Gauge theory for embedded surfaces, I

P. B. Kronheimer[1] *and T. S. Mrowka*[2]

[1] Merton College, Oxford OX1 4JD, England
[2] California Institute of Technology, Pasadena, California 91125, USA

1. Introduction

(i) Topology of embedded surfaces.

Let X be a smooth, simply-connected 4-manifold, and ξ a 2-dimensional homology class in X. One of the features of topology in dimension 4 is the fact that, although one may always represent ξ as the fundamental class of some smoothly embedded surface, it is not always possible to take this surface to be a sphere. The purpose of this paper and its sequel [KrM] is to establish a lower bound for the genus of the surface, in terms of the self-intersection number of the class. The result applies to those 4-manifolds for which the *polynomial invariants* [D3] are defined and not all zero. As a corollary, we shall answer a question raised by Milnor [Mi] concerning the unknotting number of algebraic knots. This first paper contains statements of the results and most of the technical material; the sequel provides the more geometrical parts of the proof.

Obstructions to embedding spheres arise in relation to the known obstructions to realizing a quadratic form as the intersection form of a smooth 4-manifold. Specifically, there are obstructions which are related to Rohlin's theorem [R1,KeM], and to Donaldson's theorem on definite quadratic forms [D1,Ku,FS]. In the case that the given homology class is not primitive (so is an integer multiple of another class), there are obstructions also to embedding surfaces of higher genus: the results of [HS] and [R2] give a lower bound for the genus of a non-primitive surface in terms of the self-intersection number $\Sigma \cdot \Sigma$ and the invariants of X. In the case of the 4-manifold $\mathbb{C}\mathrm{P}^2$, similar bounds are obtained in [Tr].

More recent results [MMR] have established further obstructions to embedding spheres and tori in 4-manifolds for which the polynomial invariants are non-zero; these can be seen as extensions of the theorem of Donaldson on the indecomposability of complex surfaces [D3]. The main result which we announce here is in the same spirit, and extends the result of [MMR] to surfaces of higher genus, though the scheme of the proof is rather different. In the statement below,

[2] Partially supported by NSF grant number DMS 88-09085.

b^+ is, as usual, the dimension of a maximal positive subspace for the intersection form on the second homology.

Theorem 1.1. *Let X be a smooth, closed, oriented 4-manifold which is simply connected, has b^+ odd and not less than 3, and has non-trivial polynomial invariants. Then the genus of any orientable, smoothly embedded surface Σ, other than a sphere of self-intersection -1 or an inessential sphere of self-intersection 0, satisfies the inequality $2g - 2 \geq \Sigma \cdot \Sigma$.*

Remarks. For the moment, the polynomial invariants we have in mind are those coming from the SU(2) moduli spaces [D3], though the theory could be extended to include SO(3). The non-vanishing of the polynomials is certainly an essential hypothesis, and the restriction on b^+ is there because this condition is necessary to ensure that the polynomial invariants are defined. The orientation of X is important also: the polynomial invariants and the self-intersection number are dependent on the orientation, and the same orientation must be used for both. Indeed, the theorem really has no content if the self-intersection number is negative, for the inequality is then satisfied automatically except in the one case that is explicitly excluded. Note finally that it is a consequence of Freedman's theorem [F] that the smoothness assumption cannot be weakened to allow locally flat, topologically embedded surfaces; indeed, if X is spin for example, then every primitive class in $H_2(X)$ is represented by a locally flat sphere (see [LW]).

The inequality in the theorem says that the sum of the Euler numbers of Σ and its normal bundle is not greater than zero. Alternatively, we can think of the decomposition of TX into a sum of oriented 2-plane bundles along Σ as determining an almost complex structure on the 4-manifold in the neighbourhood of the surface; the theorem then says that the first Chern class of TX is not positive on Σ. From this description we see that, if X is the underlying 4-manifold of a complex surface with $c_1 = 0$, and if Σ is a smooth holomorphic curve, then the inequality of the theorem is sharp (this is a special case of the adjunction formula [GH]).

The simply-connected complex surfaces with $c_1 = 0$ are the $K3$ surfaces, and it is one of the main results of [D3] that the hypothesis of Theorem 1.1 on the non-vanishing of the invariants is satisfied by $K3$ and by any simply-connected complex surface with b^+ not less than 3. As a consequence of the theorem we therefore have

Corollary 1.2. *If C is a smooth complex curve in a $K3$ surface and Σ is a smoothly embedded orientable 2-manifold in the same homology class, then the genus of Σ is not smaller than the genus of C.* □

In the case of $K3$, if one uses the fact that the diffeomorphism group acts transitively on the primitive classes of any given self-intersection number [LP,Ma], it is not hard to show that the inequality of Theorem 1.1 is sharp for all homology classes of non-negative self-intersection. This is just a matter of finding, for each

even integer $2d$, a smooth, primitive surface in $K3$, having self-intersection number $2d$ and genus $d+1$. One way to obtain such a surface is to realize the $K3$ surface as an elliptic fibration and then take the sum of a section of the fibration and $d+1$ smooth fibres; smooth the intersection points to obtain the required Σ.

