Mathematical Research Letters **1**, 809–822 (1994)

THE SEIBERG-WITTEN INVARIANTS AND SYMPLECTIC FORMS

Clifford Henry Taubes

Recently, Seiberg and Witten (see [SW1], [SW2], [W]) introduced a remarkable new equation which gives differential-topological invariants for a compact, oriented 4-manifold with a distinguished integral cohomology class. A brief mathematical description of these new invariants is given in the recent preprint [KM].

My purpose here is to prove the following theorem:

Main Theorem. *Let X be a compact, oriented, 4 dimensional manifold with $b^{2+} \geq 2$. Let ω be a symplectic form on X with $\omega \wedge \omega$ giving the orientation. Then the first Chern class of the associated almost complex structure on X has Seiberg-Witten invariant equal to ± 1.*

(Note: There are no symplectic forms on X unless b^{2+} and the first Betti number of X have opposite parity.)

In a subsequent article with joint authors, a vanishing theorem will be proved for the Seiberg-Witten invariants of a manifold X, as in the theorem, which can be split by an embedded 3-sphere as $X_- \cup X_+$ where neither X_- nor X_+ have negative definite intersection forms. Thus, no such manifold admits a symplectic form. That is,

Corollary. *Connect sums of 4-manifolds with non-negative definite intersection forms do not admit symplectic forms which are compatible with the given orientation. For example, when $n > 1$ and $m \geq 0$, then $({}_{\#n}\mathbb{CP}^2)\#({}_{\#m}\underline{\mathbb{CP}}^2)$ has no symplectic form which defines the given orientation.*

The Main Theorem also implies that the Seiberg-Witten invariant for the canonical class of a complex surface with $b^{2+} \geq 3$ is equal to ± 1. However, this result is easy to prove directly, as there is just one nondegenerate solution in this case.

Received November 14, 1994.

Supported in part by the National Science Foundation.

There are two steps to the proof of the main theorem. The first step uses the symplectic form to construct a canonical solution to a 1-parameter family of perturbations of the Seiberg-Witten equation. This family is parameterized by $r \in [0, \infty)$. The second step of the proof argues that canonical solution to the perturbed equation is the only solution if the perturbation parameter r is sufficiently large. One must also prove that this canonical solution is nondegenerate when r is large. The arguments for the latter assertion are easy versions of those for the former and are omitted.

1. The $\mathbf{r = 0}$ equation

Fix a metric on X which makes the form ω self dual with length squared equaling $1/2$. Use K to denote the canonical bundle of the almost complex structure. (Thus, the bundle Λ_+ of self dual 2-forms splits (after tensoring with $\mathbb{C}$) as $\mathbb{C} \cdot \omega \oplus K \oplus K^{-1}$.)

The bundle K defines a $Spin_{\mathbb{C}}$ bundle $S = S_+ \oplus S_- \to X$ and the complex 2-plane bundle S_+ splits as $S_+ \approx \mathbb{I} \oplus K^{-1}$, where $\mathbb{I}$ is the trivial, complex line bundle. The form ω defines the splitting; the summand $\mathbb{I}$ is an eigenspace for Clifford multiplication by ω with eigenvalue $-i$. The summand K^{-1} has eigenvalue $+i$. The aforementioned splitting of S_+ will be implicit in much of what follows.

There is a unique connection A_0 (up to gauge) on K^{-1} whose induced covariant derivative $\boldsymbol{\nabla}_{A_0}$ on S_+ has $\nabla_{A_0} \cong 2^{-1}(1 + i \cdot \omega)\boldsymbol{\nabla}_{A_0}$ equal to the product covariant derivative d on the summand $\mathbb{I}$ in S_+. In particular, for this connection A_0, there is a nontrivial section, u_0, of $\mathbb{I}$ which is annihilated by ∇_{A_0}. This section u_0 has constant length and should be normalized to have length equal to 1. If the manifold X were honestly Kähler, with ω the Kähler form, then u_0 would be covariantly constant for $\boldsymbol{\nabla}_{A_0}$ as well. Instead, one has

$$\boldsymbol{\nabla}_{A_0} u_0 = b \tag{1}$$

where b is a section of $K^{-1} \otimes T^*_{\mathbb{C}}X$. Below, I will sometimes introduce a local orthonormal section u_1 of the K^{-1} summand in S_+, in which case I will write $b = \underline{b} \otimes u_1$ where $\underline{b}$ is, locally, a complex valued 1-form. (This b is essentially the torsion of the almost complex structure that is defined by the metric and ω.)

