NIGEL J. HITCHIN

The Yang-Mills equations and the topology of 4-manifolds

Astérisque, tome 105-106 (1983), Séminaire Bourbaki, exp. n° 606, p. 167-178

§ 1. The result

(1.1) THEOREM (S.K. Donaldson [8]).— Let $X$ be a compact, smooth, simply connected, oriented $4$-manifold such that the intersection form $Q$ on $H^2(X,\mathbb{Z})$ is positive definite. Then there exists an integral basis for $H^2(X,\mathbb{Z})$ such that

$$Q(u,u) = u_1^2 + u_2^2 + \ldots + u_r^2.$$ 

This theorem should be contrasted with

(1.2) THEOREM (M.H. Freedman [9]).— Let $Q$ be any unimodular quadratic form over $\mathbb{Z}$. Then there exists a compact, simply connected, topological $4$-manifold $X$ such that $Q$ is equivalent to the intersection form on $H^2(X,\mathbb{Z})$.

There are sufficient examples of definite unimodular forms (see [17]) to see that Donaldson's theorem imposes strong restrictions on smooth $4$-manifolds.

(1.3) Proof of Theorem (1.1)

Let $r = \text{rank } H^2(X,\mathbb{Z})$ and $2n = \# \{u \in H^2(X,\mathbb{Z}) \mid Q(u,u) = 1\}$. The proof consists of constructing (as in §3–§7) an oriented cobordism between $X$ and $n$ copies of $\mathbb{CP}^2$. Let $p$ of these have the canonical orientation of the complex structure and $q = n - p$ the opposite orientation. Then

(i) By the cobordism invariance of signature,

$$r = \text{Sign } X = (p - q) \text{ Sign } \mathbb{CP}^2 = p - q \leq n.$$ 

(ii) Let $\{*x_1,*x_2,*\ldots,*x_n\} = \{u \in H^2(X,\mathbb{Z}) \mid Q(u,u) = 1\}$, then $Q(x_i,x_j) \in \mathbb{Z}$ but by the Cauchy-Schwarz inequality $|Q(x_i,x_j)| < 1$ if $i \neq j$. Hence $\{x_1,*\ldots,x_n\}$ is orthonormal and $n \leq r$.

(iii) From (1) and (ii) $n = r$ and $\{x_1,*\ldots,x_n\}$ is an orthonormal basis for $H^2(X,\mathbb{R})$. Thus for $u \in H^2(X,\mathbb{Z})$, $u = \sum_{i=1}^n Q(u,x_i)x_i = \sum_{i=1}^n u_ix_i$ with $u_i \in \mathbb{Z}$ and $\{x_1,*\ldots,x_n\}$ is a basis for $H^2(X,\mathbb{Z})$. Hence $Q(u,u) = \sum_{i=1}^n u_i^2$.
§ 2. Background

(2.1) Let $X$ be an oriented riemannian 4-manifold. A 2-form $\alpha \in \Omega^2$ is said to be self-dual (resp. anti-self-dual) if $\ast \alpha = \alpha$ (resp. $\ast \alpha = -\alpha$) where $\ast : \Omega^2 \to \Omega^2$ denotes the Hodge star operator.

Let $G$ be a compact Lie group and $P$ a principal $G$-bundle over $X$. A connection $A$ on $P$ has curvature $F(A) \in \Omega^2(g)$ where $g$ denotes the vector bundle associated to $P$ by the adjoint representation. For any bundle $V$ associated to $P$, a connection $A$ defines a differential operator $d_A : \Omega^p(V) \to \Omega^{p+1}(V)$. The metric on $X$ defines the formal adjoint $d_A^* : \Omega^{p+1}(V) \to \Omega^p(V)$. The Bianchi identity, satisfied by all connections, is $d_A F(A) = 0$. The Yang-Mills equations are $D_A F(A) = 0$.

A connection $A$ on $P$ is said to be self-dual if $F(A) = \ast F(A)$. In this case $d_A^* F(A) = *d_A *F(A) = *d_A F(A) = 0$ by the Bianchi identity, so a self-dual connection automatically satisfies the Yang-Mills equations.

The Yang-Mills equations describe the critical points for the Yang-Mills functional (or action).

$$\|F(A)\|_{L^2}^2 = \int_X |F(A)|^2 d\mu.$$  

The self-dual connections give the absolute minimum for compact $X$ which, if $G = SU(2)$, may be expressed via the Chern-Weil theorem as $-8\pi^2 c_2(P)$ where $c_2(P)$ is the 2nd Chern class of the associated rank 2 vector bundle.

