ISSN 1364-0380 (on line) 1465-3060 (printed)

Geometry & Topology
Volume 7 (2003) 965–999
Published: 21 December 2003

A non-abelian Seiberg–Witten invariant for integral homology 3–spheres

Yuhan Lim

Dept of Mathematics, Univ of California
Santa Cruz, CA 95064, USA

Email: ylim@math.ucsc.edu

Abstract

A new diffeomorphism invariant of integral homology 3–spheres is defined using a non-abelian "quaternionic" version of the Seiberg–Witten equations.

AMS Classification numbers Primary: 57R57

Secondary: 57M27

Keywords: Seiberg–Witten, 3–manifolds

Proposed: Ronald Stern
Seconded: Ronald Fintushel, John Morgan

Received: 9 January 2003
Revised: 10 December 2003

1 Introduction

The Seiberg–Witten equations when applied to the study of oriented integral homology 3–spheres yield an invariant which was shown in [9] to coincide with Casson's invariant. In [3], Boden and Herald introduced a generalization of Casson's invariant from $SU(2)$ to the higher structure group $SU(3)$ based on the gauge theory approach of Taubes [12]. This $SU(3)$–Casson invariant utilizes values of the Chern–Simons function which makes it a real valued invariant rather than an integral one. In the present article we define a non-Abelian version of the Seiberg–Witten equations which we call *quaternionic* and construct a topological invariant of integral homology 3–spheres in a manner parallel to the $SU(3)$–Casson invariant. This new invariant has the property that it is independent of orientation of the 3–manifold and a linear combination with the $SU(3)$–Casson invariant gives a $\mathbf{Z}$ mod $4\mathbf{Z}$ invariant for unoriented integral homology 3–spheres.

The contents of this article are as follows. In section 2 we introduce the generalization of the SW–equations we use. The technical issue of admissible perturbations is also discussed. We use the novel approach of non-gradient perturbations. Section 3 gives the main results which are Theorems 3.7 and 3.8. The remaining sections take up the proofs. We assume the reader has some familiarity with [3], [9] and [12].

2 Quaternionic gauge theory in 3–dimensions

Standing Convention *Throughout this article Y will denote an oriented closed integral homology 3–sphere (ZHS). Y will also be assumed to have a fixed Riemannian metric g.*

The aim is to introduce a quaternionic setting in which the Seiberg–Witten equations will make sense. Since Y is a ZHS it has a unique spin structure, up to equivalence. With respect to g this is given by a principal $spin(3) \cong SU(2)$ bundle $P \to Y$. In the (real) Clifford bundle $CL(T^*Y) \cong CL(Y)$ the volume form ω_Y has the property that $\omega_Y^2 = 1$. The action of ω_Y on $CL(Y)$ induces a splitting into ± 1 eigenbundles $CL^+ \oplus CL^-$. Both CL^+ and CL^- are bundles of algebras over Y with each fibre isomorphic, as an algebra, to the quaternions $\mathbf{H}$.

Let $S \to Y$ be the complex spinor bundle on which CL^+ acts non-trivially. This is a rank 2 complex Hermitian vector bundle. Since the fibres of CL^+

are quaternionic vector spaces, S possesses an additional action by $\mathbf{H}$ which commutes with the Clifford action (see [8]); we may take this to be a right action $S \times \mathbf{H} \to S$.

Suppose now that $E \to Y$ is a given fixed rank one metric quaternionic vector bundle — we assume the action by $\mathbf{H}$ is a left action. E also has a description as a complex Hermitian rank 2 vector with trivial determinant, i.e. with structure group $SU(2)$. We can twist the spinor bundle S by tensoring with E over the quaternions to form the bundle $S \otimes_{\mathbf{H}} E$. This is a *real* rank 4 Riemannian vector bundle and does not naturally inherit a complex structure from S or E.

Given an $SU(2)$–connection A on E (henceforth any connection on E mentioned will be assumed to be such type) we may construct using the canonical Riemiannian connection on S, a metric (i.e. $SO(4)$) connection on $S \otimes_{\mathbf{H}} E$. This then defines in the usual way a Dirac operator

$$D_A = \sum_{i=1}^{3} e_i \cdot \nabla^A_{e_i}.$$

Here the e_i are an orthonormal frame and ∇^A is the connection on $S \otimes_{\mathbf{H}} E$ mentioned above. We emphasize that D_A is in general only a real linear operator on $S \otimes_{\mathbf{H}} E$.

Lemma 2.1 *The complexification of $S \otimes_{\mathbf{H}} E$ is naturally isomorphic as a complex Clifford module with $S \otimes_{\mathbf{C}} E$. Under this isomorphism the complexification $D_A \otimes \mathbf{C}$ corresponds to the complex Dirac operator $D_A^{\mathbf{C}}$.*

Proof Introduce the notation $\overline{\otimes}$ to denote the tensor product of elements in $S \otimes_{\mathbf{H}} E$ and $\otimes$ the complex tensor product in $S \otimes_{\mathbf{C}} E$. Define the vector bundle map h from $S \otimes_{\mathbf{C}} E$ to $(S \otimes_{\mathbf{H}} E) \otimes \mathbf{C}$ by

$$h(e \otimes f) = e \overline{\otimes} f - \sqrt{-1}(ei \overline{\otimes} f).$$

One checks directly that this map is a complex isomorphism and commutes with Clifford multiplication. □

Since the real two forms Λ^2 naturally include in $CL(Y)$ we have by Clifford mutiplication the action of Λ^2 on S. This representation of Λ^2 on S is well-known to be injective and with image the adjoint bundle $\mathrm{ad}S$, the bundle of skew-Hermitian transformations of S. The bundle $\mathrm{ad}E$ acts on E from the left. Define an action of $\Lambda^2 \otimes \mathrm{ad}E$ on $S \otimes_{\mathbf{H}} E$ by the rule

$$(\omega \otimes l) \cdot (\phi \otimes e) := (\omega \cdot \phi) \otimes l(e).$$

This is well-defined since the actions of Λ^2 and $\mathrm{ad}E$ commute with the quaternionic structures.

Remark 2.2 The Clifford action of $\beta \in \Lambda^2$ is the same as the action of $-*\beta \in \Lambda^1$ on S since the volume form ω_Y acts by the identity. Thus we may equivalently work (up to multiplication by -1) with the action of $\Lambda^1 \otimes \mathrm{ad}E$ on $S \otimes_{\mathbf{H}} E$.

Lemma 2.3 *The representation $\Lambda^2 \otimes \mathrm{ad}E \to \mathrm{End}_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$ above is injective and has image the subbundle $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$ of trace zero real symmetric transformations of $S \otimes_{\mathbf{H}} E$.*

Proof That the representation of the lemma is injective is easily verified. We may rewrite the action of $\Lambda^2 \otimes \mathrm{ad}E$ as $(\omega \otimes l) \cdot (\phi \otimes e) = -(i\omega \cdot \phi) \otimes il(e)$. Since $i\mathrm{ad}S$ is exactly the trace zero Hermitian symmetric bundle endomorphisms of S, and similiarly for $i\mathrm{ad}E$, the image of the representation clearly lies in the trace zero real symmetric endomorphisms of $S \otimes_{\mathbf{H}} E$. That it is onto follows by a dimension count giving both $\Lambda^2 \otimes \mathrm{ad}E$ and $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$ real vector bundles of rank 9. □

The above lemma shows that we may regard the bundle $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$ as identical to $\Lambda^2(Y) \otimes \mathrm{ad}E$. Thus whenever convenient we can think of a trace zero real symmetric endomorphism of $S \otimes_{\mathbf{H}} E$ as a twisted 2–form with values in $\mathrm{ad}E$.

Lemma 2.4 *There is a unique fibrewise symmetric bilinear form $\{\cdot\}_0$ on $S \otimes_{\mathbf{H}} E$ with values in $\Lambda^2 \otimes \mathrm{ad}E$ determined by the rule that*

$$\langle \omega, \{\phi \cdot \psi\}_0 \rangle = \langle \omega \cdot \psi, \phi \rangle = \langle \omega \cdot \phi, \psi \rangle$$

holds for all sections ω of $\Lambda^2 \otimes \mathrm{ad}E$. As a section of $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$, $\{\phi \cdot \psi\}_0$ is given by the expression

$$\{\phi \cdot \psi\}_0 = \frac{1}{2}\left(\phi \otimes \psi^* + \psi \otimes \phi^* - \frac{1}{2}\langle \phi, \psi \rangle \mathrm{I}\right).$$

Here $\phi \otimes \psi^*(\nu) = \phi \langle \nu, \psi \rangle$ and similiarly for $\psi \otimes \phi^*$.

Proof Let $\{\phi_i\}$, $\{\omega_j\}$ be a local orthonormal frames for $S \otimes_{\mathbf{H}} E$, $\Lambda^2 \otimes \mathrm{ad}E$ respectively. Let $\{\phi_i \cdot \phi_j\}_0 = c^k_{i,j}\omega_k$. Then we see that $c^k_{i,j} = \langle \omega_k \cdot \psi_i, \psi_j \rangle = \langle \omega_k \cdot \psi_j, \psi_i \rangle = c^k_{j,i}$ determines $\{\cdot\}_0$. Identify $\Lambda^2 \otimes \mathrm{ad}E$ with $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$. In a local trivialization we may regard sections of $\mathrm{Sym}^0_{\mathbf{R}}(S \otimes_{\mathbf{H}} E)$ as functions with values in $\mathrm{Sym}^0_{\mathbf{R}}(\mathbf{R}^4)$, the 4×4 real symmetric matrices, and sections of $S \otimes_{\mathbf{H}} E$ as $\mathbf{R}^4$–valued functions. As such the inner product in $\mathrm{Sym}^0_{\mathbf{R}}(\mathbf{R}^4)$ is given by $\langle M,N\rangle = \mathrm{Tr}(MN)$. The right side of the defining equation for $\{\cdot\}_0$ can be expressed locally as

$$\begin{aligned}\frac{1}{2}\Big(\mathrm{Tr}(W\Phi\Psi^T)+\mathrm{Tr}(W\Psi\Phi^T)\Big) &= \frac{1}{2}\mathrm{Tr}(W(\Phi\Psi^T+\Psi\Phi^T)) \\ &= \langle W, \frac{1}{2}(\Phi\Psi^T+\Psi\Phi^T)\rangle.\end{aligned}$$

The claimed expression for $\{\cdot\}_0$ is exactly the trace-free component of the symmetric expression $\frac{1}{2}(\Phi\Psi^T+\Psi\Phi^T)$. □

The *configuration space* $\mathcal{C}$ is the space of all pairs (A,Φ) consisting of an $SU(2)$–connection A on E and a section Φ of $S\otimes_{\mathbf{H}}E$. As usual we should work within the framework of a certain functional space; for us choose A and Φ to be of class L^2_2 (for A this means $A-A_0$ is L^2_2 where A_0 is a fixed C^∞–connection). $\mathcal{C}$ is an affine space modelled on the Hilbert space

$$L^2_2(\Lambda^1\otimes \mathrm{ad}E)\times L^2_2(S\otimes_{\mathbf{H}}E).$$

The *gauge automorphism group* $\mathcal{G}$ in this case will consist of the L^2_3–bundle automorphisms which preserve the quaternionic structure of E, or equivalently the L^2_3–sections of $\mathrm{Ad}E$. Since $L^2_2\subset C^0$ in dimension 3, $\mathcal{C}$ and $\mathcal{G}$ consists of continuous objects. $\mathcal{G}$ acts on $\mathcal{C}$ by $g\cdot(A,\Phi)=(g(A),g^{-1}\Phi)$. This action is differentiable and the quotient we denote by $\mathcal{B}$. Our convention is that $g(A)$ is the pull-back of A by g.

We have the following observation: the stabilizer

$$\mathrm{stab}(A,\Phi)=\begin{cases}\{1\} & \text{if } \Phi\neq 0\\ \mathrm{stab}(A) & \text{if } \Phi=0.\end{cases}$$

The possible choices for $\mathrm{stab}(A)$ are $\{\pm1\}$, $U(1)$ or $SU(2)$. Note that in the last possibility A is necessarily a trivial connection. The pair (A,Φ) is *irreducible* if $\Phi\neq 0$ and *reducible* otherwise. Thus $\mathcal{G}$ acts freely on $\mathcal{C}^*$, the irreducible portion of $\mathcal{C}$ and the quotient $\mathcal{C}^*$ by $\mathcal{G}$ is denoted $\mathcal{B}^*$.

$\mathcal{G}$ is a Hilbert Lie group with tangent space at the identity $T_e\mathcal{G}=L^2_3(\mathrm{ad}E)$. Let $\mathcal{G}\to\mathcal{C}$, $g\mapsto(g(A),g^{-1}\Phi)$ be the map which is the orbit of (A,Φ) under the action of $\mathcal{G}$. The derivative at the identity is the map

$$\begin{gathered}\delta^0_{A,\Phi}\colon L^2_3(\mathrm{ad}E)\to L^2_2(\Lambda^1\otimes\mathrm{ad}E)\oplus L^2_2(S\otimes_{\mathbf{H}}E),\\ \delta^0_{A,\Phi}(\gamma)=(d_A\gamma,-\gamma(\Phi)).\end{gathered}\tag{2.1}$$

A slice for the action of $\mathcal{G}$ on $\mathcal{C}$ at (A,Φ) is given by $(A,\Phi)+X_{A,\Phi}$ where $X_{A,\Phi}$ is the *slice space* which is the L^2–orthogonal complement in $L^2_2(\Lambda^1\otimes\mathrm{ad}E)\oplus$

$L^2_2(S \otimes_{\mathbf{H}} E)$ of the image of $\delta^0_{A,\Phi}$. We may also regard $X_{A,\Phi}$ as the tangent space to $\mathcal{B}^*$ at an irreducible orbit $[A,\Phi]$.

Define a bilinear product B: $(S \otimes_{\mathbf{H}} E) \otimes (S \otimes_{\mathbf{H}} E) \to \mathrm{ad}E$ by the rule that $\langle \gamma(\phi), \psi \rangle = \langle \gamma, B(\psi,\phi) \rangle$ holds for all $\gamma \in \mathrm{ad}E$. Then $X_{A,\Phi}$ has the description as the subspace of $L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2_2(S \otimes_{\mathbf{H}} E)$ defined by the equation

$$\delta^{0*}_{A,\Phi}(a,\psi) = 0 \quad \Longleftrightarrow \quad d^*_A a - B(\Phi,\psi) = 0. \tag{2.2}$$

A reducible we will often simply denote by A instead of $(A,0)$. Corresponding reducible subspaces of $\mathcal{C}$ and $\mathcal{B}$ are denoted $\mathcal{A}$ and $\mathcal{B}_{\mathcal{A}}$. At a reducible A the slice X_A splits into a product $X^r_A \times L^2_2(S \otimes_{\mathbf{H}} E)$ where X^r_A is the slice for the action of $\mathcal{G}$ on $\mathcal{A}$. Then the normal space to $\mathcal{B}_{\mathcal{A}}$ in $\mathcal{B}$ near $[A]$ is modelled on

$$L^2_2(S \otimes E)/\mathrm{stab}(A).$$

For instance if A is irreducible as a connection then this normal space is a cone on the quotient of the unit sphere in a separable Hilbert space by the antipodal map $v \mapsto -v$.