The spinor u_0 solves the Dirac equation when $\underline{b} \cdot u_1 = 0$; where $\underline{b}$ is considered to act by Clifford multiplication. If one introduces local orthonormal frame $\{e^\nu\}$ for T^*X, then b expands as $b_\nu \otimes e^\nu$ with b_ν being a local section of K^{-1}. The fact that u_0 solves the Dirac equation is also expressed as the condition that $e^\nu b_\nu = 0$, where here, e^ν acts on the spin bundle as Clifford multiplication.

All of the above makes sense whether or note the form ω is closed. Now consider:

Lemma 1. *The form ω is closed if and only if u_0 solves the Dirac equation.*

Proof. Use the Dirac operator on the equation $\omega\cdot u_0 = -i\cdot u_0$. The resulting equation reads:

$$z \cdot (*d\omega) \cdot u_0 + e^\nu \cdot \omega \cdot b = -ie^\nu \cdot b. \tag{2}$$

Here, z is a nonzero integer. This last equation uses (1). Since ω acts as multiplication by i on K^{-1}, equation (2) asserts that $z\cdot(*d\omega)\cdot u_0 = -2ie^\nu\cdot b$, which gives the lemma.

Note that Clifford multiplication by a real form on S has no kernel; but if v is Clifford multiplication by a complex valued 1-form, the condition that v annihilate the K^{-1} summand requires only that v have coefficients, $\{v_\nu\}$ (with respect to a frame $\{e^\nu\}$ for T^*X for which $\omega = 2^{-1}(e^1e^2 + e^3e^4)$), which obey

$$v_1 - iv_2 = 0 \quad \text{and} \quad v_3 - iv_4 = 0. \tag{3}$$

That is, v should be a section of $T^{*0,1}$.

The Seiberg-Witten equation for K^{-1} is an equation for a pair of connection A on K^{-1} and section $\psi \equiv \alpha \cdot u_0 + \beta$ of $S_+ = \mathbb{I} \oplus K^{-1}$. (Thus, β is thought of as a section of K^{-1}.) This equation reads

$$D_A\psi = 0.$$

$$P_+F_A = i \cdot (|\alpha|^2 - |\beta|^2) \cdot \omega - i \cdot (\alpha^*\beta + \alpha\beta^*). \tag{4}$$

Here D_A is the Dirac operator as defined with the covariant derivative $\boldsymbol{\nabla}_A$ on S_+ as defined using the connection A.

With this last equation understood, consider instead the perturbed Seiberg-Witten equation which reads

$$D_A\psi = 0.$$

$$P_+F_A = P_+F_{A_0} + i \cdot (|\alpha|^2 - |\beta|^2 - 1) \cdot \omega - i \cdot (\alpha^*\beta + \alpha\beta^*). \tag{5}$$

This last equation differs from (4) by a relatively benign perturbation, and so the Seiberg-Witten invariant of X as defined by the solutions to (4) is the same as that which is defined by the solutions to (5).

The advantage of (5) over (4) is that the pair (A_0, u_0) is, by construction, a solution to (5). Equation (5) is the $r = 0$ version of the 1-parameter family of perturbation that is alluded to in the introduction to the proof of the main theorem.

2. The family of perturbed equations

A connection A on K^{-1} can be written uniquely as $A_0 + i \cdot a$, where a is a real valued 1-form on X. (On the gauge orbit through A, there is unique connection of the form $A_0 + i \cdot a$ with $d * a = 0$.)

Given $r \geq 0$, here is an equation for a connection $A \equiv A_0 + i \cdot a$ on K^{-1} and a section $\psi = (\alpha, \beta)$ of $S_+ = \mathbb{I} \oplus K^{-1}$:

$$D_A \psi = 0.$$

$$\begin{aligned} P_+ F_A =& P_+ F_{A_0} + i \cdot (|\alpha|^2 - |\beta|^2 - 1) \cdot \omega - i \cdot (\alpha^* \beta + \alpha \beta^*) \\ & - i \cdot 4 \cdot r (1 + r \cdot |\alpha|^2)^{-1} \cdot (\alpha^* \langle b, \nabla_a \alpha \rangle - \alpha \cdot \langle b, \nabla_a \alpha \rangle^*). \end{aligned} \tag{6}$$

Here, $\langle \, , \rangle$ signifies the $\mathbb{C}$-bilinear extension to $T^*_{\mathbb{C}} X$ of the metric inner product on $T^* X$. In (6), b is the section of $K^{-1} \oplus T^*_{\mathbb{C}} X$ which is defined in (1). Also, $\nabla_a \equiv d + i \cdot a$ is a covariant derivative on complex valued functions on X.

