Mathematical Research Letters **2**, 221–238 (1995)

THE SEIBERG-WITTEN AND GROMOV INVARIANTS

Clifford Henry Taubes

0. Introduction

The purpose of this article is to announce an equivalence between the Seiberg-Witten gauge theory invariants of a symplectic 4-manifold and a certain set of Gromov invariants as defined using pseudo-holomorphic 2-submanifolds (Theorem 4.1). The theorem below lists some consequences of this equivalence:

Theorem A. *Let X be a compact, oriented 4-manifold with $b_2^+ > 1$ and with a symplectic form ω. Then*

(1) *The Poincaré dual to the canonical bundle K of X is represented by the fundamental class of an embedded, symplectic curve.*
(2) *A homology class in X which is represented by the fundamental class of an embedded 2-sphere that has self-intersection number -1, and that pairs nontrivially with $c_1(K)$ has pairing ± 1 with $c_1(K)$ and is represented by the fundamental class of an embedded, symplectic 2-sphere.*
(3) *If $c_1(K) \cdot c_1(K) < 0$, then X can be symplectically blown down along a symplectic sphere of self-intersection -1.*
(4) *Suppose that X cannot be symplectically blown down along a symplectic sphere of self-intersection -1; and suppose that $c_1(K)$ is not rationally trivial. Then, the signature of the intersection form of X is no smaller than $-\frac{4}{3}(1 - b_1) - \frac{2}{3} b_2$. (The b_i's are the Betti numbers.)*

One can also deduce constraints on the Seiberg-Witten invariants of Theorem A's manifold X using known facts about pseudo-holomorphic curves. These are detailed in Section 4, below.

There are also consequences for manifolds with $b_2^+ = 1$. In particular, one obtains (with the help of a theorem of Gromov [G])

Received March 1, 1995.
Research supported in part by the National Science Foundation.

Theorem B. *The manifold $\mathbb{CP}^2$ has a unique (up to symplectomorphism) symplectic structure.*

This theorem and other consequences of the stated equivalence are discussed below. Except for the briefest outlines, the proofs of these two theorems and the assertions below will appear in [T1].

The organization of the remainder of this paper is as follows. The first section reviews the definition of the Seiberg-Witten invariant in a fairly general context. Section 2 specializes to a discussion of the Seiberg-Witten invariant for symplectic manifolds. Section 3 defines the relevant Gromov invariant of a symplectic manifold. (This invariant is defined by Ruan in [R].) The fourth section states the main theorem—that the Seiberg-Witten invariant and the Gromov invariant agree. This section also details some corollaries of the main theorem. The fifth section outlines the strategy of the proof of the main theorem, and the last section contains some brief speculations about the meaning of the Seiberg-Witten invariant in the non-symplectic world.

1. The Seiberg-Witten invariants

The Seiberg-Witten invariants are defined *a priori* for a compact, oriented 4-manifold X with the characteristic number $b_2^+ > 1$. (There is a more complicated structure in the case where $b_2^+ = 1$.) Here, b_2^+ is equal to the number of $+1$ eigenvalues of the intersection form on the (rational) second homology. (Thus, $b_2^+ = 2^{-1}(\text{rank}(H_2(X) + \text{sign}(X))$, where $\text{sign}(X)$ is the signature of X.) These invariants were first described in [W], and the reader is also referred to [KM] and [T2]. (A more complete description is in preparation with multiple authors.) What follows is a brief description of the invariants. (Note: these invariants vanish when $b_2^+ + b_1$ is even.)

The Seiberg-Witten invariants of X constitute a map from the set of equivalence classes of $Spin_{\mathbb{C}}$ structures on X (covering the frame bundle) to the integers. (These invariants are defined by counting (in a suitable sense) the solutions to a natural system of differential equations on X which are defined using a $Spin_{\mathbb{C}}$ structure.)

Here is a five-part digression to review some crucial 4-dimensional geometry.

Part 1: Note that

$$SO(4) = (SU(2) \times SU(2))/\{\pm 1\} \tag{1.1}$$

and

$$Spin_{\mathbb{C}}(4) = (U(1) \times SU(2) \times SU(2))/\{\pm 1\}, \tag{1.2}$$

where $\{\pm 1\}$ acts on all factors in both cases in the obvious way.

Part 2: Fix a Riemannian metric on X. The metric defines the principal $SO(4)$ bundle of orthonormal frames on X. A $Spin_{\mathbb{C}}$ structure (denoted by $\mathcal{L}$) is simply a lift of this $SO(4)$ principal bundle to a $Spin_{\mathbb{C}}(4)$ principal bundle. The set of equivalence classes of such lifts has, in a natural way, the structure of a principal $H^2(X;\mathbb{Z})$ bundle over a point. This principal $H^2(X;\mathbb{Z})$ bundle, *Spin*, is canonically defined and independent of the original choice of metric on X. Thus, one should think of the Seiberg-Witten invariant as a map from *Spin* to $\mathbb{Z}$.

Part 3: The group $SO(4)$ has two evident representations into $SO(3) = SU(2)/\{\pm 1\}$, which will be called λ_+ and λ_-. They are distinguished by the convention that the associated $\mathbb{R}^3$ bundles to the frame bundle of X be isomorphic to the bundles Λ_+ of self-dual 2-forms and Λ_- of anti-self-dual 2-forms, respectively.

Likewise, the group $Spin_{\mathbb{C}}(4)$ has two evident representations, s_+ and s_-, into $U(2) = (U(1) \times SU(2))/\{\pm 1\}$. The convention is that the composition of s_+ with the quotient homomorphism $U(2) \to U(2)/\text{Center} = SO(3)$ factors through $SO(4)$ via λ_+. With the preceding understood, given a $Spin_{\mathbb{C}}$ structure $\mathcal{L}$ on X, introduce the $\mathbb{C}^2$-vector bundles

$$S_+, S_- \longrightarrow X \tag{1.3}$$

which are associated to $\mathcal{L}$ via the representations s_+ and s_-, respectively. (These bundles inherit natural fiber metrics.)

