Matemática Contemporânea, Vol. 2, 157–194
http://doi.org/10.21711/231766361992/rmc26

GAUGE THEORY AND THE TOPOLOGY OF 4-MANIFOLDS

J. D. S. Jones

Introduction

In these notes I will outline some of the key ideas in Donaldson's work concerning the question of which symmetric unimodular forms over the integers can arise as the intersection forms of smooth closed 4-manifolds. The original papers are [3, 4, 5] and the results are also proved in the book by Donaldson and Kronheimer [6]. I will also discuss how some of the results can be interpereted in homotopy theoretic terms, more precisely in terms of the cohomology of mapping spaces.

The first section contains a summary of results on the classification of 4-manifolds. I have not tried to make this a complete survey of the subject; rather my aim has been to state some of Donaldson's theorems and to set them in context. The articles [14, 9] are more thorough surveys of the theory of 4-manifolds; see also [6, Chapter 1].

The second section contains an outline of some of the essential background from gauge theory. This material is by now well-understood and details can be found in the main references [3, 4, 5, 6, 7, 13, 19] and the discussion here is very brief.

In [3] Donaldson proves that if the intersection form of a smooth 4-manifold is definite then it is diagonal. In [4] he goes on to study the question of which indefinite forms can arise as the intersection forms of smooth 4-manifold. The known results concerning this question are described in §1 but the complete answer is not known; there is a conjectural answer, stated in §1, but as yet this conjecture has not been settled. In §3 and §4 I will give an outline of Donaldson's argument which proves the best non-existence results which are currently known. The basic argument is very geometrical; it depends in a subtle way on the structure of the ends of the moduli spaces of anti-self-dual connections and therefore on Uhlenbeck's weak compactness theorem and Taubes's gluing construction for anti-self-dual connections. The methods described in §3 and §4 are important techniques, used by Donaldson, for doing computations.

The arguments in §3 and §4 are based on a study of the relation between the cohomology of the various moduli spaces which arise in gauge theory and the cohomology of mapping spaces. In §5 I will describe some of the results due to Masbaum [15] and Mielke [16] on the cohomology of the appropriate mapping spaces. In addition §5 contains a discussion of how the results of §3 can be re-interpreted in terms of the cohomology of mapping spaces and how the cohomology of these mapping spaces is related to the cohomology of the moduli spaces.

These notes are based on a series of lectures given at the workshop on the geometry and topology of gauge fields held at Campinas in April 1991. It is a great pleasure to thank the organisers of this meeting, in particular A. Rigas, F. Mercuri, and C. Negreiros for their hospitality and for creating such a stimulating atmosphere.

1. 4-manifolds

Throughout the term closed manifold will mean one which is compact and has no boundary. Let X be a simply-connected, closed, oriented 4-manifold; the term simply-connected 4-manifold will, unless specified otherwise, mean a 4-manifold which satisfies these hypotheses. Associated to X is a basic invariant, its intersection form

$$Q = Q_X : H_2(X; \mathbf{Z}) \times H_2(X; \mathbf{Z}) \to \mathbf{Z}.$$

This is a bilinear form defined over $\mathbf{Z}$ which is symmetric and unimodular: symmetric means that $Q(x,y) = Q(y,x)$ for all $x, y \in H_2(X;\mathbf{Z})$ and unimodular means that if we choose a basis $e_1, \dots, e_r$ for the free abelian group $H_2(X;\mathbf{Z})$ and express Q as the symmetric matrix

$$A = (a_{ij}), \qquad a_{ij} = Q(e_i, e_j)$$

then $\det A = \pm 1$. The fact that $H_2(X;\mathbf{Z})$ is a free abelian group follows from the hypotheses on X and Poincaré duality shows that Q is unimodular.

Two natural questions immediately present themselves:

The Realisation Question Given a symmetric unimodular form Q is it the intersection form of some simply connected 4-manifold?

The Classification Question Classify 4-manifolds with given intersection form.

The first classification theorem is the following result of [17].

Theorem 1.1 *Let X and Y be simply-connected 4-manifolds. Then if $Q_X \cong Q_Y$ it follows that X and Y are homotopy equivalent.*

This is the crudest possible classification of 4-manifolds and we would like more refined results which classify topological 4-manifolds up to homeomorphism and smooth 4-manifolds up to diffeomorphism. It is interesting to see that Milnor points out in [17] that there are significant differences between the classification of 4-manifolds and the analogous classification problem in higher dimensions.

Let us briefly digress to introduce some of the terminology of bilinear forms Q over $\mathbf{Z}$ and their invariants. More details can be found in [18] and [22]

(1) The rank of Q is the rank of the group on which Q is defined. In terms of a matrix representation of Q it is the size of the matrix.

(2) The form Q can be diagonalised over $\mathbf{R}$ and we define $b^+ = b^+(Q)$ to be the number of positive entries which occur when Q is diagonalised over $\mathbf{R}$ and $b^- = b^-(Q)$ to be the number of negative entries.

(3) The signature of Q is defined by

$$\sigma(Q) = b^+ - b^-.$$

(4) The type of Q is defined as follows: Q has type II, or is even, if $Q(x,x)$ is always even; otherwise Q has type I, or is odd.

(5) If $Q(x,x) \geq 0$ for all x, and $Q(x,x) = 0$ if and only if $x = 0$, we say that Q is positive definite; Q is negative definite if $-Q$ is positive definite; Q is definite if it is either positive definite or negative definite.

There is a basic algebraic fact about even definite forms, see [18] or [22].

Lemma 1.2 *Suppose Q is an even definite symmetric unimodular form over $\mathbf{Z}$; then $\sigma(Q)$ is divisible by 8.*

There is an even definite symmetric unimodular form over $\mathbf{Z}$ with signature 8; this is the form E_8 defined by the following matrix

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

The form E_8 has rank 8 and signature 8. It cannot be diagonalised over the integers, see for example [18, 19, 22].

The second classical theorem about the intersection forms of 4-manifolds is Rohlin's theorem.

Theorem 1.3 *Let X be a smooth simply connected 4-manifold; then $\sigma(Q_X)$ is divisible by 16.*

This is first proved in [20] see also [8]. Rohlin's theorem shows that there are genuine restrictions on the intersection forms of smooth 4-manifolds. For example E_8 cannot be the intersection form of a smooth simply connected 4-manifold. The significance of Rohlin's theorem for the classification of 4-manifolds is discussed in Milnor's paper [17]. In particular Milnor points out that even though Rohlin's theorem shows there is no smooth 4-manifold with intersection form E_8 there could possibly be a topological 4-manifold with intersection form E_8.

Before describing some of answers to the general realisation and classification problems let us discuss other classical invariants of smooth simply connected 4-manifolds. Since we are now assuming the manifold is smooth the other source of invariants of X is the tangent bundle T_X and, in particular, its characteristic classes. The tangent bundle has two basic characteristic classes, the Stiefel-Whitney class

$$w_2 \in H^2(X; \mathbf{Z}/2)$$

and the Pontryagin class

$$p_1 \in H^4(X; \mathbf{Z}).$$

These can be computed from Q_X as follows.

Let us continue to write Q_X for the intersection form on $H^2(X; \mathbf{Z}/2)$, cohomology modulo 2. Then the function

$$H^2(X; \mathbf{Z}/2) \to \mathbf{Z}/2, \qquad x \mapsto Q_X(x, x)$$

is linear and so, by the unimodularity of Q, it is given by

$$x \mapsto Q_X(c, x) \bmod 2$$

for some $c \in H^2(X; \mathbf{Z}/2)$. Then it is easy to check that

$$c = w_2.$$

Note that this shows that X is spin, that is $w_2 = 0$, if and only if the intersection form Q_X is even.

By applying the Hirzebruch signature theorem we deduce that

$$\langle p_1, [X]\rangle = 3(b^+ - b^-)$$

where $\langle , \rangle$ is the pairing between cohomology and homology and $[X] \in H^4(X; \mathbf{Z})$ is the fundamental class of the oriented 4-manifold X. Since $H^4(X;\mathbf{Z})$ is isomorphic to $\mathbf{Z}$ and the isomorphism is given by $x \mapsto \langle x, [X]\rangle$ it follows that p_1 is determined by the intersection form.

Thus we see that the classical invariants of X are all determined by the intersection form and if we follow the analogy with the classification of manifolds of dimension 5 or more it should now follow that the intersection form of X essentially determines X. Indeed if we classify 4-manifolds up to homeomorphism this is indeed true. The main theorem in the purely topological study of simply connected 4-manifolds is the following result due to Freedman [10].