On $\mathcal{C}$ we have the Chern–Simons–Dirac function csd: $\mathcal{C} \to \mathbf{R}$ (with respect to a choice of trivial connection Θ say) given by

$$\mathrm{csd}(A,\Phi) = \frac{1}{8\pi^2}\int_Y \mathrm{Tr}\Big(a \wedge d_\Theta a + \frac{2}{3} a \wedge a \wedge a\Big) - \int_Y \langle D_A\Phi, \Phi\rangle, \quad a = A - \Theta.$$

A direct computation gives

$$\begin{aligned} d\mathrm{csd}_{A,\Phi}(a,\phi) &= \int_Y \mathrm{Tr}(F_A \wedge a) + \int_Y \langle *a \cdot \Phi, \Phi\rangle - \int_Y \langle D_A\Phi, \phi\rangle \\ &= \int_Y \langle - * F_A + *\{\Phi \cdot \Phi\}_0, a\rangle - \int_Y \langle D_A\Phi, \phi\rangle. \end{aligned}$$

Thus the negative of the L^2–gradient of csd is the 'L^2_1–vector field' on $\mathcal{C}$

$$\mathcal{X}(A,\Phi) \stackrel{\mathrm{def}}{=} (*F_A - *\{\Phi\cdot\Phi\}_0, D_A\Phi) \in L^2_1. \tag{2.3}$$

By this we mean that $\mathcal{X}$ is a section of the L^2_1–version of the tangent bundle to $\mathcal{C}$. The *Quaternionic Seiberg–Witten equation* is the equation for the zeros of $\mathcal{X}$, i.e. the critical points of csd.

Definition 2.5 The *Quaternionic Seiberg–Witten equation* is the equation defined for a pair (A,Φ) consisting of a connection on E and a section Φ ('*spinor*') of $S \otimes_{\mathbf{H}} E$. The equation reads:

$$\begin{cases} F_A - \{\Phi \cdot \Phi\}_0 = 0 \\ D_A\Phi = 0 \end{cases} \tag{2.4}$$

where F_A is the curvature of A, and since A is an $SU(2)$–connection, a section of $\Lambda^2 \otimes \mathrm{ad}E$. D_A is the Dirac operator on $S \otimes_{\mathbf{H}} E$ and $\{\ \}_0$ denotes the quadratic form of Lemma 2.4.

If g is gauge transformation then $\mathrm{csd}(g(A), g^{-1}\Phi) = \mathrm{csd}(g,\Phi) \pm \deg(g)$, so csd descends to an $\mathbf{R}/\mathbf{Z}$–valued function on $\mathcal{B}$. This implies that $\mathcal{X}(A,\Phi) \in X_{A,\Phi} \cap L^2_1$ and the portion of $\mathcal{X}$ over $\mathcal{C}^*$ descends to a 'L^2_1–vector field' $\widehat{\mathcal{X}}$ over $\mathcal{B}^*$.

Definition 2.6 The *moduli space* of solutions to (2.4) we denote by

$$\mathcal{M} \stackrel{\mathrm{def}}{=} \{(A,\Phi) \text{ solving } (2.4)\}/\mathcal{G} \subset \mathcal{B}.$$

$\mathcal{M}^*$ will denote irreducible and $\mathcal{M}^r$ will denote the reducible portion of $\mathcal{M}$ respectively.

Thus $\mathcal{M}^*$ is the zeros of $\widehat{\mathcal{X}}$ and following Taubes, will be the basis for defining a Poincare–Hopf index for $\mathcal{B}^*$.

Remark 2.7 In our Quaternionic SW–theory the reducible portion $\mathcal{M}^r$ of $\mathcal{M}$ is just the moduli space of flat $SU(2)$–connections on Y. This is the space dealt with by Taubes [12] in the gauge theory approach to Casson's invariant.

We need to now address the issue of an admissible class of perturbations which will make $\mathcal{M}$ a finite number of *non-degenerate points* (made precise below) to apply the idea of a Poincare–Hopf index. Unlike the holonomy perturbations used by Taubes and Boden–Herald which are gradient perturbations we elect to perturb $\mathcal{X}$ directly rather than csd; i.e. at the level of vector fields, for this avoids a number of technical problems which the author has presently no satisfactory solution. This approach will be adequate for defining a Poincare–Hopf index but not a Floer type homology theory where gradient perturbations are required.

Definition 2.8 An *admissible perturbation* π consists of a differentiable $\mathcal{G}$–equivariant map of the form $(*k, l)\colon \mathcal{C} \to L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \times L^2_2(S \otimes_{\mathbf{H}} E)$ where

(i) $\pi_{A,\Phi} = (*k_{A,\Phi}, l_{A,\Phi}) \in X_{A,\Phi}$

(ii) the linearization of $(*k, l)$ at (A,Φ) is a bounded linear operator

$$(L\pi)_{A,\Phi}\colon L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2_2(S \otimes_{\mathbf{H}} E) \to L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2_2(S \otimes_{\mathbf{H}} E)$$

(iii) there is a uniform bound

$$\|\pi_{A,\Phi}\|_{L^2_{2,A}} \stackrel{\text{def}}{=} \sum_{i=0}^{2} \|(\nabla^A)^i \pi_{A,\Phi}\|_{L^2} \leq C.$$

Remark 2.9 In the unperturbed case, $\mathcal{M}$ can be easily shown to be compact. The preceding uniform L^2_2–type bound requirement on the perturbation is crucial to retain compactness of the moduli space for the perturbed equation below. This is a gauge invariant bound.

Definition 2.10 The *perturbed Quaternionic Seiberg–Witten equations* are the equations

$$\begin{cases} F_A - \{\Phi \cdot \Phi\}_0 + k_{A,\Phi} &= 0 \\ D_A\Phi + l_{A,\Phi} &= 0. \end{cases}$$

The corresponding moduli space is denoted $\mathcal{M}_\pi$, the irreducible portion $\mathcal{M}^*_\pi$ and the reducible portion $\mathcal{M}^r_{\overline{\pi}}$ where $\overline{\pi}$ is the restriction to $\mathcal{A}$ or equivalently the k–component of π. Note that when $\Phi = 0$, stab(A)–invariance forces $l_{A,0} = 0$ and the only effective portion of π on $\mathcal{A}$ is the k–component.

Let $\mathcal{X}_\pi = \mathcal{X} + \pi$, the perturbation of $\mathcal{X}$. The linearization at a zero (A,Φ) is a map

$$(L\mathcal{X}_\pi)_{A,\Phi}\colon L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2_2(S \otimes_{\mathbf{H}} E) \to X_{A,\Phi} \cap L^2_1 \tag{2.5}$$
$$(L\mathcal{X}_\pi)_{A,\Phi}(a,\phi) = (*d_A a - *\{\phi \cdot \Phi\}_0, D_A\phi + a \cdot \Phi) + (L\pi)_{A,\Phi}(a,\phi).$$

Definition 2.11 Call (A,Φ) or $[A,\Phi]$ *non-degenerate* if $L\mathcal{X}_\pi$ is surjective at (A,Φ). $\mathcal{M}_\pi$ is *non-degenerate* if it consists entirely of non-degenerate points. In this instance we also call π *non-degenerate*. The standard Kuranishi local model argument shows that a non-degenerate point is isolated in $\mathcal{B}$. (This includes reducible points.)

Fix a connection ∇^0 and let L^2_2 denote the Sobolev norm with respect to ∇^0. A metric on $\mathcal{B}$ is defined by the rule

$$d([A,\Phi],[A',\Phi']) = \inf_{g \in \mathcal{G}} \left\{ \|(A - g(A'), \Phi - g^{-1}\Phi')\|_{L^2_2} \right\}. \tag{2.6}$$

Proposition 2.12 *For any admissible perturbation $\mathcal{M}_\pi$ is a compact subspace of $\mathcal{B}$. Furthermore there is an $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$, if $\|\pi_{A,\Phi}\|_{L^2_{2,A}} < \varepsilon$ uniformly then given any $[A,\Phi] \in \mathcal{M}_\pi$ there is a $[A',\Phi'] \in \mathcal{M}$ such that $d([A,\Phi],[A',\Phi']) < \varepsilon$.*

Proposition 2.13 *There exists non-degenerate admissible perturbations. Furthermore such a perturbation π may be choosen so that $\|\pi_{A,\Phi}\|_{L^2_{2,A}}$ is arbitrarily small (uniformly) and π vanishes on any given closed subset of $\mathcal{C}$ which is disjoint from the subspace of unperturbed SW–solutions.*

The proofs are in sections 4 and 5.

3 Spectral flow and definition of the invariant

Fix a perturbation π (not necessarily non-degenerate). Regard the image of $\mathcal{X}_\pi$ as lying in the larger space $L^2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2(S \otimes_{\mathbf{H}} E)$ since $X_{A,\Phi} \cap L^2_1$ is a subspace of the former. The analog of the operator used by Taubes to define relative signs between non-degenerate zeros of $\widehat{\mathcal{X}_\pi}$ is the unbounded operator on $L^2(\mathrm{ad}E) \oplus L^2(\Lambda^1 \otimes \mathrm{ad}E) \oplus L^2(S \otimes_{\mathbf{H}} E)$ given in block matrix form:

$$L^\pi_{A,\Phi} = \begin{pmatrix} 0 & \delta^{0*}_{A,\Phi} \\ \delta^0_{A,\Phi} & (L\mathcal{X}_\pi)_{A,\Phi} \end{pmatrix}.$$

Here $\delta^{0*}_{A,\Phi}$ is the formal L^2–adjoint of $\delta^0_{A,\Phi}$ and the splitting used above is the (first) $\oplus$ (2nd and 3rd factors). $L^\pi_{A,\Phi}$ has dense domain the subspace of L^2_2–sections. The ellipticity of $L^\pi_{A,\Phi}$ implies that it is closed and unbounded as an operator on L^2. In general $L^\pi_{A,\Phi}$ will not have a real spectrum, due to the non-gradient perturbations we are using.

Remark 3.1 If it were the case that $L^\pi_{A,\Phi}$ is formally self-adjoint (i.e. on smooth sections) then it is well-known that $L^\pi_{A,\Phi}$ has only a discrete real spectrum which is unbounded in both directions in $\mathbf{R}$ and is without any accumulation points. It can be shown in general that since $L^\pi_{A,\Phi}$ is a $L_{A,\Phi}$–compact perturbation of $L_{A,\Phi}$ on L^2 the spectrum continues to be discrete, the real part of the spectrum is also unbounded in both directions in $\mathbf{R}$ and is without any accumulation points, see [7].

Let us consider the behaviour of $L^\pi_{A,\Phi}$ along the reducible stratum $\mathcal{A} \subset \mathcal{C}$. Since $\Phi = 0$ we abbreviate the operator to L^π_A. This has a natural splitting

$$L^\pi_A = K^{\overline{\pi}}_A \oplus D^\pi_A \tag{3.1}$$

corresponding to the splitting $(\Lambda^{0+1} \otimes \mathrm{ad}E) \oplus (S \otimes_{\mathbf{H}} E)$. We call $K^{\overline{\pi}}_A$ the *tangential operator* (the dependence on only the restriction $\overline{\pi}$ of π will be clear below) and D_A^π the *normal operator*. Explicitly

$$K_A^{\overline{\pi}} = \begin{pmatrix} 0 & d_A^* \\ d_A & *d_A + (L\overline{\pi})_A \end{pmatrix} \quad \text{and} \quad D_A^\pi = D_A + (L^\nu\pi)_A.$$

Here $(L^\nu\pi)_A$ denotes $(L\pi)_{A,0}$ restricted to the normal space $L_2^2(S \otimes_{\mathbf{H}} E)$ followed by projection onto the normal space again. The normal space is naturally acted on by $\mathrm{stab}(A)$ and D_A^π commutes with this action. If A is irreducible as a connection then $\mathrm{stab}(A) = \{\pm 1\}$ and D_A^π remains real linear but when $A = \Theta$ a trivial connection, $\mathrm{stab}(A) \cong SU(2)$ and it is quaternionic linear. (There is a $\mathrm{stab}(A) \cong U(1)$ case but this will not play a role so we will omit discussing it.)

A fact established in section 5 is:

Lemma 3.2 *The operator $(L^\nu\pi)_A$ is multiplication by a real function $f_A \in L_2^2$. Thus $D_A^\pi = D_A + f_A$ extends to an unbounded self-adjoint operator on $L^2(S \otimes_{\mathbf{H}} E)$ and spectral flow is defined for this operator.*

To define relative signs between non-degenerate zeros of $\widehat{\mathcal{X}_\pi}$ one usually uses the mod2 spectral flow of $L_{A,\Phi}^\pi$ when this operator is self-adjoint. In the general case we use the determinant line $\mathrm{detind}\, L^\pi$ regarding $L_{A,\Phi}^\pi$ as a family parameterized by $(A,\Phi) \in \mathcal{C}$. This is equivalent to the spectral flow definition in the self-adjoint case. $\mathrm{detind}\, L^\pi$ descends to a line bundle over $\mathcal{B}$ which we also denote by the same notation. However we note:

Lemma 3.3 $\mathrm{detind}\, L^\pi$ *is non-orientable over $\mathcal{B}$, i.e. there exists closed loops $\gamma\colon S^1 \to \mathcal{B}$ such that $\gamma^*(\mathrm{detind}\, L^\pi)$ is a non-trivial line bundle over S^1.*

Proof It suffices to consider the determinant index $\mathrm{detind}\, L$ of the unperturbed family over $\mathcal{B}_A \subset \mathcal{B}$. Then $L_A = K_A \oplus D_A$ where K_A is essentially the boundary of the (twisted) Self-dual operator in dimension 4. Spectral flow around closed loops for K_A is equivalent to the index of the (twisted) Self-dual operator over $Y \times S^1$. The latter index is well-known to be $\equiv 0 \bmod 8$. By Lemma 2.1 the spectral flow for D_A around closed loops is equivalent to the index of the twisted complex Dirac operator over $Y \times S^1$, where the twisting bundle E is rank 2 complex. According to the Atiyah–Singer Index Theorem this index is the negative of 2nd Chern class of E evaluated over the fundamental class of $Y \times S^1$. We may choose any closed loop so that this is ± 1. (See the proof of Lemma 3.5 for more details on this part of the calculation.) □

In particular if we have two non-degenerate zeros of $\widehat{\mathcal{X}}_\pi$ then the Lemma asserts that it is impossible in this scheme to define a relative i.e. mod2 sign between non-degenerate zeros. Thus as far as defining an invariant goes we can only work with the cardinality

$$\sum_{[A]\in\mathcal{M}^*_\pi} 1 \mod 2$$

for π non-degenerate.

Assume now that π is non-degenerate. We define counter-terms associated to $\mathcal{M}^r_{\overline{\pi}}$ to make $\sum_{[A]\in\mathcal{M}^*_\pi} 1 \bmod 2$ a well-defined invariant. These counter-terms will depend on the normal operator D^π_A, the Chern–Simons function and spectral invariants.