Let $\mathcal{L}$ and $\mathcal{L} \cdot e$ be elements in *Spin*, where $e \in H^2(X;\mathbb{Z})$. Then, the bundles S_+ for these two $Spin_{\mathbb{C}}$ structures are related by $S_+(\mathcal{L} \cdot e) \simeq S_+(\mathcal{L}) \otimes E$, where E is a complex line bundle with first Chern class $c_1(E) = e$.

Part 4: Clifford multiplication, c, maps T^*X into the skew adjoint endomorphisms of $S_+ \oplus S_-$; it is characterized by the fact that $c(v)^2$ is multiplication by $-|v|^2$. In particular, c induces maps

$$\sigma : S_+ \otimes T^*X \to S_- \tag{1.4}$$

(by duality) and also $c_+ : \Lambda_+ \to \text{End}(S_+)$. The adjoint of the latter will be denoted by

$$\tau : \text{End}(S_+) \longrightarrow \Lambda_+ \otimes \mathbb{C}; \tag{1.5}$$

it maps a self-adjoint endomorphism into an imaginary-valued form.

Part 5: Let A be a connection on $L \equiv \det(S_+)$. The connection A with the Levi-Civita connection on T^*X induces a covariant derivative on S_+. This maps sections of S_+ into sections of $S_+ \otimes T^*X$. The composition of this last map with σ in (1.4) defines the Dirac operator D_A, a first order, elliptic operator mapping sections of S_+ to sections of S_-.

End the digression.

With the preceding understood, notice that the Seiberg-Witten equations are equations for a pair (A, ψ), where A is a connection on $L = \det(S_+)$ and where ψ is a section of S_+. These equations read

$$D_A\psi = 0 \quad \text{and} \quad P_+F_A = \frac{1}{4}\tau(\psi \otimes \psi^*), \tag{1.6}$$

where $P_+ : \Lambda^2T^*X \to \Lambda_+$ is the orthogonal projection. It proves useful at times to consider perturbations of (1.6) which have the form

$$D_A\psi = 0 \quad \text{and} \quad P_+F_A = \frac{1}{4}\tau(\psi \otimes \psi^*) + \mu, \tag{1.7}$$

where μ is a fixed, imaginary-valued, self-dual 2-form on X.

The Seiberg-Witten invariant for the given $Spin_{\mathbb{C}}$ structure $\mathcal{L} \in Spin$ is obtained by making a suitable count of solutions of (1.6) or (1.7). Remark here that the group $C^\infty(X; S^1)$ acts on the space of solutions to (1.7); a map φ sends (A, ψ) to $(A - 2\varphi^{-1}d\varphi, \varphi\psi)$. Here, S^1 is thought of as the unit sphere in $\mathbb{C}$. (This group acts freely at solutions where ψ is not identically zero.) The quotient of the space of solutions to (1.7) by $C^\infty(S^1; X)$ will be denoted by M. (The dependence on the $Spin_{\mathbb{C}}$ structure and on the choice of μ in (1.7) will usually be suppressed.)

Here are five crucial facts about M:

Fact 1: When $b_2^+ \geq 1$, the space of solutions to (1.6) and (1.7) will contain no points where $\psi \equiv 0$ for a generic metric or choice of μ as long as $c_1(L)$ is rationally nonzero. Here, generic means off of a set of codimension b_2^+. (This follows from a theorem of Uhlenbeck in [FU].)

Fact 2: The space M has naturally the structure of a real analytic variety. When $b_2^+ \geq 1$, the space M will be a smooth manifold for a generic choice of μ in (1.7). (Here, generic means a Baire subset of $C^\infty(\Lambda_+)$.) The dimension of this manifold is computed with the help of the Atiyah-Singer index theorem to be

$$d = -\frac{1}{4}(2\chi(X) + 3sign(X)) + \frac{1}{4}c_1(L) \cdot c_1(L). \tag{1.8}$$

Here, χ is the Euler characteristic of X and $sign(X)$ is the signature. Also, the notation $u \cdot v$ for classes $u, v \in H^2(X; \mathbb{Z})$ denotes the evaluation of their cup product on the fundamental class of X.

Fact 3: A choice of orientation for the line

$$\det(H^0(X; \mathbb{R})) \otimes \det(H^1(X; \mathbb{R})) \otimes \det(H^{2+}(X; \mathbb{R})) \tag{1.9}$$

serves to orient M. (The orientation of a point is a choice of ± 1 to assign to said point.)

Fact 4: Fix a base point in X and let $C_0^\infty(S^1; X)$ denote the subset of maps which map said base point to 1. Let M^0 denote the quotient of the space of solutions to (1.7) by the latter group. Where M is a smooth manifold, the projection $M^0 \to M$ defines a principal S^1 bundle.

Fact 5: The space M is compact.

With the preceding understood, here is the definition of the Seiberg-Witten invariant:

Definition 1.1. Let X be a compact, oriented 4-manifold with $b_2^+ > 1$ and let $\mathcal{L} \in$ *Spin* be a $Spin_{\mathbb{C}}$ structure on X. Choose an orientation for (1.9). The Seiberg-Witten invariant $SW(\mathcal{L})$ for $\mathcal{L}$ is defined as follows:

a) When $d < 0$ in (1.8), the invariant is defined to be zero.
b) When $d = 0$ in (1.8), choose μ in (1.7) to make M a smooth manifold. Then this M is a finite union of signed points and the Seiberg-Witten invariant is the sum over these points of the corresponding ± 1's.
c) When $d > 0$ in (1.8), choose μ in (1.7) to make M a smooth manifold. This M is compact and oriented, so it has a fundamental class. The Seiberg-Witten invariant is obtained by pairing this fundamental class with the maximal cup product of the first Chern class of the line bundle $M^0 \times_{S^1} \mathbb{C}$.

Note that the dimension of M and $b_2^+ + b_1$ have opposite parity; thus, the Seiberg-Witten invariants are zero when $b_2^+ + b_1$ is even.