Theorem 1.4

(1) *Suppose X and Y are smooth 4-manifolds such that $Q_X \cong Q_Y$; then X and Y are homeomorphic.*

(2) *Let Q be a symmetric unimodular form over $\mathbf{Z}$; then there is a topological 4-manifold X with $Q_X \cong Q$.*

(3) *Suppose X and Y are topological 4-manifolds with $Q_X \cong Q_Y \cong Q$. If Q has type II then X and Y are homeomorphic. If Q has type I then there are precisely two topological manifolds, up to homeomorphism, with intersection form Q.*

In part (3) of Freedman's theorem the two manifolds with the same intersection form are distinguished by their Kirby-Siebenmann invariant; this is an invariant $k(X) \in \mathbf{Z}/2$ and it vanishes if and only if $X \times S^1$ has a smooth structure. In particular if X is smooth $k(X) = 0$ and we see the relation between part (1) and part (3). This result completely settles the realisation and classification questions for simply connected topological 4-manifolds up to homeomorphism.

Let us now turn to smooth manifolds and, therefore, to Donaldson's theorems. I will divide Donaldson's work into three parts.

Definite forms. Donaldson proves the following theorem which gives very dramatic restrictions on the possible definite forms which arise as the intersection forms of simply-connected smooth 4-manifolds.

Theorem 1.5 *Suppose that X is a smooth simply connected 4-manifold such that Q_X is definite; then Q_X is diagonal.*

The original reference is [3] and the theorem is discussed, very carefully, in [6]. It should be contrasted with Freedman's theorem which tells us that, given a symmetric unimodular form defined over $\mathbf{Z}$, there always exists a simply-connected topological 4-manifold with this intersection form. Donaldson's theorem tells us that if the form is definite and not diagonal then the manifold given by Freedman's theorem cannot be smooth. The theory of definite symmetric unimodular forms is a difficult part of classical number theory, see for example [18] and [22] but Donaldson's theorem tells us that none of these forms, apart from the simple diagonal forms, can occur as the intersection forms of smooth 4-manifolds.

One of the consequences of the combination of Donaldson's theorem (1.5) and Freedman's theorem (1.4) is that there must exist a fake $\mathbf{R}^4$ – this is a smooth manifold which is homeomorphic to $\mathbf{R}^4$ but not diffeomorphic to $\mathbf{R}^4$. Two different ways of proving there exists a fake $\mathbf{R}^4$ are described in [7] and [13]. It is proved in [11] that there are at least 3 fake $\mathbf{R}^4$'s. It is now known that there are uncountable families of fake $\mathbf{R}^4$'s see [12], [24]. The existence of fake $\mathbf{R}^4$'s is proved by an implicit argument – the only way to account for the fact that Freedman's methods must break down in the smooth category is that there is a fake $\mathbf{R}^4$ – and there is no known way of constructing a fake $\mathbf{R}^4$ directly.

Donaldson's theorem proves, for example, that

$$nE_8 = E_8 \oplus \cdots \oplus E_8$$

(where there are n-summands) cannot be the intersection form of a smooth simply connected 4-manifold. Note that if n is odd then this also follows from Rohlin's theorem.

Now let K be the K3 surface

$$K = \{[z_0, z_1, z_2, z_3] : z_0^4 + z_1^4 + z_2^4 + z_3^4 = 0\} \subset \mathbf{CP}^3$$

where $[z_0, z_1, z_2, z_3]$ are the homogeneous coordinates of a point in 3-dimensional complex projective space $\mathbf{CP}^3$. Then K is a smooth 4-manifold; Milnor shows in [17] that K is simply-connected and

$$Q_K = -E_8 \oplus -E_8 \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

On the other hand we know from Donaldson's theorem (1.5) that $-E_8 \oplus -E_8$ cannot be the intersection form of a smooth simply-connected 4-manifold. It is natural to look for the dividing line between the non-existence results and the intersection form of K. For this we need to study indefinite forms.

Indefinite forms. There is a classification of indefinite symmetric unimodular forms over $\mathbf{Z}$ given for example in [18, 22]. Such forms are classified by their rank $r = b^+ + b^-$, signature $\sigma = b^+ - b^-$, and type. There are three building blocks,

$$(1), \qquad H, \qquad E_8.$$

The first of these three is the 1×1 matrix (1), the second is

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and the third is the matrix E_8 given above. Every indefinite form fits in to one of three distinct families:

(1) type I,

$$n(1) \oplus m(-1), \qquad n, m \geq 1$$

where, $r = n + m$, $\sigma = n - m$, $b^+ = n$, $b^- = m$;

(2) type II, $\sigma \leq 0$

$$-nE_8 \oplus mH, \qquad n \geq 0, m \geq 1$$

where, $r = 8n + 2m$, $\sigma = -8n$, $b^+ = m$, $b^- = 8n + m$;

(3) type II, $\sigma > 0$

$$nE_8 \oplus mH, \qquad n, m \geq 1$$

where, $r = 8n + 2m$, $\sigma = 8n$, $b^+ = 8n + m$, $b^- = m$.

By $-nE_8$ we mean the direct sum of n copies of $-E_8$. Note that $H \cong -H$ over $\mathbf{Z}$ and so $-H$ never occurs in this list. The assumptions on n and m ensure that the forms are indefinite.

For our purposes the distinction between type II, $\sigma < 0$ and type II, $\sigma > 0$ is one of orientation convention. It fits best with the natural orientations of 4-manifolds like the K3 surface K if we arrange conventions in the type II case so that $\sigma \leq 0$.

Each of the type I forms is the intersection form of a smooth 4-manifold. The intersection form of $\mathbf{CP}^2$ with it usual orientation is just (1) and the intersection form of $\overline{\mathbf{CP}}^2$, by which we mean $\mathbf{CP}^2$ with the opposite orientation, is (-1). Now by taking the connected sum of n copies of $\mathbf{CP}^2$ with m copies of $\overline{\mathbf{CP}}^2$ we get intersection form $n(1) \oplus m(-1)$.

We are left to deal with the type II forms. The case mH is easy to handle. Let S be the product $S^2 \times S^2$ so that

$$Q_S = H.$$

By taking the connected sum of m copies of S we can realise mH as the intersection form of a smooth 4-manifold. The cases $-nE_8 \oplus mH$ are more subtle however.

Theorem 1.6 *Suppose X is a simply-connected smooth 4-manifold with even indefinite intersection form; then*

$$\begin{aligned} b^+ = 1 &\implies Q_X = H \\ b^+ = 2 &\implies Q_X = H \oplus H. \end{aligned}$$

The original reference is [4] and the theorem is discussed in careful detail in [6]. Thus it follows, by combining this theorem with Rohlin's theorem, that in the family of forms $-nE_8 \oplus mH$, $n, m \geq 1$ the minimal, in the obvious sense, form with non-zero n which can occur as the intersection form of a smooth simply-connected 4-manifold is

$$Q_K = -2E_8 \oplus 3H$$

Thus K is indecomposable and it is tempting to believe it is one of the basic building blocks of smooth 4-manifolds in the sense of the following conjecture; compare [9].

Conjecture 1.7 *The only even indefinite unimodular forms defined over* $\mathbf{Z}$ *which can be the intersection forms of smooth simply-connected 4-manifolds are*

$$pQ_K \oplus qQ_S.$$

If this conjecture is true then we get a complete answer to the realisation question for smooth simply-connected 4-manifolds. There are four indecomposable pieces

$$S, \quad \mathbf{CP}^2, \quad \overline{\mathbf{CP}}^2, \quad K$$

and every smooth 4-manifold is homeomorphic to a connected sum of these indecomposable pieces. The only intersection forms which can occur are given by direct sums of

$$Q_S, \quad Q_{\mathbf{CP}^2}, \quad Q_{\overline{\mathbf{CP}}^2}, \quad Q_K.$$

It is a very pleasant exercise to use Freedman's theorem (1.4) and the classification of indefinite forms to find relations amongst the 4-manifolds formed by taking connected sums of the four indecomposable pieces. Now let us turn our attention to the classification question for smooth simply-connected 4-manifolds.

Polynomial Invariants. The classification question for smooth 4-manifolds is to classify smooth 4-manifolds up to diffeomorphism. We will see that it is considerably more complicated than the classification up to homeomorphism given by Freedman's theorem. Indeed one of the conclusions of Donaldson theory is that in many cases there are an infinite number of smooth manifolds with a fixed intersection form. In view of Freedman's theorem we can express this by saying that in many cases there are an infinite number of smooth manifolds within each homeomorphism class. To distinguish between these smooth manifolds we need more invariants and these are provided by Donaldson's polynomial invariants.