The following lemmas summarize the important features of (6). They have one immediate corollary, namely that for any $r \geq 0$, one can use (6) to compute the Seiberg-Witten invariant for X and the line bundle K^{-1}.

Lemma 2. *Let* $\ker(d*)$ *denote the subspace of co-closed 1-forms on* X *and let* $\Gamma(S_+)$ *denote the space of sections of* S_+. *A 1-form,* a, *in* $\ker(d^*)$ *defines a connection on* K^{-1} *by the setting* $A = A_0 + i \cdot a$. *With this understood, then for any* $r \geq 0$, *Equation (6) defines a smooth map* $\Psi \colon \ker(d^*) \times C^\infty(S_+) \to C^\infty(\Lambda_+ \oplus S_-)$. *This map depends smoothly on the parameter* $r \geq 0$ *and it extends as smooth map from the* L^2_1 *Sobolev space completion of the range into the* L^2 *Sobolev space completion of the domain. Furthermore, the differential of* Ψ *at any* (a, ψ) *defines an elliptic operator which is Fredholm when the domain is completed using the* L^2_1 *Sobolev topology and the range is completed with the* L^2 *Sobolev topology. Thus, the index of the differential does not depend on the value of the parameter* $r \in [0, \infty)$.

Proof. The only serious issue here is to verify that (6) linearizes to an elliptic operator from $\ker(d^*) \oplus C^\infty(S_+)$ into $C^\infty(\Lambda_+ \oplus S_-)$. Written in block diagonal form, this linearization has the form

$$\begin{pmatrix} P_+ d & r \cdot v\imath d \\ 0 & D \end{pmatrix} + \text{ zero'th order terms,} \tag{7}$$

where $v\imath$ is (locally) contraction with a complex valued vector field and D is the Dirac operator. It is clear from (7) that the linearization is elliptic for any r as claimed.

In the next lemma, and subsequently, the symbol ∇'_A will be used below to denote the induced covariant derivative on the K^{-1} summand of S_+; that is, $\nabla'_A \equiv 2^{-1} \cdot (1 - i \cdot \omega) \cdot \boldsymbol{\nabla}_A$. (Of course, this equals $\nabla'_{A_0} + i \cdot a$.)

Lemma 3. *Let $\{r_i\}$ be a countable sequence in $[0,\infty)$ and let $\{(A_i,\psi_i)\}$ be a sequence of solutions to (6) for the corresponding r_i. Then there is an infinite subsequence of $\{r_i,(A_i,\psi_i)\}$ which has the following property:*

(1) *The subsequences $\{|\psi_i|\}$ and $\{|\boldsymbol{\nabla}_{A_i}\psi_i|\}$ converge strongly in the L^2_1 topology and the L^2 topology, respectively. Furthermore, the former is also uniformly bounded in the supremum norm.*

(2) *Writing $\psi_i=(\alpha_i,\beta_i)$, then the subsequences $\{|\alpha_i|\}$ and $\{|\beta_i|\}$ converge strongly in the L^2_1 topology, and the subsequences $\{|\nabla_{a_i}\alpha_i|\}$ and $\{|\nabla'_{A_i}\beta_i|\}$ converge strongly in the L^2 topology.*

(3) *If the subsequence $\{r_i\}$ converges to some $r_0<\infty$, then the subsequence $\{(A_i,\psi_i)\}$ is gauge equivalent to a sequence which converges in the C^∞ topology to a solution to (6) for the parameter value $r\equiv r_0$.*

Proof. The key step to proving the first assertion consists of a digression first to obtain some *a priori* estimates for solutions to (6). Take the approach which Witten [W] found for (4); that is, compute the Weitzenboch formula for $D_AD_A\psi$. The formula reads:

$$D_A^2\psi=\boldsymbol{\nabla}_A{}^*\boldsymbol{\nabla}_A\psi+s\cdot\psi+2^{-1}(P_+F_A)\cdot\psi. \tag{8}$$

Here, (P_+F_A) is an imaginary valued 2-form which acts by Clifford multiplication on S_+. Also, s is proportional (with a positive constant) to the scalar curvature of the metric on X. In the present case, $D_A\psi$ is assumed to vanish. Input this information into (8) and take the inner product of the resulting equation with ψ to find that

$$2^{-1}\cdot d^*d|\psi|^2+|\boldsymbol{\nabla}_A\psi|^2+s|\psi|^2+\langle\psi,2^{-1}(P_+F_A)\psi\rangle=0. \tag{9}$$