The preceding definition is justified by

Proposition 1.2. *Let X be a compact, oriented, connected 4-manifold with $b_2^+ > 1$. Then SW defines a map from Spin to $\mathbb{Z}$ which depends only on the underlying smooth structure of X. That is, the value of $SW(\mathcal{L})$ is independent of the choice of metric and perturbing form μ in (1.7). It depends only on $\mathcal{L}$ up to isomorphism. Furthermore, the assignment of SW to a $Spin_{\mathbb{C}}$ structure is invariant under self-diffeomorphisms of X in the following sense: if φ is a diffeomorphism of X, then the value of SW on $\varphi^*\mathcal{L}$ is, up to sign, the same as the value of SW on $\mathcal{L}$.*

2. Symplectic manifolds

A 2-form ω on an oriented 4-manifold X is symplectic when

$$d\omega = 0 \quad \text{and} \quad \omega \wedge \omega \neq 0 \tag{2.1}$$

everywhere. Furthermore, the 4-form $\omega \wedge \omega$ will be required to orient X. A 4-manifold X with a symplectic form ω will be called a symplectic 4-manifold.

Every symplectic manifold has a canonical complex line bundle K; called the *canonical bundle*. Fix a Riemannian metric on X and then K can be identified as the orthogonal 2-plane bundle to the projection of ω into Λ_+. (The fact that $\omega \wedge \omega \neq 0$ implies that this last projection is nowhere zero.) Alternately, K can be defined by choosing an almost-complex structure for TX which is compatible (in the sense of Gromov [G]) for ω (see below). In this case, K is $\det(T^{1,0}X)$. The specification of such an almost-complex structure is equivalent to the specification of a metric on X for which ω is self-dual. Note that when $t \to \omega_t$ is a continuous, 1-parameter family of symplectic forms on X, then the canonical bundles for (X, ω_0) and (X, ω_1) will be isomorphic.

A symplectic manifold also has a canonical $Spin_{\mathbb{C}}$ structure (see [T1]). Indeed, use a metric for which ω is self-dual with length $\sqrt{2}$. For such a metric, the canonical $Spin_{\mathbb{C}}$ structure is characterized by the fact that its associated bundle S_+ is naturally isomorphic to $\mathbb{I} \oplus K^{-1}$, where $\mathbb{I}$ is the trivial complex line bundle. Here, ω acts by Clifford multiplication on the $\mathbb{I}$ summand with eigenvalue $-2i$, and it acts on K^{-1} summand with eigenvalue $+2i$. (When $t \to \omega_t$ is a continuous, 1-parameter family of symplectic forms on X, then the canonical $Spin_{\mathbb{C}}$ structures for ω_0 and ω_1 can be naturally identified.)

The defining of this canonical $Spin_{\mathbb{C}}$ structure allows one to identify *Spin*, the set of equivalence classes of $Spin_{\mathbb{C}}$ structures on X, with the set of equivalence classes of complex line bundles over X. (The latter is, of course, the same as $H^2(X; \mathbb{Z})$.) The $Spin_{\mathbb{C}}$ structure which corresponds to a given complex line bundle E is characterized by the fact that its bundle S_+ is given by

$$S_+ = E \oplus (K^{-1} \otimes E). \tag{2.2}$$

Note that Clifford multiplication by ω on S_+ in (2.2) preserves the splitting with the summand E having eigenvalue $-2i$.

For a symplectic manifold with $b_2^+ > 1$, [T2] has a proof that the Seiberg-Witten invariants are ± 1 for the cases $E = \mathbb{I}$, K in (2.2). (In general, the Seiberg-Witten invariant for E and for $K \otimes E^{-1}$ are equal in absolute magnitude. This follows from the quaternionic nature of the $Spin_{\mathbb{C}}$ representations [W].) It is proved in [T3] that a $Spin_{\mathbb{C}}$ structure as in (2.2) has nonzero Seiberg-Witten invariants only if

$$0 \leq [\omega] \cdot c_1(E) \leq [\omega] \cdot c_1[K], \tag{2.3}$$

where $[\omega]$ is the cohomology class of the symplectic form. (It is also proved in [T3] that left hand equality in (2.3) is achieved only by $E = \mathbb{I}$ and that the right hand equality is only achieved by $E = K$.)

By the way, because the line bundle L for (2.2) is $K^{-1} \otimes E^2$, the dimension d of M as given in (1.8) is simply

$$d = -c_1(K) \cdot c_1(E) + c_1(E) \cdot c_1(E). \tag{2.4}$$

The final remark of this section is to point out that the line in (1.9) has a canonical orientation on a symplectic manifold; the line in (1.9) is naturally isomorphic to the determinant line bundle of the elliptic operator $\delta \equiv (P_+ d, d^*) : C^\infty(T^*X) \to C^\infty(\mathbb{I}_\mathbb{R} \oplus \Lambda_+)$, where $\mathbb{I}_\mathbb{R}$ is the trivial real line bundle. The latter operator differs by a canonical zero'th order term from an operator, δ_ω, which intertwines the action of J on T^*X with the action of a square -1 endomorphism J' of $\mathbb{I}_\mathbb{R} \oplus \Lambda_+$. Because of this intertwining property, both the kernel and cokernel of δ_ω will be complex, and so its index bundle will have a natural orientation. (The endomorphism J' is defined as follows: write $\zeta \in \mathbb{I}_\mathbb{R} \oplus \Lambda_+$ as $\zeta_0 + \zeta_1 \cdot \underline{\omega} + \rho$, where ρ is orthogonal to ω in Λ_+ and where $\zeta_{0,1}$ are real numbers. Here, $\underline{\omega}$ is the unit length section $\omega/\sqrt{2}$. Then, set $J'(\zeta) \equiv -\zeta_1 + \zeta_0 \cdot \underline{\omega} + \underline{\omega} \times \rho$, where $\underline{\omega} \times (\cdot)$ at a point $p \in X$, is given by the cross product on $\Lambda_+\big|_p = \mathbb{R}^3$ by the unit vector defined $\underline{\omega}$. With J' understood, set $\delta_\omega \equiv \frac{1}{2} \cdot (\delta - J' \delta J)$. Note that δ_ω is the same as δ when X is complex and ω is the Kähler form.)