The Yang-Mills functional and Yang-Mills equations are invariant under (i) conformal changes of the metric on $X$ (ii) automorphisms of the principal bundle $P$ ("gauge transformations").

(2.2) The initial mathematical development of the study of self-dual connections, motivated by the interest of mathematical physicists, concentrated on the case $X = S^4$ and an explicit description of all solutions was possible [2] using the twistor approach of R. Penrose and R.S. Ward [6] which converted the problem into one of holomorphic bundles on $\mathbb{C}P^3$.

More recently the self-duality equations have been studied on more general 4-manifolds. There are three major lines of thought which have spurred this progress:

(2.3) If $X$ is a Kähler manifold, the space of anti-self-dual 2-forms $\Omega^2 = \Omega^1_{\omega,1}$, the space of primitive 2-forms of type $(1,1)$. A vector bundle with an anti-self-dual connection is then automatically endowed with a holomorphic structure (see [3]) and is moreover stable in the sense of Mumford and Takemoto (see [8], [11]). Converse results have been conjectured and in some cases proved ([13], [8]).

(2.4) The analysis of self-dual connections has been pushed forward by the funda-
mental results of K.K. Uhlenbeck ([20], [21]). Amongst these is the following removable singularity theorem: If $A$ is an $SU(2)$ connection (on the trivial bundle) over the punctured ball $B^4 \setminus \{0\}$, self-dual with respect to some smooth Riemannian metric on $B^4$ and with finite action; then there is a bundle automorphism $g : B^4 \setminus \{0\} \to SU(2)$ such that $g(A)$ extends smoothly over $B^4$.

(2.5) The existence of self-dual connections is assured under very general circumstances by a theorem of C.H. Taubes [19]: Let $X$ be a compact, oriented, Riemannian 4-manifold with positive definite intersection form $Q$, and let $P$ be a principal $SU(2)$ bundle over $X$ with $c_2(P) \leq 0$. Then $P$ admits an irreducible self-dual connection. Taubes' construction makes use of an implicit function theorem which involves $L^p$ estimates on curvature. It should be noted that anti-self-dual harmonic 2-forms may certainly obstruct the existence of self-dual connections, as can be seen by considering $\mathbb{CP}^2$ with opposite orientation. There are no stable rank 2 bundles on $\mathbb{CP}^2$ with $c_2(P) = 1$ [16] and hence by (2.3) no anti-self-dual connections. Taubes' hypotheses and result are the starting point for Donaldson's theorem.

(2.6) As an example of a self-dual connection, take $X = \mathbb{R}^4$ and $G = SU(2)$. Then in terms of a quaternionic coordinate $x \in \mathbb{H} = \mathbb{R}^4$ and using the isomorphism $SU(2) \cong \text{Im} \mathbb{H}$ the 1-instanton [7] solution of the self-duality equations is given by

$$A_\lambda = \text{Im}\left(\frac{xd\overline{x}}{\lambda^2 + |x|^2}\right) \quad \text{with} \quad F(A_\lambda) = \frac{\lambda^2 dx \wedge d\overline{x}}{(\lambda^2 + |x|^2)^2}$$

and action $8\pi^2$.

(2.7) PROPOSITION.—Let $A$ be a self-dual $SU(2)$ connection on $\mathbb{R}^4$ with action $8\pi^2$. Then up to a gauge transformation and a translation of $\mathbb{R}^4$, $A$ is equal to $A_\lambda$ for some $\lambda \in \mathbb{R}$.

Proof.—By conformal invariance and stereographic projection $A$ is defined on $S^4 \setminus \{x\}$, and by the removable singularity theorem is defined on a bundle $P \to S^4$. Now use [3] § 9 or [2] or [6].

§ 3. The moduli space

(3.1) The cobordism in the proof of (1.1) is modelled on a moduli space of self-dual connections whose general structure is described next.

Let $X$ be as in Theorem (1.1), and given a Riemannian metric. Let $P$ be a principal $SU(2)$ bundle over $X$ with $c_2(P) = -1$. Using the covariant derivative of a fixed smooth connection $A_0$ on $P$, one may define Sobolev spaces $L^p_q(V)$ of sections of any associated vector bundle $V$.