It is an outstanding open question whether the statement of Corollary 1.2 continues to hold for complex surfaces other than $K3$, and in particular for the complex projective plane. For complex surfaces X with $b^+ \geq 3$ and positive canonical class K, the inequality of Theorem 1.1 is weaker than the inequality which a generalization of (1.2) would predict. Thus, for the homology class which is d times the canonical class, the lower bound for $(2g-2)$ given in the theorem is $d^2(K\cdot K)$, whereas one can conjecture that $d(d+1)(K\cdot K)$ is the best possible.

Theorem 1.1 also implies results about the slice genus of algebraic knots and links. (The slice genus of a knot is the smallest genus of any oriented surface in the 4-ball whose boundary is the given knot in S^3). Let C be a smooth algebraic curve in $\mathbb{C}^2$ and B^4 an embedded 4-ball whose boundary S^3 meets C transversely in a knot K. Then we have:

Corollary 1.3. *If $\Sigma \subset B^4$ is a smooth orientable surface in the 4-ball having boundary K, then the genus of Σ is not less than the genus of the algebraic curve $C \cap B^4$. That is, the algebraic curve realizes the slice genus of the knot.*

Proof. Let $\hat{C}$ be the closure of C in $\mathbb{C}\mathrm{P}^2$. By deforming C slightly if necessary, we can arrange that this closure is a smooth algebraic curve. Take a smooth curve D of degree six in $\mathbb{C}\mathrm{P}^2$ which meets $\hat{C}$ transversely and is disjoint from the ball B^4. The branched double cover of $\mathbb{C}\mathrm{P}^2$, branched along D, is a $K3$ surface, and the inverse image of $\hat{C}$ is a smooth complex curve $\tilde{C}$. If a 2-manifold Σ existed which contradicted (1.3) then we could modify $\hat{C}$ inside the ball B^4 so as to reduce its genus without affecting its intersection with D. The inverse image of the new 2-manifold would be a closed surface in the $K3$, with the same homology class but smaller genus than $\tilde{C}$, contradicting the previous corollary. □

This statement is sometimes called the 'local Thom conjecture'. An equivalent formulation is to say that Corollary 1.2 holds for the complex projective plane provided one imposes an additional constraint on Σ, that it coincides with C outside some compact domain in $\mathbb{C}^2$. An attractive application of (1.3) is to the case in which C is obtained from a singular curve C_0 by a small deformation and B^4 is a small, standard ball centred at the singular point. The intersection $C \cap B^4$ is then the Milnor fibre of the singularity. For example, the torus knot $K_{p,q}$, with $(p,q)=1$, arises from the singular curve $x^p = y^q$, whose Milnor fibre has genus $\frac{1}{2}(p-1)(q-1)$. This number is therefore the slice genus of the knot.

In general, the slice genus is a lower bound for the *unknotting number* (also called the *gordian number*) of a knot: this is the least number of times that the string must be allowed to pass through itself if the knot is to be changed to the unknot. It is also known that, for the knots which arise from singularities in algebraic curves, the genus of the Milnor fibre is an upper bound for the un-

knotting number [BW] (for example, in the case of $K_{p,q}$, a little experimentation shows that passing the curve through itself $\frac{1}{2}(p-1)(q-1)$ times is sufficient). It therefore follows from Corollary 1.3 that, for such knots, the unknotting number and the genus of the Milnor fibre are equal, which confirms the conjecture made in [Mi]. A useful survey of conjectures related to (1.3) is given in [BW]; see also [Ru].

Apart from the challenging problem of extending (1.2) to other complex surfaces, there are some modest extensions one could envisage making to Theorem 1.1. For example, the condition on the fundamental group could probably be relaxed, and (perhaps with very little change in the proof) one should expect to prove the same inequality for 4-manifolds such as the Barlow surface, using the invariant developed by Kotschick [Ko].

(ii) Twisted connections.

A natural strategy for the class of problems discussed above is to study gauge theory (connections and the anti-self-duality equations) on the complement of the surface. A choice must be made at the outset here, since the anti-self-duality equations depend on a Riemannian metric or a conformal structure, and whereas any two choices will give equivalent results on a closed manifold, this is not the case for $X \setminus \Sigma$.