3. The Gromov invariants

Let X be a symplectic manifold. Then the tangent bundle to X admits an almost-complex structure, a homomorphism $J : TX \to TX$ with $J^2 = -1$. As with a genuinely complex manifold, the almost-complex structure J defines a splitting $T^*X \otimes \mathbb{C} \cong T^{1,0} \oplus T^{0,1}$, where J acts on T' as multiplication by i. The canonical bundle of the almost-complex structure is defined to be $K \equiv \det(T^{1,0})$.

An almost-complex structure J is deemed compatible with ω when

$$\begin{aligned} &\text{a)} \quad \omega(w, Jv) = -\omega(Jw, v). \\ &\text{b)} \quad \omega(v, Jv) \geq 0 \text{ with equality if and only if } v = 0. \end{aligned} \tag{3.1}$$

Here, ω is thought of as a skew symmetric form on TX. Since the space of compatible almost-complex structures is contractible, one can think of K as being associated to the symplectic structure ω. (Note that the assignment of (v, w) in TX to $\omega(v, Jw)$ defines a metric on TX. For such a metric, the form ω is self-dual. Correspondingly, any metric for which ω is self-dual can be used to construct an almost-complex structure J which satisfies (3.1).)

Let Σ be a compact, embedded 2-manifold of X. Call Σ a *pseudo-holomorphic submanifold* of X when J maps the tangent space of Σ to itself. (This Σ need not be connected.) Notice that if Σ is a pseudo-holomorphic submanifold, then J defines a complex structure on Σ and then the identity map of Σ (as a map from the complex curve Σ into X) is pseudo-holomorphic in the sense of Gromov [G]. See also [McS], [R]. (A map from a complex curve to X is called pseudo-holomorphic by Gromov when the differential intertwines the complex structure on the curve with the given J.)

The restriction of the symplectic form ω to a pseudo-holomorphic submanifold Σ is always symplectic, and so orients Σ. With this orientation understood, then each pseudo-holomorphic submanifold of X has a fundamental class which is *a priori* nontrivial in $H_2(X;\mathbb{Z})$. (It is nontrivial since the integral of ω over Σ is positive.) Use $e \equiv e(\Sigma) \in H^2(X;\mathbb{Z})$ to denote the Poincaré dual of this fundamental class.

An important feature of a connected pseudo-holomorphic submanifold Σ is that the class e determines the genus of Σ through the adjunction formula:

$$2 - 2g + e \cdot e = -c_1(K) \cdot e. \tag{3.2}$$

(Note that $e \cdot e$ is the degree of the normal bundle in X to Σ.)

Let $e \in H^2(X;\mathbb{Z})$ be a nontrivial cohomology class, and let $H \equiv H(e)$ denote the space of pseudo-holomorphic submanifolds of X whose fundamental class is Poincaré dual to e. Here, submanifolds count with multiplicity 1 except that tori with zero self-intersection can have multiplicity 1 or more. Here is the first structure theorem for H: (In the lemmas below, the term "generic J" signifies that J is to be chosen from a certain Baire subset of the space of smooth almost-complex structures J which obey (3.1).)

Lemma 3.1. *For a generic choice of J, this space H is a smooth manifold of dimension*

$$d = -c_1(K) \cdot e + e \cdot e \tag{3.3}$$

Furthermore, H has a canonical orientation.

(The tangent space to H at a given Σ can be naturally identified with the kernel of a certain first order elliptic operator on Σ. In fact, with the complex structure J on Σ, this operator is a zero'th order perturbation of the $\overline{\partial}$ operator coupled to the normal bundle of Σ. Since the components of Σ are assumed to be disjoint and embedded, the sum of the degrees of their normal bundles is equal to $e \cdot e$.)

See [T1] for the proofs of this lemma and of the subsequent assertions in this subsection. See also [R].

Note, by the way, that $c_1(K)$ reduces mod(2) to the 2nd Stieffel-Whitney class of X. This implies that (3.3) is always an even number. If d in (3.3) is positive, choose a set of $d/2$ distinct points Ω in X and define H_Ω as the subspace of submanifolds $\Sigma \in H$ which contain all points in Ω. Here is the structure theorem for H_Ω:

Lemma 3.2. *For a generic choice of J and points Ω, this $H_\Omega \subset H$ will be a canonically oriented 0-submanifold.*

(Remember that an oriented point is a point with a choice of ± 1.)

The preceding structure lemma, coupled with the following compactness lemma are the crucial ingredients for the definition of the Gromov invariants of the manifold X.

Lemma 3.3. *In the case $d = 0$ of (3.3), the space H is compact for a generic choice of J. When $d > 0$, the space H_Ω is compact for a generic choice of J and of the points Ω.*

Here is a remark concerning the compactness proof for the case where $d = 0$. (The general case is mostly similar.) A sequence in H can fail to converge by limiting to a multi-component, immersed surface, or, more generally, to a multi-component, immersed surface with singular points. With the help of the Smale-Sard theorem and dimension-counting arguments, one can prove that such limits are absent when the almost-complex structure J is suitably generic. Essentially, one argues that the possible limiting sets form H-like moduli spaces which are locally defined by equations with Fredholm linearizations whose indices are at least two less than the dimension d in (3.3). Then, because $d = 0$, these moduli spaces will have negative dimension and so the Smale-Sard theorem will guarantee their absence for the generic J.

For example, if the limit of a sequence in H is the image of a connected, immersed curve, then this curve will be pseudo-holomorphic with some number n of double points, all with positive intersection number (see, e.g. [McS]). (The local model for the degeneration is the zero locus of $z_1 z_2 = \epsilon$ in $\mathbb{C}^2$ as $\epsilon \to 0$.) The image in $H_2(X;\mathbb{Z})$ of the fundamental class of this curve will still represent the Poincaré dual to e.