These polynomial invariants are defined for any smooth, oriented, simply-connected, 4-manifold X with b^+ odd and $b^+ \geq 3$. The invariants are multilinear functions

$$\Phi_k = \Phi_k(X) : H_2(X; \mathbf{Z}) \times \ldots \times H_2(X; \mathbf{Z}) \to \mathbf{Z}$$

where there are

$$d = 4k - \frac{3(b^+ + 1)}{2}$$

factors $H_2(X; \mathbf{Z})$. They are defined for sufficiently large k. To say that they are invariants means that if $f : X \to Y$ is an orientation preserving diffeomorphism and

$$f_* : H_2(X; \mathbf{Z}) \to H_2(Y; \mathbf{Z})$$

is the induced isomorphism on homology, then

$$\Phi_k(Y)(f_*(x_1), \ldots, f_*(x_d)) = \Phi_k(X)(x_1, \ldots, x_d).$$

Theorem 1.8 *Suppose that $X = X_1 \# X_2$ is a smooth oriented connected sum where $b^+(X_1), b^+(X_2) > 0$ and $b^+(X) = b^+(X_1) + b^+(X_2)$ is odd; then $\Phi_k(X) = 0$ for all sufficiently large k.*

Theorem 1.9 *If Z is an algebraic surface with b^+ odd and $b^+ \geq 3$; then, for sufficently large k, $\Phi_k(Z)$ is non-zero.*

The original reference is [5] and the construction of the polynomial invariants and the proof of these theorems is very carefully discussed in [6].

Here is an example. Let S_d be a smooth algebraic surface in $\mathbf{CP}^3$ of degree d. So S_d is the zero set of a homogeneous polynomial in 4 variables of degree d. Then by repeating the method Milnor used to compute the intersection form of the K3 surface K we deduce

$$\begin{aligned} b^+ &= \alpha_d = \tfrac{1}{3}(d-1)(d-2)(d-3) \\ b^- &= \beta_d = \tfrac{2}{3}(d-1)(2d^2-4d+3). \end{aligned}$$

If d is odd it follows that the intersection form of S_d has type I. Thus if d is large enough it follows that the form is indefinite and type I and so, by the classification of such forms, it must be isomorphic to

$$\alpha_d(1) \oplus \beta_d(-1).$$

By using the Lefshetz Hyperplane Theorem it follows that S_d is simply connected and so by Freedman's Theorem (1.4), it follows that S_d is homeomorphic to a connected sum of α_d copies of $\mathbf{CP}^2$ and β_d copies of $\overline{\mathbf{CP}}^2$. By Theorem (1.8), provided d is large enough, all the polynomial invariants of this connected sum vanish. Since S_d is an algebraic surface, Theorem (1.9), shows the polynomial invariants of S_d do not all vanish. Therefore it follows that S_d cannot be diffeomorphic to a connected sum of α_d copies of $\mathbf{CP}^2$ and β_d copies of $\overline{\mathbf{CP}}^2$. This shows that, provided d is large enough, there are at least two smooth manifolds with intersection form

$$\alpha_d(1) \oplus \beta_d(-1).$$

In fact it can be shown that, up to diffeomorphism, there are an infinite number of smooth manifolds homeomorphic to a connected sum of one copy of $\mathbf{CP}^2$ and nine copies of $\overline{\mathbf{CP}}^2$. This result is discussed in [6] which contains the references to the original papers.

2. Gauge theory

The theme running through Donaldson's work is to use the spaces of solutions of the Yang-Mills equations to construct invariants of the underlying smooth 4-manifold X. From now on we will assume that X is a smooth simply-connected 4-manifold equipped with a Riemannian metric.

Let P be a principal $SU(2)$ bundle over X. Such principal bundles are classified by their Chern class $c_2(P) \in H^4(X)$. In this section H^p will denote integral homology. Since X is closed and oriented, $H^4(X) \cong \mathbf{Z}$ so we can identify the Chern class $c_2(P)$ with an integer. Thus we write

$$k = k(P) = \langle c_2(P), [X] \rangle$$

where $[X] \in H_4(X)$ is the fundamental class. Now let P_k be a bundle with Chern class k. Let A be a connection on P_k. Thus locally, on an open set U in X on which the bundle is trivialised, such a connection is given by

$$A_U = A_1(x)dx_1 + A_2(x)dx_2 + A_3(x)dx_3 + A_4(x)dx_4$$

where the A_i are functions on U which take their values in $\mathfrak{su}(2)$, the Lie algebra of $SU(2)$. The Lie algebra $\mathfrak{su}(2)$ is the space of skew-adjoint 2×2 complex matrices with trace zero and so the A_i are matrix valued functions. On $U \cap V$, A_U and A_V are related by

$$A_U = g^{-1} A_V g + g^{-1} dg$$

where $g : U \cap V \to SU(2)$ is the transition function of the bundle P_k. To make sense of this equation remember that both g and the A_i are matrix valued functions.

Let $\mathcal{A}_k$ be the space of connections on P_k and let $\mathcal{G}_k$ be the group of automorphisms of P_k which cover the identity on the base space X. This group $\mathcal{G}_k$ is the group of gauge transformations of P_k and it acts on $\mathcal{A}_k$ by pull-back of connections

$$\mathcal{G}_k \times \mathcal{A}_k \to \mathcal{A}_k, \qquad (g, A) \mapsto g^*(A).$$

Locally this is given by

$$g^*(A) = g^{-1}Ag + g^{-1}dg$$

where, since we are working locally, the gauge transformation g becomes a function with values in $SU(2)$.

Let A be a connection on P_k and let Γ_A be the isotropy group of the action of $\mathcal{G}_k$ on A,

$$\Gamma_A = \{g \in \mathcal{G}_k : g^*(A) = A\}.$$

Then Γ_A always contains the constant gauge transformations ± 1. Recall that a connection on P_k is irreducible if and only if

$$\Gamma_A = \{\pm 1\}.$$

If A is reducible then Γ_A is isomorphic to a circle. We would like to form the quotient space of the action of $\mathcal{G}_k$ on $\mathcal{A}_k$ but the presence of the reducible connections causes difficulties. There are two ways around this problem.

The first is to consider the space $\mathcal{A}_k^*$ of irreducible connections on P_k and then form the quotient

$$\mathcal{B}_k^* = \mathcal{A}_k^*/\mathcal{G}_k.$$

The action has local slices and the projection

$$\mathcal{A}_k^* \to \mathcal{B}_k^*$$

is a locally trivial principal budle with group

$$\bar{\mathcal{G}}_k = \mathcal{G}_k/\{\pm 1\}.$$

The second is to work instead with the subgroup $\mathcal{G}_k^0$ of gauge transformations which are the identity at a chosen base point in X. Now the group $\mathcal{G}_k^0$ acts freely and the action has local slices and so we can form the quotient

$$\mathcal{B}_k^0 = \mathcal{A}_k / \mathcal{G}_k^0.$$

This time the projection

$$\mathcal{A}_k \to \mathcal{B}_k^0$$

is a locally trivial principle bundle with group $\mathcal{G}_k^0$.

Now we introduce the self-duality equations and the Yang-Mills moduli space. Given a connection A on P_k we can form the curvature $F_A \in \Omega^2(X; \mathfrak{su}(P_k))$. Here $\Omega^2(X; \mathfrak{su}(P_k))$ is the space of 2-forms on X with values in the bundle $\mathfrak{su}(P_k)$ defined by

$$\mathfrak{su}(P_k) = P_k \times_{SU(2)} \mathfrak{su}(2)$$

where $SU(2)$ acts on $\mathfrak{su}(2)$ by the adjoint representation. Locally the curvature is given by the formula

$$F_A = dA + A \wedge A.$$

In this local formula remember that A is a matrix of 1-forms so dA is the matrix of 2-forms obtained by applying the exterior derivative d to each of the entries of A, and $A \wedge A$ is defined by the combination of matrix multiplication and the exterior product of forms.

Now suppose that X has a metric. Then the metric and the orientation define the Hodge star operator

$$* : \Omega^2(X; \mathfrak{su}(P_k)) \to \Omega^2(X; \mathfrak{su}(P_k)).$$

On $\mathbf{R}^4$ with its usual metric and orientation $*$ is given by

$$*(dx_i \wedge dx_j) = \pm dx_k \wedge dx_l$$

where $\{i, j, k, l\} = \{1, 2, 3, 4\}$ and the sign is $+1$ if $(1, 2, 3, 4) \mapsto (i, j, k, l)$ is an even permutation and -1 if it is odd. This operator is extended to matrix valued forms by applying it to each entry of the matrix. The Yang-Mills equations are as follows.

