1
Introduction

In 1985 Andreas Floer discovered new topological invariants of certain 3-manifolds, the 'Floer homology groups'. This book originated from a series of seminars on this subject held in Oxford in 1988, the manuscript for the book being written sporadically over the intervening 12 years. The original plan of the project has been modified over time, but the basic aims have remained largely the same: these are, first, to give a thorough exposition of Floer's original work, and, second, to develop some further aspects of the theory which have not appeared in detail in the literature before. The author can only apologise for the long delay in completing this project.

Floer's original motivation for introducing his groups – beyond the intrinsic interest and beauty of the construction – seems to have been largely as a source of new invariants in 3-manifold theory, refining the Casson invariant which had been discovered shortly before. It was soon realised however that Floer's conception fitted in perfectly with the 'instanton invariants' of 4-dimensional manifolds, which date from much the same period. Roughly speaking, the Floer groups are the data required to extend this theory from closed 4-manifolds to manifolds with boundary. From another point of view the Floer groups appear, formally, as the homology groups in the 'middle dimension' of an infinite-dimensional space (the space of connections modulo equivalence) associated to a 3-manifold. This picture is obtained by carrying certain aspects of the Morse theory description of the homology of a finite-dimensional manifold over to infinite dimensions. All of this is closely related to ideas from quantum field theory – indeed, one of Floer's starting points was the renowned paper of Witten, [49], which *inter alia* forged a link between quantum mechanics and Morse theory

– and the connection with mathematical physics permeates the whole subject.

The formal properties of the Floer groups, and their relation with invariants in four dimensions, fit into a general conceptual framework of 'topological quantum field theories' which was propounded in the late 1980s by Segal, Atiyah, Witten and others. We recall from [2] that a topological field theory, in $d+1$ dimensions, consists of two functors on manifolds. The first assigns to each closed, oriented, d-manifold Y a vector space $\mathcal{H}(Y)$ (over, say, the complex numbers). The second assigns to each compact, oriented $(d+1)$-dimensional manifold X with boundary Y a vector

$$Z(X) \in \mathcal{H}(Y).$$

These are required to satisfy three axioms:

(1) The vector space assigned to a disjoint union $Y_1 \cup Y_2$ is the tensor product

$$\mathcal{H}(Y_1 \cup Y_2) = \mathcal{H}(Y_1) \otimes \mathcal{H}(Y_2).$$

(2) $\mathcal{H}(\overline{Y}) = \mathcal{H}(Y)^*$, where $\overline{Y}$ is Y with the reversed orientation.

(3) Suppose X is a $(d+1)$-manifold with boundary (which may be disconnected), and that X contains Y and $\overline{Y}$ as two of its boundary components. Let $X^\sharp$ be the oriented manifold obtained from X by identifying these two boundary components. Then we require that

$$Z(X^\sharp) = c(Z(X)),$$

where the contraction $c : \mathcal{H}(\partial X) \to \mathcal{H}(\partial X^\sharp)$ is induced from the dual pairing $\mathcal{H}(Y) \otimes \mathcal{H}(\overline{Y}) \to \mathbf{C}$ and the decomposition

$$\mathcal{H}(\partial X) = \mathcal{H}(Y) \otimes \mathcal{H}(\overline{Y}) \otimes \mathcal{H}(\partial X^\sharp).$$

These axioms have some simple consequences. First, Axiom 1 implies that if $Y = \emptyset$ is the empty d-manifold then $\mathcal{H}(\emptyset)$ is canonically isomorphic to $\mathbf{C}$. Thus if X is a *closed* $(d+1)$-manifold the vector $Z(X)$ is a numerical invariant of X. Second, suppose that a $(d+1)$-manifold U is a cobordism from Y_1 to Y_2, so the oriented boundary of U is a disjoint union $\overline{Y}_1 \cup Y_2$. Then, by Axioms 1 and 2, $Z(U)$ is an element of $\mathcal{H}(Y_1)^* \otimes \mathcal{H}(Y_2)$ and hence gives a linear map

$$\zeta_U : \mathcal{H}(Y_1) \to \mathcal{H}(Y_2).$$

If V is a cobordism from Y_2 to a third manifold Y_3 then Axiom 3 states that

$$\zeta_{V\circ U} = \zeta_V \circ \zeta_U : \mathcal{H}(Y_1) \to \mathcal{H}(Y_3),$$

where $V \circ U$ is the obvious composite cobordism. So we obtain a functor from the category of d-manifolds, with morphisms defined by cobordisms, to the category of vector spaces and linear maps.