Let $\mathcal{E}$ denote the affine space of connections on $P$ differing from $A_0$ by
an element of $L^2_2(\Omega^1(g))$, and let $\mathcal{G}$ denote the group of $L^2_4$ sections of $\mathcal{P}$ (c End $V$ for some faithful representation). Then $\mathcal{G}$ is a Banach Lie group of gauge transformations acting smoothly on $\mathcal{G}$ by $g(A) = A - (d_A g) g^{-1}$. Let $\mathcal{B}$ denote the quotient space with projection $p: \mathcal{N} \to \mathcal{B}$, and $p(A) = [A]$.

(3.2) Recall that a connection on $P$ is reducible if its holonomy group is a proper subgroup of $SU(2)$. Since $X$ is simply-connected and $P$ is topologically non-trivial, the only possible reduction is to $U(1) \subset SU(2)$. Let $\Gamma_A \subset \mathcal{G}$ denote the subgroup of covariant constant sections with respect to the connection $A$. Then $A$ is reducible iff $\Gamma_A \cong U(1)$. The equivalence classes of irreducible connections form an open subset $\mathcal{B}^* \subset \mathcal{B}$.

(3.3) PROPOSITION. (i) $\mathcal{B}$ is a Hausdorff space in the quotient topology.
(ii) $\mathcal{B}^*$ is a Banach manifold with charts constructed from the slices $T_A, \epsilon = \{A + a | d_A a = 0, \|a\|_{L^2_3} < \epsilon\}$ of the action of $\mathcal{G}$.
(iii) $p: p^{-1}(\mathcal{B}^*) \to \mathcal{B}^*$ is a principal $\mathcal{G}/\mathbb{Z}$ bundle with a connection defined by the slices.
(iv) If $A$ is reducible, $\Gamma_A$ acts on $T_A, \epsilon$ and the map $T_A, \epsilon / \Gamma_A \to \mathcal{B}^*$ is a homeomorphism to a neighbourhood of $[A] \in \mathcal{B}$, smooth away from the fixed point set.

Proof. Standard methods (see [3], [12], [14]) using Banach space inverse and implicit function theorems.

(3.4) Let $\mathcal{M} \subset \mathcal{B}$ denote the subspace of equivalence classes of self-dual connections on $P$. $\mathcal{M}$ is the moduli space. If $A \in \mathcal{B}$ is reduced to a connection on a principal $U(1)$ bundle $Q \subset P$, then (since $\pi_1(X) = 0$) its equivalence class is determined by its curvature $F(A) \in \Omega^2$. If $A$ is self-dual, $F(A)$ is a self-dual closed 2-form, hence harmonic. By Hodge theory $F(A)$ is determined by its cohomology class $2\pi ic_1(Q)$. The reduction to $U(1)$ is well-defined modulo the Weyl group, so $[A] \in \mathcal{M}$ is determined by $\pm c_1(Q)$. Since $c_2(P) = -c_1(Q)^2 = -1$ there are $n$ distinguished points in $\mathcal{M}$ representing the reducible self-dual connections, where $2n = \# \{u \in H^2(X,\mathbb{Z}) | Q(u,u) = 1\}$. From (2.5) there are also irreducible connections.

(3.5) If $A$ is a self-dual connection on $P$, then there exists an elliptic complex $[3]$

\[ \Omega^0(g) \xrightarrow{d_A} \Omega^1(g) \xrightarrow{d_A} \Omega^2(g) \]

where $d_A^*$ is the projection of $d_A$ onto the anti-self-dual 2-forms. Let $H^P_A (0 \leq p \leq 2)$ denote the associated harmonic spaces, then by the Atiyah-Singer index theorem (see [3])

\[ -\sum_{p=0}^2 (-1)^p \dim H^P_A = 8|c_2(P)| - \frac{3}{2}(\chi(X) - \text{Sign}(X)) = 5. \]
(3.6) PROPOSITION.—Let $A$ be a self-dual connection on $P$.

Then there exists a neighbourhood $U$ of $0 \in H^1_A$ and a smooth map $\phi : U \to H^2_A$ such that:

(i) if $A$ is irreducible, a neighbourhood of $[A] \in \mathcal{M}$ is diffeomorphic to

$$\phi^{-1}(0) \cong H^1_A .$$

(ii) if $A$ is reducible, a neighbourhood of $[A] \in \mathcal{M}$ is diffeomorphic to $\phi^{-1}(0)/\Gamma_A$.