Now one should substitute for P_+F_A from (6) into the last term above. The key point to observe here is that Clifford multiplication by the r-dependent term in the second equation in (6) has no diagonal entries when written in 2×2 block form with respect to the decomposition of S_+ as $\mathbb{I}\oplus K^{-1}$. This has the following consequence: The norm of this part of the curvature has the following *a priori* r-independent bound:

$$z\cdot\frac{r|\alpha|}{(1+r|\alpha|^2)}\cdot|\nabla_a\alpha|\le z\cdot|\alpha|^{-1}\cdot|\nabla_a\alpha|. \tag{10}$$

Here, z is determined by the size of the torsion b and is independent of the parameter r. (Here and below, z will be used indiscriminantly to denote a universal constant which is, in particular, r-independent. Its precise value from line to line may change.)

In (9), this part of the curvature appears with a factor of either $\alpha\beta^*$ or $\alpha^*\beta$. Thus, the r-dependent part of (6) contributes a term in (9) which is *a priori* bounded by

$$z \cdot |\psi| \cdot |\nabla_a \alpha| \le z \cdot |\psi| \cdot (|\boldsymbol{\nabla}_A \psi| + |\psi|) \tag{11}$$

for any value of $r \ge 0$.

The first two factors in (6) for $P_+ F_A$ contribute to (9) as described by Seiberg and Witten (see also [KM]), and allow one to write (9) as

$$2^{-1} d^* d|\psi|^2 + |\boldsymbol{\nabla}_A \psi|^2 + |\psi|^4 = R(\psi, \boldsymbol{\nabla}_A \psi) \tag{12}$$

where

$$|R(\psi, \boldsymbol{\nabla}_A \psi)| \le z \cdot (|\psi|^2 + |\psi| \cdot |\boldsymbol{\nabla}_A \psi|). \tag{13}$$

With (12) and (13) understood, uniform L^2 bounds for ψ and $|\boldsymbol{\nabla}_A \psi|$ and uniform L^∞ bounds for ψ follow by standard techniques. Also, remember that $|d|\psi|\,| \le |\boldsymbol{\nabla}_A \psi|$ so that $|\psi|$ is uniformly bounded in L^2_1.

Finally, note that these uniform bounds for $|\psi|$ and for $|\boldsymbol{\nabla}_A \psi|$ provide the analogous uniform bounds for $|\alpha|$ and $|\nabla_a \alpha|$, and also for $|\beta|$ and $|\nabla'_A \beta|$. End the digression.

Now, reconsider the sequence $\{r_i, (A_i, \psi_i)\}$. It follows from the preceding that the sequence $\{|\psi_i|\}$ is uniformly bounded in L^2_1 and hence it has a weakly convergent subsequence which strongly converges in L^2. The limit of this subsequence will be denoted by f;, an L^2_1 function on X. Note that f is in L^∞ since the sequence of ψ's is uniformly bounded in L^∞. In fact, $\{\psi_i\}$ converges strongly to f in L^p for any $p < \infty$.

Meanwhile, the sequence $\{|\boldsymbol{\nabla}_{A_i} \psi_i|\}$ has a weakly convergent subsequence in L^2 which can be assumed to coincide with the convergent subsequence of $\{|\psi_i|\}$. The limit of the weakly convergent subsequence of $\{|\boldsymbol{\nabla}_{A_i} \psi_i|\}$ will be called g.

One can now use the strong L^p convergence of the $|\psi_i|$'s to f to conclude that the relevant subsequence of $\{|\boldsymbol{\nabla}_{A_i} \psi_i|\}$ converges *strongly* to g in L^2, and that the subsequence of $\{|\psi_i|\}$ converges strongly to f in L^2_1. Indeed, for the former, one need only show that the relevant subsequence of L^2 norms of $\{|\boldsymbol{\nabla}_{A_i} \psi_i|\}$ converges to the L^2 norm of g, and this follows by integrating both sides of (9) over X and using the already established facts about the strong convergence of the ψ_i's in L^p. A similar argument proves that the relevant subsequence of $\{d|\psi_i|\}$ converges strongly in L^2.

This last point establishes Assertion 1 of Lemma 3. A similar argument gives Assertion 2. Assertion 3 is proved by using the usual elliptic regularity arguments.

Lemma 4. *For no $r \geq 0$ does* (6) *have a solution with $\psi = 0$.*

Proof. Such a solution would have $P_+ F_A = P_+ F_{A_0} - i \cdot \omega$. If such were the case, then F_A and F_{A_0} would not be cohomologous, and so A would not be a connection on K^{-1}.

These last three lemmas have as corollary:

Proposition 5. *For any $r \geq 0$, the solutions to* (6) *can be used to compute the Seiberg-Witten invariant for the pair (X, K^{-1}) just as well as the original equation* (4).