The original motivation which led Segal and others to develop this kind of axiomatic picture was to abstract in a tidy mathematical form the basic structure of quantum field theories (more precisely, of conformal field theory on Riemann surfaces). The theories which are usually considered in physics differ from the set-up considered above in that they operate on manifolds with some additional differential-geometric structure, for example a Riemannian metric or a conformal structure. It is precisely the absence of these geometric structures in our set-up which leads to the designation *topological* quantum field theories, and which means that we obtain topological (or, more precisely, differential-topological) invariants of manifolds. In a typical physical set-up the corresponding space $\mathcal{H}(Y)$ would be an infinite-dimensional Hilbert space defined, at least schematically, by associating to Y a space of 'fields' $\mathcal{C}(Y)$ (an element of $\mathcal{C}(Y)$ might be a tensor field over Y), and then letting $\mathcal{H}(Y)$ be a space of L^2 *functions* on $\mathcal{C}(Y)$. The vector $Z(X)$ is obtained by functional integration over a space of fields on X, with given boundary value on Y.

The Yang–Mills invariants, and Floer groups, fit into this general scheme, with $d = 3$. In outline, for a 3-manifold Y, we take the Floer groups (with complex co-efficients say)

$$\mathcal{H}(Y) = HF_*(Y).$$

For a closed 4-manifold X the Yang–Mills instantons define a numerical invariant $Z(X)$, and for a 4-manifold with boundary we obtain invariants with values in the Floer homology of the boundary. Actually, as we shall see, the simple axioms above need to be modified slightly to apply to the Yang–Mills set-up and the theory has a number of special features. For example, the invariants of a closed 4-manifold are not in general just numbers but functions on the homology of the manifold – so we might regard the functor as being defined on a category of 4-manifolds containing preferred homology classes. Nevertheless these axioms capture the essence of the matter. Contrasting with the physical set-up outlined above, we can say very roughly that in place of the infinite-dimensional space $\mathcal{C}(Y)$ of *all* fields (i.e. connections) we restrict in this topological theory to the *finite* set of flat connections (modulo equivalence) over Y, and we restrict to 'instanton' connections over 4-manifolds, so that in place of the functional integration over connections we now have merely to *count* the instantons with given flat boundary values. To make rigorous sense of this, a key step is to add half-infinite tubes to our 4-manifolds, so that we have a picture in which the boundary is 'at infinity'.

An important goal then of this book is to develop this picture, of the Floer groups as part of a topological field theory, in detail. It is important to emphasise at the outset that, even after all this time, we are not able to complete this task. On the one hand there are, as we shall see, rather fundamental technical reasons why one cannot expect to have this simple picture without imposing some restrictions on the manifolds which are considered. On the other hand, even within the confines of the theory that one might reasonably hope for, there are crucial technical difficulties, arising from the non-compactness of instanton moduli spaces which – despite much labour by many mathematicians – have not yet been fully overcome. Failing, therefore, a definitive treatment we round off the book, in Chapter 8, by seeking to explain the problems that remain, and further developments one may expect in the future. We shall see that – far from being dull, technical matters – these difficulties lead to striking and unexpected formulae involving classical special functions.

Throughout the early 1990s an important motivation for the development of Floer theory was the hope that this might lead to new calculations of 4-manifold invariants, via cutting and pasting techniques. It has to be said that, at least on a narrow interpretation, this programme did not yield as much fruit as one might have hoped, and its goals have been to a large extent overtaken by events. The main lines of progress in this area (aside from algebro-geometrical techniques) came roughly thus. Firstly, through work of Mrowka and others involving cutting and pasting along 3-tori which, while it probably could be incorporated in a suitable generalisation of Floer theory, was not formulated explicitly in these terms. Secondly, through work of Kronheimer and Mrowka using singular connections (although again a version of Floer theory appeared in their arguments). Thirdly, and most decisively, through the introduction in 1994 of the Seiberg–Witten invariants. Leaving aside the well-known issue of the equivalence of the two theories, this last gives a more economical and powerful basis for the entire subject and makes the older instanton theory largely redundant as far as applications to 4-manifold topology go.