An immersed curve has a well defined normal bundle with some degree ν. Here, ν is given by

$$\nu = e \cdot e - 2n. \tag{3.4}$$

Also, there is the adjunction formula analog of (3.2) which reads

$$2 - 2g' + \nu = -c_1(K) \cdot e, \tag{3.5}$$

where g' is the genus of the immersed curve. Furthermore, the immersed curve sits in a moduli space which is defined like H except that the sets in question are no longer submanifolds but images of immersions. For a suitably generic J, such a moduli space will be a smooth manifold of dimension $d' = 2 - 2g' + 2\nu$. With this understood, it follows from (3.4) and (3.5) that

$$d' = -c_1(K) \cdot e + e \cdot e - 2n \tag{3.6}$$

which is $2n$ less than the dimension d in (3.3). Thus, if d in (3.3) is zero, then d' in (3.6) will be negative and the generic J will admit no connected, immersed curves whose fundamental class represents the Poincaré dual to e.

Here is a definition of Gromov invariants for a symplectic 4-manifold:

Definition 3.4. Let X be a compact, symplectic 4-manifold. Define an invariant $Gr : H^2(X;\mathbb{Z}) \to \mathbb{Z}$ by assigning to a class e the following:

a) When $d < 0$ in (3.3), set $Gr(e) = 0$.
b) When $d = 0$ in (3.3), then H, as defined for a generic J, is a finite set of signed points, and $Gr(e)$ is the sum over the points in H of the corresponding ± 1's.
c) When $d > 0$ in (3.3), then H_Ω, as defined for a generic J and Ω, is a finite set of signed points, and $Gr(e)$ is the sum over the points in H of the corresponding ± 1's.

This definition is justified by the following proposition:

Proposition 3.5. *The invariant $Gr : H_2(X;\mathbb{Z}) \to \mathbb{Z}$ depends only on the pair (X, ω) of a compact, oriented 4-manifold and a symplectic form ω. That is, the value of Gr is independent of the choice of almost-complex structure J and, in the $d > 0$ case, of the set Ω. Furthermore, if $\{\omega_t : t \in [0,1]\}$ is a continuous family of symplectic forms on X, then the invariants Gr as defined for ω_0 and for ω_1 agree. Finally, Gr is invariant under self-diffeomorphisms of X in the following sense: if $\varphi : X \to X$ is a self-diffeomorphism, then the value of Gr (as defined by form ω) on a class e is, up to sign, the same as that of Gr (as defined by the pulled back form $\varphi^*\omega$) on the class $\varphi^* e$.*

4. The main theorems

With the Seiberg-Witten invariants and the Gromov invariants now defined, here is the main theorem:

Theorem 4.1. *Let X be a compact, oriented, symplectic 4-manifold. Let E be a nontrivial, complex line bundle over X and use E to define a $Spin_{\mathbb{C}}$ structure $\mathcal{L} \in Spin$ with S_+ as in (2.2). Then $SW(\mathcal{L}) = \pm Gr(c_1(E))$.*

(An outline of the proof is given in the next section; see [T1] for the details.)

There are some simple corollaries concerning the Seiberg-Witten invariants and symplectic curves which follow from this theorem. For example, the first application is simply an existence theorem for symplectic curves:

Proposition 4.2. *Let X be as in Theorem 4.1 and let E be a nontrivial complex line bundle for which the associated $Spin_{\mathbb{C}}$ structure (as in (2.2)) has nonzero Seiberg-Witten invariant. Then the Poincaré dual to $c_1(E)$ is represented by the fundamental class of an embedded, symplectic curve which consists of some number N of components. Let Σ be any such component, let $g = genus(\Sigma)$ and let $e \in H^2(X;\mathbb{Z})$ be the Poincaré dual to the fundamental class of Σ. Then $g = 1 + e\cdot e$. In particular, the Poincaré dual to $c_1(K)$ is represented by a symplectic curve. Also, if X has no embedded spheres of self-intersection -1, then $c_1(K)\cdot c_1(K) \geq 0$.*

Proposition 4.3. *Symplectic manifolds as in Theorem 4.1 have "simple type" in the sense that the nonzero Seiberg-Witten invariants are found only for $Spin_{\mathbb{C}}$ structures $\mathcal{L}$ for which the dimension d in (1.8) (or, equivalently, (2.4)) is equal to zero.*

The proof that d in (2.4) is zero for line bundles E with $c_1(E)\cdot c_1(E) \geq 0$ and SW nonzero employs the usual adjunction formula of (3.2) along with the generalized adjunction inequalities from [KM] and [MST]. The latter assert the following: let X be a compact, oriented 4-manifold with $b_2^+ > 1$ and let $e' \in H^2(X;\mathbb{Z})$ have nonnegative $e'\cdot e'$. Define g' to be the minimal genus of an embedded surface whose fundamental class represents the Poincaré dual to e'. Then $2-2g'+e'\cdot e' \leq -|c_1(L)\cdot e'|$ where $L = \det(S_+)$ for any $Spin_{\mathbb{C}}$ structure with nonzero SW invariant.

Proposition 4.4. *Let X be as in Theorem 4.1, but suppose that X admits no symplectically embedded 2-spheres with self-intersection number -1. Then the signature of the intersection form of X and the Betti numbers of X are constrained by the inequality $-\frac{4}{3}(1-b_1) - \frac{2}{3}b_2 \leq sign(X)$.*

This last proposition simply combines the previous two with the dimension formula for d in (1.8).

Proposition 4.5. *Let X be as in Theorem 4.1 and suppose that X admits an embedded 2-sphere whose fundamental class has Poincaré dual e obeying $e \cdot e = -1$ and $c_1(K) \cdot e \neq 0$. Then X admits a symplectically embedded 2-sphere in the same homology class as the original 2-sphere.*

The proof of this last proposition uses the fact that when X admits an embedded sphere of self-intersection -1, there is a diffeomorphism of X which acts on $H_2(X;\mathbb{Z})$ as reflection in the hyperplane orthogonal to the class of the given 2-sphere.