One possibility is to consider a complete metric which is asymptotically cylindrical, so the end of the manifold is modeled on $\mathbb{R}^+ \times Y$, where Y is the boundary of the tubular neighbourhood of Σ, a circle bundle over the surface. In the case of S^2, this is conformally equivalent to "blowing down" the surface to obtain a 4-dimensional orbifold with a single quotient singularity. For genus zero or one, several results have been proved using this line (in particular, the results of [MMR]), and the analysis of the moduli spaces which arise has been developed in some generality.

A different approach is to take a smooth Riemannian metric on X and consider its restriction as an incomplete metric on $X \setminus \Sigma$. The corresponding moduli spaces will not be the same as those which result from a cylindrical end, and one of the first problems is to develop the usual tools of gauge theory in this setting. In an earlier paper [Kr], the first author wrote down some conjectures about the features of the moduli spaces for the incomplete metric and outlined a scheme whereby these might be applied to prove results such as Theorem 1.1. The main purpose of the present paper is to develop the machinery to confirm the conjectured properties.

The new feature of gauge theory on $X \setminus \Sigma$ is that a connection which is flat, or has small curvature, can still be locally non-trivial near Σ on account of having non-trivial holonomy on the small circles linking the surface. When the structure group is SU(2) for example, we can have a connection A on $X \setminus \Sigma$ which is represented on each normal plane to Σ by a connection matrix which looks like

$$i \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} d\theta + (\text{lower terms}), \tag{1.4}$$

where r and θ are polar coordinates in the normal plane. The size of the connection matrix is $O(1/r)$, since this is the size of $d\theta$, so the connection appears singular along the surface. The holonomy of this connection on the positively-oriented small circles of constant r is approximately

$$\exp 2\pi i \begin{pmatrix} -\alpha & 0 \\ 0 & \alpha \end{pmatrix}. \tag{1.5}$$

Since only the conjugacy class of the holonomy has any invariant meaning, we may suppose that α lies in the interval $[\,0, \frac{1}{2}\,]$, since the matrices (1.5) then run through each conjugacy class just once.

When $\alpha = 0$ the holonomy will be trivial and, if the phrase "lower terms" is suitably defined, we are just considering ordinary connections on X. Also when $\alpha = \frac{1}{2}$, since the holonomy is -1, the associated SO(3) bundle has trivial holonomy and, with this twist, we can consider these as connections on X again. The interesting new phenomena occur when α lies in the interval $(0, \frac{1}{2})$, as it will throughout this paper.

In section 2 we introduce a space $\mathcal{A}^\alpha$ of SU(2) connections modeled on (1.4) by choosing an initial model connection A^α and defining $\mathcal{A}^\alpha$ as the set of all $A^\alpha + a$, where a is in a suitably defined function space. We shall refer to the elements of $\mathcal{A}^\alpha$ as *twisted* connections, or as α-twisted connections when the holonomy parameter needs to be mentioned. This terminology is not ideal; the connections are not really twisted in any sense.

One soon realizes that there are two topological quantities determined by A^α, in addition to the holonomy parameter α. There is an integer k which, as in the usual set-up of Yang-Mills theory, measures the second Chern class of the bundle on X, and there is another integer l which measures the degree of the reduction of the bundle near Σ determined by the eigenspaces of the holonomy (1.5). We call these the instanton number and the monopole number.

Choosing a metric on X, we define a moduli space $M^\alpha_{k,l}$ as the space of gauge equivalence classes of anti-self-dual connections in $\mathcal{A}^\alpha$. We shall develop the usual framework of Fredholm theory and transversality results, to show that, with the usual provisos about reducible and flat solutions, the moduli spaces are generically smooth, finite-dimensional manifolds. We can also consider the extra parameter α as part of the moduli space, so as to obtain a smooth space of dimension one greater, with a smooth map to the interval $(0, \frac{1}{2})$. We shall prove a weak compactness theorem for these moduli spaces and a theorem on the removal of point singularities.