Note that in a ZHS the trivial orbit $\{[\Theta]\}$ is always a point in $\mathcal{M}^r_{\overline{\pi}}$ for every perturbation. In the unperturbed case this is clear. In the presence of a perturbation invariance by the stabilizer action at Θ forces $\overline{\pi}_\Theta = 0$.

When $A = \Theta$, the Dirac operator D_Θ can be identified with the canonical quaternionic linear Dirac operator on S which we denote as D. The operator K_A (presently take $\pi = 0$) is the boundary B of the 4–dimensional signature operator, after identifying $\Lambda^2 \cong \Lambda^1$ by the Hodge $*$–operator. To these two operators D and B we can associate the APS–spectral invariants [2]:

$$\eta(B), \quad \xi = \frac{1}{2}\Big(\eta(D) + \dim_{\mathbf{C}}\ker D\Big).$$

If X is compact oriented spin 4–manifold with oriented boundary Y then an application of the APS index theorems to X shows that

$$\xi + \frac{1}{8}\eta(B) = -\operatorname{Index} D^{(4)} - \frac{1}{8}\operatorname{sign} X. \tag{3.2}$$

Here $D^{(4)}$ is the Dirac operator on X and $\operatorname{sign} X$ the signature. Thus we see that the left-side of (3.2) is always an integer. As an aside, the $\bmod 2$ reduction of the right-side only involves the signature term (since in four dimensions the Dirac operator is quaternionic linear and so its index is even) and therefore is just the *Rokhlin invariant* $\mu(Y)$. Given a perturbation π now set

$$c(g,\pi) = \xi + \frac{1}{8}\eta(B) + \Big(\mathbf{C}\text{–spectral flow of } \{(1-t)D_\Theta + tD^\pi_\Theta\}^1_{t=0}\Big) \in \mathbf{Z}.$$

In the spectral-flow term D_Θ, D^π_Θ are quaternionic linear and thus $c(g,\pi) \equiv \mu(Y) \bmod 2$ continues to be true. $c(g,\pi)$ is our counter-term associated to $\{[\Theta]\}$.

Remark 3.4 Our convention for spectral flow is the the number of eigenvalues (counted algebraically) crossing $-\varepsilon$ for $\varepsilon > 0$ sufficiently small.

In order to define the counter-terms associated with points in $\mathcal{M}^{r*}_{\pi}$ we shall need two preliminaries. Firstly, consider the *normal spectral flow* of L^{π}_{A} along a path γ in $\mathcal{A}$ i.e.

$$\mathrm{SF}^{\nu}(\gamma) \quad = \quad \text{spectral flow of } D^{\pi}_{A} \text{ along } \gamma$$

which is defined because of Lemma 3.2. On the reducible stratum $\mathcal{A} \subset \mathcal{C}$, the Chern–Simons–Dirac function reduces to the Chern–Simons function which we denote as cs. We remind the reader that cs depends on a basepoint which we choose to be a trivial connection Θ (which we fix once and for all).

Lemma 3.5 *Let $[x]$ be a point in $\mathcal{B}_{\mathcal{A}}$ and $[\gamma(t)]$, $t \in [0,1]$ a closed differentiable loop in $\mathcal{B}_{\mathcal{A}}$ based at $[x]$. Then*

$$\mathrm{SF}^{\nu}(\gamma) \quad = \quad \mathrm{cs}(\gamma(1)) - \mathrm{cs}(\gamma(0)) \in \mathbf{Z}.$$

Proof First we invoke Lemma 2.1 which says we only need to compute the complex spectral flow for the complex Dirac operator $D^{\mathbf{C}}_{A}$ on $S \otimes E = S \otimes_{\mathbf{C}} E$. According to [2] this spectral flow coincides with the index of the four-dimensional Dirac operator $D^{(4)}_{\widehat{A}}$ on the pull-back $\widehat{S} \otimes \widehat{E} \to Y \times [0,1]$ of $S \otimes E \to Y$ with $\widehat{A}$ interpolating between $\gamma(0)$ at $Y \times \{0\}$ and $\gamma(1)$ at $Y \times \{1\}$. Since the initial and final connections are gauge equivalent, the boundary terms cancel in the application of the APS index theorem and we are left with

$$\begin{aligned}
\operatorname{Index} D^{(4)}_{\widehat{A}} &= -\int_X c_2(\widehat{A}) \quad (\text{Chern form}) = -\frac{1}{8\pi^2}\int_{Y \times [0,1]} \mathrm{Tr}(F^2_{\widehat{A}}) \\
&= -\frac{1}{8\pi^2}\int_{Y \times [0,1]} d\,\mathrm{Tr}\Big(a \wedge d_{\Theta} a + \frac{2}{3} a \wedge a \wedge a\Big), \quad a = \widehat{A} - \Theta \\
&= -\Big(\mathrm{cs}(\gamma(0)) - \mathrm{cs}(\gamma(1))\Big).
\end{aligned}$$

□

The second preliminary: cs descends to a function $\overline{\mathrm{cs}}\colon \mathcal{B}_{\mathcal{A}} \to \mathbf{R}/\mathbf{Z}$ on the quotient space. Since the value of $\overline{\mathrm{cs}}$ is constant on components of $\mathcal{M}^{r}$, the image set $\overline{\mathrm{cs}}(\mathcal{M}^{r})$ is a finite number of values $c_1, \dots, c_m$ in $\mathbf{R}/\mathbf{Z}$. Let $\varepsilon_1 > 0$ be the smallest distance between pairwise distinct c_i's where $\mathbf{R}/\mathbf{Z}$ has the distance inherited from $\mathbf{R}$. Let $\varepsilon_2 > 0$ be the constant which is the smallest distance between pairwise distinct components of $\mathcal{M}^{r}$, in the metric (2.6).

Definition 3.6 Call a perturbation π *small* if $\overline{\mathrm{cs}}(\mathcal{M}^{r*}_{\pi})$ is within an $\varepsilon_1/3$–neighbourhood of $\overline{\mathrm{cs}}(\mathcal{M}^{r*})$, and $\mathcal{M}^{r*}_{\pi}$ is within an $\varepsilon_2/3$–neighbourhood of $\mathcal{M}^{r*}$.

Assume π to be small and non-degenerate in the sense of the preceding. This can be done by making $\|\pi_{A,\Phi}\|_{L^2_{2,A}}$ sufficiently small, by Proposition 2.12. Write

$$\mathcal{M}^{r*} = \mathcal{K}^1 \cup \ldots \cup \mathcal{K}^n$$

as the union of connected components. Then given any $[A] \in \mathcal{M}^{r*}_{\pi}$ there is a unique component $\mathcal{K}^i$ which is within $\varepsilon_1/3$ of $[A]$. Denote by $\mathcal{N}^i$ the intersection of the $\varepsilon_2/3$–neighbourhood of $\mathcal{K}^i$ and the preimage under $\overline{\mathrm{cs}}$ of the $\varepsilon_1/3$–neighbourhood of $\overline{\mathrm{cs}}(\mathcal{M}^{r*})$ in $\mathbf{R}/\mathbf{Z}$. Let $[\gamma]$ be any path from $[\Theta]$ to $[A]$. Let $[\overline{\gamma}]$ be any other path from $[\Theta]$ to $\mathcal{K}^i$ with the property that $[\gamma]$ and $[\overline{\gamma}]$ are homotopic relative to $\mathcal{N}^i \cup \{[\Theta]\}$. Then the expression

$$\kappa[A] = \mathrm{SF}^{\nu}(\gamma) + \mathrm{cs}(\overline{\gamma}(0)) - \mathrm{cs}(\overline{\gamma}(1)) \in \mathbf{R}$$

is well-defined and independent of choice of γ and $\overline{\gamma}$, by Lemma 3.5.

Over $\mathcal{B}_{\mathcal{A}}$ we have the line bundle $\det\mathrm{ind}\, K^{\overline{\pi}}$ of the family of tangential operators $\{K^{\overline{\pi}}_A\}$. In contrast to $\det\mathrm{ind}\, L^{\pi}$ this is an orientable line bundle. This is basically the Taubes' orientation of $\mathcal{M}^{r*}_{\overline{\pi}}$ in [12]. We fix the overall orientation by specifying $\det\mathrm{ind}\, K^{\overline{\pi}}$ at $[\Theta]$ by the following rule. The kernel and cokernel of $K^{\overline{\pi}}_{\Theta}$ are $\cong \mathrm{su}(2)$, the constant sections of $\mathrm{ad}E$, after $\mathrm{ad}E$ is trivialized as $Y \times \mathrm{su}(2)$ by Θ. Orient $\det\mathrm{ind}\, K^{\overline{\pi}}_{\Theta}$ as $o(\mathrm{su}(2)) \wedge o(\mathrm{su}(2)^*)$ where $o(\mathrm{su}(2))$ is a any chosen orientation and $o(\mathrm{su}(2)^*)$ is the dual orientation. We denote the induced oriention at $[A] \in \mathcal{M}^{r*}_{\overline{\pi}}$ by $\varepsilon[A] \in \{\pm 1\}$.

Theorem 3.7 *Let Y be an oriented closed integral homology 3–sphere with Riemannian metric g. Let π be a non-degenerate and small admissible perturbation for $\mathcal{M}_{g,\pi}$, the perturbed Quaternionic Seiberg–Witten moduli space with respect to g. The terms $c(g,\pi)$, $\varepsilon[A]$ and $\kappa[A]$ as above are well-defined and the sum*

$$\tau(Y) = \sum_{[A]\in\mathcal{M}^*_{\pi}} 1 + \frac{1}{2}c(g,\pi) + \sum_{[A]\in\mathcal{M}^{r*}_{\overline{\pi}}} \varepsilon[A]\Big(\kappa[A] + \frac{1}{4}\Big) \in \mathbf{R} \bmod 2\mathbf{Z}$$

is independent of both g and π chosen. Furthermore $\tau(Y)$ does not depend on the orientation of Y and therefore τ defines an unoriented diffeomorphism invariant for integral homology 3–spheres.

The extra term $1/4$ in the sum is inserted to make the invariant independent of the orientation of Y.

Let $\lambda_{SU(3)}(Y)$ be the $SU(3)$–Casson invariant of Boden–Herald. The definition of $\tau(Y)$ is modelled on $\lambda_{SU(3)}(Y)$ and both suffer from the defect that no multiple is obviously integral valued. This is due to the usage of the Chern–Simons function. (Boden–Herald–Kirk [4] have devised an integer version of $SU(3)$–Casson that gets around the usage of Chern–Simons by an ad-hoc device. It is not a completely natural definition.) However we have the following.

Theorem 3.8 *Let Y be an integral homology 3–sphere and $\lambda_{SU(3)}(Y)$ be the $SU(3)$–Casson invariant for Y. Then*

$$\lambda_{SU(3)}(Y) + 2\tau(Y)$$

is a $\mathbf{Z} \bmod 4\mathbf{Z}$–valued invariant of the unoriented diffeomorphism type of Y.

The assertion of this theorem is that we have a cancellation of the Chern–Simons terms, leaving only an integral expression. Our contention is that combining $SU(3)$–Casson with an $SU(2)$–version of Seiberg–Witten is the natural way of presenting the topological information contained in the two theories. This will be worked out in greater detail in a further article where a unified approach to the two theories and an integer valued Seiberg–Witten/Casson invariant is defined.

The proof of the Theorem 3.7 is in section 6 and Theorem 3.8 in section 7.

4 Compactness

In this section we prove Proposition 2.12. Recall that our 3–manifold Y is assumed to be Riemannian with metric g. We shall need to vary g at two points in this article. In the present section we shall utilize rescaling g to establish compactness of the moduli space. In section 6 we shall analyse the change in the moduli space as g varies in a 1–parameter family.

We set-up a framework for comparing the SW–equation for different metrics. Spinors and in particular the Dirac operator are not canonically associated objects to a Riemannian structure.

The first task is to fix a model for the spin structure and spinors. Our metric g shall be taken as the reference. On a compact 3–manifold we can always find a smooth nowhere vanishing vector field, let us denote this as e_1. Additionally assume it is of unit length with respect to g. By working perpendicular to e_1 we can complete this to a global orthonormal frame (e_1, e_2, e_3). Assume the orientation $e_1 \wedge e_2 \wedge e_3$ coincides with the orientation on Y. This global frame defines a trivialization $Y \times SO(3)$ of the (positively) oriented orthonormal frame bundle of Y.

Let $sp(1) = \mathrm{spin}(3) \subset CL(Y)$ denote the unit quaternions and fix a group homomorphism $sp(1) \to SO(3)$ which is the 2–fold covering map. Then we fix the spin structure on Y (with respect to g) by the projection

$$P = Y \times sp(1) \to Y \times SO(3).$$

The spinor bundle S is then given by $P \times_\varrho \mathbf{H}$ where ϱ is the fundamental representation of $sp(1)$ on $\mathbf{H}$. Since $[(y,h),q] = [(y,1),hq]$, S has a natural trivialization as $Y \times \mathbf{H}$ and sections of S are simply the $\mathbf{H}$–valued functions on Y. Notice that the quaternionic structure on S is exactly right multiplication on the $\mathbf{H}$ factor of $Y \times \mathbf{H}$.

The trivialization $Y \times SO(3)$ of TY also induces a trivialization $Y \times CL(\mathbf{R}^3)$ of the Clifford bundle $CL(Y)$ with the constant section $\widehat{e}_1 = (1,0,0)$ corresponding to the vector field e_1, $\widehat{e}_2 = (0,1,0)$ to e_2 etc. Fix the (left) Clifford representation on $\mathbf{H}$ of the Clifford algebra $CL(\mathbf{R}^3)$ by mapping

$$\widehat{e}_1 \mapsto i, \quad \widehat{e}_2 \mapsto j, \quad \widehat{e}_3 \mapsto k.$$

That is to say, $\widehat{e}_i \cdot h = ih$ etc. On $S \otimes_{\mathbf{H}} E$ the Dirac operator now takes the form

$$D_A = (i \otimes 1)\nabla^A_{e_1} + (j \otimes 1)\nabla^A_{e_2} + (k \otimes 1)\nabla^A_{e_3}.$$

Suppose now we want to change the metric from g. This is achieved by pulling back the metric g by an automorphism h of TY. Using the frame (e_1, e_2, e_3) as a basis can conveniently think of h as a smooth map $h\colon Y \to GL(3)$. The global frame (e_1, e_2, e_3) is pulled back to a global frame $(h^{-1}(e_1), h^{-1}(e_2), h^{-1}(e_3))$ for the pulled back metric. In the same way as above this global frame defines a trivialization $Y \times SO(3)$ of the oriented orthonormal frame bundle in the pulled back metric and we may proceed with the spin structure, spinors etc. as constructed before. In particular we notice that the model for the spinor bundle as $\mathbf{H}$–valued functions on Y remains the same in the pulled back metric but the Clifford mutiplication changes and is now defined by

$$\widehat{h^{-1}(e_1)} \mapsto i, \quad \widehat{h^{-1}(e_2)} \mapsto j, \quad \widehat{h^{-1}(e_3)} \mapsto k.$$

If h is actually an isometry with respect to g then we are merely changing the trivialization of S.