Proposition 4.6. *Let X be as in Theorem 4.1 and suppose that X admits no symplectically embedded 2-spheres with self-intersection number -1. Suppose, as well, that $c_1(K) \cdot c_1(K) = 0$. Then any nontrivial line bundle E which corresponds (via (2.2)) to a $Spin_{\mathbb{C}}$ structure $\mathcal{L} \in Spin$ having nonzero Seiberg-Witten invariant is Poincaré dual to a union of disjoint, embedded, symplectic tori (possibly with multiplicities) with zero self-intersection.*

The proof of this last proposition also uses the generalized adjunction inequalities from [KM] and [MST]. Indeed, $c_1(K)$ is Poincaré dual to the sum of the fundamental classes of a disjoint union of embedded, symplectic tori of square zero. Apply the generalized adjunction formula to each of these tori for the line bundle $L = K^{-1} \otimes E^2$ to prove that $c_1(K) \cdot c_1(E) = 0$. This implies (via Proposition 4.3) that $c_1(E) \cdot c_1(E) = 0$ and thus, that $c_1(E)$ is also Poincaré dual to a disjoint union of embedded, symplectic tori of square zero. (Use Proposition 4.2.)

As will be outlined in the next subsection, the proofs of the preceding theorems use a 1-parameter deformation of the Seiberg-Witten equations (as in (1.7)) which produces, in the limit, a pseudo-holomorphic curve from the projection of ψ onto the E summand in (2.2). This existence construction works for $b_2^+ = 1$ symplectic manifolds also. However, in this case, the value of the Seiberg-Witten invariant can change along the deformation. However, the amount of change can be analyzed.

In particular, when this deformation is applied to the case where X is $\mathbb{CP}^2$ with a possibly nonstandard symplectic form, the result is a construction of a pseudo-holomorphic sphere whose fundamental class generates the second homology. With this understood, Theorem B of the introduction is obtained by referring to a theorem of Gromov in [G] which asserts that any such X must be diffeomorphic to the standard $\mathbb{CP}^2$ by a diffeomorphism which takes the given symplectic form to a multiple of the Kähler form. The details appear in [T1].

This section ends with the remark that Theorem 4.1 has an analog which can be used to analyze the Seiberg-Witten Floer cohomology of a compact, oriented 3-manifold M which fibers over the circle. Indeed, there is an equivalence between this Floer theory and a certain symplectic Floer cohomology which can also be defined in this context [F]. The results are reminiscent of the description in [DS] of the $SO(3)$ gauge theoretic Floer cohomology of the same sort of manifold. Anyway, the details here will be described elsewhere [T4].

5. An outline of the proof of Theorem 4.1

There are roughly three parts to the proof of Theorem 4.1. The first part produces pseudo-holomorphic submanifolds of X from solutions to the Seiberg-Witten equations in (1.7). The second part (an inverse of sorts to the first) produces solutions to a Seiberg-Witten equation (as in (1.7)) from a pseudo-holomorphic submanifold. The third part combines the constructions in Parts 1 and 2 to produce a demonstration of the stated equivalence between the Seiberg-Witten and Gromov invariants.

Of the three parts above, the first part is technically the hardest by a good margin, while the second part is less difficult. The strategies for the first two parts of the proof are presented below. Part 3 of the proof consists of fairly formal corollaries to the propositions of Parts 1 and 2. Thus, this part of the proof will not be discussed here.

The strategy for Part 1: As remarked earlier, pseudo-holomorphic submanifolds of X are obtained by considering the behavior of solutions to (1.7) where the perturbation form μ becomes singular in a controllable way as a parameter r gets large. The analysis here has six steps.

Step 1 (The definition of μ): Use a metric on X where the symplectic form ω is self-dual with norm $\sqrt{2}$. With this understood, recall that the canonical $Spin_{\mathbb{C}}$ structure has its bundle S_+ given by

$$S_+ = \mathbb{I} \oplus K^{-1}. \tag{5.1}$$

And, recall from [T2] that the bundle K^{-1} has a unique connection (up to gauge equivalence), A_0, which is characterized as follows: the composition of the spin covariant derivative with orthogonal projection onto the $\mathbb{I}$ summand in (5.1) defines a covariant derivative, ∇_{A_0}, on said summand which annihilates a nowhere-vanishing section. This section, u_0, can be taken to have unit length.

With the preceding understood, take $r > 0$ and consider now the $Spin_{\mathbb{C}}$ structure in (2.2) and the perturbed Seiberg-Witten equation for the pair (A, ψ) where A is a connection on $L = K^{-1} \otimes E^2$ and ψ is a section of $S_+ = E \oplus (K^{-1} \otimes E)$:

$$D_A\psi = 0 \quad \text{and} \quad P_+F_A = \frac{1}{4}\tau(\psi \otimes \psi^*) + P_+F_{A_0} - \frac{ir}{4} \cdot \omega. \tag{5.2}$$

In analyzing the behavior of the solutions to (5.2) for r large, it proves convenient to rewrite ψ as

$$\psi = r^{1/2} \cdot (\alpha \cdot u_0 + \beta), \tag{5.3}$$

where α is a section of E and where β is a section of $K^{-1} \otimes E$. Noting that u_0 is a solution to the Dirac equation $D_{A_0} u_0 = 0$, one can rewrite the first equation in (5.2) as

$$\sigma(u_0 \otimes \nabla_a \alpha) + D_A \beta = 0, \tag{5.4}$$

where ∇_a is the induced covariant derivative which is induced on E from the covariant derivatives ∇_{A_0} on K^{-1} and ∇_A on $K^{-1} \otimes E^2$. (Note that $F_a = 2^{-1} \cdot (F_A - F_{A_0})$.) Here, σ is the Clifford multiplication map in (1.4).

Step 2 (L^2-estimates on β): Act on both sides of (5.4) with D_A and commute derivatives to obtain an equality between the covariant Laplacian of the spinor ψ and zero'th order terms which are linear and cubic in ψ. (Use (5.2) here to express the bundle curvature in terms of ψ itself.)