Proof.—The connection $A + a$ is self-dual iff

$$\Phi(A + a) = \Phi(A) + d_A^{-}a + \frac{1}{2}[a, a] = 0 \in L^2_2(\Omega^2(g)) .$$

Restricted to a slice $T_{A(x)} \in$ the derivative $d\Phi_A$ of $\Phi$ at $A$ is the Fredholm operator

$$d_A^- : \ker d_{A} \subset L^2_2(\Omega^1(g)) \to L^2_2(\Omega^2(g)) ,$$

and so $\Phi$ is a Fredholm map ([11], [18]). After a local diffeomorphism $\Phi$ may be represented as $\Phi(x) = (d\Phi_A)_x + \phi(x)$.

The argument is analogous to the methods applied to moduli of complex structures [10].

(3.7) As a consequence of (3.5) and (3.6), if $A$ is irreducible and $H^2_A = 0$, then

$\mathcal{M}$ is a smooth 5-manifold in a neighbourhood of $[A]$. A particular case when this

holds for all irreducible $A$ is when the underlying metric on $X$ is self-dual

with positive scalar curvature (see [3]). Note that if $A$ is reducible, $\Gamma_A$ acts

on $H^1_A$ by complex multiplication ($b_1(X) = 0$) so that if $H^2_A = 0$, $H^1_A/\Gamma_A \cong \mathbb{C}^3/S^1$

from the index theorem and $\dim H^0_A = \dim \Gamma_A = 1$.

§ 4. A key result

(4.1) An important tool in understanding the global structure of the moduli space

is the following : (see also [15]).

(4.2) PROPOSITION.—Let $\mathcal{A}_i \in \mathcal{M}$ be a sequence of self-dual connections on $P$. Then

there is a subsequence such that either:

(i) each $\mathcal{A}_i$ is gauge equivalent to $A_1 \in \mathcal{M}$ converging in $C^\infty$ to a self-dual connection $A_\infty$ on $P$, and hence $[\mathcal{A}_i] \to [A_\infty] \in \mathcal{M}$.

or

(ii) there is a point $x \in X$ and trivializations $\rho_i$ of $\mathcal{P}_x$ on the complement $K$ of

any geodesic ball about $x$ such that $\rho_i^* \mathcal{A}_i \to \emptyset$ (the trivial flat connection) in

$C^\infty(K)$.

Proof.—The proof uses two lemmas:

(4.3) Lemma.—Given $L, C > 0$ let $\{f_1\}$ be a sequence of integrable functions on $X$

with $f_1 \geq 0$ and $\int_X f_1 \, du \leq L$. Then there exists a subsequence, a finite set

$\{x_1, \ldots, x_L\} \subset X$ and a countable collection $\{B_\alpha\}$ of geodesic balls in $X$

such that the half-sized balls cover $X\setminus\{x_1, \ldots, x_L\}$ and for each $\alpha$, lim sup $\int_{B_\alpha} f_1 \, du < C$.

Proof.—Elementary : the $x_i$'s are characterized by the property that each lies in
(4.4) Lemma.— Let $h_i$ be a sequence of metrics on $B^n$, sufficiently close to the Euclidean metric, and converging in $C^\infty(B^n)$ to $h_\infty$. Let $\tilde{A}_i$ be a sequence of connections on the trivial bundle over $B^n$ with $\tilde{A}_i$ self-dual with respect to $h_i$. Then there is a constant $C$ (independent of $h_i$ and $\tilde{A}_i$) such that if $\int_{B^n} |F(\tilde{A}_i)|^2 d\mu \leq C$, there is a subsequence such that $A_i$ (gauge equivalent to $\tilde{A}_i$) converges in $C^\infty(\frac{1}{2}B^n)$ to $A_\infty$, a connection which is self-dual with respect to $h_\infty$.

Proof.— Consequence of ([21] Theorem (1.3)).

(4.5) To obtain (4.2) first consider a geodesic coordinate system $\chi$ on a geodesic ball $B \subset X$ of radius $r$. Thus $\chi$ defines a diffeomorphism $\chi : B_r^n \to B$ from the Euclidean ball of radius $r$ to $B$. Pulling back the metric $h$, and putting it on the Euclidean unit ball by dilation gives a metric

$$h_r = \chi^* h(\chi x) = r^2 (\delta_{ij} + r^2 g_0(y^2)) dy_i dy_j .$$

Choose $r$ small enough that the metric $r^{-2}h_r$ on $B^n$ satisfies the condition for (4.4). By conformal invariance each $\tilde{A}_i$ is self-dual with respect to $h_r$.