The self-duality equations.

$$*F_A = F_A$$

The anti-self-duality equations.

$$*F_A = -F_A$$

To understand these equations better it is a very good exercise work them out explicitly in terms of the above local description of connections, curvature, and the Hodge star operator. They are first order non-linear equations for the connection A. It is straightforward to check that if A satisfies one of these equations then so does $g^*(A)$ where $g \in \mathcal{G}_k$. Self-duality or anti-self-duality is a matter of orientation conventions. Here I will follow Donaldson and concentrate on the ASD (anti-self-dual) equations and refer to a connection whose curvature satisfies the ASD equations as an ASD connection. Now we define the moduli space of ASD connections

$$\mathcal{M}_k = \frac{\text{ASD connections}}{\mathcal{G}_k}.$$

We use the obvious notation $\mathcal{M}_k^*$ for the moduli space of irreducible ASD connections and $\mathcal{M}_k^0$ for the quotient of the space of ASD connections by the group $\mathcal{G}_k^0$.

The structure of the moduli spaces. In general there are singularities in $\mathcal{M}_k$ corresponding to reducible ASD connections. However it is possible to analyse the local structure of $\mathcal{M}_k$ in a neighbourhood of these singularities, see [6, 7], and [13]. There are two important special cases where there are no reducible ASD connections. For the proof of the following theorem see [6] or [5] and also [7, 13, 19].

Theorem 2.1 *Suppose that either*

(1) *the intersection form Q_X is indefinite, or*

(2) Q_X *is even and* $k = 1$.

Then for a generic metric on X*, there are no reducible ASD connections on* X *and* $\mathcal{M}_k$ *is a smooth manifold of dimension*

$$8k - 3(1 + b^+).$$

This gives us a complete description of the local structure of $\mathcal{M}_k$ so we now look at its global structure. The moduli space $\mathcal{M}_k$ is not compact so we should analyse what happens as we "go off to infinity" in $\mathcal{M}_k$. To deal with this precisely, following [6], introduce the following definition.

Definition 2.2 An ideal ASD connection with Chern number k consists of a pair

$$([A]; \{x_1, \ldots, x_l\})$$

where $[A] \in \mathcal{M}_{k-l}$ and $\{x_1, \ldots, x_l\}$ is an unordered l-tuple of points in X. The curvature of the ideal connection $([A]; \{x_1, \ldots, x_l\})$ is the measure

$$|F_A|^2 + 8\pi^2 \sum_{i=1}^{l} \delta_{x_i}$$

where $|F_A|^2$ is the pointwise norm of the curvature F_A.

Here $|F_A|^2 + 8\pi^2 \sum \delta_{x_i}$ is the measure which, for any continuous function f on X, gives the integral

$$\int_X f|F_A|^2 d\mu + 8\pi^2 \sum_{i=1}^{l} f(x_i)$$

where $d\mu$ is the measure on X defined by the metric. Note that we allow the possibility that $l = 0$, in which case we have a genuine ASD connection. We also allow the possibility $l = k$, in which case we have a flat connection on the product bundle on X and a set of k points in X; since X is simply connected it must follow that the flat connection is the trivial connection and we simply identify the ideal ASD connection with the set of points $\{x_1, \dots, x_k\}$.

Definition 2.3 A sequence of ASD connections $[A_\alpha]$ converges weakly to the ideal ASD connection $([A]; \{x_1, \dots, x_l\})$ if

(1) The sequence $|F_{A_\alpha}|^2$ converges to $|F_A|^2 + 8\pi^2 \sum \delta_{x_i}$ as measures.

(2) There are bundle isomorphisms

$$\rho_\alpha : P_l|_{X_0} \to P_k|_{X_0},$$

where $X_0 = X \setminus \{x_1, \dots, x_l\}$, such that the sequence of connections $\rho_\alpha^* A_\alpha$ converges to A in the C^∞ topology on compact sets.

Here part (1) means that for each continuous function f on X

$$\int_X f|F_{A_\alpha}|^2 d\mu \to \int_X f|F_A|^2 d\mu + 8\pi^2 \sum_{i=1}^{l} f(x_i).$$

Now we have the following version of Uhlenbeck's weak compactness theorem.

Theorem 2.4 *Let $[A_\alpha]$ be a sequence of ASD connections. Then there is a subsequence which converges weakly to an ideal ASD connection.*

The proof of this theorem is given in each of the main references. There is a very simple analogy which may help to understand ideal ASD connections and the weak compactness theorem. Let Rat_k be the space of meromorphic functions of degree k on the Riemann sphere $S^2 = \mathbf{C} \cup \infty$; equivalently the space of holomorphic maps $S^2 \to S^2$. Then such a function is completely determined, up to a constant, by its zeroes $\{z_1, \dots, z_k\}$ and its poles $\{p_1, \dots, p_k\}$. We can examine the behaviour of a sequence of such functions f_α where the poles $\{p_1, \dots, p_k\}$ remain constant, one zero, say $z_1(\alpha)$, converges to one of the poles, say p_1, and the other zeroes $\{z_2, \dots, z_k\}$ remain constant. Then this sequence does not converge to an element of Rat_k; rather it converges weakly, in exactly the sense described above, to the "ideal rational function" $(f; \{p_1\})$ where the zeroes and poles of f are

$$\{z_2, \ldots, z_k\}, \qquad \{p_2, \ldots, p_k\}.$$

Here the role of the curvature is played by the energy density $|df|^2$.

The weak compactness theorem is used in many places in the theory. One immediate application is that it gives a compactification of the moduli spaces $\mathcal{M}_k$ as follows. Define $SP^l(X)$, the l-th symmetric product of X, to be the space

$$SP^l(X) = X^l/\Sigma_l$$

where X^l is the l-fold Cartesian product of X and the symmetric group Σ_l acts on X^l by permuting factors. Now define the space of ideal ASD connections to be

$$I\mathcal{M}_k = \bigcup_{l=0}^{k} \mathcal{M}_{k-l} \times SP^l(X)$$

topologised so that sequences converge if and only if they converge weakly in the sense of the above definition. The weak compactness theorem tells us that the space $I\mathcal{M}_k$ is compact. Now define the compactified moduli space $\bar{\mathcal{M}}_k$ to be the closure of $\mathcal{M}_k$ in $I\mathcal{M}_k$.

3. Connections, mapping spaces, and cohomology

Now we start to analyse the relation between the cohomology of the Yang-Mills moduli space $\mathcal{M}_k$ and the cohomology of the space $\mathcal{B}_k$ of all connections modulo gauge equivalence. The first step in this process is to understand the homotopy type of the space $\mathcal{B}_k^0$ of connections modulo pointed gauge transformations.

Theorem 3.1 *There is a homotopy equivalence*

$$\mathcal{B}_k^0 \simeq \mathrm{Map}_k^0(X, \mathbf{HP}^\infty).$$

In the statement of the lemma $\mathbf{HP}^\infty$ is infinite dimensional quaternionic projective space and Map_k^0 means the space of base point preserving maps

$f : X \to \mathbf{HP}^\infty$ such that the induced homomorphism

$$f_* : H_4(X) \cong \mathbf{Z} \to H_4(\mathbf{HP}^\infty) \cong \mathbf{Z}$$

is multiplication by k. For a proof see [1].

There is a natural principal $SU(2)$-bundle

$$\mathcal{P}_k \to \mathcal{B}_k^0$$

defined as follows. The group $\mathcal{G}_k^0$ acts freely on $\mathcal{A}_k$; it also acts on P_k since, by definition it is a group of automorphisms of P_k. Thus we may form the quotient

$$\mathcal{P}_k = \mathcal{A}_k \times_{\mathcal{G}_k^0} P_k.$$

Since P_k is a principal $SU(2)$ bundle over X it follows that $\mathcal{P}_k$ is a principal $SU(2)$ bundle over

$$\mathcal{A}_k \times_{\mathcal{G}_k^0} X = \mathcal{B}_k^0 \times X$$

where the last equality follows from the fact that $\mathcal{G}_k^0$ acts trivially on X.

In terms of mapping spaces we can describe this bundle $\mathcal{P}_k$ as follows. There is a natural evaluation map

$$\mathrm{Map}_k^0(X, \mathbf{HP}^\infty) \times X \to \mathbf{HP}^\infty$$

and $\mathcal{P}_k$ is the bundle over $\mathcal{B}_k^0 \times X \simeq \mathrm{Map}_k(X, \mathbf{HP}^\infty) \times X$ induced from the universal principal $SU(2)$ bundle over $\mathbf{HP}^\infty$ by this map.