Take this latter equation and project onto the $K^{-1} \otimes E^2$ summand of (2.2). Then take the L^2-inner product of the result with β. An application of integration by parts will yield, with little difficulty, a lemma to the effect that the L^2 norms of $\nabla_A \beta$ and of $r^{1/2} \cdot \beta$ are bounded from above by a uniform multiple of $r^{-1/2}$ times the L^2 norm of $\nabla_a \alpha$.

These last facts imply that α is almost-"holomorphic" for large r in the sense that the L^2 norm of the projection of $\nabla_a \alpha$ onto $E \otimes T^{0,1}$ is bounded from above by a uniform multiple of $r^{-1/2}$ times the projection of $\nabla_a \alpha$ on $E \otimes T^{1,0}$. The point is that for large r, the section α of E wants to be holomorphic and the section β of $K^{-1} \otimes E^2$ wants to be zero.

Step 3 (L^2 estimates on α): Take the equation for the covariant Laplacian of ψ and project onto the E summand in (2.2). Take the L^2 inner product of both sides of this identity with $\alpha \cdot u_0$ and integrate by parts. The result is a uniform bound, independent of r, of

$$\int \{ \frac{4}{r} |P_+ F_a|^2 + |\nabla_a \alpha|^2 + \frac{r}{8} (1 - |\alpha|^2)^2 \} \tag{5.5}$$

by a multiple of $c_1(K) \cdot c_1(E)$. (The existence of such a bound gives another proof of the results in [T2] and [T3].) In fact, the uniform bound on (5.5) is part of a "Bogomolny" energy bound for solutions to (5.2) which is a generalization to (5.2) of an energy bound which was introduced by Witten in [W] for solutions to (1.6).

As r gets large, the uniform bound on (5.5) gives a uniform bound on $r^{1/2} \cdot (1 - |\alpha|^2)$ and also $\nabla_a \alpha$ in the L^2 sense. Indeed, for large r, $|\alpha|^2$ would like to equal one everywhere, and $\nabla_a \alpha$ would like to equal zero everywhere. (One can prove that the latter is small on balls where the former is near 1.) However, both are prevented from vanishing everywhere by the assumption that the bundle E is nontrivial.

In any event, one can conclude that the measure of the set where $|\alpha|^2 < 1/2$ is bounded from above by a multiple of r^{-1} for large r. (Note that a maximum principle argument which is akin to one in [KM] proves that $|\alpha|$ is uniformly bounded by $1 + O(r^{-1})$.)

Step 4 (Control of $\alpha^{-1}(0)$): The goal here is to obtain a reasonable limit for the sequence of zero sets (i.e., $\alpha^{-1}(0)$) for a sequence of α's given by different values of r as r tends to ∞. The key to controlling this limit is a monotonicity formula for the square of the L^2 norm of $r^{1/2} \cdot (1 - |\alpha|^2)$ and also of $\nabla_a \alpha$ in balls of radius $s > 0$. The aim here is to prove that this local energy grows at least as fast as s^2 when α vanishes at the center of the ball. (The estimates must be uniform in r.) The existence of such a monotonicity formula readily implies that the sequence of zero sets for a sequence of α's as $r \to \infty$ converges nicely on X. Indeed, the zero sets are constrained as r gets large for the following reason. The monotonicity estimate and the energy bound for (5.5) imply the following: there is a constant z (which is independent of r) such that for any $\delta > 0$, the zero set of α is contained in some number $N(\delta) \le z \cdot \delta^{-2}$ balls of radius δ whenever r is sufficiently small.

By the way, the sort of monotonicity formula needed above exists for many types of nonlinear equations. For example, a similar formula plays a key step in the approach of Schoen and Uhlenbeck [SU] to harmonic map regularity. Also, a similar monotonicity formula plays the key role in Uhlenbeck's removable singularity theorem for solutions to the 4-dimensional Yang-Mills equations [U]. In these last examples, the monotonicity formulae are used to control the behavior of singular sets of the relevant nonlinear differential equation.

Step 5 (Pointwise bounds on F_a): In the present context, the proof of the required monotonicity formula requires some extremely delicate *a priori* bounds for the curvature F_a. The required bounds for $P_+ F_a$ are direct consequences of the second equation in (5.2). However, $P_- F_a$ does not directly appear in (5.2); its bounds must come via some other argument.

Miraculously, the exact bound required can be derived via the maximum principle and the equation for $d^* d(P_- F_a)$. In this regard, note that $d^* dP_- F_a = -d^* dP_+ F_a$ from the Bianchi identity ($dF_a = 0$) so (5.2) can be used to control $P_- F_a$ also. Via a standard Bochner-Weitzenbock formula, the equation for $d^* d(P_- F_a)$ gives an inequality for the Laplacian of $|P_- F_a|$. It is the latter inequality which implies, via the maximum principle, the required pointwise bound,

$$|P_-F_a| \leq \frac{r}{4\sqrt{2}} \cdot (1 - |\alpha|^2) + \text{ small, irrelevant term} \tag{5.6}$$

Step 6 (The pseudo-holomorphic limit): With the bounds on F_a coming from (5.2) and (5.6), the monotonicity formula can be proved. As remarked, the latter gives a limit for the zero sets of an $r \to \infty$ sequence of α. It remains yet to prove that this limiting set is pseudo-holomorphic. A dilation argument is used to accomplish this last goal. Indeed, dilation to unit radius of the radius $r^{1/2}$ ball about a point in this limit set rescales r out of the picture. The aforementioned *a priori* estimates for F_a show, via reasonably standard elliptic estimates, that the sequence of rescaled fields (corresponding to the unbounded sequence or r values) has a subsequence which converges in the C^∞ topology on compact neighborhoods of the given point. Because of the aforementioned small size of $\nabla_a\alpha$'s projection into $E \otimes T^{0,1}$, the limiting α can be shown to be holomorphic on the unit ball in $\mathbb{R}^4 = \mathbb{C}^2$. This last fact can be parlayed into a proof that the limiting zero set has the asserted pseudo-holomorphicity.