While it cannot be denied that the material in this book is less topical now than a decade ago (and at some points the text may have a slightly dated air, reflecting the long period over which it has been written) the author hopes that it is still worthwhile to present this material. We mention three grounds for this hope. First, the main thrust of the first part of the book is to develop certain differential-geometric and analytical techniques which apply to a wide range of problems, going beyond Yang–Mills theory (for example to the analogous symplectic Floer theory, to the Seiberg–Witten version of the Floer theory, to gluing problems for other structures such as self-dual metrics and metrics of special holonomy). Second, Floer's fundamental idea of defining 'middle-dimensional homology' for suitable infinite-dimensional manifolds is such an appealing one, and again one which in principle could appear in many different contexts, that it seems to deserve a thorough treatment. Third, while, as we have said above, some of the original motivation for the theory *vis-à-vis* 4-manifold topology is now reduced, there are intriguing questions which remain to be settled in setting up the Floer theory and understanding the whole relation between the instanton invariants and the Seiberg–Witten invariants. Some of these, in particular the appearance of modular forms, are touched on in Chapter 8. The Seiberg–Witten version of the Floer theory is a topic which is being very actively developed at the time of writing and, in conjunction with Floer's original groups, is expected to have important consequences in 3-manifold theory.

There are many topics omitted from this book. (In some cases these are things which we had hoped to include, in earlier and more ambitious plans, but found the energy wanting when it came to the point.) There are absolutely no examples: this is an entirely 'theoretical' treatment. We do not discuss the Casson invariant of homology spheres [46], or Floer's exact surgery sequence [8]. We do not mention Fukaya's extension of Floer's homology groups [9]. We do not have anything to say explicitly about the related theories developed by Taubes [47] and Morgan, Mrowka and Rubermann [36]. We do not say anything about the various interesting links between Floer's theory and the moduli spaces of flat connections over surfaces, and with algebraic geometry. Finally we say nothing about many of the deeper and more recent developments, connected with the Seiberg–Witten theory, such as the work of Muñoz [37] and Froyshov [25] on the 'finite type' condition. We do not discuss the Seiberg–Witten equations [51], [13], and the variant of the Floer theory they define. Except for the discussion of Fintushel and Stern's work of 1993 in Chapter 8 we have confined ourselves to an exposition of ideas that were current *circa* 1990.

On the other hand, we do digress from the narrow goal of setting up the Floer theory at a number of points. Thus, for example, we develop some of the main analytical results (in Chapter 4) in more generality than we need, because the ideas seem interesting and useful in other applications. We attempt to say a little about the background in mathematical physics, and the analogy with the symplectic theory.

The general scheme of the book is as follows. The first part (Chapters 2–5) aims to give a complete definition of the Floer groups of a homology 3-sphere: essentially following Floer's original paper. Chapter 6 develops the basic connection with 4-manifold invariants. The thrust of the first part is towards the geometrical and analytical techniques: at the beginning of Chapter 6 we step back to discuss the overall conceptual picture. Some readers may wish to look at the beginning of Chapter 6 at an earlier stage. Chapter 7 is devoted to refinements of the theory, mainly involving ideas from algebraic topology. This sets the stage for Chapter 8 in which, as we have mentioned, we discuss open problems and likely further developments.

The author expresses thanks to all the colleagues who have provided both help with this work and encouragement to complete the task. Two people should be mentioned in particular. Dieter Kotschick made notes of the original course of seminars which was the starting point for this project, and provided a great deal of help with the early development of the manuscript. Mikio Furuta was a participant in the seminar and contributed some invaluable ideas and drafts covering the less standard material in Chapter 7. In particular the formulation of the 'category of chain complexes' in that Chapter is due to Furuta.

2
Basic material

2.1 Yang–Mills theory over compact manifolds

In this Section we recall the rudiments of Yang–Mills theory in the standard situation – treated in numerous references – of a compact base manifold. (In general in this Chapter we follow the notation of [17].) So let V be a compact, connected, smooth manifold of dimension n, G be a compact Lie group and $P \to V$ be a principal G bundle over V. The gauge group $\mathcal{G}$ of automorphisms of P, covering the identity on V, acts on the space $\mathcal{A}$ of all connections on P by