Now let

$$c = c_2(\mathcal{P}_k) \in H^4(\mathcal{B}_k^0 \times X)$$

be the second Chern class of $\mathcal{P}_k$. We can use the Künneth theorem (together with our standing hypotheses on X) to decompose $H^4(\mathcal{B}_k^0 \times X)$ as

$$H^0(\mathcal{B}_k^0) \otimes H^4(X) \oplus H^2(\mathcal{B}_k^0) \otimes H^2(X) \oplus H^4(\mathcal{B}_k^0) \otimes H^0(X).$$

With respect to this decomposition write

$$c^{2,2} \in H^2(\mathcal{B}_k^0) \otimes H^2(X)$$

for the appropriate component of c. Now we use $c^{2,2}$ to define a homomorphism

$$\mu_0 : H_2(X) \to H^2(\mathcal{B}_k^0)$$

in the natural way. There is a pairing

$$H^2(X) \otimes H_2(X) \to \mathbf{Z}$$

and this gives a pairing

$$H^2(\mathcal{B}_k^0) \otimes H^2(X) \otimes H_2(X) \to H^2(\mathcal{B}_k^0)$$

which we denote by $\langle , \rangle$. Then μ_0 is defined by

$$\mu_0(u) = \langle c^{2,2}, u \rangle.$$

In fact this homomorphism μ_0 descends to a homomorphism

$$\mu : H_2(X) \to H^2(\mathcal{B}_k^*)$$

The relation between $\mathcal{B}_k$ and $\mathcal{B}_k^*$ is as follows. By definition $\mathcal{B}_k^{*,0}$ is a subspace of $\mathcal{B}_k^0$ and $\mathcal{B}_k^{*,0}$ is the total space of a principal $SO(3)$ bundle over $\mathcal{B}_k^*$,

$$SO(3) \to \mathcal{B}_k^{*,0} \xrightarrow{\pi} \mathcal{B}_k^*.$$

By arguing directly with this bundle it is possible to prove that there is a commutative diagram

$$\begin{array}{ccc} H_2(X) & \xrightarrow{\mu_0} & H^2(\mathcal{B}_k^0) \\ \mu \downarrow & & \downarrow \\ H^2(\mathcal{B}_k^*) & \xrightarrow{\pi^*} & H^2(\mathcal{B}_k^{*,0}), \end{array}$$

compare [6].

Now we can compose the maps μ and μ_0 with the homomorphism of cohomology induced by the inclusion of the moduli spaces to get corresponding homomorphisms

$$\mu_0 : H_2(X) \to H^2(\mathcal{M}_k^0), \qquad \mu : H_2(X) \to H^2(\mathcal{M}_k^*).$$

These are very important ingredients in the theory.

To simplify matters let us assume for the rest of this section that the hypotheses of Theorem (2.1) are satisfied and we have chosen a generic metric on X. Thus there are no reducible ASD connections, $\mathcal{M}_k = \mathcal{M}_k^*$, and $\mathcal{M}_k$ is a smooth manifold. Since our main objective is to discuss the proof of Theorem (1.6) there is no loss in this assumption.

Now we look for geometric representatives for the cohomology classes $\mu(u) \in H^2(\mathcal{M}_k)$. Recall that, geometrically, p-dimensional closed submanifolds of a manifold M define p-dimensional homology classes in M. On the other hand, codimension q submanifolds, which must have no boundary but need not be compact, define q-dimensional cohomology classes in M. Each 2 dimensional homology class u in the 4-manifold X can be represented by a 2-dimensional surface $\Sigma_u \subset X$ and we now describe how to represent the cohomology class $\mu(u) \in H^2(\mathcal{M}_k)$ by a codimension 2 submanifold $V_u \subset \mathcal{M}_k$ and how this submanifold V_u is related to Σ_u. The following result is one of the main techniques for computing with the map μ.

Theorem 3.2 *Let $\Sigma \subset X$ be a compact orientable surface with no boundary and let $u \in H_2(X)$ be the homology class represented by Σ. Let N_Σ be a sufficiently small tubular neighbourhood of Σ. Then, for $k \geq 1$, we can find a smooth codimension 2 submanifold $V_\Sigma^{(k)} \subset \mathcal{M}_k$ with the following properties:*

(1) *The submanifold $V_\Sigma^{(k)} \subset \mathcal{M}_k$ represents the cohomology class $\mu(u) \in H^2(\mathcal{M}_k)$.*

(2) *Given surfaces $\Sigma_1, \ldots, \Sigma_r \subset X$ in general position, the submanifolds $V_{\Sigma_i}^{(k)} \subset \mathcal{M}_k$ are in general position.*

(3) *Let $[A_\alpha]$ be a sequence of connections in $V_\Sigma^{(k)}$ which converges to an ideal connection $([A]; \{x_1, \ldots, x_l\})$. Then either one of the points x_i must lie in the tubular neighbourhood N_Σ or the connection $[A]$ lies in $V_\Sigma^{(k-l)} \subset \mathcal{M}_{k-l}$.*

This result is proved in [4] and [6]. Note how it relates the homomorphism μ to Uhlenbeck's weak compactness theorem, Theorem (2.4). We now go on to outline the proof of Theorem (1.6) using Theorem (3.2) to do computations with the homomorphism μ.

4. Even intersection forms

Let X be a smooth, simply connected 4-manifold with even intersection form Q_X. We now describe how to prove the following results.

(4.1) If Q_X is definite then $H_2(X) = 0$.

(4.2) If Q_X is indefinite and $b^+ = 1$; then

$$Q_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

(4.3) If Q_X is indefinite and $b^+ = 2$; then

$$Q_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Notice how (4.1) is implied by Donaldson's theorem (1.5) on definite intersection forms; the form is definite so by (1.5) it must be diagonal, however it is even and there are no non-trivial even diagonal forms. Note also that (4.2) and (4.3) are restatements of Theorem (1.6).

Proof of (4.1) Our assumptions are that Q_X is definite and even. By changing the orientation of X if necessary we can assume that the intersection form of X is negative definite and so $b^+ = 0$. Then, for a generic metric on X, there are no irreducible ASD connections and so the moduli space $\mathcal{M}_1$ is a smooth 5-dimensional manifold.

Pick two surfaces $\Sigma_1, \Sigma_2 \subset X$ in general position which represent homology classes $u_1, u_2 \in H_2(X)$. Pick suitably small tubular neighbourhoods N_i of the surfaces Σ_i. Now by Theorem (3.2) we can find codimension 2 submanifolds

$V_1, V_2 \subset \mathcal{M}_1$ which represent the classes $\mu(u_1), \mu(u_2) \in H^2(\mathcal{M}_1)$ and are in general position. Let

$$L = V_1 \cap V_2$$

so, since V_1 and V_2 are both 3-dimensional submanifolds of a 5-dimensional manifold, it follows that L has dimension 1. Now we count the number of ends of L.

Recall the definition of an end of a topological space Y. Intuitively the ends of Y are the components of $Y \setminus C$ where C is a sufficiently large compact set. The precise definition is as follows. If C, D are compact sets with $D \subset C$ we get an inclusion

$$Y \setminus C \subset Y \setminus D$$

and this inclusion induces a map

$$Y \setminus C \to Y \setminus D.$$

The number of ends of Y is the inverse limit

$$\varprojlim \pi_0(Y \setminus C)$$

and an end of Y is a component of the topological space

$$\varprojlim Y \setminus C.$$

If we take a sequence $[A_\alpha]$ of connections in L which converges to an ideal connection then, since $k = 1$, the only possibility is that it converges to the ideal connection given by the trivial connection on the product bundle and a single point in X. In view of part (3) of Theorem (3.2) this point must lie in $N_1 \cap N_2$. Now a direct construction proves the following lemma.

Lemma 4.4 *There is precisely one end of L for each component of* $N_1 \cap N_2$.

To prove this lemma, more generally to analyse the ends of the moduli spaces $\mathcal{M}_k$, it is necessary to use the glueing construction due to Taubes. We will not go into this construction in detail, see Taubes's paper [23] and the basic references [6, 7], and [13] for details. The proof of the above Lemma (4.4) is given in [4] and [6].