(Of course, one would have to count degenerate solutions to (6) with the appropriate multiplicity. These multiplicities can be determined by perturbing the second equation of (6) with the addition of some small self dual 1-form to the right side to resolve the degeneracy. This sort of perturbation is also considered in [W] and [KM].)

3. A vanishing theorem for large r

As just remarked, the solutions to (6) can be used to compute the Seiberg-Witten invariant. For any $r \geq 0$, (6) has one obvious solution, namely $A = A_0$ and $\psi = u_0$. The purpose of this section is to prove that this solution is the only solution to (6) when r is sufficiently large. Furthermore the same arguments (but linearized) will show that (A_0, u_0) is a nondegenerate solution to (6) when r is sufficiently large.

The proof that (6) has a unique solution for large r will be had by modifying the proof for a Kähler manifold that K^{-1} has just one Seiberg-Witten solution. (See e.g. [W].) There are three steps in this process.

In Step 1, one supposes that there exists an increasing, unbounded sequence $\{r_m\}$ of parameter value for which (6) has a solution $(A_m, \psi_m \equiv (\alpha_m, \beta_m))$. One then establishes

Lemma 6. *For a sequence as above, zero is the limit as m tends to ∞ of the measure of the set of points in X where $|\alpha_m| < 1/2$.*

(Here, one could just as well take the set of points in X where $|\alpha_m|$ is less than $1 - \delta$ for any $\delta > 0$.)

In Step 2, one establishes

Lemma 7. *For a sequence as above, let $V(m)$ denote the set of points in X where $|\alpha_m| < 1/2$. For each m, the integral over x of $|\nabla_{A_m} \beta_m|^2 + |\beta_m|^2$ is bounded from above by an (r, m)-independent multiple of the integral over $V(m)$ of $|\beta_m|^2$.*

Step 3 is short, as it uses a standard Sobolev inequality to establish a contradiction from Lemma 6 and 7.

Step 1: To begin the task of proving Lemma 6, project (8) onto the two summands of $S_+ = \mathbb{I} \oplus K^{-1}$ so as to get a pair of coupled equations for α and β. To begin start with the Dirac equation which is written out below with the help of a local orthonormal frame $\{e^\alpha\}$ for T^*X which is chosen so that

$$\omega = 2^{-1}(e^1 e^2 + e^3 e^4). \tag{14}$$

Here is the equation $D_A\psi = 0$:

$$(\nabla_a \alpha)_\nu \cdot e^\nu \cdot u_0 + e^\nu (\nabla'_A \beta)_\nu = 0. \tag{15}$$

Here, and below, $(\nabla_a\alpha)_\nu$ denotes $\langle e^\nu, \nabla_a \alpha\rangle$ (with $\nabla_a \equiv d + ia$); there is a similar formula for $(\nabla'_A\beta)_\nu$. Note that the torsion b does not appear in (15). Acting on both sides of (15) by D_A yields

$$\begin{aligned}((\nabla_a\nabla_a\alpha)_{\nu\mu} e^\nu e^\mu)u_0 &+ (\nabla_a\alpha)_\mu e^\nu e^\mu b_\nu \\ &+ (e^\nu e^\mu)(\nabla'_A\nabla'_A\beta)_{\nu\mu} - b^*_\nu (\nabla'_A\beta)_\mu e^\nu e^\mu u_0 = 0\end{aligned} \tag{16}$$

This equation can be simplified by using the Clifford Multiplication rule $e^\nu e^\mu + e^\mu e^\nu = -2\delta^{\nu\mu}$. Also, the equation can be simplified if one uses the identities $e^\nu b_\nu = 0$ and also $b^*_\nu e^\nu u_0 = 0$ which come from the fact that u_0 is annihilated by D_{A_0}. With the preceding understood, (16) simplifies to

$$(\nabla_a\nabla_a\alpha)_{\nu\mu} e^\nu e^\mu u_0 - 2\langle\nabla_a\alpha, b\rangle + e^\nu e^\nu (\nabla'_A\nabla'_A\beta)_{\nu\mu} + 2\langle b^*, \nabla'_A\beta\rangle u_0 = 0. \tag{17}$$

(Remember that $\langle , \rangle$ has been defined as the $\mathbb{C}$-bilinear extension to $T^*_{\mathbb{C}}X$ of the Riemannian inner product on X; thus, the b^* in the last term of (17).)