Now in Lemma (4.3) take the constant $C$ from (4.4), $f_i = |F(\tilde{A}_i)|^2$ and $L = 8\pi^2$. Thus from (4.4) on each ball $B_\alpha$ some subsequence converges (after gauge transformations) to $A_\infty(\alpha)$. By a diagonal argument the convergence may be achieved simultaneously for all $\alpha$.

The gauge transformations introduced in the above process give rise to connection matrices $A_i(\alpha) \to A_\infty(\alpha)$ in $C^\infty(\frac{1}{2}B_\alpha)$ and transition functions $g_i(\alpha, \beta) : \frac{1}{2}B_\alpha \cap \frac{1}{2}B_\beta \to SU(2)$ satisfying:

(4.6) $A_i(\alpha) = -dg_i(\alpha, \beta)g_i(\alpha, \beta)^{-1} + g_i(\alpha, \beta)A_i(\beta)g_i(\alpha, \beta)^{-1} .$

The compactness of $SU(2)$ gives a uniform bound to $dg_i$ in (4.6) and so one can find a uniformly convergent subsequence. Repeatedly applying (4.6) gives convergence in $C^\infty$, and using a diagonal argument one obtains a subsequence $A_i(\alpha, g_i(\alpha, \beta)) \to (A_\infty(\alpha), g_\infty(\alpha, \beta))$ for all $(\alpha, \beta)$ simultaneously. This represents a self-dual connection on a bundle $Q$ over $X \setminus \{x_1, \ldots, x_\ell\}$. Furthermore, if $K \subset X \setminus \{x_1, \ldots, x_\ell\}$ is compact then by induction on the number of balls $\frac{1}{2}B_\alpha$ covering $K$ (see [21] Sect. 3) one obtains isomorphisms $\rho_1 : Q|K \to P|K$ such that $\rho_1^* : A_i \to A_\infty$ in $C^\infty(K)$.

(4.7) Let $B_j'$ be a small punctured ball centred on $x_j$ ($1 \leq j \leq \ell$). Since

$$\int_{B_j'} |F(\tilde{A}_i)|^2 d\mu \leq 8\pi^2,$$

by Fatou's lemma

$$\int_{B_j} |F(A_\infty)|^2 d\mu \leq 8\pi^2 .$$

Hence by the removable singularity theorem (2.4) the connection $A_\infty$ and bundle $Q$ extend over $X$.

By the definition of $x_j$, $\lim_{j \to \infty} \int_{B_j} |F(\tilde{A}_i)|^2 d\mu > \frac{1}{2}C$ for all balls $B_j$; hence for a sufficiently small ball

no ball with $\limsup_{B_j} f x d\mu \leq \frac{1}{2}C$.
On the other hand, since all connections are self-dual these integrands are Chern forms. They may therefore be evaluated mod. $8\pi^2\mathbb{Z}$ by boundary integrals (Chern-Simons invariants). Hence by uniform convergence on the boundary $\partial B_j$, 

$$\int_{B_j} |F(A_\infty)|^2 d\mu = \lim_{j \to \infty} \int_{B_j} |F(\tilde{A}_j)|^2 d\mu \quad \text{mod. } 8\pi^2\mathbb{Z}.$$ 

(4.9) However, since $\int_{B_j} |F(A_\infty)|^2 d\mu \geq 0$ and $\int_{B_j} |F(\tilde{A}_j)|^2 d\mu \leq 8\pi^2$ the only possibilities from (4.7) and (4.8) are:

(i) $\lambda = 0$ or

(ii) $\lim_{j \to \infty} \int_{B_j} |F(\tilde{A}_j)|^2 d\mu = 8\pi^2$ and $\int_X |F(A_\infty)|^2 d\mu < 8\pi^2$ and hence $Q$ is trivial and $A_\infty$ flat. Thus Proposition (4.2) follows.

(4.10) The proposition shows that a self-dual connection on $P$ can only degenerate by having its curvature concentrate in the neighbourhood of a point. An example is the instanton $A_\lambda$ in (2.6) as $\lambda \to 0$.

