floer, *A relative Morse index for the symplectic action*, Comm. Pure Appl. Math. **41** (1988), no. 4, 393–407.

[Flo88c] ———, *The unregularized gradient flow of the symplectic action*, Comm. Pure Appl. Math. **41** (1988), no. 6, 775–813.

[FOOO09] K. Fukaya, Y-G. Oh, H. Ohta, and K. Ono, *Lagrangian intersection Floer theory: anomaly and obstruction. Part I*, AMS/IP Studies in Advanced Mathematics, vol. 46, American Mathematical Society, Providence, RI; International Press, Somerville, MA, 2009.

[Fre95] Daniel S. Freed, *Classical Chern-Simons theory. I.*, Adv. Math. **113** (1995), no. 2, 237–303 (English).

[FU84] Daniel S. Freed and Karen K. Uhlenbeck, *Instantons and four-manifolds*, Mathematical Sciences Research Institute Publications, vol. 1, Springer-Verlag, New York, 1984.

[Fuk96] K. Fukaya, *Floer homology of connected sum of homology* 3*-spheres*, Topology **35** (1996), no. 1, 89–136.

[Fuk97] ______, *Floer homology for 3-manifolds with boundary*, 1997, never published: available at Fukaya's webpage.

[Fuk98] ______, *Anti-self-dual equation on* 4*-manifolds with degenerate metric*, Geom. Funct. Anal. **8** (1998), no. 3, 466–528.

[Fuk15] ______, $SO(3)$*-Floer homology of 3-manifolds with boundary 1*, preprint: arXiv:1506.01435v1, 2015.

[Gol84] William M. Goldman, *The symplectic nature of fundamental groups of surfaces*, Adv. in Math. **54** (1984), no. 2, 200–225.

[Hat02] Allen Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[Heb99] Emmanuel Hebey, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*, Courant Lecture Notes in Mathematics, vol. 5, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.

[HKK$^+$03] Kentaro Hori, Sheldon Katz, Albrecht Klemm, Rahul Pandharipande, Richard Thomas, Cumrun Vafa, Ravi Vakil, and Eric Zaslow, *Mirror symmetry*, Clay Mathematics Monographs, vol. 1, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2003, With a preface by Vafa.

[HTWZ95] H. Hofer, C. H. Taubes, A. Weinstein, and E. Zehnder (eds.), *The Floer memorial volume*, Progress in Mathematics, vol. 133, Birkhäuser Verlag, Basel, 1995.

[Jos11] Jürgen Jost, *Riemannian geometry and geometric analysis. 6th ed.*, 6th ed. ed., Berlin: Springer, 2011 (English).

[KN69] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol. I*, Interscience Tracts in Pure and Applied Mathematics, No. 15 Vol. II, Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney, 1969.

[Lan72] Serge Lang, *Differential manifolds.*, Addison-Wesley Series in Mathematics. Reading, Mass. etc.: Addison-Wesley Publishing Company, Inc. IX, 230 p. 5.60 (1972)., 1972.

[Lip14] Max Lipyanskiy, *Gromov-Uhlenbeck compactness*, preprint: arXiv:1409.1129v1, 2014.

[LM89] H. Blaine Lawson, Jr. and Marie-Louise Michelsohn, *Spin geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, Princeton, NJ, 1989.

[Lur09] J. Lurie, *On the classification of topological field theories*, Current developments in mathematics, 2008, Int. Press, Somerville, MA, 2009, pp. 129–280.

[Mer14] Will J. Merry, *Floer theory*, 2014, Lecture notes from a course at ETH Zürich. Available at the author's webpage.

[Mor98] John W. Morgan, *An introduction to gauge theory.*, Gauge theory and the topology of four-manifolds. Lectures of the graduate summer school, Providence, RI: American Mathematical Society, 1998, pp. 53–143 (English).

[Mro10] Tomasz Mrowka, *Introduction to instanton Floer homology*, 2010, Homology Theories of Knots and Links, MSRI Workshop: recording of talk available at: www.msri.org/workshops/511.

[MS12] D. McDuff and D. Salamon, J*-holomorphic curves and symplectic topology*, second ed., American Mathematical Society Colloquium Publications, vol. 52, American Mathematical Society, Providence, RI, 2012.

[Nic07] Liviu I. Nicolaescu, *Lectures on the geometry of manifolds*, second ed., World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[Oh93] Yong-Geun Oh, *Floer cohomology of Lagrangian intersections and pseudo-holomorphic disks. I*, Comm. Pure Appl. Math. **46** (1993), no. 7, 949–993.

[Oh15] ______, *Symplectic topology and Floer homology. Vol. 1*, New Mathematical Monographs, vol. 28, Cambridge University Press, Cambridge, 2015, Symplectic geometry and pseudoholomorphic curves.

[Ohm08] Toshiaki Ohmori, *Bubbling phenomena for sequences of harmonic maps from surfaces*, Master's thesis, Tohoku University, 2008.

[Oht01] Hiroshi Ohta, *Obstruction to and deformation of Lagrangian intersection Floer cohomology*, Symplectic geometry and mirror symmetry (Seoul, 2000), World Sci. Publ., River Edge, NJ, 2001, pp. 281–309.

[Pol09] Alexander Polishchuk, *Field guide to A_∞ sign conventions*, unpublished: available at the author's webpage., 2009.

[PW93] Thomas H. Parker and Jon G. Wolfson, *Pseudo-holomorphic maps and bubble trees*, J. Geom. Anal. **3** (1993), no. 1, 63–98.

[Sal97] Dietmar Salamon, *Lectures on Floer homology*, unpublished: available at author's webpage., 1997.

[Sch93] M. Schwarz, *Morse homology*, Progress in Mathematics, vol. 111, Birkhäuser Verlag, Basel, 1993.

[Sei08] Paul Seidel, *Fukaya categories and Picard-Lefschetz theory*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2008.

[Sik10] Adam S. Sikora, *Character varieties*, preprint: arXiv:0902.2589v3, 2010.

[SW08] D. Salamon and K. Wehrheim, *Instanton Floer homology with Lagrangian boundary conditions*, Geom. Topol. **12** (2008), no. 2, 747–918.

[Tau11] Clifford Henry Taubes, *Differential geometry*, Oxford Graduate Texts in Mathematics, vol. 23, Oxford University Press, Oxford, 2011, Bundles, connections, metrics and curvature.

[Weh05a] K. Wehrheim, *Anti-self-dual instantons with Lagrangian boundary conditions. I. Elliptic theory*, Comm. Math. Phys. **254** (2005), no. 1, 45–89.

[Weh05b] ______, *Lagrangian boundary conditions for anti-self-dual instantons and the Atiyah-Floer conjecture*, J. Symplectic Geom. **3** (2005), no. 4, 703–747, Conference on Symplectic Topology.

[Wen14] Chris Wendl, *Lectures on holomorphic curves in symplectic and contact geometry*, preprint: arXiv:1011.1690v2, 2014.

[Wit82] E. Witten, *Supersymmetry and Morse theory*, J. Differential Geom. **17** (1982), no. 4, 661–692.

INDEX