Consider the projection of (17) onto K^{-1}. The result is

$$4^{-1}\alpha(1 - i\omega)\cdot(F_A - F_{A_0})\cdot u_0 - 2\langle\nabla_a\alpha, b\rangle + 2^{-1}(1 - i\omega)\cdot D_A D_A \beta = 0. \tag{18}$$

Here, $F_A - F_{A_0}$ is acting as Clifford multiplication by an imaginary valued 2-form. After substituting from (6) for $F_A - F_{A_0}$, this last equation becomes

$$\frac{1}{2}|\alpha|^2\beta + 2\frac{r|\alpha|^2}{(1 + r|\alpha|^2)}\langle\nabla_a\alpha, b\rangle - 2\langle\nabla_a\alpha, b\rangle + \frac{1}{2}(1 - i\omega)D_A D_A\beta = 0. \tag{19}$$

Finally, (19) can be rearranged to read

$$\frac{1}{2}|\alpha|^2\beta - 2\frac{1}{(1 + r|\alpha|^2)}\langle\nabla_a\alpha, b\rangle + \frac{1}{2}(1 - i\omega)D_A D_A\beta = 0. \tag{20}$$

To make hay from (20), take the inner product of both sides with β and integrate the result over X. After an integration by parts in the last term of (20) and some rearragnements, one finds

$$\int \left(|D_A\beta|^2 + |\alpha|^2\beta|^2\right) = 2\int \frac{1}{(1+r|\alpha|^2)}\langle\nabla_a\alpha, \beta^* b\rangle \tag{21}$$

The reader should note here that the terms on the left side of (21) are nonnegative. And it is crucial to note that the term on the right side has r appearing in the denominator.

Equation (21) will be used twice. For the first application, integrate by parts on the right side and then take absolute values to find

$$\begin{aligned}\int &(D_A\beta|^2 + |\alpha|^2\beta|^2)\\ &\leq z\int \frac{1}{(1+r|\alpha|^2)}(|\alpha||\nabla'_A\beta| + |\alpha|\cdot|\beta| + |\beta||d|\alpha||).\end{aligned} \tag{22}$$

There are three terms on the right side of (22). The first two are smaller than a multiple of $r^{-1/2}$; this follows from Lemma 3 and the fact that the function $h(s) = s(1+rs^2)^{-1}$ is bounded by $r^{-1/2}\sqrt{2/3}$.

To analyze the last term on the right hand side of (22), introduce $V(r) = V(r, |\alpha|)$ to denote the subset of X where

$$|\alpha| \leq r^{-1/4}. \tag{23}$$

Split the integral of the last term on the right in (22) into two pieces, the integration over $V(r)$ and the integration over the compliment of $V(r)$. On the latter, the function $(1+r|\alpha|^2)^{-1}$ is small, smaller than $(2r)^{-1/2}$. And so it follows with Lemma 3 that the integral of the last term in (22) over the compliment in $V(r)$ is also bounded by a multiple of $r^{-1/2}$. In summary, the left side of (2) is seen bounded by

$$\int (|D_A\beta|^2 + |\alpha|^2|\beta|^2) \leq z(\frac{1}{r^{1/2}} + \int_{V(r)} |\beta|\frac{|d|\alpha||}{(1+r|\alpha|^2)}). \tag{24}$$

One can now deduce the following lemma from (24):

Lemma 8. *Let $\{r_m, (A_m, \psi_m)\}$ be a sequence of solutions to (6) where the sequence $\{r_m\}$ is increasing without bound. Assume that the sequences $\{|\psi_m|\}$ and $\{|\boldsymbol{\nabla}_{A_m}\psi_m|\}$ converge in L^2_1 and L^2, respectively. Consecutively,* *take (α, β) equal to (α_m, β_m) in the right side of (24). Then, the resulting sequence of numbers converges to zero as m tends to infinity.*

Proof. The first term in (24) evidently converges to zero. Argue that the second term converges to zero as follows: As remarked in Lemma 3, the sequence of $|\alpha_m|$'s converges strongly in L^2_1 to some function f: with this understood, one can replace $|d|\alpha||$ in (24) with $|df|$ with small error at large m. And, since the $|\beta_m|$'s are uniformly bounded, the issue is simply whether

$$\int_{V(r)} \frac{|df|}{(1 + r|\alpha|^2)} \tag{25}$$

converges to zero when $r = r_m$ and $\alpha = \alpha_m$ and m tends to infinity. To settle the issue, compare the integral in (25) with

$$\int_{V(r)} \frac{|df|}{(1 + rf^2)} \tag{26}$$

The strategy is to first prove that (26) tends to zero as r tends to infinity, and then prove that the difference between (25) and (26) tends to zero. Here is a proof that (26) tends to zero as r tends to ∞: First, (26) is bounded by the integral of the same integrand, but over X, Second, for each $n = 1, 2, \ldots$, consider breaking the integral (over X) into two integrals, the former over the domain where $f^2 > 1/n$ and the latter over the region where $f^2 \leq 1/n$. The integral over the former region is bounded by

$$z\frac{n}{r}\|df\|_{L^2}, \tag{27}$$

which tends to zero for fixed n as r tends to ∞. The integral over the latter region tends to zero by appeal to