§ 5. The boundary of $\mathcal{M}$

(5.1) Let $\beta : \mathbb{R} \to \mathbb{R}$ be a bump function approximating and dominated by $\chi_{[-1,1]}$ and set $R_A(x,s) = \int_0^s \beta(d(x,y)/s) |F(A)|^2 d\mu$, where $d(x,y)$ is the geodesic distance in $X$. Then define

$$\lambda(A) = K^{-1}\min\{s \mid \exists x \text{ with } R_A(x,s) = 4\pi^2\}$$

where $K$ is chosen so that $\lambda(A_1) = 1$ for the instanton $A_1$. Donaldson introduces this convenient but ad hoc function as a measure of the concentration of curvature: if $\beta$ is replaced by $\chi_{[-1,1]}$ then $\lambda(A)$ becomes the radius of the smallest ball containing half the action. In any case a ball of radius $\lambda(A)$ contains more than half the action and hence any sequence $[A_i] \subset [A]$ without convergent subsequences has $\lambda(A_i) \to 0$ from (4.2). It is thus a measure of the distance from the boundary.

(5.3) PROPOSITION.—There exists $\lambda_0 > 0$ such that if $A$ is a self-dual connection on $P$ with $\lambda(A) < \lambda_0$, then the minimum in (5.2) is attained at a unique point $x(A) \in X$.

Proof.—Take a small geodesic ball of radius $r$ centred on a minimum $x$ for $A$, and pull back the metric and connection as in (4.5) to the Euclidean ball of radius $r/\lambda(A)$. For each sequence of connections with $\lambda(A_i) \to 0$, the pulled-back connections $\tilde{A}_i$ satisfy $\lambda(\tilde{A}_i) = 1$ by construction and applying (4.4) and (4.2) there is a subsequence converging to a self-dual connection on $R^n$. From the classification (2.7) and normalization this is the instanton $A_\lambda$. Since $\lambda(\tilde{A}_i) = 1$, from
every subsequence converges and since the limit is unique, \( \Lambda_1 \to \Lambda_1 \) as \( \lambda(\Lambda_1) \to 0 \). Now the function \( R_{\Lambda_1} \) has a unique non-degenerate minimum so for sufficiently small \( \lambda(\Lambda) \), so will \( R_{\Lambda} \). Any two minima for \( \Lambda \) must however be separated by a distance of at most \( 2\lambda(\Lambda) \), since the ball of radius \( \lambda(\Lambda) \) about each contains more than half the action, thus a unique minimum for \( R_{\Lambda} \) implies a unique one for \( R_{\Lambda_1} \).

Note how the connectedness of the moduli space for \( R^n \) is essential for this argument.

(5.4) Let \( \mathcal{M}_{\lambda_0} = \{[\Lambda] \in \mathcal{M} \mid \lambda(\Lambda) < \lambda_0 \} \), and define \( p : \mathcal{M}_{\lambda_0} \to X \times (0, \lambda_0) \) by \( p(\Lambda) = (x(\Lambda), \lambda(\Lambda)) \).

(5.5) **Proposition.**—(i) \( \mathcal{M}_{\lambda_0} \) is compact.
(ii) \( \mathcal{M}_{\lambda_0} \) is a smooth manifold.
(iii) \( p \) is a smooth covering map.

**Proof.**—(i) Immediate from Proposition (4.2).
(ii) As \( \lambda(\Lambda) \to 0 \), \([\Lambda] \to \varnothing \) in \( C^\infty(X, B(x(\Lambda), r)) \) from (4.2). Then using an argument of Taubes [19], \( H^2_\Lambda = 0 \). The result follows from (3.6).
(iii) \( p \) is smooth because the minimum of \( R_{\Lambda} \) is non-degenerate, and proper by (4.2). Thus one only needs to check that the derivative of \( p \) is an isomorphism. Taubes' implicit function theorem provides an inverse.

(5.6) **Proposition.**—\( p \) is a diffeomorphism.

**Proof.**—This is the most technical part of Donaldson's proof, and involves delicate curvature estimates. The idea is to show that any two self-dual connections \( \Lambda \), \( B \) with \( x(\Lambda) = x(B) \) and \( \lambda(\Lambda) = \lambda(B) \) sufficiently small may be joined by a short path in \( \mathcal{M} \) (see [8]).

§ 6. Perturbation of \( \mathcal{M} \)

(6.1) If \( H^2_\Lambda = 0 \) for all self-dual connections then \( \mathcal{M} \) is a smooth manifold except at the \( n(Q) \) points corresponding to the reducible connections. This may not be true in general and there may be a subset \( K \subset \mathcal{M} \) (compact from (5.5)) for which \( H^2_\Lambda \neq 0 \). A perturbation of \( \mathcal{M} \) is then necessary to obtain a manifold.