Let g' denote the new metric defined by h and $\nabla^{g'}$ the spin connection on S. Then the Dirac operator coupled to A with respect to g' is given by

$$D^{g'}_A = (i \otimes 1)\nabla^{g',A}_{h^{-1}(e_1)} + (j \otimes 1)\nabla^{g',A}_{h^{-1}(e_2)} + (k \otimes 1)\nabla^{g',A}_{h^{-1}(e_3)}.$$

Similiarly one may obtain expressions for the bilinear forms $\{\cdot\}_0$ and B with respect to g' in terms of h.

Consider now the special case when g is rescaled as $g_\lambda = \lambda^2 g$ where $\lambda > 0$ is a constant. Clearly g_λ is induced by $h = \lambda \mathrm{Id}$ so $h^{-1}(e_i) = e_i/\lambda$. Under the above model for the spinors, the Hermitian metric on S is fixed. However, we may choose to vary this with λ. In the present case, for g_λ we may set

$$\langle \cdot, \cdot \rangle_\lambda = \lambda^\alpha \langle \cdot, \cdot \rangle \tag{4.1}$$

where the right-hand inner product is the original one on S. A good choice for α will be made later. In the next lemma, a 'λ' superscript means an object taken with respect to the metric g_λ. Unmarked objects are taken with respect to g.

Lemma 4.1 *Fix the model for spinor bundle S by g, and use the spinor metric given by (4.1) in the Riemannian metric g_λ. Then the following hold.*

(i) $D^\lambda_A = (1/\lambda) D_A$

(ii) $\{\cdot\}^\lambda_0 = \lambda^{2+\alpha}\{\cdot\}_0$

(iii) $d^{*,\lambda}_A = (1/\lambda^2) d^*_A$ *on* $\Omega^1 \otimes \mathrm{ad}E$

(iv) $B^\lambda = \lambda^\alpha B$

(v) $*^\lambda = (1/\lambda)*$ *on* Λ^2

Proof For (i) recall that the Levi–Civita connection is invariant under rescaling the metric by a constant. This leaves the connection term $\nabla^{g',A} = \nabla^{g,A}$. The formula now follows from $h^{-1}(e_i) = e_i/\lambda$. For (ii) establish the rule $\omega \cdot_\lambda \phi = \frac{1}{\lambda^2}\omega \cdot \phi$ and $\alpha \cdot_\lambda \phi = \frac{1}{\lambda}\alpha \cdot \phi$ where ω is a 2–form and α a 1–form. The new coframe $e^{\lambda *}_i = \lambda e^*_i$ and so the action of e^*_i with respect to g_λ is $1/\lambda$ of the action with respect to g. For (iii) in the defining equation $\int \langle d_A \gamma, a\rangle_\lambda dg_\lambda = \int \langle \gamma, d^{*,\lambda}_A a\rangle_\lambda dg_\lambda$, we have $\langle \gamma, d^{*,\lambda}_A a\rangle_\lambda = \lambda^{-2}\langle \gamma, d^{*,\lambda}_A a\rangle$. For (iv) the defining equation is $\langle \gamma(\phi), \psi\rangle_\lambda = \lambda^\alpha \langle \gamma(\phi), \psi\rangle = \lambda^\alpha \langle \gamma, B(\phi,\psi)\rangle = \langle \gamma, B^\lambda(\psi,\phi)\rangle$. (v): $*(e^*_1 \wedge e^*_2) = e^*_3$ and $*^\lambda(\lambda e^*_1 \wedge \lambda e^*_2) = \lambda e^*_3$, etc. □

The preceding lemma easily implies the following principle result we need on rescaling the metric:

Proposition 4.2 *Fix the model for the spinor bundle S by g, and let S have the metric (4.1) with respect to g_λ where $\alpha = -2$. Then the perturbed SW–equation (2.10) with respect to g is equivalent to the following equation with respect to g_λ:*

$$\left\{ \begin{array}{lcl} F_A - \{\Phi \cdot \Phi\}^\lambda_0 + k_{A,\Phi} & = & 0 \\ D^\lambda_A \Phi + \dfrac{1}{\lambda} l_{A,\Phi} & = & 0. \end{array} \right. \tag{4.2}$$

Furthermore the perturbation $\pi^\lambda(A,\Phi) = (*^\lambda k_{A,\Phi}, (1/\lambda) l_{A,\Phi})$ is an admissible perturbation with respect to g_λ.

The scheme of the proof of the compactness of the moduli space rests on a Bochner argument to get a L^4–bound on the spinors, Uhlenbeck's Theorem [13] and as mentioned above, rescaling. In the 4–dimensional context such an argument is presented in Feehan–Leness [6]. The basic input is contained in the following two lemmas.

Lemma 4.3 *Let (A,Φ) be a solution of the perturbed SW–equation (4.2), defined on Y with respect to the metric $g_\lambda = \lambda^2 g$. Let s denote the scalar curvature of Y with respect to g. Then*

$$\begin{aligned}
\int_Y |\Phi|_\lambda^4 dg_\lambda &\leq \frac{8}{\lambda}\int_Y \frac{s^2}{16} + |k_{A,\Phi}|^2 + |l_{A,\Phi}|^2 dg \\
\int_Y |k_{A,\Phi}|_\lambda^2 dg_\lambda &\leq \frac{1}{\lambda}\int_Y |k_{A,\Phi}|^2 dg \\
\int_Y \left|\frac{1}{\lambda} l_{A,\Phi}\right|_\lambda^2 dg_\lambda &\leq \frac{1}{\lambda}\int_Y |l_{A,\Phi}|^2 dg.
\end{aligned}$$

The spinor metric (4.1) on the left-side is taken with $\alpha = -2$.

Proof This is a straightforward manipulation involving the Bochner formula for the Dirac operator which reads:

$$(D_A^\lambda)^* D_A^\lambda \Phi = (\nabla_A^\lambda)^* \nabla_A^\lambda \Phi + \frac{1}{4} s^\lambda \Phi + F_A \cdot_\lambda \Phi.$$

Here and below a 'λ' subscript or superscript indicates the object taken with respect to g_λ. Unscripted objects are taken with respect to g. Taking the inner product with Φ and integrating gives

$$\int_Y |D_A^\lambda \Phi|_\lambda^2 \, dg_\lambda = \int_Y |\nabla_A^\lambda \Phi|_\lambda^2 \, dg_\lambda + \int_Y \frac{1}{4} s^\lambda |\Phi|_\lambda^2 \, dg_\lambda + \int_Y \langle F_A, \{\Phi \cdot \Phi\}_0^\lambda \rangle_\lambda \, dg_\lambda.$$

Applying the SW–equation (4.2) and after some manipulation we obtain

$$\int_Y \left(\frac{s^\lambda}{4} - |k_{A,\Phi}|_\lambda\right) |\Phi|_\lambda^2 \, dg_\lambda + \frac{1}{2}\int_Y |\Phi|_\lambda^4 \, dg_\lambda \leq \frac{2}{\lambda^2}\int_Y |l_{A,\Phi}|_\lambda^2 \, dg_\lambda.$$

This in turn implies

$$\int_Y |\Phi|_\lambda^4 \, dg_\lambda \leq 2\Gamma_\lambda \left(\int_Y |\Phi|_\lambda^4 \, dg_\lambda\right)^{1/2} + \frac{4}{\lambda^2}\int_Y |l_{A,\Phi}|_\lambda^2 \, dg_\lambda$$

where $\Gamma_\lambda \geq 0$ is given by

$$\Gamma_\lambda^2 = \int_Y \left(\frac{s^\lambda}{4} - |k_{A,\Phi}|_\lambda\right)^2 dg_\lambda.$$

Therefore

$$\int_Y |\Phi|_\lambda^4 \, dg_\lambda \le 4\Gamma_\lambda^2 + \frac{8}{\lambda^2}\int_Y |l_{A,\Phi}|_\lambda^2 \, dg_\lambda. \tag{4.3}$$

Under rescaling the metric from g to $g_\lambda = \lambda^2 g$ we have $dg_\lambda = \lambda^3 dg$ and the following relations hold:

$$\begin{aligned} \int_Y (s^\lambda)^2 \, dg_\lambda &= \int_Y (\lambda^{-2}s)^2 \lambda^3 \, dg \\ \int_Y |k_{A,\Phi}|_\lambda^2 \, dg_\lambda &= \int_Y \lambda^{-4} |k_{A,\Phi}|^2 \lambda^3 \, dg \\ \int_Y |l_{A,\Phi}|_\lambda^2 \, dg_\lambda &= \int_Y \lambda^{-2} |l_{A,\Phi}|^2 \lambda^3 \, dg. \end{aligned} \tag{4.4}$$

Hence

$$\Gamma_\lambda^2 \le \frac{4}{\lambda}\int_Y \Big(\frac{s^2}{16} + |k_{A,\Phi}|^2\Big) dg$$

Together with (4.3) and (4.4) we get the desired bounds. □

Introduce the notation B_r for the closed Euclidean ball of radius r in $\mathbf{R}^3$. Fix a model for the spinors S on B_1 with respect to the Euclidean metric as in the preceding and let $E_0 = B_1 \times \mathbf{C}^2$ denote the trivial $SU(2)$–bundle. This trivialization defines the canonical trivial connection d on E.

Lemma 4.4 *Allow any metric on B_1. Let the pair $(A = d + a, \Phi) \in L_2^2$ be defined on $E_0 \to B_1$. Assume that (a) $d^*a = 0$ (b) $\|a\|_{L_1^2} \le C_1$, $\|\Phi\|_{L^4} \le C_2$ (c) (A, Φ) satisfies a perturbed SW–equation of the form (2.10) on E with $\pi_{A,\Phi} = (*k, l)_{A,\Phi} \in L_2^2$, and (d) $\|\pi_{A,\Phi}\|_{L_2^2} \le C_3$. Then $\|a\|_{L_3^2(B_{1/2})}$, $\|\Phi\|_{L_3^2(B_{1/2})}$ are uniformly bounded independent of a and Φ.*

Proof We may rewrite the equations both a and Φ satisfy as

$$\begin{aligned} (d + d^*)a &= -a \wedge a + \{\Phi \cdot \Phi\}_0 - k_{A,\Phi} \\ D\Phi &= -a \cdot \Phi - l_{A,\Phi}. \end{aligned}$$

Here D is the canonical Dirac operator associated with B_1 tensored with the trivial factor $\mathbf{C}^2$. $\|a\|_{L^4}$ is uniformly bounded by the Sobolev embedding $L_1^2 \subset L^4$ and condition (a). The terms $k_{A,\Phi}$, $l_{A,\Phi}$ being uniformly bounded in L_2^2 are uniformly bounded in C^0. Since $\|a \cdot \Phi\|_{L^2} \le \|a\|_{L^4}\|\Phi\|_{L^4}$ we see that $D\Phi$ is uniformly bounded in L^2. The basic elliptic inequality $\|\Phi\|_{L_{k+1}^p(B_{r'})} \le \text{const.}(\|D\Phi\|_{L_k^p(B_r)} + \|\Phi\|_{L^p(B_r)})$, $r' \le r$ for D forces Φ to be uniformly bounded in L_1^2 over B_{r_1}, $r_1 < 1$. The embedding $L_1^2 \subset L^6$ now makes both a and Φ uniformly bounded in L^6 over B_{r_1}. The bound

$\|a\cdot\Phi\|_{L^3(B_{r_1})} \leq \|a\|_{L^6(B_{r_1})}\|\Phi\|_{L^6(B_{r_1})}$ now makes $D\Phi$ uniformly bounded in $L^3(B_{r_1})$ and thus Φ is uniformly bounded in $L^3_1(B_{r_2})$, $r_2 < r_1$ and therefore $L^p(B_{r_2})$, $2 \leq p < \infty$. Now observe $\|a\cdot\Phi\|_{L^4(B_{r_2})} \leq \|a\|_{L^6(B_{r_2})}\|\Phi\|_{L^{12}(B_{r_2})}$ and by repeating the argument we get Φ uniformly bounded in $L^4_1(B_{r_3})$, $r_3 < r_2$. A similiar type of argument using the elliptic estimate for $d+d^*$ also establishes that a is uniformly bounded in $L^4_1(B_{r_3})$.

To obtain uniform bounds for a and Φ in $L^2_2(B_{r_4})$, $r_4 < r_3$ we need to obtain uniform bounds for the quadratic terms $a\wedge a$, $\{\Phi\cdot\Phi\}_0$ and $a\cdot\Phi$ in $L^2_1(B_{r_3})$. However this follows from the continuous multiplication $L^4_1(B_{r_3})\times L^4_1(B_{r_3}) \to L^2_1(B_{r_3})$. Finally this puts a and Φ in the continuous range for Sobolev multiplication and from this a uniform bound in $L^2_3(B_{r_5})$, $r_5 < r_4$ is obtained. □

Proposition 4.5 *$\mathcal{M}_\pi$ is a compact subspace of $\mathcal{B}$ where π an admissible perturbation. That is to say, given any sequence (A_i,Φ_i) of L^2_2–solutions to (2.10) there is a subsequence $\{i'\}\subset\{i\}$ and L^2_3–gauge transformations $g_{i'}$ such that $g_{i'}(A_{i'},\Phi_{i'})$ converges in L^2_2 to a solution of the π–perturbed SW–equations.*

Proof By Proposition 4.2 a solution (A,Φ) of (2.10) is equivalent to a solution of (4.2), the SW–equation with respect to g_λ and with perturbation π^λ. Thus it suffices to prove compactness of the moduli space $\mathcal{M}_{g_\lambda,\pi^\lambda}$ of solutions of (4.2) for any $\lambda > 0$.

Choose λ large such that any geodesic ball B of unit radius in Y is sufficiently close to the Euclidean metric in C^3, so that Uhlenbeck's Theorem [13] applies over B. Let $\varepsilon_0 > 0$ be the constant in Uhlenbeck's Theorem such that if any L^2_1 connection A on $E|_B$ satisfies $\|F_A\|_{L^2(B)} < \varepsilon_0$ then there is a gauge transformation $g\in L^2_2(B)$ which changes A so that $g(A) = d+a$ is in Coloumb gauge $d^*a = 0$ and $\|a\|_{L^2_1(B)} \leq c\|F_A\|_{L^2(B)}$. Here we use a fixed trivialization $E|_B \cong B\times\mathbf{C}^2$ with trivial connection ∇ or d.

Assume that (A,Φ) is a solution of (4.2). The proof of Lemma 4.4 gives us an additional fact. It shows that a is of class $L^2_2(B_{1/2})$ and by a straightforward bootstrapping argument we see that g is actually in $L^2_3(B_{1/2})$.