Let us now complete the proof of (4.1). The surfaces $\Sigma_1, \Sigma_2 \subset X$ are in general position so they meet in a finite number of points. Since Σ_i represents $u_i \in H^2(X)$ it follows that

$$Q_X(u_1, u_2) = |\Sigma_1 \cap \Sigma_2| \mod 2$$

where $|\Sigma_1 \cap \Sigma_2|$ is the number of points in the finite set $\Sigma_1 \cap \Sigma_2$. The neighbourhoods N_1 and N_2 can be chosen small enough so that the number of components of $N_1 \cap N_2$ is equal to the number of points of intersection of Σ_1 and Σ_2. The number of components of $N_1 \cap N_2$ is equal to the number of ends of L and since L is 1-dimensional it must have an even number of ends. Putting these facts together leads to the following conclusion: for all $u_1, u_2 \in H^2(X)$

$$Q_X(u_1, u_2) = 0 \mod 2.$$

Notice that the assumption is that Q_X is even, that is

$$Q_X(u, u) = 0 \bmod 2, \qquad \text{for all } u \in H^2(X),$$

and the conclusion is that

$$Q_X(u_1, u_2) = 0 \bmod 2, \qquad \text{for all } u_1, u_2 \in H^2(X).$$

Now suppose that $H_2(X) \neq 0$ and pick a non-zero $u \in H_2(X)$. Then since Q_X is unimodular there must exist another element $v \in H_2(X)$ such that $Q_X(u, v) = 1$. But we have just established that $Q_X(u, v)$ is even and this contradiction shows that $H_2(X) = 0$.

Notice how the above argument contains three main steps:

(1) Use the given information about Q_X to determine the structure of the moduli space.

(2) Now look at the intersection L of codimension 2 submanifolds of the form V_Σ and count the number of ends of L geometrically.

(3) Finally count the number of ends of L algebraically.

There is a certain amount of fine tuning involved in choosing which moduli space to use and the number of codimension 2 submanifolds. We now outline how to prove (4.2) by repeating the above steps.

Proof of (4.2) We are now assuming that Q_X is even and indefinite, and $b^+ = 1$. In this case we use $\mathcal{M}_2$, which is a smooth manifold of dimension 10, and consider the intersections of codimension 2 submanifolds V_Σ. The contradiction comes from looking at four surfaces $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4 \subset X$ in general position and the corresponding codimension 2 submanifolds $V_1, V_2, V_3, V_4 \subset \mathcal{M}_2$ in general position. So we analyse the ends of

$$L = V_1 \cap V_2 \cap V_3 \cap V_4.$$

Let $[A_\alpha]$ be a sequence of connections in L which converges to an ideal connection. Since $k = 2$ there are two possibilities to consider:

(1) The limit ideal ASD connection is of the form $([A]; \{x\})$ with $[A] \in \mathcal{M}_1$ and $x \in X$.

(2) The limit is the product connection on the trivial bundle and a set two points $x, y \in X$.

We now use part (3) of Theorem (3.2) to show that the first case cannot happen. Since the surfaces Σ_i are in general position no three of them intersect and we can assume that the tubular neighbourhoods N_i are chosen sufficiently small so that no three of the N_i intersect. Thus the point x can lie in at most two of the N_i. For convenience let us suppose that x does not lie in N_3 nor in N_4. Now part (3) Theorem (3.2) shows that, using the obvious notation, the connection $[A]$ must lie in

$$V_3^{(1)} \cap V_4^{(1)} \subset \mathcal{M}_1.$$

But now we count dimensions; the dimension of $\mathcal{M}_1$ is 2 and so $V_3^{(1)}$ and $V_4^{(1)}$ are codimension 2 submanifolds of a 2-dimensional manifold which are in general position. Therefore

$$V_3^{(1)} \cap V_4^{(1)} = \emptyset$$

and so the first possibility cannot happen.

Thus the sequence $[A_\alpha]$ must converge to a pair of points $x, y \in X$. Where can the points x, y lie? Since the surfaces are in general position no three of them intersect. We can suppose the neighbourhoods N_i are chosen small enough so that no three of them intersect and, for $i \neq j$, the number of components of $N_i \cap N_j$ is the same as the number of points of intersection of Σ_i and Σ_j. In this case Theorem (3.2) shows that each of the N_i must contain one of the points and we have just shown that the intersection of any three of the N_i must be empty. We can assume, by interchanging x and y if necessary, that $x \in N_1$ and then one of the following possibilities must hold:

(1) $x \in N_1 \cap N_2$, $y \in N_3 \cap N_4$

(2) $x \in N_1 \cap N_3$, $y \in N_2 \cap N_4$

(3) $x \in N_1 \cap N_4$, $y \in N_2 \cap N_3$.

Note that from the choice of the neighbourhoods N_i it follows that there are exactly

$$|\Sigma_1 \cap \Sigma_2|\ |\Sigma_3 \cap \Sigma_4| + |\Sigma_1 \cap \Sigma_3|\ |\Sigma_2 \cap \Sigma_4| + |\Sigma_1 \cap \Sigma_4|\ |\Sigma_2 \cap \Sigma_3|$$

such possibilities. Next we must prove that each of these possibilities does in fact occur and so we must analyse the ends of L geometrically. The following lemma is the analogue of Lemma (4.4) in the present situation and the proof is very similar; see [6].

Lemma 4.5 *There is precisely one end of L for each (unordered) pair $\{C, D\}$ where C is a component of $N_i \cap N_j$, D is a component of $N_k \cap N_l$, and $\{i, j, k, l\} = \{1, 2, 3, 4\}$. Furthermore there is a compact set $K \subset L$ and a homeomorphism*

$$L \setminus K \to (0, 1) \times \coprod \Lambda_{C,D}$$

where the disjoint union is taken over all (unordered) pairs $\{C, D\}$ and each $\Lambda_{C,D}$ is a compact 1-manifold.

Now we need a result, proved in [6], which allows us to compute the number of ends of L homologically.

Lemma 4.6 *There is a cohomology class $w_1 \in H^1(\mathcal{M}_2; \mathbf{Z}/2)$ such that*

$$\langle w_1, [\Lambda_{C,D}] \rangle = 1$$

where $\Lambda_{C,D}$ is as in Lemma (4.5) and $[\Lambda_{C,D}]$ is the homology class defined by the compact 1-manifold $\Lambda_{C,D}$.

Now using Lemma (4.5) we can truncate the space L by removing open cylinders around the ends $\Lambda_{C,D}$ to produce a compact 2-manifold N with boundary such that

$$\partial N = \coprod \Lambda_{C,D}.$$

Thus our geometric analysis shows two things:

(1) The number of ends of L is equal, modulo 2, to

$$Q_X(u_1, u_2)Q_X(u_3, u_4) + Q_X(u_1, u_3)Q_X(u_2, u_4) + Q_X(u_1, u_4)Q_X(u_2, u_3)$$

where $u_i \in H_2(X)$ is the homology class represented by the surface $\Sigma_i \subset X$.

(2) There is a cohomology class $w_1 \in H^1(\mathcal{M}_2; \mathbf{Z}/2)$ such that

$$\langle w_1, \partial N \rangle = \sum \langle w_1, [\Lambda_{C,D}] \rangle$$

and thus $\langle w_1, \partial N \rangle$ is equal, modulo 2, to the number of ends of L.

The first of these follows from Lemma (4.5) and the second from Lemma (4.6). But, necessarily,

$$\langle w_1, \partial N \rangle = 0$$

and so we conclude that

$$Q_X(u_1, u_2)Q_X(u_3, u_4) + Q_X(u_1, u_3)Q_X(u_2, u_4) + Q_X(u_1, u_4)Q_X(u_2, u_3) = 0 \bmod 2.$$

Now suppose that Q_X has rank $r > 2$. Our hypothesis is that Q_X is indefinite and even and it follows that we can find elements $u_1, u_2, u_3, u_4 \in$

$H_2(X)$ such that

$$Q_X(u_1, u_2) = Q_X(u_3, u_4) = 1 \bmod 2$$
$$Q_X(u_1, u_3) = Q_X(u_2, u_4) = Q_X(u_1, u_4) = Q_X(u_2, u_3) = 0 \bmod 2$$

One (rather crude) way to see this is to use the classification of even indefinite forms. Another, more direct, way is to work modulo 2 and prove that any non-singular symmetric bilinear form Q over $\mathbf{Z}/2$ must have even rank, say $2r$, over $\mathbf{Z}/2$ and we can choose a basis $\alpha_1, \ldots, \alpha_r$, $\beta_1, \ldots, \beta_r$ such that

$$\begin{aligned} Q(\alpha_i, \beta_j) &= \delta^i_j \\ Q(\alpha_i, \alpha_j) &= 0 \\ Q(\beta_i, \beta_j) &= 0 \end{aligned}$$

where δ^i_j is the Kronecker δ.

Thus if the rank of Q_X is different from 2 we have a contradiction and, since Q_X is even and indefinite it follows that

$$Q_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This proves the result in the case $b^+ = 1$.

Notice how the basic argument shows that if we had used five surfaces then, using obvious notation,

$$V_1 \cap V_2 \cap V_3 \cap V_4 \cap V_5$$

has no ends. Therefore this intersection is a compact 0-dimensional submanifold of $\mathcal{M}_2$ and so consists of a finite number of points. This fact leads to the definition of the Donaldson polynomials.