$$g(A) = A - (d_A g)g^{-1}.$$

In Yang–Mills theory one needs to work with connections modulo gauge equivalence, i.e. modulo the action of $\mathcal{G}$, and to do this one can form the quotient spaces $\mathcal{B} = \mathcal{A}/\mathcal{G}$. This quotient is made more complicated by the possible existence of reducible connections, by which we mean connections A whose stabilisers Γ_A in $\mathcal{G}$ are larger than the centre $C(G)$ of G. (The stabiliser Γ_A is always a compact Lie group – the centraliser of the holonomy group of the connection A.) To avoid these complications one can restrict to the subset

$$\mathcal{B}^* = \{[A] \in \mathcal{B} : \Gamma_A \cong C(G)\}.$$

This is an open, dense subset of $\mathcal{B}$ (so long as the dimension n is greater than 1). We can make $\mathcal{B}^*$ into a smooth infinite-dimensional Banach manifold if we complete our spaces in suitable Sobolev norms. For example, we can take connections of class L^p_{k-1}, acted on by gauge transformations of class L^p_k (i.e. k derivatives in L^p). If the indices k and p satisfy the inequality $k - (n/p) > 0$ the L^p_k gauge transformations are *continuous* and the completion is naturally a Banach Lie group.

For the rest of this Chapter we shall denote by $\mathcal{G}$ and $\mathcal{A}$ these Sobolev completions. Thus $\mathcal{G}$ is a Banach Lie group acting smoothly on the Banach manifold $\mathcal{A}$.

To see the manifold structure of $\mathcal{B}^*$ explicitly we have to find slices for the action of $\mathcal{G}$. Fixing a background connection A_0 we have

$$\mathcal{A} = A_0 + \Omega^1(\mathfrak{g}_P)$$

where $\mathfrak{g}_P$ is the bundle of Lie algebras associated to P by the adjoint action of G. The tangent space to the orbit $\mathcal{G}(A_0)$ at A_0 is the image of the covariant derivative

$$d_{A_0} : \Omega^0(\mathfrak{g}_P) \to \Omega^1(\mathfrak{g}_P).$$

If V is equipped with a Riemannian metric then the space of connections becomes an infinite-dimensional, affine, Euclidean space, with the $\mathcal{G}$-invariant metric inherited from the standard L^2 metric on $\Omega^1(\mathfrak{g}_P)$. (This is not, of course, the same as an L^p_{k-1} metric used in completing $\mathcal{A}$.) There is then a standard choice of complementary subspace, namely the L^2 orthogonal complement. By Hodge theory, this is just the kernel of the formal adjoint operator

$$d^*_{A_0} : \Omega^1(\mathfrak{g}_P) \to \Omega^0(\mathfrak{g}_P).$$

The set of connections $A_0 + a$, for small a and with $d^*_{A_0} a = 0$, forms a local slice for the action of $\mathcal{G}$, and these slices give charts for $\mathcal{B}^*$. At the linear level we can identify the tangent space

$$T_{[A_0]}\mathcal{B}^* = \Omega^1(\mathfrak{g}_P)/\operatorname{Im} d_{A_0} = \ker d^*_{A_0}. \tag{2.1}$$

The *curvature* F_A or $F(A)$ of an L^p_{k-1} connection lies in L^p_{k-2}, so long as the inequality $k > n/p$ holds. The curvature can be regarded as a $\mathcal{G}$-equivariant map

$$F : \mathcal{A} \to \Omega^2(\mathfrak{g}_P),$$

whose derivative at a connection A_0 is the coupled exterior derivative

$$d_{A_0} : \Omega^1(\mathfrak{g}_P) \to \Omega^2(\mathfrak{g}_P).$$

This is obtained by linearising the formula

$$F_{A+a} = F_A + d_A a + a \wedge a. \tag{2.2}$$

Down on $\mathcal{B}^*$ we can think of the curvature as a section of a bundle of Banach spaces, the bundle over $\mathcal{B}^*$ associated to the action of $\mathcal{G}$ on $\Omega^2(\mathfrak{g}_P)$.