The strategy for Part 2: Part 2 of the proof of Theorem 4.1 gives the details for constructing a solution to (5.2) for large r from a pseudo-holomorphic submanifold. The idea here is to "graft" a rescaled (by a factor of $r^{1/2}$), canonical solution of vortex equation on $\mathbb{R}^2 = \mathbb{C}$ into the normal bundle of the pseudo-holomorphic submanifold. After grafting, the resulting (A, ψ) will almost solve (5.2). (The error in solving (5.2) vanishes in the limit as r tends to ∞.)

For large enough r, an honest solution to (5.2) can be obtained by correcting the grafted solution by solving a certain first order, inhomogeneous differential equation with small inhomogeneous part. For sufficiently large r, a contraction mapping argument (perturbation theory) finds a small solution to this inhomogeneous differential equation provided that the original pseudo-holomorphic submanifold was a smooth point in H. (The difference between the honest solution and the grafted solution will tend to zero as r tends to ∞.) The strategy here is much like that in [T5] for finding connections on principle $SU(2)$ bundles with self-dual curvature forms.

The canonical vortex equation on $\mathbb{C}$ is an equation for a connection, A, (on the trivial bundle) and a complex-valued function α. The equations read

$$\overline{\partial}_A\alpha = 0 \quad \text{and} \quad *\,iF_a = \frac{1}{8}(1 - |\alpha|^2). \tag{5.7}$$

These equations are supplemented with the boundary conditions:

$$\begin{aligned} &1)\quad |\alpha| \to 1 \text{ as } |z| \to \infty. \\ &2)\quad \text{The integral of } iF_a \text{ over } \mathbb{C} \text{ has value } 2\pi. \end{aligned} \tag{5.8}$$

(A part of the proof of Theorem 4.1 requires considering (5.8.2) with $2\pi n$ for $n > 1$ also.)

6. Further remarks

It is observed in [KMT] that there are manifolds with nonzero Seiberg-Witten invariants which do not admit symplectic forms. (The examples in [KMT] are all non-simply connected and they are connect sums of symplectic manifolds with negative definite manifolds.) With this understood, one is led to ask whether there is any sort of "Gromov invariant" interpretation for the Seiberg-Witten invariants in the nonsymplectic world. The proof of Theorem 4.1 does suggest such an interpretation. Indeed, let X be a compact, oriented 4-manifold with $b_2^+ > 0$ and let ω be a nontrivial, anti-self-dual harmonic 2-form. If X is not symplectic, then this form ω must vanish at points of X. Also, it is not hard to prove that the set where ω vanishes is, for a sufficiently generic metric, a disjoint union of embedded circles. Anyway, ω can be thought of as a degenerate symplectic form.

Choose $\mu = -i \cdot r \cdot \omega$ in (1.7). Some of the estimates for Theorem 4.1 can still be derived in this case. These estimates suggest that there is a theory of pseudo-holomorphic submanifolds (in dimension 4) for degenerate symplectic forms. However, at present, the very existence of such a theory, never mind its properties, has yet to be established.

The final remark concerns a different subject—supersymmetric quantum field theories. The Seiberg-Witten equations were discovered by Seiberg and Witten while investigating the physical predictions of a certain postulated supersymmetric quantized Yang-Mills theory [SW1], [SW2]. The analysis here suggests that the relevant quantized gauge field theory is a deformation of some superstring quantum field theory.

Acknowledgement

The author wishes to acknowledge his extreme gratitude for the advice, suggestions and support of Tomasz Mrowka.

References

[DS] S. Dostoglu and D. Salamon, *Self dual instantons and holomorphic curves*, Ann. Math. **139** (1994), 581–640.

[F] A. Floer, *Symplectic fixed points and holomorphic spheres*, Commun. Math. Phys. **120** (1989), 575–611.

[FU] D. Freed and K. K. Uhlenbeck, *Instantons and four manifolds*, Springer-Verlag New York, 1984.

[G] M. Gromov, *Pseudo-holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), 307–347.

[KM] P. B. Kronheimer and T. S. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Letters **1** (1994), 119–124.

[McS] D. McDuff and D. Salamon, *J-holomorphic curves and quantum cohomolgy*, preprint (1994).

[MST] J. Morgan, Z. Szabo and C. H. Taubes, *The generalized Thom conjecture*, in preparation.

[KMT] D. Kotschick, J. Morgan, C. H. Taubes, *Four manifolds without symplectic structures but with nontrivial Seiberg-Witten invariants*, Math. Res. Letters **2** (1995), 797–808.

[R] Y. Ruan, *Symplectic topology and complex surfaces*, in Geometry and topology of complex surfaces (1994), World Scientific.

[SU] R. Schoen and K. K. Uhlenbeck, *A regularity theory for harmonic maps*, J. Diff. Geo. **17** (1982), 307–335.

[SW1] N. Seiberg and E. Witten, *Electromagnetic duality, monopole condensation and confinement in $N = 2$ supersymmetric Yang-Mills theory*, Nucl. Phys. **B426** (1994), 19–52.

[SW2] ———, *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*, Nucl. Phys. **B431** (1994), 581–640.

[T1] C. H. Taubes, in preparation.

[T2] ———, *The Seiberg-Witten invariants and symplectic forms*, Math. Res. Letters **1** (1994), 809–822.

[T3] ———, *More constraints on symplectic manifolds from Seiberg-Witten equations*, Math. Res. Letters **2** (1995), 9–14.

[T4] ———, in preparation.

[T5] ———, *Self-dual connections on non-self-dual 4-manifolds*, J. Diff. Geo. **17** (1982), 139–170.

[U] K. K. Uhlenbeck, *Removable singularities in Yang-Mills fields*, Commun. Math. Phys. **83** (1982), 11–29.

[W] E. Witten, *Monopoles and 4-manifolds*, Math. Res. Letters **1** (1994), 769–796.

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