Lemma 9. *Let f be a non-negative, L^2_1 function over a compact manifold. For $\varepsilon > 0$, let θ_ε denote the integral of $|df|$ over the region where f is less than ε. Then the sequence of θ_ε's has limit zero as ε tends to zero.*

This lemma is proved below.

With the preceding understood, the comparison of (25) with (26) considers the integral over $V(r)$ of

$$|df|\;|(1 + rf^2)^{-1} - (1 + r|\alpha|^2)^{-1}| = |df|r\frac{||\alpha|^2 - f^2|}{(1 + rf^2)(1 + r|\alpha|^2)}. \tag{28}$$

Break this integral up into two pieces, the first where $f^2 > 1/n$, and the second where $f^2 \leq 1/n$. By Lemma 9, the integral over the latter part is bounded by a sequence, $\{\theta_\varepsilon : \varepsilon = n^{-1/2}\}$ of r-independent numbers which go to zero as n tends to ∞. Meanwhile, for fixed n, the integral over the former tends to zero as r gets large because the measure of the set where $|\alpha|^2$ differs from f^2 by $(2n)^{-1}$ vanishes as r tends to infinity since $|\alpha|^2$ converges to f^2 in this limit. With this understood, one compares (25) to (26) by fixing n so that the integral of (28) where $f^2 < n^{-1]}$ is small, and then one takes r large to make the remainder small.

Thus, the proof of Lemma 8 has been reduced to a proof of Lemma 9. The following proof was suggested by David Jerison.

Proof of Lemma 9. Let $U_\varepsilon \subset X$ denote the set of points x where $f(x) \leq \varepsilon$. One has $U_{\varepsilon_1} \subset U_{\varepsilon_2}$ when $\varepsilon_1 < \varepsilon_2$ and $\cap_\varepsilon U_\varepsilon \equiv W$ is the set of points x where $f(x) = 0$. If W has measure zero, then the sequence of measures of U_ε tends to zero as ε tends to zero and so the lemma is immediate. So consider the case where W has positive measure. To handle this case, one must claim that

$$df = 0 \text{ almost everywhere on } W. \tag{29}$$

Granted (29), the integral of $|df|$ over U_ε is the same as the integral of $|df|$ over $U'_\varepsilon = U_\varepsilon - W$. The measure of $\cap_\varepsilon U'_\varepsilon$ is zero which means that the measures of the U'_ε tend to zero as ε tends to zero. This implies the lemma in the case where W has positive measure.

As for (29), the reader is referred to Theorem 1 on page 242 of [S], the remark labled (ii) on page 247 of [S], and then Theorem 2 of page 249 of [S]. (All of this takes place in Chapter 8 of [S].)

With Lemma 8 understood, let us agree to relable the original sequence $\{r_m, (A_m, \psi_m)\}$ so that the left hand side of (24) is less than m^{-1} when $r = r_m$ and $(A, \psi) = (A_m, \psi_m)$.

Turn next to the projection of (17) along u_0. The resulting equation is

$$\nabla_a^* \nabla_a \alpha + (|\alpha|^2 - |\beta|^2 - 1)\alpha + \langle u_0, D_A^2 \beta \rangle = 0. \tag{30}$$

Multiply both sides of this equation with α^* and integrate the resulting equation over X. After an integration by parts, the resulting equation reads:

$$\int (|\nabla_a \alpha|^2 + (|\alpha|)^2 - 1 - |\beta|^2)(|\alpha|^2 - 1) = \int \langle D_A(\alpha u_0), D_A \beta \rangle. \tag{31}$$

Here, I have used the fact that

$$\int (1 - |\alpha|^2 + |\beta|^2) = 0; \tag{32}$$

an expression of the fact that the integral over X of $F_A \wedge \omega$ is equal to the integral over X of $F_{A_0} \wedge \omega$. (This is because F_A and F_{A_0} are closed forms which are cohomologous.)