In the definition of an admissible perturbation $\|\pi^\lambda_{A,\Phi}\|_{L^2_{2,A}}$ is uniformly bounded for every $\lambda > 0$. In order to apply Lemma 4.4 we need to deduce a uniformly bound for $\|\pi^\lambda_{A,\Phi}\|_{L^2_2(B)}$. The covariant derivatives ∇ and ∇^A upto second order are related by

$$\nabla\omega \;=\; \nabla^A\omega - a(\omega)$$

$$\begin{aligned}\nabla^2\omega &= (\nabla^A - a)(\nabla^A\omega - a(\omega)) \\ &= (\nabla^A)^2\omega + (\nabla a)(\omega) + 2a(a(\omega)).\end{aligned}$$

Utilizing the embedding $L^2_2(B) \subset C^0(B)$ and $L^2_1(B) \subset L^4(B)$ we obtain

$$\begin{aligned}\|\nabla\omega\|_{L^2(B)} &\leq \text{const.}\Big(\|\nabla^A\omega\|_{L^2(B)} + \|a\|_{L^2(B)}\|\omega\|_{L^2_2(B)}\Big) \\ \|\nabla^2\omega\|_{L^2(B)} &\leq \text{const.}\Big(\|(\nabla^A)^2\omega\|_{L^2(B)} + \|a\|_{L^2_1(B)}\|\omega\|_{L^2_2(B)} \\ &\qquad + \|a\|^2_{L^2_1(B)}\|\omega\|_{L^2_2(B)}\Big).\end{aligned}$$

Choose $\varepsilon_1 \leq \varepsilon_0$ so that $\|F_A\|_{L^2(B)} < \varepsilon_1$ forces $\|a\|_{L^2_1(B)}$ to be very small; then the error terms $\|a\|_{L^2(B)}\|\omega\|_{L^2_2(B)}$, $\|a\|_{L^2_1(B)}\|\omega\|_{L^2_2(B)}$ and $\|a\|^2_{L^2_1(B)}\|\omega\|_{L^2_2(B)}$ are $\ll \|\omega\|_{L^2_2(B)}$ and we get a uniform estimate $\|\omega\|_{L^2_2(B)} \leq \text{const.}\|\omega\|_{L^2_{2,A}(B)}$.

Lemma 4.3 shows that $\|\Phi\|_{L^4}$ is uniformly bounded with respect to g_λ and $\|F_A\|_{L^2} \to 0$ as $\lambda \to \infty$. Increase λ if necessary so that $\|F_A\|_{L^2(B)} < \varepsilon_1$ for all B. Suppose now that (A_i, Φ_i) is a sequence of solutions of (4.2). Denote by $B_{1/2}$ the geodesic ball with the same center as B but half the radius. Uhlenbeck's Theorem and the uniform bounds of Lemma 4.4 finds L^2_3 gauge transformations g_i over $B_{1/2}$ such that after passing to a subsequence, $g_i(A_i, \Phi_i)$ converges in $L^2_2(B_{1/2})$ to a SW–solution (4.2) over B. Now the standard covering argument in [5, section 4.4.2] (also see [6]) shows that after global gauge transformations and passing to subsequences, (A_i, Φ_i) can be made to converge in L^2_2 over all of Y. □

The preceding proof also shows:

Corollary 4.6 *Let (A, Φ) be a perturbed SW–solution (2.10). There is an L^2_3 gauge transformation g such that $g(A, \Phi)$ is in L^2_3.*

Corollary 4.7 *There is an $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$, if $\|\pi_{A,\Phi}\|_{L^2_{2,A}} < \varepsilon$ uniformly then given any $[A, \Phi] \in \mathcal{M}_\pi$ there is a $[A', \Phi'] \in \mathcal{M}$ such that $d([A, \Phi], [A', \Phi']) < \varepsilon$, d being the metric (2.6).*

Proof Suppose false. Then there exists sequences $\{\pi_i\}$ and $\{(A_i, \Phi_i)\}$ with $[A_i, \Phi_i] \in \mathcal{M}_{\pi_i}$ such that $\|(\pi_i)_{A_i,\Phi_i}\|_{L^2_{2,A_i}} \to 0$ but with $d([A_i, \Phi_i], [A', \Phi'])$ bounded away from zero over $[A', \Phi'] \in \mathcal{M}$. The sequence also satisfies

$$\|F_{A_i} - \{\Phi_i \cdot \Phi_i\}_0\|_{L^2} + \|D_{A_i}\Phi_i\|_{L^2} \to 0. \tag{4.5}$$

The proof of Proposition 4.5 shows that after gauge transformations and passing to a subsequence which we shall also denote as (A_i, Φ_i), (A_i, Φ_i) converges in L^2_2 and the limit, by (4.5) is necessarily a unperturbed SW–solution. This is a contradiction. □

5 Construction of perturbations

In this section we prove Proposition 2.13. Introduce the notation $B(\varepsilon)$ for the ε–ball in the slice space $X_{A,\Phi}$. (Recall this is a Hilbert space in an L^2_2–Sobolev norm.) Denote by $\beta\colon X_{A,\Phi} \to [0,1]$ a smooth cut-off function with support in $B(\varepsilon)$.

Lemma 5.1 *Fix (A,Φ). There is an $\varepsilon > 0$ and a differentiable function $\xi\colon B(\varepsilon) \times X_{A,\Phi} \to (\ker \delta^0_{A,\Phi})^\perp \subset L^2_3(\mathrm{ad}E)$ such that given any $(a,\phi;b,\psi) \in B(\varepsilon) \times X_{A,\Phi}$, the relation*

$$(b,\psi) + \delta^0_{A,\Phi}\xi(a,\phi;b,\psi) \in X_{A+a,\Phi+\phi} \tag{5.1}$$

holds. Here $(\ker \delta^0_{A,\Phi})^\perp$ denotes the L^2–orthogonal complement.

Proof Apply the Implicit Function theorem to the map

$$H(\xi,(a,\phi),(b,\psi)) = \delta^{0*}_{A+a,\Phi+\phi}\delta^0_{A,\Phi}(\xi) + \delta^{0*}_{A+a,\Phi+\phi}(b,\psi)$$

from $(\ker \delta^0_{A,\Phi})^\perp \times B(\varepsilon) \times X_{A,\Phi} \to (\ker \delta^0_{A,\Phi})^\perp \cap L^2_1$. The linearization of H at $(0,0,0)$ restricted to $(\ker \delta^0_{A,\Phi})^\perp$ is an isomorphism. This establishes the existence of the function $\xi = \xi(a,\phi;b,\psi)$ but only for (a,ϕ) and (b,ψ) defined in sufficiently small neighbourhoods of zero. However notice that if (b,ψ) satisfies (5.1) then for any real constant c, $c(b,\psi)$ satisfies the same equation but with ξ replaced by $c\xi$. That is we can allow the (b,ψ) to be defined in ξ for all $X_{A,\Phi}$ by extending ξ linearly in that factor. □

Let us now assume $\Phi \neq 0$. Set $\varepsilon > 0$ to be less than the constant in Lemma 5.1 and also such that $B(\varepsilon)$ injects into $\mathcal{B}$. Assume $\operatorname{supp}\beta \subset B(\varepsilon)$. Fix $(b,\psi) \in X_{A,\Phi}$. Define a function $(A,\Phi) + X_{A,\Phi} \to L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \times L^2_2(S \otimes_{\mathbf{H}} E)$ by the rule

$$\pi_{A+a,\Phi+\phi} = \beta(a,\phi)(b,\psi) + \delta^0_{A,\Phi}\xi(a,\phi;\beta(a,\phi)(b,\psi)) \tag{5.2}$$

for $(a,\phi) \in B(\varepsilon)$. By construction π has support in $B(\varepsilon)$. Extend π to $\mathcal{C}$ by $\mathcal{G}$–equivariance. Clearly $\pi_{A+a,\Phi+\phi} \in X_{A+a,\Phi+\phi}$ and $\pi_{A,\Phi} = (b,\psi)$.

Lemma 5.2 *For $\varepsilon > 0$ sufficiently small, the perturbation π in (5.2) satisfies a uniform bound $\|\pi_{A,\Phi}\|_{L^2_{2,A}} \leq C$.*

Proof ξ satisfies $H(\xi,(a,\phi),\beta.(b,\psi)) = 0$. Thus

$$\Delta_{A,\Phi}\xi + N_1(a,\phi)\xi + N_2(a,\phi)(b,\psi) + \delta^{0*}_{A,\Phi}(\beta.(b,\psi)) = 0 \tag{5.3}$$

where $\Delta_{A,\Phi}$ is a second order elliptic operator with coefficients depending on (A,Φ) and N_1 and N_2 are lower order terms. N_1 is a bilinear expression in (a,ϕ) and $\delta^0_{A,\Phi}(\xi)$. N_2 is a bilinear expression in (a,ϕ) and (b,ψ). After some calculation it is seen that N_1, N_2 satisfy, by Sobolev theorems

$$\begin{aligned} \|N_1(a,\phi)\xi\|_{L^2_1} &\leq \text{const.}\|(a,\phi)\|_{L^2_2}\|\xi\|_{L^2_3} \\ \|N_2(a,\phi)(b,\psi)\|_{L^2_1} &\leq \text{const.}\|(a,\phi)\|_{L^2_2}\|(b,\psi)\|_{L^2_2}. \end{aligned} \tag{5.4}$$

On the other hand since $\Delta_{A,\Phi}$ is invertible on $(\ker\delta^0_{A,\Phi})^\perp$,

$$\|\xi\|_{L^2_3} \leq \text{const.}\|\Delta_{A,\Phi}\xi\|_{L^2_1}. \tag{5.5}$$

Now make $\varepsilon > 0$ sufficiently small so that $\|(a,\phi)\|_{L^2_2}$ is correspondingly small. Then (5.3), (5.4) and (5.5) give $\|\xi\|_{L^2_3} \leq \text{const.}\|(b,\psi)\|_{L^2_2}$. Thus by (5.2) we have a uniform bound

$$\|\pi_{A+a,\Phi+\phi}\|_{L^2_2} \leq \text{const.}\|(b,\psi)\|_{L^2_2} \leq C.$$

In the above the Sobolev norms were taken with respect to some fixed connection A_0, which is commensurate to the Sobolev norm taken to say A. If $\|a\|_{L^2_2}$ is sufficiently small then

$$\begin{aligned} \|\nabla^{A+a}\pi_{A+a,\Phi+\phi}\|_{L^2} &\leq \text{const.}\|\nabla^A\pi_{A+a,\Phi+\phi}\|_{L^2}, \\ \|\nabla^{A+a}\nabla^{A+a}\pi_{A+a,\Phi+\phi}\|_{L^2} &\leq \text{const.}\|\nabla^A\nabla^A\pi_{A+a,\Phi+\phi}\|_{L^2} \end{aligned}$$

uniformly. By reducing ε again if necessary, the bound $\|\pi_{A,\Phi}\|_{L^2_{2,A}} \leq C$ is established. □

This lemma directly shows

Proposition 5.3 *Assume $\Phi \neq 0$. Given any $(b,\psi) \in X_{A,\Phi}$ there is an admissible perturbation π such that $\pi_{A,\Phi} = (b,\psi)$. Furthermore the support of π may be chosen to be contained in an arbitarily small $\mathcal{G}$–invariant neighbourhood of the orbit $\mathcal{G}\cdot(A,\Phi)$.*

The slice at a reducible $(A,0)$ has a natural splitting $X_{A,0} = X^r_A \times L^2_2(S\otimes_{\mathbf{H}} E)$. The stabilizer of $(A,0)$ (which is $\{\pm 1\}$, $U(1)$ or $SU(2)$) acts diagonally on both of the factors X^r_A and $L^2_2(S\otimes_{\mathbf{H}} E)$. If π is a perturbation then the stabilizer action forces the normal or spinor component of $\pi_{A,0}$ to be zero, since π is required to be $\mathcal{G}$–equivariant.

Assume the case that A is irreducible as a connection. Then the stabilizer of $(A,0)$ is $\{\pm 1\}$ and this acts on the $L^2_2(S\otimes_{\mathbf{H}} E)$ factor only, by multiplication. Let $b \in X^r_A$ and set

$$\pi'_{A+a,\Phi+\phi} = \beta(a,\phi)(b,0) + \delta^0_{A,\Phi}\xi(a,\phi;\beta(a,\phi)(b,0)).$$

Then in the same manner as Lemma 5.2 π' is admissible provided the support of β is small, and by construction $\pi'_{A,0} = (b, 0)$. This defines perturbations in the connection irreducible portion $\mathcal{A}^*$ of the reducible strata $\mathcal{A} \subset \mathcal{C}$.

Let us now consider the normal direction linearization $(L^\nu \pi)_A$ of any perturbation π at $A \in \mathcal{A}$. In preparation for this we need a little technical result:

Lemma 5.4 *Let $V \to Y$ be a trivial real vector bundle of rank ≥ 2 and let $L\colon L^2_2(V) \to L^2_2(V)$ be a bounded linear operator. Regard $L^2_2(V) \subset C^0(V)$. Suppose that $L(\sigma)(x) \in \langle \sigma(x) \rangle$ wherever $\sigma(x) \neq 0$. Then there exists a real function $f \in L^2_2(Y)$ such that $L(\sigma) = f\sigma$ for all σ.*

Proof Let σ be a nowhere zero section. Then $L(\sigma) = f\sigma$ for some $f \in C^0(Y)$. Let σ_1 be a section which is pointwise linearly independent to σ wherever it is non-zero. Then $L(\sigma_1)(x) = f_1(x)\sigma_1(x)$ for some f_1 at such points. However it must also be the case that

$$L(\sigma + \sigma_1)(x) = h(x)(\sigma + \sigma_1)(x)$$

for some h. If $\sigma_1(x) \neq 0$ this leads to the relation

$$(f(x) - h(x))\sigma(x) = (h(x) - f_1(x))\sigma_1(x)$$

which forces $f(x) = h(x) = f_1(x)$. On the other hand if $\sigma_1(x) = 0$ then we obtain

$$L(\sigma_1)(x) = (h(x) - f(x))\sigma(x).$$

Since we have the freedom to make other choices for σ the only possibility is that $h(x) = f(x)$ and so $L(\sigma_1)(x) = 0$ wherever $\sigma_1(x) = 0$. Thus $L(\sigma_1) = f\sigma_1$ i.e., $L(\sigma_1)(x) = f(x)\sigma_1(x)$ *for all* x.