(6.2) Perturbation around the reducible connections is dealt with in a straightforward manner: the finite-dimensional map \( \phi(x) \) in the decomposition \( \Phi(x) = (D\Phi)(x + \phi(x)) \) is modified by a nearby map with surjective derivative. Then, as in (3.6) a neighbourhood of \([\Lambda]\) is diffeomorphic to \( \mathbb{E}^3/S^1 \) - a cone on \( \mathbb{E}P^2 \). One may assume, then, that \( K \subset \mathcal{M} \cap \mathbb{D}^* \).

(6.3) The group \( \mathbb{G}/\mathbb{Z} \) acts on the Banach spaces \( L^2_\mathbb{G}(\Omega^2(\mathbb{G})) \) and \( L^2_\mathbb{G}(\Omega^2(\mathbb{G})) \) and
associated to the principal $G/S^1$ bundle $p^{-1}(E)$ over $B^*$ one obtains vector bundles $\mathcal{E}_3 \subset \mathcal{E}_2$ with norms and connections. There is a canonical section $\Phi = F_\pi(A)$ of $\mathcal{E}_2$ and one seeks perturbations $\sigma \in \mathcal{C}_c^\infty(B^*, \mathcal{E}_3)$, such that $\Phi + \sigma$ vanishes non-degenerately.

(6.4) **PROPOSITION.**—There exists $\sigma \in \mathcal{C}_c^\infty(B^*, \mathcal{E}_3)$ supported in a neighbourhood of $K$, such that $(\Phi + \sigma)^{-1}(0)$ is a smooth 5-manifold.

**Proof.**—Covering $K$ with a finite number of slices $T_A \in \mathcal{E}$ and shrinking, take open sets $U_1$, $U_2$ with $K \subset U_1$ and $\overline{U}_1 \subset U_2$, and let $\sigma$ be a bounded section of $\mathcal{E}_3$ supported in $U_2$. Then $K = \{[A] \in \overline{U}_1 : \|\Phi(\sigma)(A)\|_{L^2} \leq R\}$ is compact. This follows from the fact that $U_2$ is covered by a finite number of slices and on each one $\Phi(A) = d_A^*a + \frac{1}{2}[a, a] + \sigma(A)$ with $d_A^*a = 0$ and $\|a\|_{L^2} < \varepsilon$. Thus $L^3_2$ bounds on $\sigma(A)$, $a$ and $(\Phi + \sigma)(A)$ give an $L^3_2$ bound on $(d_A^- + d_A^*)a$ and so by ellipticity an $L^4_2$ bound on $a$. Since $L^4_2 \subset L^3_2$ is compact the statement follows. Thus if $\Phi + \sigma$ vanishes non-degenerately in $\overline{U}_1$, so do nearby sections $\Phi + \sigma'$ in the topology of uniform convergence of $\sigma$ and its derivative on compact sets.

The space of such non-degenerate perturbations is also dense: at each point take a slice on which there is a decomposition $\Phi + \sigma = L + \phi$ where $L$ is linear and $\phi$ finite dimensional. By compactness, take a finite subcovering and modify $\Phi + \sigma$ by subtracting a regular value of $\phi$, extended by a bump function. By Sard’s theorem such perturbations can be made arbitrarily close in $L^2_3$ norm to $\Phi + \sigma$.

The section $\Phi$, itself vanishes non-degenerately outside $\overline{U}_1$. By the density argument choose a perturbation $\sigma$ sufficiently small that $\Phi + \sigma$ (by the openness argument on $U_2 \setminus \overline{U}_1$) vanishes non-degenerately on $U_2 \setminus \overline{U}_1$. Then $\Phi + \sigma$ is non-degenerate everywhere. Let $\mathcal{H}_\sigma = (\Phi + \sigma)^{-1}(0)$, a 5-manifold with $n$ quotient singularities $\mathcal{E}_3/S^1$ and boundary $X$.

§ 7. Orientability of $\mathcal{H}_\sigma$

(7.1) On the manifold $\mathcal{H}_\sigma \cap B^*$ one must consider the Stiefel-Whitney class $w_1(Ker V(\Phi + \sigma))$. The singular points can be avoided by using the gauge transformations $\mathcal{G}_0 \subset \mathcal{G}$ which are the identity at a fixed point $x_0 \in X$. These then act freely on $\mathcal{H}$ to give quotient $\mathcal{B} \xrightarrow{\pi} \mathcal{B}$. Over $B^*$, $\pi$ gives a principal $SO(3)$ bundle, so $\mathcal{H}_\sigma \cap B^*$ is orientable iff its pull back to $\pi^{-1}(\mathcal{H}_\sigma \cap B^*)$ is.