2.2 The case of a compact 4-manifold

Now we specialise to the case when $V = X^4$ is an oriented Riemannian 4-manifold. On X we have the Hodge $*$-operator, which acts on (bundle-valued) 2-forms, with square 1. Decomposing the curvature F_A of a connection A according to the decomposition $\Omega^2 = \Omega^+ \oplus \Omega^-$ of the 2-forms into *self-dual* and *anti-self-dual* parts (the ± 1 eigenspaces of $*$) we write $F_A = F_A^+ + F_A^- \in \Omega^+(\mathfrak{g}_P) \oplus \Omega^-(\mathfrak{g}_P)$. The *instanton* or *anti-self-dual (ASD) Yang–Mills* equation for a connection over any oriented Riemannian 4-manifold is the equation

$$F^+(A) = 0.$$

We refer to the solutions as 'instantons' or 'ASD connections'. Note that the instanton equation is *conformally invariant.*

The linearisation of the ASD equation about a given solution A is obtained by taking the self-dual part of Equation 2.2. We have

$$F^+_{A+a} = d_A^+ a + (a \wedge a)^+,$$

where d_A^+ is the projection of the exterior derivative to $\Omega^+(\mathfrak{g}_P)$. We get a complex

$$\Omega^0(\mathfrak{g}_P) \xrightarrow{d_A} \Omega^1(\mathfrak{g}_P) \xrightarrow{d_A^+} \Omega^+(\mathfrak{g}_P). \tag{2.3}$$

Notice that these operators are defined for any connection A over X, not just the instantons, and in general the composite $d_A^+ \circ d_A$ is given by the algebraic action of F_A^+.

Plainly the linearisation of the instanton equation is the equation, for $a \in \Omega^1(\mathfrak{g}_P)$,

$$d_A^+ a = 0.$$

The instanton equation is gauge-invariant, so to study the solutions near A we may as well restrict to the slice defined by Equation 2.1. Thus the linearised equation modulo gauge equivalence can be written as the single equation

$$D_A(a) = 0, \tag{2.4}$$

where $D_A = -d_A^* \oplus d_A^+ : \Omega^1(\mathfrak{g}_P) \to \Omega^0(\mathfrak{g}_P) \oplus \Omega^+(\mathfrak{g}_P)$. The operator d_A^+ is not elliptic by itself but the gauge-fixing condition built into D_A makes this latter operator elliptic. (Thus the instanton equation, viewed modulo gauge transformations, is a non-linear elliptic PDE.) Like any elliptic operator over a compact manifold, D_A has a Fredholm index:

$$\operatorname{ind} D_A = \dim\ker D_A - \dim\ker D_A^*.$$

The Atiyah–Singer index theorem gives a topological formula for this index which takes the form

$$\operatorname{ind} D_A = c(G)\kappa(P) - \dim G(1 - b_1 + b^+).$$

Here $c(G)$ is a normalising constant, $\kappa(P)$ is a characteristic number of P obtained by evaluating a 4-dimensional characteristic class on the fundamental cycle $[X]$, b_1 is the first Betti number of X and b^+ is the rank of a maximal positive subspace for the *intersection form* on $H_2(X)$. We now focus on the case when $G = SU(2)$ and we can take κ to be the second Chern class $c_2(P)$. Then the index formula becomes

$$8c_2(P) - 3(1 - b_1 + b^+). \tag{2.5}$$

(In Chapter 5 we will discuss the case of $U(2)$ and $SO(3)$ connections.)

Chern–Weil theory expresses the topological characteristic number $\kappa(P)$ as a curvature integral. Specialising again to the case of the group $SU(2)$ where $\kappa = c_2$ we have

$$\kappa(P) = \frac{1}{8\pi^2}\int_X \operatorname{Tr}\left(F_A^2\right). \tag{2.6}$$

This applies, of course, to any connection A on P. The wedge product form is equal to the square of the norm on self-dual 2-forms and opposite on the anti-self-dual forms, so we have the fundamental equation

$$\kappa(P) = \frac{1}{8\pi^2}\int_X \left(\left|F_A^-\right|^2 - \left|F_A^+\right|^2\right) d\mu. \tag{2.7}$$

So a connection is an instanton if and only if

$$\kappa(P) = \frac{1}{8\pi^2}\int_X |F|^2 \, d\mu. \tag{2.8}$$

This shows, in particular, that $\kappa(P) \geq 0$ if P supports an ASD connection. (And if $\kappa(P) = 0$ the connection must be *flat* – associated to a representation of $\pi_1(X)$.)

2.3 Technical results

We will now recall briefly the main theorems about Yang–Mills instantons, from the point of view of applications to 4-manifold differential topology. These will be used in Chapters 4 and 5 when we extend the theory to certain non-compact base manifolds. We refer to [17] for proofs.