The proof of (4.3) is a similar; $b^+ = 2$ and we argue with the $k = 3$ moduli space and six codimension 2 submanifolds V_Σ. If we now try the argument in the case $b^+ = 3$ with the $k = 4$ moduli space and eight codimension 2 submanifolds V_Σ the argument breaks down. In this case, if we take a sequence of connections $[A_\alpha]$ in L, the intersection of the eight codimension 2 submanifolds, which converges to an ideal ASD connection we can no longer conclude that the only possibility is that the limiting ideal ASD connection consists of four points in X and the trivial flat connection on the product bundle. Of course the argument must break down because of the existence of the K3 surface K described in §1.

Donaldson points out in [4] that even though the argument breaks down in the case $b^+ = 3$ and we use eight codimension 2 submanifolds of $\mathcal{M}_4$ it is still possible to extract some information form this case. This suggestion is followed up by Ruan, [21], who shows that the argument gives a relation between Donaldson polynomials modulo 2.

5 The cohomology of mapping spaces

In Donaldson's use of gauge theory to prove theorems about 4-manifolds the map $\mu : H_2(X) \to H^2(\mathcal{B}_k)$ plays a critical role. In the proof of Theorem (1.6) outlined in the previous section we have seen the main techniques Donaldson uses to calculate with this map. Motivated by this we now begin to study this map and the spaces involved in more detail using techniques from algebraic topology. From the point of view of algebraic topology it is more convenient to have base point conditions and so we study the spaces

$$\mathcal{B}_k^0 \simeq \mathrm{Map}_k^0(X, \mathbf{HP}^\infty)$$

and the moduli spaces

$$\mathcal{M}_k^0 \subset \mathcal{B}_k^0 \simeq \mathrm{Map}_k^0(X, \mathbf{HP}^\infty).$$

For the rest of this section we use the notation

$$\mathrm{Map}^0(Y) = \mathrm{Map}^0(Y, \mathbf{HP}^\infty)$$

for the based mapping space.

To study $\mathrm{Map}^0(X)$ recall that up to homotopy

$$X \simeq W \cup_f e^4.$$

Here W is X with an open disc removed and therefore

$$W \simeq \bigvee_1^r S^2$$

where $r = \operatorname{rank} H_2(X;\mathbf{Z})$. The homotopy class of the attaching map

$$f : S^3 \to W \simeq \bigvee_1^r S^2$$

is determined by the following procedure. Pick a basis $\alpha_1, \ldots, \alpha_r$ for $H_2(X;\mathbf{Z})$; this determines a homotopy equivalence

$$\alpha : W \to \bigvee_1^r S^2.$$

Now recall that $\pi_3(\bigvee_1^r S^2)$ is the free abelian group generated by homotopy classes

$$\begin{aligned} &\eta_i, && 1 \leq i \leq r \\ &w_{ij}, && 1 \leq i < j \leq r. \end{aligned}$$

To describe these homotopy classes explicitly let $\iota_i : S^2 \to \bigvee_1^r S^2$ be the inclusion of the i-th factor and let $\eta : S^3 \to S^2$ be the Hopf map; then

$$\eta_i = \iota_i \circ \eta, \qquad w_{ij} = [\iota_i, \iota_j]$$

where $[,]$ is the Whitehead product. Now using the chosen basis $\alpha_1, \ldots, \alpha_r$ for $H_2(X;\mathbf{Z})$ we can represent the intersection form Q_X of X by a matrix of integers

$$(a_{ij}) = (Q_X(\alpha_i, \alpha_j))$$

and then

$$f \simeq \sum_{i=1}^r a_{ii}\eta_i + \sum_{1 \leq i < j \leq r} a_{ij} w_{ij}.$$

Note how we have essentially proved the Theorem (1.1).

Now consider the cofibration sequence

$$S^3 \to W \to X \to S^4.$$

Applying Map^0 to this sequence gives a fibration sequence

$$\mathrm{Map}^0(S^4) \to \mathrm{Map}^0(X) \to \mathrm{Map}^0(W) \to \mathrm{Map}^0(S^3);$$

The term fibration sequence means that any three consecutive maps form a fibration. Now

$$\mathrm{Map}^0(S^2) \simeq \Omega S^3$$

and using the homotopy equivalence $\alpha : W \to \bigvee_1^r S^2$ we get a map

$$\alpha : \mathrm{Map}^0(X) \to \mathrm{Map}^0(W) \simeq \prod_1^r \Omega S^3.$$

The cohomology of ΩS^3 is well-known;

$$\begin{aligned} H^{2n}(\Omega S^3, \mathbf{Z}) &= \mathbf{Z} \\ H^{2n+1}(\Omega S^3, \mathbf{Z}) &= 0. \end{aligned}$$

Write a_n for the generator of $H^{2n}(\Omega S^3, \mathbf{Z})$ and then products are given by

$$a_n a_m = \frac{(n+m)!}{n!\, m!} a_{n+m}.$$

Thus $H^*(\Omega S^3; \mathbf{Z})$ is a divided power algebra. More generally

$$H^*(\prod_1^r \Omega S^3; \mathbf{Z}) = \Gamma(H^2(X; \mathbf{Z}))$$

is the divided power algebra generated by $H^2(X; \mathbf{Z})$ and we identify $H^2(X; \mathbf{Z})$ with a subspace of $H^2(\prod \Omega S^3; \mathbf{Z})$ using this isomorphism. Now we get a homomorphism

$$H^2(X; \mathbf{Z}) \subset H^2(\prod_1^r \Omega S^3; \mathbf{Z}) \xrightarrow{\alpha^*} H^2(\mathrm{Map}^0(X); \mathbf{Z}).$$

It is straighforward to check that the composite

$$H_2(X; \mathbf{Z}) \cong H^2(X; \mathbf{Z}) \to H^2(\mathrm{Map}^0(X); \mathbf{Z})$$

is μ^0 where the first isomorphism is given by Poincaré duality.

The cohomology of $\mathrm{Map}^0(X)$ has been studied first by Masbaum [15] and later by Mielke [16] using standard spectral sequences. Masbaum uses the Serre spectral sequence of the fibration

$$\mathrm{Map}^0(S^4) \to \mathrm{Map}^0(X) \to \mathrm{Map}^0(W)$$

and Mielke studies the Eilenberg-Moore spectral sequence of the fibration

$$\mathrm{Map}^0(X) \to \mathrm{Map}^0(W) \to \mathrm{Map}^0(S^3).$$

Masbaum gets results for any X and cohomology with coefficients in $\mathbf{Z}/p$ where p is an odd prime. In the case of $\mathbf{Z}/2$-cohomology he gets results provided the intersection form of X is even. By using the Eilenberg-Moore spectral sequence Mielke is able to complete the computations of the cohomology of $\mathrm{Map}^0(X)$ by computing $\mathbf{Z}/2$-cohomology in the case where the intersection form of X is odd. For the precise statements of the general results see [15], [16].

The case which occurs in §4 is when X is a spin manifold and cohomology has $\mathbf{Z}/2$-coefficients. In this case the Serre spectral sequence of the fibration

$$\mathrm{Map}^0(S^4) \to \mathrm{Map}^0(X) \to \mathrm{Map}^0(W) \simeq \prod_1^r \Omega S^3$$

collapses at the E_2 term and there is an isomorphism

$$H^*(\mathrm{Map}^0(X); \mathbf{Z}/2) = H^*(\Omega^3 S^3; \mathbf{Z}/2) \otimes \bigotimes_{i=1}^{r} H^*(\Omega S^3; \mathbf{Z}/2).$$

Now suppose X is a smooth spin manifold and let $\mathcal{M}_k^0$ be the moduli space of ASD connections on X. Instead of compactifying $\mathcal{M}_k^0$ as in §4 we form the truncated moduli space $\mathcal{N}_k^0$ given by removing a small neighbourhood of the ends of $\mathcal{M}_k^0$. Since it is a subspace of $\mathcal{B}_k^0$ we get a map

$$i_k : \mathcal{N}_k^0 \to \mathrm{Map}^0(X).$$

using the equivalence of $\mathcal{B}_k^0$ with the mapping space. The truncated moduli space $\mathcal{N}_k^0$ is a smooth manifold with boundary and we can consider the homology class

$$i_*[\partial \mathcal{N}_k] \in H^*(\mathrm{Map}(X)).$$

This homology class is zero since it is obviously a boundary. However in favourable circumstances we can explicitly compute this homology class, or at least enough of it, to get non-trivial conclusions from the fact that it is zero.

A systematic homotopy theoretic analysis of the homology classes $i_*[\partial\mathcal{N}_k]$ is work in progress and more details will be given in future publications. For the moment however we can at least rephrase the the results of the computations of the previous section in these terms.