Choose σ_1 to be nowhere vanishing and reverse the roles of σ and σ_1 above. Then we obtain $L(k\sigma) = f(k\sigma)$ for any function $k \in L^2_2(Y)$. Finally, given *any* section σ' we may write this as a sum $\sigma' = k\sigma + \sigma_1$ where σ and σ_1 are as in the preceding paragraph. Then

$$L(\sigma') = L(k\sigma + \sigma_1) = f(k\sigma) + f\sigma_1 = f\sigma'.$$

If $L(\sigma) = f\sigma \in L^2_2$ for all $\sigma \in L^2_2$ then it must be the case that $f \in L^2_2$ as well. □

The next results limits the possibilities for the normal linearization of a perturbation which in turn forces it to be self-adjoint:

Lemma 5.5 *Given any admissible perturbation π and $A \in \mathcal{A}$ there is a real function $f_A \in L^2_2$ on Y such that $(L^\nu \pi)_A(\delta\phi) = f_A \delta\phi$ for all $\delta\phi \in L^2_2(S \otimes_{\mathbf{H}} E)$. It follows that $(L^\nu\pi)_A$ is L^2–self-adjoint on $L^2_2 \subset L^2(S \otimes_{\mathbf{H}} E)$.*

Proof Assume an admissible perturbation π is given. Then

$$\langle (d_A\gamma, -\gamma(\phi)), \pi_{A,\phi} \rangle_{L^2} = 0 \quad \text{for all } \gamma \in L^2_3(\mathrm{ad}E).$$

Performing a variation $\phi \mapsto \phi + \delta\phi$ at $\phi = 0$ gives $\langle \gamma(\delta\phi), (L^\nu\pi)_A(\delta\phi) \rangle_{L^2} = 0$ for all γ. Write $(L^\nu\pi)_A(\delta\phi) = \delta\psi$. By assumption $\delta\phi, \delta\psi \in L^2_2 \subset C^0$ so we can consider them as continuous sections. Then pointwise we have $\langle \gamma(\delta\phi), \delta\psi \rangle_x = 0$. A local model for the fibre of $S \otimes_{\mathbf{H}} E$ is just $\mathbf{H}$ and with the action of γ as multiplication by $\mathrm{Im}\,\mathbf{H}$. Thus we see that $\delta\psi(x) = (L^\nu\pi)_A(\delta\phi)(x) \in \langle \delta\phi(x) \rangle$ at all points x where $\delta\phi(x) \neq 0$. The proof is completed by Lemma 5.4. □

Let us now construct perturbations normal to $\mathcal{A} \subset \mathcal{C}$. Assume the cutoff β on $X_{A,0}$ is invariant under the stabilizer action. Let f_A be a real L^2_2 function on Y. Set

$$\begin{aligned} \pi''_{A+a,\Phi+\phi} &= \beta(a,\phi)(0, f_A\phi) + \delta^0_{A,\Phi}\xi(a,\phi;\beta(a,\phi)(0,f_A\phi)) \\ &= \beta(a,\phi) f_A \phi \in X_{A+a,\Phi+\phi}. \end{aligned}$$

This is again an admissible perturbation for $\mathrm{supp}\,\beta$ small and the linearization of π'' in a normal direction $\delta\phi$ at $(A,0)$ is $(L^\nu\pi'')_A(\delta\phi) = f_A\delta\phi$.

Thus we have:

Proposition 5.6 *If A is irreducible then given any $b \in X^r_A$ there is an admissible perturbation π' such that $\pi'_{A,0} = (b,0) \in X_{A,0} = X^r_A \times L^2_2(S \otimes_{\mathbf{H}} E)$. On the other hand for any A there exists an admissible perturbation π'' such that $\overline{\pi}'' = 0$ and $(L^\nu\pi'')_A(\delta\phi) = f_A\delta\phi$ given any real function $f_A \in L^2_2(Y)$. Furthermore the support of π' and π'' may be chosen to be contained in an arbitarily small $\mathcal{G}$–invariant neighbourhood of the orbit $\mathcal{G}\cdot(A,0)$ in $\mathcal{C}$.*

Proof of Proposition 2.13

Let $\mathcal{X}^r_{\overline{\pi}}(A) = *F_A + *\overline{\pi}_A \in X^r_A$. Then $(\mathcal{X}^r_{\overline{\pi}})^{-1}(0)/\mathcal{G}_E = \mathcal{M}^r_{\overline{\pi}}$. Let H^τ_A denote the cokernel of $(L\mathcal{X}^r_\pi)_A$: $L^2_2(\Lambda^1 \otimes \mathrm{ad}E) \to X^r_A \cap L^2_1$. Then at the reducible $A = (A,0)$ the cokernel H^2_A of $(L\mathcal{X}_\pi)_{A,0}$ splits as $H^\tau_A \oplus H^\nu_A$. H^ν_A is the cokernel of the normal operator D^π_A. (Recall the map $\mathcal{X}_\pi = \mathcal{X} + \pi$ of (2.3) and its linearization (2.5).)

Step 1 For a ZHS the orbit of the trivial connection $[\Theta] \in \mathcal{M}^r$ is already isolated in $\mathcal{B}_\mathcal{A}$ since $H^\tau_\Theta \cong H^1(Y) = \{0\}$. By Proposition 5.6 and the compactness of $\mathcal{M}^{r*}$, we can find a finite set of perturbations $\{\pi^{(i)}\}$ with support away from $\{[\Theta]\}$ such that if $v \in H^\tau_A$, $A \in (\mathcal{X}^r)^{-1}(0) \cap \mathcal{A}^*$, is L^2–orthogonal to each $\pi^{(i)}_A$ then $v = 0$. Thus by Sard–Smale there is a perturbation, call it π_1 so that $(\mathcal{X}^r_{\pi_1})^{-1}(0)$ is cut out equivariantly transversely over $\mathcal{A}^*$, i.e. $H^\tau_A = \{0\}$ for every $A \in (\mathcal{X}^r_{\pi_1})^{-1}(0) \cap \mathcal{A}^*$. Hence $\mathcal{M}^r_{\pi_1}$ is, by the local Kuranishi model, a finite set of points which are *non-degenerate within* $\mathcal{B}_A$.

Step 2 Let $[A] \in \mathcal{M}^r_{\pi_1}$. The normal operator $D^{\pi_1}_A$ at A is of the form $D_A + f_A$, by Lemma 5.5. This operator is self-adjoint Fredholm and therefore has discrete spectrum. Let π_2 be a perturbation with the property that $\overline{\pi}_2 = 0$ and $(L^\nu \pi_2)_A \delta\phi = \mu_A \delta\phi$, $\mu_A \in \mathbf{R}$ where $|\mu_A|$ is less than the distance of the closest non-zero eigenvalue of $D^{\pi_1}_A$ from zero. Then $D^{\pi_1+\pi_2}_A$ has trivial kernel and $[A]$ is a non-degenerate point in $\mathcal{M}_{\pi_1+\pi_2}$. π_2 can be chosen to have support in an arbitarily small $\mathcal{G}$–invariant neighbourhood of the orbit of A. Repeating this procedure for every $[A]$ we can find a perturbation π' such that $\mathcal{M}^r_{\pi'}$ consists entirely of non-degenerate points within $\mathcal{B}$.

Step 3 After the preceding steps, $\mathcal{M}^r_{\pi'}$ is isolated in $\mathcal{M}_{\pi'}$. By Proposition 5.3 and the compactness of $\mathcal{M}^*_{\pi'}$ we can find a finite set of perturbations $\{\pi^{(j)}\}$ supported away from $\mathcal{M}^r_{\pi'}$ such that if $v \in H^2_x$, $x \in \mathcal{X}^{-1}_{\pi'}(0) \cap \mathcal{C}^*$, is L^2–orthogonal to every $\pi^{(j)}$, then $v = 0$. Thus by Sard–Smale there exists a π'_1 which is an arbitarily small linear combination of these $\pi^{(j)}$'s, such that $\mathcal{X}^{-1}_{\pi_1+\pi'_1}(0)$ is cutout equivariantly transversely over $\mathcal{C}^*$, i.e. $H^2_x = \{0\}$ for every $x \in \mathcal{X}^{-1}_{\pi_1+\pi'_1}(0) \cap \mathcal{C}^*$. Note that π'_1 is supported away from $\mathcal{M}^r_{\pi'}$. Choosing our final perturbation π to be $\pi' + \pi'_1$ we get $\mathcal{M}_\pi$ non-degenerate.

At every stage in Steps 1, 2 and 3 we can make the chosen perturbation as small as we like in the uniform norm

$$\|\pi'\|_{\mathcal{B}} = \sup_{A,\Phi}\Big\{ \|\pi'_{A,\Phi}\|_{L^2_{2,A}} \Big\}.$$

This completes the proof of Proposition 2.13.

6 Proof of Theorem 3.7

Let (g_0, π_0) and (g_1, π_1) be given. Assume that π_i is non-degenerate with respect to g_i. In order to compare the moduli spaces for different metrics we may assume, as in section 4 a fixed model for the spinor bundle with respect to g_0. Then we have a SW–equation depending smoothly on the parameter t corresponding to the metric $g_t = (1-t)g_0 + tg_1$ and with perturbation $\pi_t = (1-t)\pi_0 + t\pi_1$. In this section we shall assume objects sub- or superscripted with 't' are with respect to g_t.

To the family $\{(g_t,\pi_t)\}$ we have a *parameterized moduli space*

$$Z = \bigcup_t \mathcal{M}_{g_t,\pi_t} \times \{t\} \subset \mathcal{C} \times [0,1].$$

As in [3] and [9] to prove invariance of $\tau(Y)$ we need to show that Z is, after suitable perturbation, a compact 1–dimensional cobordism with the appropriate singularities. The counter-terms in the definition of $\tau(Y)$ are due to these singularities.

In our analysis of Z we work first with the reducible strata Z^r. In the following the notation Z^{r*} denotes the connection irreducible portion of Z^r. In the parameterized context an *admissible time-dependent perturbation* σ is one which is a finite sum $\sum_i \varrho_i(t)\pi^{(i)}$ where $\pi^{(i)}$ is admissible and ϱ_i has support in $[0,1]$. Z_σ, Z^r_σ etc. shall denote perturbed parameterized moduli spaces. Recall the uniform norm, for non-time-dependent perturbations,

$$\|\pi\|_{\mathcal{B}} = \sup_{A,\Phi}\Big\{ \|\pi_{A,\Phi}\|_{L^2_{2,A}} \Big\}.$$

Lemma 6.1 *There exists an admissible time-dependent perturbation σ such that the perturbed parameterized reducible moduli space Z^r_σ is non-degenerate as a subspace of $\mathcal{B}_A \times [0,1]$. Furthermore if $\|\pi_{0,1}\|_{\mathcal{B}} < \delta_0$ then we can assume $\|\pi_t + \sigma(t)\|_{\mathcal{B}} < 2\delta_0$.*

Z^r_σ can be regarded as the $\mathcal{G}$–quotient of the zeros of the map $\mathcal{X}^r_\sigma(A,t) = *_t F_A + \overline{\sigma}(t)_A \in X^r_A$. The proof of the Lemma follows easily from constructing and applying time-dependent perturbations to $\mathcal{X}^r_\sigma$ supported away from $\{[\Theta]\}$ in the manner of section 5. In this way the strata corresponding to the trivial connection is isolated in Z^r_σ and is the product $\{[\Theta]\} \times [0,1]$. The irreducible portion of Z^r_σ with the choice of σ in Lemma 6.1 is a compact corbordism between $\mathcal{M}^{r*}_{g_0,\pi_0}$ and $\mathcal{M}^{r*}_{g_1,\pi_1}$.

Assume now Z^r_σ as in Lemma 6.1. The normal operator at A with respect to $(g_t, \pi_t + \sigma(t))$ will be denoted by $D^{t,\sigma}_A$ (Eq. (3.1)). The kernel of this operator (= cokernel by Lemma 5.5) is the *normal cohomology* $H^\nu_{A,t}$.

Let $u \mapsto ([A(u)], t(u))$, $|u| < \varepsilon$ be a 1–1 parameterization of an open subset $\mathcal{J}$ of Z^r_σ. Let $\mathbf{K} = \mathbf{R}$ if $\mathcal{J}$ is in the connection-irreducible strata and $\mathbf{K} = \mathbf{H}$ if $\mathcal{J}$ is in the connection-trivial strata.

Definition 6.2 Call Z^r_σ *normally transverse* along $\mathcal{J}$ if the family $\{D^{t(u),\sigma}_{A(u)}\}$ has transverse spectral flow as $\mathbf{K}$–linear operators. (Recall that transverse spectral flow is the situation of simple eigenvalues, with respect to $\mathbf{K}$, crossing zero transversely.) Call Z^r_σ *normally tranverse* if it is normally transverse in a neighbourhood of every point.

In terms of local models, let $A = A(0)$, $t_0 = t(0)$ and $\mathcal{U}$ be a sufficiently small stab(A_0)–invariant neighbourhood of (A_0, t_0) in the slice $(A_0, t_0) + X^r_A \times [0,1]$. $\mathcal{U}$ can be identified with a neighbourhood of $([A_0], t_0)$ in $\mathcal{B}_A \times [0,1]$. To simplify notation henceforth denote $H^\nu_{A_0,t_0}$ by H^ν_0. Assume H^ν_0 is non-trivial, otherwise $\mathcal{U}$ can be chosen such that $D^{t,\sigma}_A$ is invertible in $\mathcal{U}$. Consider the restriction of $D^{t,\sigma}_A$ to the normal cohomology H^ν_0 followed by L^2–projection $\prod$ back onto H^ν_0. This determines, for each $(A,t) \in \mathcal{U}$ a symmetric operator $T(A,t)$ acting on H^ν_0. The latter space is endowed with the natural L^2–inner product. Then the kernel (cokernel) of $D^{t,\sigma}_A$ is exactly modelled by the kernel (cokernel) of $T(A,t)$. Denote the symmetric operators which commute with $\mathbf{K}$ by $\mathrm{Sym}_{\mathbf{K}}(H^\nu_0)$. Let $u \mapsto (A(u), t(u)) \in X^r_A \times (0,1)$, $|u| < \varepsilon$ be a 1–1 parameterization of an open subset $\mathcal{J} = \mathcal{U} \cap (\mathcal{X}^r_\sigma)^{-1}(0)$ of Z^r_σ. Then the condition of being normally transverse along $\mathcal{J}$ translates as (i) $H^\nu_0 \cong \mathbf{K}$ and (ii) the path $u \mapsto T(A(u), t(u)) \in \mathrm{Sym}_{\mathbf{K}}(H^\nu_0) \cong \mathbf{R}$ is transverse to $\{0\}$.

Lemma 6.3 *Assume σ as in Lemma 6.1. There exist an admissible time-dependent perturbation σ' such that $Z^r_{\sigma+\sigma'} \simeq Z^r_\sigma$ and $Z^r_{\sigma+\sigma'}$ is normally transverse. Furthermore if $\|\pi_{0,1}\|_{\mathcal{B}} < \delta_0$ we can assume $\|\pi_t + \sigma(t) + \sigma'(t)\|_{\mathcal{B}} < 3\delta_0$.*

Proof We divide the argument into the separate cases of the irreducible and trivial strata of Z^r_σ. No matter what perturbation σ is chosen the trivial strata is always $\{[\Theta]\} \times [0,1]$. However changing σ can change Z^{r*}_σ. The space $\mathcal{S}$ of admissible perturbations is a normed linear space, with the norm $\|\cdot\|_{\mathcal{B}}$. Since Z^{r*}_σ is already non-degenerate as a subspace of $\mathcal{B}^*_A \times [0,1]$ i.e. $(\mathcal{X}^r_\sigma)^{-1}(0) \cap \mathcal{A}^* \times [0,1]$ is cutout equivariantly transversely, it follows that for any sufficiently uniformly small $\sigma' \in \mathcal{S}$, Z^{r*}_σ and $Z^{r*}_{\sigma+\sigma'}$ are related by a cobordism which is a product and thus are diffeomorphic spaces. In fact the transverse condition means that the normal bundle to Z^{r*}_σ in $\mathcal{B}_A \times [0,1]$ at any point is isomorphic to $\mathcal{S}/\mathcal{S}_0$ where $\mathcal{S}_0$ is the subspace of those π such that $\overline{\pi} = 0$.

Case 1: Irreducible strata Fix $([A_0], t_0) \in Z^{r*}_\sigma$, $t \neq 0,1$ and let $\mathcal{U}$ be a sufficiently small $\mathcal{G}$–invariant neighbourhood of (A_0, t_0) in $(A_0, t_0) + X^r_A \times [0,1]$ such that a local model for $D^{t,\sigma}_A$ as above exists in $\mathcal{U}$. Assume H^ν_0 is non-trivial. We examine the effect of a perturbation on the family $D^{t,\sigma}_A$ along Z^{r*}_σ.