(7.2) The vector bundle $Ker V(\Phi + \sigma)$ restricted to any compact subset $Y \subset \pi^{-1}(\mathcal{H}_\sigma \cap B^*)$ defines an element of $KO(Y)$. This is the index class [5] of the family of Fredholm operators $d_A^* + d_A^- + (\nabla\theta)A$, which by considering the deformation $d_A^* + d_A^- + t(\nabla\theta)A$, $0 \leq t \leq 1$, is independent of $\sigma$. Since $w_1$ factors...
through KO, the orientability can be decided by considering $\text{ind}(d^*_A + d^-_A) \in KO(Y)$ where $Y$ is a loop. Since this is now defined for all equivalence classes of connections, the loop may be deformed in $\mathcal{B}$.

(7.3) If SU(2) is embedded in SU(3) in the standard way, the Lie algebra bundle $\tilde{g}$ of the associated SU(3) connection $\tilde{A}$ splits as $\tilde{g} = g \oplus \mathbb{R} \oplus V$ where $V$ is a complex rank 2 bundle and $\mathbb{R}$ a trivial bundle, all preserved by the connection. Thus $\omega_1(\text{ind}(d^*_A + d^-_A)) = \omega_1(\text{ind}(d^*_A + d^-_A))$ and so the loop may be deformed in the space $\mathcal{B}_3$ of equivalence classes of SU(3) connections.

(7.4) PROPOSITION. $\pi_1(\mathcal{B}_3) = 0$.

Proof. Since the group $\mathcal{G}_0$ of SU(3) gauge transformations preserving $x_0$ acts freely, $\pi_1(\mathcal{B}_3) \cong \pi_0(\mathcal{G}_0)$. The principal bundle $P$ is trivial on the complement of a point and in particular on the 2-skeleton of $X$. Since $\pi_2(SU(3)) = 0$ any element of $\mathcal{G}_0$ can be deformed to one which is the identity on the 2-skeleton. Collapsing the 2-skeleton of $X$ gives a sphere $S^4$. The homotopy type of $\mathcal{G}_0$ on $S^4$ is independent of $c_2(P)$ (see [4]), so the question reduces to the trivial bundle. But $\pi_4(SU(3)) = 0$, so $\mathcal{G}_0$ is connected.

Thus $\mathcal{M}_0 \cap \mathcal{B}^*$ is an oriented 5-manifold which, putting in the boundaries of the quotient singularities provides the cobordism of Theorem (1.1).

§ 8. Examples

(8.1) Let $X = S^4$, with the canonical metric. Then any self-dual connection on $P$ is gauge equivalent to $f^*A$ where $f : S^4 \longrightarrow \mathbb{HP}^1$ is a conformal map and $A$ is the canonical connection on the quaternionic Hopf bundle. Since isometries of $\mathbb{HP}^1$ preserve $A$, the moduli space is $SO(5,1)/SO(5) \cong$ hyperbolic 5-space. This is the ball $B^5$ with boundary $S^4$. There are many ways of proving this ([2], [3], [6]).

(8.2) Let $X = \mathbb{CP}^2$ with its canonical metric. In the non-compact component of $\mathcal{M}_0$, any connection is gauge equivalent to $f^*A$ where $f : \mathbb{CP}^2 \longrightarrow \mathbb{HP}^2$ is equivalent under the action of SU(3) on $\mathbb{CP}^2$ to a map of the form

$$(z, w) \longmapsto \left( z, \frac{a \overline{z} + w}{\sqrt{1 - a^2}} \right), \quad a \in [0, 1)$$

in affine coordinates. When $a = 0$ this is the standard embedding $\mathbb{CP}^2 \subset \mathbb{HP}^2$ and gives the reducible connection. The moduli space is a cone on $\mathbb{CP}^2$ where $a$ is essentially the distance from the vertex. This was proved by Donaldson (unpublished) using the algebraic geometry of the flag manifold $F_3$, and the Penrose/Ward approach.
§ 9. References


Nigel J. Hitchin
St. Catherine's College
GB - OXFORD OX1 3UJ