First consider the case where X is spin and has definite intersection form. Then $\mathcal{N}_1^0$ has dimension 8 and its boundary has dimension 7. Then it can be proved that there is a class $y_3 \in H^3(\Omega^3 S^3; \mathbf{Z}/2)$ such that

$$\langle i_1^*(y_3 a_1 a_2), \partial\mathcal{N}_1^0\rangle = Q_X(a_1, a_2) \bmod 2$$

for $a_1, a_2 \in H^2(X; \mathbf{Z}/2)$. Here we are using the above isomorphism

$$H^*(\mathrm{Map}^0(X); \mathbf{Z}/2) = H^*(\Omega^3 S^3; \mathbf{Z}/2) \otimes \bigotimes_{i=1}^{r} H^*(\Omega S^3; \mathbf{Z}/2).$$

and the identifcation of $H^2(X; \mathbf{Z}/2)$ with a subspace of

$$H^2(\prod \Omega S^3; \mathbf{Z}/2) = \bigotimes H^*(\Omega S^3; \mathbf{Z}/2).$$

Notice how the fact that $[\partial\mathcal{N}_1^0] = 0$ shows that is must follow that $H_2(X; \mathbf{Z}/2) = 0$.

Now consider the case where X is spin, its intersection form is indefinite, and $b^+ = 1$. In this case $\mathcal{N}_2^0$ has dimension 13 and its boundary has dimension 12. This time we can find a 4-dimensional class $y_4 \in H^4(\Omega^3 S^3; \mathbf{Z}/2)$ such that

$$\langle i_1^*(y_4 a_1 a_2 a_3 a_4), \partial\mathcal{N}_1^0\rangle = Q_X^{(2)}(a_1, a_2, a_3, a_4) \bmod 2$$

where $a_1, a_2, a_3, a_4 \in H^2(X; \mathbf{Z}/2)$. In this formula

$$\begin{aligned} Q_X^{(2)}(a_1, a_2, a_3, a_4) &= Q_X(a_1 a_2) Q_X(a_3 a_4) + Q_X(a_1, a_3) Q_X(a_2, a_4) \\ &+ Q_X(a_1, a_4) Q_X(a_2, a_3). \end{aligned}$$

More generally it is possible to push these computations to get some information in the case where X is spin, its intersection form is indefinite, $b^+ = k-1$ and we consider the truncated moduli space $\mathcal{N}_k^0$ which has dimension $5k+3$. In this case there is a class $y_{k+3} \in H^{k+3}(\Omega^3 S^3; \mathbf{Z}/2)$ such that, modulo 2

$$\langle i_k^*(y_{k+3} a_1 \ldots a_2 k), \partial\mathcal{N}_k^0\rangle = \langle Q_X^{(k)}(a_1, a_2, a_3, a_4)(a_1, \ldots, a_{2k}), [X]\rangle + \text{other terms}.$$

Here $Q_X^{(k)}$ is the obvious extension of $Q_X^{(2)}$. This is of course related to the fact that the argument outlined in §4 allows us to calculate the number of ends of the intersection L of $2k$ codimension 2 submanifolds in $\mathcal{M}_k$ which correspond to sequences of connections converging to an ideal ASD connection of the form $([A], \{x_1, \ldots, x_{2k}\})$, where A is the product connection on the trivial bundle. In the cases $k = 1, 2, 3$ these are the only ends of L and there are no other terms. In the general case there must be other ends and correspondingly there must be other terms. In [21] Ruan shows how these extra terms can be used to find relations modulo 2 between Donaldson polynomials. In general it seems that the extra terms will involve $Q_X^{(j)}$ for $j < k$ and Donaldson polynomials but the full formula is not known.

Since we are discussing the relation between the homology of the space $\mathrm{Map}^0(X)$ and the homology of the moduli space it is worth mentioning the conjecture from [1] which has come to be known as the Atiyah-Jones conjecture.

Conjecture. *For any closed 4-manifold X the map*

$$i_k : \mathcal{M}_k^0 \to \mathrm{Map}_k^0(X)$$

induces an isomorphism in homology

$$(i_k)_* : H_q(\mathcal{M}_k; \mathbf{Z}) \to H_q(\mathrm{Map}_k^0(X); \mathbf{Z})$$

provided $q < q(k)$ where $q(k) \to \infty$ as $k \to \infty$.

A limiting form of this conjecture is proved in [24]. The full conjecture has been recently proved by Boyer, Hurtubise, Mann, and Milgram [2] in the case where $X = S^4$. There is some hope of extending their methods to 4-manifolds like $S^2 \times T_g$ where T_g is a Riemann surface of genus g; more generally to ruled algebraic surfaces. It is clear that these results on the homology of mapping spaces have some genuine relevance to this conjecture.

References

[1] M.F. Atiyah and J.D.S. Jones , *Topological aspects of Yang-Mills theory,* Communications in Mathematical Physics 61 (1978), 97–118.

[2] C.P. Boyer, J.C. Hurtubise, B.M. Mann, and R.J. Milgram, *The topology of instanton moduli spaces I: The Atiyah-Jones conjecture,* Preprint (1991) (to appear).

[3] S. K. Donaldson , *An application of gauge theory to four dimensional topology,* Journal of Differential Geometry 18 (1983), 279–315.

[4] S. K. Donaldson , *Connections, cohomology and the intersection forms of four manifolds,* Journal of Differential Geometry 24 (1986), 275–341.

[5] S. K. Donaldson , *Polynomial invariants for smooth 4-manifolds,* Topology 29 (1990), 257–315.

[6] S. K. Donaldson and P. B. Kronheimer , The geometry of four manifolds, Oxford University Press, Oxford, UK, 1990.

[7] D. S. Freed and K. K. Uhlenbeck , Instantons and four-manifolds, MSRI publications, Vol 1 , Springer- Verlag , New York , 1984.

[8] M. Freedman and R. Kirby , *A geometric proof of Rohlin's theorem* , Proc. Symp. Pure. Math. XXXII, part 2 , Amer. Math. Soc. , Providence R.I. (1978), 85–98.

[10] M. H. Freedman , *The topology of four-dimensional manifolds* , Journal of Differential Geometry 17 (1982), 357–454.

[9] R. Friedman and J. Morgan , *Algebraic surfaces and 4-manifolds: some conjectures and speculations* , Bulletin of the American Mathematical Society 18 (1988), 1–19.

[11] R. Gompf , *Three exotic* $\mathbf{R}^4$*s and other anomalies* , Journal of Differential Geometry 18 (1983), 317–328.

[12] R. Gompf , *An infinite set of exotic* $\mathbf{R}^4$*s* , Journal of Differential Geometry 21 (1985), 283–300.

[13] H. B. Lawson , The theory of gauge fields in four dimensions , CBMS Regional Conference Series in Mathematics , American Mathematical Society , Providence, RI 1985.

[14] R. Mandelbaum , *Four dimensional topology: an introduction* , Bulletin of the American Mathematical Society 2 (1980), 1–159.

[15] G. Masbaum , *On the cohomology of the classifying space of the gauge group over some 4-complexes* , Bulletin de la Société Mathématique de France 119 (1991), 1–31.

[16] T. Mielke , *The cohomlogy of mapping spaces as they arise in gauge theory* , M.Sc. Dissertation, University of Warwick (1990).

[17] J. Milnor , On simply connected 4-manifolds , Symposium Internacional de Topologia Algebraica , Mexico, (1958), 122–128.

[18] J. Milnor and D. Husemoller , Symmetric bilinear forms , Springer-Verlag , Berlin 1973.

[19] T. Petrie and J. Randall , Connections, definite forms, and four-manifolds, Oxford University Press, Oxford, UK 1990.

[20] V. A. Rohlin , *New results in the theory of four dimensional manifolds,* Dok. Akad. Nauk. USSR 84 (1952), 221–224.

[21] Yongbin Ruan , *A universal constraint for Donaldson invariants* , Preprint, UC-Berkeley (1991) (to appear).

[22] J.-P. Serre, A course in arithmetic , Springer-Verlag , Berlin 1973.

[23] C.H. Taubes , *Self-dual connections over non-self-dual four manifolds,* Journal of Differential Geometry 17 (1982), 139–170.

[24] C.H. Taubes , *Gauge theory on asymptotically periodic 4-manifolds* , Journal of Differential Geometry 25 (1986), 363–430.

[25] C.H. Taubes , *The stable topology of self-dual moduli spaces* , Journal of Differential Geometry 29 (1989), 163–230.

Mathematics Institute,
University of Warwick,
Coventry CV4 7AL, England
e-mail: jdsj@maths.warwick.ac.uk