Consider the parameterized local model map $P\colon \mathcal{U} \times \mathcal{S} \to \mathrm{Sym}_{\mathbf{R}}(H^\nu_0)$ based at $(A_0, t_0, 0)$ given by

$$P(A,t,\pi) = \prod \circ D^{t,\sigma+\beta(t)\pi}_A$$

where $\beta = \beta(t)$ is a cutoff function on $\mathbf{R}$ with support close to t_0. Note that $D^{t,\sigma+\pi}_A = D^t_A + (L^\nu\sigma(t) + L^\nu\pi)_A$. By Lemma 5.5, $(L^\nu\sigma(t))_A\phi = f_{A,t}\phi$ and

$(L^\nu \pi)_A\phi = h_A\phi$ for some functions $f_{A,t}$ and h_A on Y. Since $\mathcal{S}$ is a linear space, we may identify tangent vectors $\delta\pi$ with elements π in $\mathcal{S}$. We then have, for the derivative of P at $(A_0, t_0, 0)$,

$$dP(\delta\pi)\phi = \prod\Big(h_{A_0}\phi\Big), \quad dP(\delta a)\phi = \prod\Big(\delta a \cdot \phi + (Lf)_{A_0,t_0}(\delta a)\phi\Big).$$

By choosing $h = h_{A_0} = 1$ we see that the image of dP includes at least the span of the identity operator in $\mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$; thus $\mathrm{rank}(dP) \geq 1$.

Claim *If* $\dim_{\mathbf{R}}(H_0^\nu) > 1$ *then* $\mathrm{rank}(dP) \geq 2$.

In order to establish the claim we invoke the unique continuation principle for H_0^ν i.e. if $\phi \in H_0^\nu$ then ϕ cannot vanish on an open set unless $\phi = 0$. Writing $A_0 = \Theta + a$ where Θ is smooth, then $\phi \in H_0^\nu$ is a solution of the perturbed smooth Dirac operator $D_\Theta\phi + a\cdot\phi + f\phi = 0$ where a and f are continuous. Unique continuation holds for such solutions.

Let $\{\phi_1, \ldots, \phi_n\}$, $n > 1$ be a $\mathbf{R}$–orthonormal basis for H_0^ν. The matrix of $dP(\delta\pi)$ with respect to this basis is $(\langle h\phi_i, \phi_j\rangle_{L^2})$. Assume the rank of dP is unity. This implies that $\langle h\phi_i, \phi_j\rangle_{L^2} = 0$ for all h and $i \neq j$. This in turn implies the pointwise orthogonal condition $\langle \phi_i, \phi_j\rangle_y = 0$, $i \neq j$ for all $y \in Y$. It then follows that $\langle dP(\delta a)\phi_i, \phi_j\rangle_{L^2} = \langle \delta a\cdot\phi_i, \phi_j\rangle_{L^2}$, $i \neq j$. However the Clifford action of $\Lambda^1 \otimes \mathrm{ad}E$ on $S \otimes_{\mathbf{H}} E$ is fibrewise transitive. Thus we can find a δa such that $\langle \delta a \cdot \phi_i, \phi_j\rangle_{L^2} \neq 0$, $i \neq j$. This proves that the image of dP is not contained in the span of the identity in $\mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$. Therefore dP is at least rank two and the claim is proven.

Let $u \mapsto (A(u), t(u)) \in X_A^r \times (0,1)$, $|u| < \varepsilon$ be a 1–1 parameterization of $\mathcal{U} \cap (\mathcal{X}_\sigma^r)^{-1}(0)$ with $A(0) = A_0$ and $t(0) = t_0$. Let $T_\sigma(u) = P(A(u), t(u), 0)$ be the local model for $D_{A(u)}^{t(u),\sigma}$ on $\mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$. By construction, $P(A_0, t_0, 0)$ is the zero operator on $\mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$. In $\mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$ the space of invertible operators is a codimension one real variety $\mathcal{V}$. Any point which is not the zero operator in this variety represents an operator of non-trivial rank.

Let X^r be the vector bundle over $\mathcal{A}^*$ whose fiber at A is the slice space X_A^r and let $Q\colon \mathcal{U} \times \mathcal{S} \to X^r \times \mathrm{Sym}_{\mathbf{R}}(H_0^\nu)$ be given by

$$Q(A, t, \pi) = (\mathcal{X}^r_{\sigma+\beta(t)\pi}(A), P(A, t, \pi)).$$

Since this is a submersion onto the first factor along $(\mathcal{X}_\sigma^r)^{-1}(0) \cap \mathcal{A}^* \times [0,1]$ (the transversality condition) and $\mathrm{rank}(dP) \geq 2$ if $\dim_{\mathbf{R}}(H_0^\nu) > 1$ and is onto if $\dim_{\mathbf{R}}(H_0^\nu) = 1$, then there is a time-dependent perturbation $\sigma_0(t)\colon = \beta(t)\pi$ such that the deformation of the family $\{T_{\sigma+s\sigma_0}\}$ at $s = 0$ is normal to the path $T_\sigma = T_\sigma(u)$. Therefore we can choose an arbitarily small σ' so that the operators $T_{\sigma+\sigma'}(u)$ have non-trivial rank for all u. (Note: at this stage we do not have sufficently many perturbations in hand to make $T_{\sigma+\sigma'}$ transverse to $\mathcal{V}$.) Thus if we work with $\sigma+\sigma'$ we find that the rank (over $\mathbf{R}$) of $H^{\nu}_{A(u),u}$ near $u=0$ drops by one if $\dim_{\mathbf{R}}(H^{\nu}_0) > 1$ and becomes transverse to $\mathcal{V}=\{0\}$ if $\dim_{\mathbf{R}}(H^{\nu}_0)=1$. To complete the argument to obtain normal transversality globally over the connection-irreducible strata, proceed by an induction argument with the overall rank of $H^{\nu}_{A,t}$ over $Z^{r*}_{\sigma+\sigma'}$ decreasing by one in each step. Letting σ' denote the final perturbation we see that over $Z^{r*}_{\sigma+\sigma'}$ there exists a finite number of points where $H^{\nu}_{A,t}$ is non-trivial and these points $H^{\nu}_{A,t} \cong \mathbf{R}$ and with $T_{\sigma+\sigma'}$ transverse to $\mathcal{V}=\{0\}$. This is equivalent to transverse spectral flow. The last assertion of the lemma in this case is a consequence of the observation that the induction is completed in a finite number of steps and in each step we may take the perturbation to be as small as we like.

Case 2: Trivial strata Let $([A_0],t_0) \in \{[\Theta]\} \times [0,1]$. Here the relevant parameterized local model map P is the same as the map P as above but with $A=\Theta$ fixed, i.e. $P\colon \mathcal{S} \to \mathrm{Sym}_{\mathbf{H}}(H^{\nu}_0)$. The argument proceeds just as before (but without the complication of the deformation in the moduli space) provided we can again establish that if $\dim_{\mathbf{H}}(H^{\nu}_0) > 1$ then $\mathrm{rank}(dP) \geq 2$. This time let $\{\phi_1,\ldots,\phi_n\}$, $n>1$ be a $\mathbf{H}$–orthonormal basis for H^{ν}_0. Again if we assume the rank of dP is unity we get the pointwise orthogonal condition $\langle \phi_i,\phi_j\rangle_y = 0$, $i \neq j$ for all $y \in Y$. However this would mean that $S \otimes_{\mathbf{H}} E$ has at least 8 pointwise orthogonal non-zero sections. This is impossible since $S \otimes_{\mathbf{H}} E$ is rank 4.

This completes the proof of the lemma. □

Remark 6.4 A more satisfactory result would be that P is a submersion onto $\mathrm{Sym}_{\mathbf{K}}(H^{\nu}_0)$ which is the situation in [3]; then transverse spectral flow follows easily by Sard–Smale. A submersion does not seem to be generally true in our and the original SW context. The same problem is encountered in [10] and [11].

Definition 6.5 Suppose Z^r_σ is normally tranverse and let $u \mapsto ([A(u)],t(u))$, $|u|<\varepsilon$ be a 1–1–parameterization of an open neighbourhood in Z^r_σ. A point in Z^r_σ which is contained in such a parameterization and where there is spectral flow for $D^{t(u),\sigma}_{A(u)}$ is called a *singular* or *bifurcation* point.

At a singular point $([A_0],t_0)$, the local model for Z_σ is the quotient by $\mathrm{stab}(A_0)$ of the zeros of a $\mathrm{stab}(A_0)$–equivariant obstruction map $\Xi\colon H^{\nu}_0 \times \mathbf{R} \to H^{\nu}_0$ of the form

$$\Xi(q,t) = qt.$$

(See [10] and [3].) This in turn implies that the a neighbourhood of $([A_0], t_0)$ is the zeros of the map $[0,\infty) \times \mathbf{R} \to \mathbf{R}$, $(r,t) \mapsto rt$ with $\{0\} \times \mathbf{R}$ corresponding to the reducible portion and $(0,\infty) \times \{0\}$ the irreducible. One other consequence of the local model in this normal transverse situation is that the points corresponding to the irreducibles sufficiently near $([A_0], t_0)$ are non-degenerate.

On the other hand, at a non-singular point $([A_0], t_0)$ of a normally transverse Z^r_σ the Kuranishi local model gives a neighbourhood of $([A_0], t_0)$ in $\mathcal{B} \times [0,1]$ an isolated open interval.

Corollary 6.6 *Assume σ as in Lemma 6.1. There exists a time-dependent admissible perturbation σ' such that (i) $Z^r_{\sigma+\sigma'} \simeq Z^r_\sigma$ (ii) $Z^r_{\sigma+\sigma'}$ is normally transverse and (iii) $Z^*_{\sigma+\sigma'}$ is non-degenerate. Furthermore if $\|\pi_{0,1}\|_{\mathcal{B}} < \delta_0$ we can assume $\|\pi_t + \sigma(t) + \sigma'(t)\|_{\mathcal{B}} < 4\delta_0$.*

Proof Run through the proof of Lemma 6.3. The comments above tell us that $Z^*_{\sigma+\sigma'}$ is non-degenerate in a neighbourhood of $Z^r_{\sigma+\sigma'}$. Now construct and apply admissible time-dependent perturbations σ'' in the manner of section 5, which can be chosen to have support away from $Z^r_{\sigma+\sigma'}$, making all of $Z^*_{\sigma+\sigma'+\sigma''}$ non-degenerate. The perturbation σ'' can be chosen arbitaily small. □

Completion of proof of Theorem 3.7 As above we have two non-degenerate metrics and perturbations (g_0, π_0) and (g_1, π_1) where π_i is small with respect to g_i.

Assume first the case that the metric $g = g_0 = g_1$ is unchanging. The condition π_0, π_1 are *small* (Definition 3.6) implies $\mathcal{M}^{r*}_{\pi_i} \subset \cup_j \mathcal{N}^j$ where the $\mathcal{N}^j$ are as in the definition of the proposed invariant. By Corollary 6.6 we can find a parameterized moduli space Z_σ such that

(i) Z^{r*}_σ is a smooth compact 1–dimensional corbodism between $\mathcal{M}^{r*}_{\pi_0}$ and $\mathcal{M}^{r*}_{\pi_1}$. Additionally we know from [12] that this is an oriented cobordism so that it's boundary is $\mathcal{M}^{r*}_{\pi_1} - \mathcal{M}^{r*}_{\pi_0}$ where $\mathcal{M}^{r*}_{\pi_{0,1}}$ are given Taubes' orientation

(ii) $Z^{r*}_\sigma \subset \cup_j \mathcal{N}^j$

(iii) $\overline{Z^*_\sigma}$ is a smooth compact 1–manifold with boundary

$$\mathcal{M}^*_{\pi_0} \cup \mathcal{M}^*_{\pi_1} \cup \{\text{singular points in } Z^r_\sigma\}.$$

Just as in [3] it is seen that

$$\begin{aligned} \sum_{[A]\in\mathcal{M}^{r*}_{\pi_1}} \varepsilon[A]\kappa[A] - \sum_{[A]\in\mathcal{M}^{r*}_{\pi_0}} \varepsilon[A]\kappa[A] \\ = \#\{\text{singular points on } Z^{r*}_\sigma\} \bmod 2. \end{aligned} \tag{6.1}$$

For completeness we give an argument. Fix a component $\mathcal{N}^j$ and consider $Z^{r*}_\sigma \cap \mathcal{N}^j$. In the definition of $\kappa[A]$ for $[A] \in \mathcal{M}^{r*}_{\pi_{0,1}} \cap \mathcal{N}^j$ choose all the paths $[\gamma]$ to be in the same homotopy class rel $\{[\Theta]\} \cup \mathcal{N}^j$. Then for these $[A]$'s the term $\mathrm{cs}(\overline{\gamma}(0)) - \mathrm{cs}(\overline{\gamma}(1))$ is the same constant. Make this choice. Then $\kappa[A]$ is the normal spectral flow $\mathrm{SF}^\nu(\gamma)$ from $[\Theta]$ to $[A]$ in the given fixed homotopy class of $[\gamma]$ plus a fixed additive constant. Notice then that $\kappa[A]$ changes exactly by the normal spectral flow as we vary $[A]$ within $\mathcal{N}^j$. Let Γ be a connected component of $Z^{r*}_\sigma \cap \mathcal{N}^j$ with non-empty boundary $\{[A],[A']\} \subset \mathcal{M}^{r*}_{\pi_{0,1}} \cap \mathcal{N}^j$. After some consideration it is seen that the three following sums compute the mod2 normal spectral flow along Γ and thus the mod2 cardinality of the singular points on Γ:

(i) $\varepsilon[A]\kappa[A] + \varepsilon[A']\kappa[A'] = \pm(\kappa[A] - \kappa[A'])$ when $[A],[A'] \in \mathcal{M}^{r*}_{\pi_1}$

(ii) $-\varepsilon[A]\kappa[A] - \varepsilon[A']\kappa[A'] = \pm(\kappa[A] - \kappa[A'])$ when $[A],[A'] \in \mathcal{M}^{r*}_{\pi_0}$

(iii) $\varepsilon[A]\kappa[A] - \varepsilon[A']\kappa[A'] = \pm(\kappa[A] - \kappa[A'])$ when $[A] \in \mathcal{M}^{r*}_{\pi_1}$, $[A'] \in \mathcal{M}^{r*}_{\pi_0}$.

On the other hand, if Γ has empty boundary then the number of singular points on Γ equals the normal spectral flow around Γ and this is zero, since it is contained within $\mathcal{N}^j$. From this it is straightforward to deduce (6.1) by rearranging the sum.