In discussing (31), introduce $w = (1 - |\alpha|^2)$. Then, (31) implies that

$$\int (|\nabla_a \alpha|^2 + w^2 + |\beta|^2) \leq z \int (|\nabla_a \alpha|\ |D_A \beta| + |\alpha|^2 \beta|^2). \tag{32}$$

With the triangle inequality, this last expression gives

$$\int (|\nabla_a \alpha|^2 + w^2 + |\beta|^2) \leq z \int (|D_a \beta|^2 + |\alpha|^2 \beta|^2). \tag{33}$$

In the case where $r = r_m$ and $(A, \psi) = (A_m, \psi_m)$, the right side of (33) is less than m^{-1}.

Fix any $\delta > 0$ and it follows from the preceding remarks that the volume of the set where $|\alpha_m|^2 < 1 - \delta$ is bounded from above by $(m\delta^2)^{-1}$. One can see this from (33) by focusing exclusively on the w^2 term. If w is bigger than δ on a set of measure greater than $(m\delta^2)^{-1}$, then the integral of the w^2 term in (33) would be larger than the integral on the right side of (33). (Being non-negative, the remaining terms can't cancel out any excess w^2 integral.) (A similar argument shows that the analogous limit of the $|\beta_m|$'s is 0.)

This gives Lemma 6 and ends Step 1.

Step 2: Return to (21). Break the integral on the right side of (21) into two parts, the first where $|\alpha| \geq 1/2$ and the second where $|\alpha| < 1/2$. Taking absolute values gives the following inequality:

$$\int (|D_a \beta|^2 + |\beta|^2 |\alpha|^2) \leq \frac{z}{r} \int |\nabla_a \alpha| \cdot |\beta| + z \int_V |\nabla_a \alpha| \cdot |\beta|. \tag{34}$$

Here V is the set of points where $|\alpha| < 1/2$. Plug this last inequality into (33) and use the triangle inequality to conclude that

$$\int (|\nabla_a \alpha|^2 + w^2 + |\beta|^2) \leq z \int_V |\beta|^2. \tag{35}$$

With (35) understood, return to (34) to conclude that

$$\int |D_A \beta|^2 \leq z \int_V |\beta|^2. \tag{36}$$

Now, write $D_A\beta = e^\nu(\nabla'_A\beta)_\nu$ to relate the integral on the right above to the integral of $|\nabla'_A\beta|^2$. The result is

$$\int |\nabla'_A\beta|^2 \le \int (|D_A\beta|^2 + |\beta|^2) \le z\int_V |\beta|^2. \tag{37}$$

Here, (6) has been used to evaluate the inner product of β with Clifford multiplied on β by the imaginary 2-form F_A to obtain the first inequality in (37). The second inequality in (37) follows from (35) and (36). In particular, this last equation implies that

$$\int (|\nabla'_A\beta|^2 + |\beta|^2) \le z\int_V |\beta|^2. \tag{38}$$

This last equation completes the proof of Lemma 7 and ends Step 2.

Step 3: Remark now that there is an inherent contradiction in (38) unless β vanishes identically for large m. Indeed,

$$\int_V |\beta|^2 \le Vol(V)^{1/2}(\int_V |\beta|^4)^{1/2} \le zVol(V)^{1/2}\int (|d|\beta||^2 + |\beta|^2) \tag{39}$$

where the last inequality is a Sobolev inequality. As $|d|\beta|\,| \le |\nabla'_A\beta|$, this last equation implies that

$$\int_V |\beta|^2 \le Vol(V)^{1/2} z\int (|\nabla'_A\beta|^2 + |\beta|^2). \tag{40}$$

Since the volume of the set V tends to zero, the inequalities in (38) and (40) imply that $\beta = \beta_m$ must vanish for all m sufficiently large. Plug this result into (35) to conclude that $\alpha = \alpha_m$ is ∇_{a_m} covariantly constant with norm 1. Thus, when m is large, (A_m, ψ_m) is gauge equivalent to (A_0, u_0) as claimed in the Main Theorem.

Acknowledgments

The author most gratefully thanks David Jerison, Tom Mrowka and Ron Stern for their comments and suggestions.

References

[KM] P. Kronheimer and T. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Research Letters **1** (1994), 797–808.

[SW1] N. Seiberg and E. Witten, *Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang-Mills theory*, preprint.

[SW2] N. Seiberg and E. Witten, *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*, preprint.

[S] E. M. Stein, *Singular Integrals and Differential Properties of Functions*, Princeton University Press 1970.

[W] E. Witten, *Lectures at MIT and Harvard, Fall 1994; and Monopoles and four-manifolds*, Math. Research Letters **1** (1994), 769–796.

Harvard University, Cambridge, MA 02138
E-mail address: chtaubes@math.harvard.edu