Next we compute that the difference

$$\begin{aligned}
&\frac{1}{2}c(g,\pi_1) - \frac{1}{2}c(g,\pi_0) \\
&= \mathbf{H}\text{–spectral flow of } \{D^{t,\sigma}_\Theta\}^1_{t=0} \\
&= \pm\#\Big\{\text{singular points on trivial strata } \{[\Theta]\} \times [0,1]\Big\}.
\end{aligned} \tag{6.2}$$

Finally we have equality of the sums

$$\sum_{\mathcal{M}^{r*}_{\pi_1}} \frac{1}{4}\varepsilon[A] = \sum_{\mathcal{M}^{r*}_{\pi_0}} \frac{1}{4}\varepsilon[A] \tag{6.3}$$

both being 1/2 of the algebraic sum which is Casson's invariant [12]. Thus from (6.1), (6.2), (6.3) we find that

$$\begin{aligned}
&\frac{1}{2}c(g,\pi_1) + \sum_{\mathcal{M}^{r*}_{\pi_1}} \varepsilon[A]\Big(\kappa[A] + \frac{1}{4}\Big) \\
&- \frac{1}{2}c(g,\pi_0) - \sum_{\mathcal{M}^{r*}_{\pi_0}} \varepsilon[A]\Big(\kappa[A] + \frac{1}{4}\Big) \\
&\equiv \#\Big\{\text{singular points on } Z^r_\sigma\Big\} \mod 2 \\
&\equiv \sum_{\mathcal{M}^*_{\pi_0}} 1 - \sum_{\mathcal{M}^*_{\pi_1}} 1 \mod 2.
\end{aligned}$$

The last line follows from $\overline{Z^*_\sigma}$ being a smooth compact 1–manifold with boundary $\mathcal{M}^*_{\pi_0} \cup \mathcal{M}^*_{\pi_1} \cup \{\text{singular points on } Z^r_\sigma\}$. Thus the independence of $\tau(Y)$ on choice of small, non-degenerate perturbation π is established.

The general case $g_0 \neq g_1$ follows an identical argument except for the following details. When varying the metric spectral flow can occur at the trivial connection Θ in $\mathrm{SF}^\nu(\gamma)$, which is the initial point of γ. However the operator D_Θ at this point is quaternionic and thus there is no change mod2. Secondly the neighbourhoods $\mathcal{N}^j$ are defined with reference to the background metric, thus we get for the different metrics g_0, g_1 two sets of neighbourhoods $\mathcal{N}^j_0$, $\mathcal{N}^j_1$. However by what we have established we can make any choice of (non-degenerate) $\pi_{0,1}$ we like. Choose $\pi_{0,1}$ sufficiently small in norm so that $\mathcal{M}^{r*}_{\pi_{0,1}} \subset \cup_j(\mathcal{N}^j_0 \cap \mathcal{N}^j_1)$. Then we may proceed with the rest of the argument as before. This proves that $\tau(Y)$ is an invariant.

Finally, let us show that $\tau(-Y) = \tau(Y)$, $-Y$ denoting Y with the reversed orientation. Reversing orientation but keeping the metric, spin structure $P \to Y$ and and spinor bundle S fixed simply changes the action of Clifford mutiplication by -1. The SW–equation of the orientation reversed structure is the same as the orginal except that the Dirac operator D_A switches to $-D_A$. If $\pi = (*k, l)$ is the non-degenerate and small perturbation used to compute $\tau(Y)$ then choose $\pi' = (*k, -l)$ for the reversed structure. Thus if (A, Φ) is a SW–solution with respect to π then $(A, -\Phi)$ is a solution of the orientation reversed situation for π'. In the following $\mathcal{M}^-$, ε^- etc. will refer to the reversed orientation structure. Thus $\mathcal{M}^-_{\pi'} = \mathcal{M}_\pi$ and π' is a non-degenerate small perturbation for $\mathcal{M}^-$.

The normal and tangential deformation operators D^π_A and $K^{\overline{\pi}}_A$ in the reversed situation are the negatives of those in the original. Then $\mathrm{SF}^{\nu,-}(\gamma) = -\mathrm{SF}^\nu(\gamma) - \dim\ker K^{\overline{\pi}}_{\gamma(0)} \equiv \mathrm{SF}^\nu(\gamma) \bmod 2$ since $K^{\overline{\pi}}_{\gamma(0)} \equiv 0 \bmod 2$. The orientation for $\mathrm{detind}(-K^{\overline{\pi}}) = \mathrm{detind}(K^{\overline{\pi}})$ on the other hand is reversed by the parity of $\dim\ker K^{\overline{\pi}}_{\gamma(0)} = 3$ as it's overall orientation is fixed by that at $[\Theta]$. The Chern–Simons functional as well as APS spectral invariants depend on the orientation of Y. Thus $\varepsilon^-[A]\kappa^-[A] = \varepsilon[A]\kappa[A]$, $c^-(g, \pi') = -c(g, \pi)$ and $\sum_{\mathcal{M}^{r*,-}_{\pi'}} \varepsilon^-[A] = -\sum_{\mathcal{M}^{r*}_\pi} \varepsilon[A]$.

Combining all of the above we obtain

$$\tau(Y) - \tau(-Y) = c(g, \pi) + \frac{1}{2} \sum_{\mathcal{M}^{r*}_\pi} \varepsilon[A].$$

Let $\lambda(Y)$ denote Casson's invariant [1]. In [12] it is established that

$$\frac{1}{2}\sum_{\mathcal{M}_{\pi}^{r*}} \varepsilon[A] = -\lambda(Y)$$

and it was proven by Casson that $\lambda(Y) \equiv \mu(Y) \bmod 2$. Since $c(g,\pi) \equiv \mu(Y) \bmod 2$ we obtain $\tau(Y) - \tau(-Y) \equiv 0 \bmod 2$. This completes the proof of Theorem 3.7.

7 Proof of Theorem 3.8

Let us begin by reviewing the $SU(3)$–Casson invariant (in our terminology). For more details refer to [3]. Denote by $\mathcal{M}^{SU(3)}$ the moduli space of flat $SU(3)$–connections on the trivial $SU(3)$ principal bundle over Y. As always Y is oriented. The reducible subspace is exactly $\mathcal{M}^{SU(2)}$, the moduli space of flat $SU(2)$–connections. This coincides with $\mathcal{M}^r$ in our SW–context. A suitable class of 'holonomy' perturbations h can be constructed so that the perturbed space $\mathcal{M}_h^{SU(3)}$ is non-degenerate. This means that it is a finite number of points. Additionally each irreducible point $[A]$ has an oriented $\widehat{\varepsilon}[A] \in \pm 1$ given by spectral flow. However the perturbed reducible portion $\mathcal{M}_h^{SU(3),r}$ does not consist of $SU(2)$–connections but essentially $U(2)$–connections. $\mathcal{M}_h^{SU(3),r}$ lies in $\mathcal{B}_{U(2)}$ the quotient space of $U(2)$–connections; as before there is a Chern–Simons function cs on connnections which descends to $\overline{\mathrm{cs}}\colon \mathcal{B}_{U(2)} \to \mathbf{R}/\mathbf{Z}$. To make an invariant out of $\sum_{\mathcal{M}_h^{SU(3)*}} \widehat{\varepsilon}[A]$ there are counter-terms associated to $\mathcal{M}_h^{SU(3)r*} = \mathcal{M}_h^{SU(3),r} - \{[\Theta]\}$. However we need to make h *small* which is the same condition used in our SW–context (and from which our definition originated). Denote by $\{\mathcal{N}_{SU(3)}^j\}$ the corresponding system of neighbourhoods of components of $\mathcal{M}^{SU(2)} - \{[\Theta]\}$ in $\mathcal{B}_{U(2)}$.

Along the reducible strata $\mathcal{B}_{U(2)}$ we have *tangential* and *normal* deformation operators giving rise to tangential and normal spectral flow quantities $\mathrm{SF}_{SU(3)}^{\tau}(\gamma)$ (real spectral flow), $\mathrm{SF}_{SU(3)}^{\nu}(\gamma)$ (complex spectral flow) along γ, respectively. The term $\mathrm{SF}_{SU(3)}^{\tau}(\gamma)$ is used to define Taubes' orientation $\varepsilon[A] = \pm 1$ for $[A] \in \mathcal{M}_h^{SU(3)r*}$. $\mathrm{SF}_{SU(3)}^{\nu}(\gamma)$ is used in the term

$$\kappa^{SU(3)}[A] = \mathrm{SF}_{SU(3)}^{\nu}(\gamma) + 2\mathrm{cs}(\overline{\gamma}(0)) - 2\mathrm{cs}(\overline{\gamma}(1)).$$

As before $[\gamma(t)]$, $0 \le t \le 1$ is a path from $[\Theta]$ to $[A] \in \mathcal{N}_{SU(3)}^j$ say and $[\overline{\gamma}]$ is the path from $[\Theta]$ to the component $\mathcal{K}^j \subset \mathcal{M}^{SU(2)}$, and homotopic to $[\gamma]$ rel

$\mathcal{N}^j_{SU(3)} \cup \{[\Theta]\}$. The value of $\kappa^{SU(3)}[A]$ does not depend on the choice of $[\gamma]$ or $[\overline{\gamma}]$. The $SU(3)$–Casson invariant is then defined as

$$\lambda_{SU(3)}(Y) = \sum_{\mathcal{M}_h^{SU(3)*}} \widehat{\varepsilon}[A] - \sum_{\mathcal{M}_h^{SU(3)r*}} \varepsilon[A](\kappa^{SU(3)}[A] + 1) \in \mathbf{R}.$$

Fix a component $\mathcal{K}^j$ and homotopy class $[\gamma_j]$ rel $\mathcal{N}^j_{SU(3)} \cup \{[\Theta]\}$ of paths from $[\Theta]$ to $\mathcal{N}^j_{SU(3)}$. For every $[A] \in \mathcal{N}^j_{SU(3)}$ define $\kappa^{SU(3)}[A]$ using a path $[\gamma]$ homotopic to $[\gamma_j]$. Then the Chern–Simons term is the same constant over all $[A] \in \mathcal{N}^j_{SU(3)}$, and the spectral flow term is well-defined (depending only on $[\gamma_j]$). We express this as

$$\kappa^{SU(3)}[A] = \mathrm{SF}^{\nu}_{SU(3)}[A] + 2\Delta\mathrm{cs}(j).$$

Thus we may rewrite the counter-term

$$\begin{aligned} \sum_{\mathcal{M}_h^{SU(3)r*}} \varepsilon[A]\kappa^{SU(3)}[A] \quad &= \quad \sum_j \sum_{\mathcal{M}_h^{SU(3)r*} \cap \mathcal{N}^j_{SU(3)}} \varepsilon[A]\,\mathrm{SF}^{\nu}_{SU(3)}[A] \\ &\quad + 2\sum_j \Big(\sum_{\mathcal{M}_h^{SU(3)r*} \cap \mathcal{N}^j_{SU(3)}} \varepsilon[A] \Big) \Delta\mathrm{cs}(j). \end{aligned} \tag{7.1}$$

The local index term

$$\iota_{U(2)}(\mathcal{K}^j) = \sum_{\mathcal{M}_h^{SU(3)r*} \cap \mathcal{N}^j_{SU(3)}} \varepsilon[A]$$

is well-defined independent of small perturbation h. Given any other small non-degenerate perturbation h' we have a parameterized moduli space which is a compact oriented cobordism between $\mathcal{M}_h^{SU(3)r*} \cap \mathcal{N}^j_{SU(3)}$ and $\mathcal{M}_{h'}^{SU(3)r*} \cap \mathcal{N}^j_{SU(3)}$.

In our SW–context make the same construction. We can identify homotopy classes $[\gamma_j]$ in our SW–context with those in the $SU(3)$–Casson by the inclusion $\mathcal{B}_A \subset \mathcal{B}_{U(2)}$ which is a homotopy equivalence. Then we have in a similiar manner

$$\begin{aligned} \sum_{\mathcal{M}_\pi^{r*}} \varepsilon[A]\,\kappa[A] \quad &= \quad \sum_j \sum_{\mathcal{M}_\pi^{r*} \cap \mathcal{N}^j} \varepsilon[A]\,\mathrm{SF}^{\nu}[A] \\ &\quad + \sum_j \Big(\sum_{\mathcal{M}_\pi^{r*} \cap \mathcal{N}^j} \varepsilon[A] \Big) \Delta\mathrm{cs}(j) \end{aligned} \tag{7.2}$$

and a local index

$$\iota_{SU(2)}(\mathcal{K}^j) = \sum_{\mathcal{M}_\pi^{r*} \cap \mathcal{N}^j} \varepsilon[A].$$

The two indices $\iota_{U(2)}$ and $\iota_{SU(2)}$ are equal. This is established by working with a restricted class of holonomy perturbations h' as in [12] or [3] which keeps $\mathcal{M}_{h'}^{SU(3)r*}$ within $SU(2)$–connections. Then it is straightforward to relate this to our space $\mathcal{M}_{\overline{\pi}}^{r*}$ by a compact oriented cobordism. The non-integral terms for $\lambda_{SU(3)}(Y)$, $2\tau(Y)$ come from (7.1), (7.2) respectively. It follows that $\lambda_{SU(3)}(Y) + 2\tau(Y) \bmod 4$ is integral. It is also independent of the orientation of Y, since $\lambda_{SU(3)}(Y)$ and $\tau(Y)$ are both independent of orientation.

References

[1] **S Akbulut**, **J. McCarthy**, *Casson's Invariant for oriented homology 3–spheres – an exposition*, Princeton Math. Notes 36, Princeton Univ. Press (1990)

[2] **M F Atiyah**, **V K Patodi**, **I M Singer**, *Spectral Asymmetry and Riemannian Geometry I, II, III*, Math. Proc. Camb. Phil. Soc. 77 (1975) no. 1; 78 (1975) no. 3; 79 (1976) no. 1

[3] **H Boden**, **C Herald**, *The* SU(3)*–Casson invariant for integral homology 3–spheres*, J. Diff. Geom. 50 (1998) 147–206

[4] **H Boden**, **C Herald**, **P Kirk**, *An integer valued $SU(3)$–Casson invariant*, Math. Res. Lett. 8 (2001) 589–603

[5] **S Donaldson**, **P Kronheimer**, *The geometry of 4–manifolds*, Clarendon Press, Oxford (1990)

[6] **P Feehan**, **T Leness**, *$PU(2)$–monopoles. I: Regularily, Uhlenbeck compactness and transversality*, J. Diff. Geom. 49 (1998) 265–410

[7] **T Kato**, *Perturbation Theory for Linear Operators*, Springer–Verlag (1980) 2nd Ed. (corrected)

[8] **H B Lawson**, **M-L Michelsohn**, *Spin Geometry*, Princeton Univ. Press (1989)

[9] **Y Lim**, *The equivalence of Seiberg–Witten and Casson invariants for homology 3–spheres*, Math. Res. Lett. 6 (1999) 631–644

[10] **Y Lim**, *Seiberg–Witten invariants for 3–manifolds in the case $b_1 = 0$ or 1*, Pac. J. Math. 195 (2000) 179–204

[11] **Y Lim**, *Seiberg–Witten moduli space for 3–manifolds with cylindrical-end $T^2 \times$* $\mathbf{R}^+$, Comm. Contemp. Math. 2 (2000) 61–509

[12] **C H Taubes**, *Casson's Invariant and Gauge Theory*, J. Diff. Geom. 31 (1990) 547–599

[13] **K K Uhlenbeck**, *Connections with L^p bounds on curvature*, Comm. Math. Phys. 83 (1982) 31–42