The two important features of these moduli spaces are their dimensions and a Chern-Weil formula. The formula for the first reads

$$\dim M^\alpha_{k,l} = 8k + 4l - 3(b^+ - b^1 + 1) - (2g - 2), \tag{1.6}$$

where b^1 and b^+ are, as usual, the invariants for the closed manifold X (*not* the complement of the surface). The Chern-Weil formula expresses the L^2 norm of the curvature (the action) of an anti-self-dual solution in terms of the topological

data: for $A \in M^{\alpha}_{k,l}$ we have

$$\frac{1}{8\pi^2}\int_{X\setminus\Sigma} \mathrm{tr}(F_A \wedge F_A) = k + 2\alpha l - \alpha^2 \Sigma\cdot\Sigma. \tag{1.7}$$

Note that, as far as the topology of the surface goes, the dimension formula involves only the genus while the Chern-Weil formula involves only the self-intersection number. It is the interplay between these two which provides the mechanism for the proof of Theorem 1.1

We close this introduction with some technical remarks. The first concerns our choice of function space for the connections in $\mathcal{A}^{\alpha}$. Our connections will be of the form $A^{\alpha} + a$, where the covariant derivative of a is in L^p; and there is not much leeway in the choice of p. On the one hand, this choice of function space means that our gauge transformations should be in a covariant L^p_2, and such gauge transformations will not form a group unless $p > 2$ because L^2_2 functions are on the borderline of the Sobolev embedding theorem in dimension 4. On the other hand p cannot be too large: the natural growth rate for the curvature of solutions near Σ appears to involve terms of size $r^{-2\alpha}$ and $r^{-1+2\alpha}$, and these will not be in L^p unless p is close to 2. Indeed, p must approach 2 as α approaches 0 and $\frac{1}{2}$. In an Appendix to this paper, we give an example of an explicit solution which illustrates this point quite clearly.

Twisted connections arise also in 2 dimensions, on a punctured surface. Suitable function spaces for this problem were introduced by Biquard in [Bi1,Bi2], and our constructions in section 3 are motivated in part by this example. The other model which guides our constructions is provided by the thesis of S. Wang [W].

Rather than use L^p spaces, one might try to use weighted Sobolev spaces based on L^2. Unfortunately, it seems that any straightforward attempt to introduce such spaces runs foul of the necessary multiplication theorems. Having settled on our definition of $\mathcal{A}^{\alpha}$, it would be aesthetically pleasing to have a theorem which stated that any smooth, anti-self-dual connection with finite action on the complement of Σ was gauge equivalent to a solution in our space; this would make the definition of our moduli spaces look less arbitrary. The authors have not proved such a theorem; a result on the lines of [SS] is what is wanted.

A second point concerns the choice of metric on X. For large parts of this paper we shall *not* use a smooth metric, but choose instead a metric with a cone-like singularity along Σ with a small cone-angle $2\pi/\nu$. The reason for this is that it simplifies the analysis while leaving unchanged the important features of the moduli spaces (such as the formulas (1.6) and (1.7)). In the first instance, if ν is a positive integer and α is a multiple of $1/\nu$, then the set-up we describe is equivalent to analysis of an orbifold connection over a space modeled locally on a cyclic quotient singularity. The elliptic theory in this case is standard: locally, solutions of the equations are just invariant solutions on a branched cover. We deduce the elliptic theory for general values of α and for the original smooth metric by regarding the relevant operators as bounded perturbations of the orbifold set-up which we understand. We take from [W] the idea that the

anti-self-duality equations for a smooth metric can be regarded as a bounded perturbation of the equations on a cone-like metric, as a special case of a device exploited in [DS].

The use of the cone-like or orbifold metrics is not just a stepping-stone to the smooth case: there are at least two technical points (one of them contained in [KrM], the other discussed below) which we can carry through only in the orbifold setting. For this reason we choose the moduli spaces associated to the orbifold metrics in order to carry through the topological applications. The analysis is easier when the cone-angle is small, because the natural decay rate for the off-diagonal terms in the connection and curvature matrices is then more rapid, and we can therefore use Sobolev spaces based on stronger norms.

The technical point referred to in the previous paragraph concerns the phenomenon of bubbling off. In general, there will be sequences of solutions in $M^{\alpha}_{k,l}$ for which part of the curvature concentrates in the neighbourhoods of isolated points of Σ, and there will be a weak limit of such a sequence in which both k and l may change. This much of the theory can be developed for either the smooth or the orbifold metrics. Such a weak compactness result is of limited value in applications however unless it is also known that the new values of k and l are such that the new solution lives in a moduli space of strictly smaller dimension than $M^{\alpha}_{k,l}$. In the usual set-up on a closed manifold this drop in dimension comes for free, because the *action* is sure to drop in the weak limit and the index formula shows that the dimension is a monotonic function of the action. We cannot make the same simple argument for the moduli spaces of twisted connections, so we must give a different proof. This is carried out in section 8, but only for the cone-like metric.

One last result which has not been proved in the generality which the authors would have liked is the correspondence between the moduli spaces we have defined and the space of stable *bundles with parabolic structure* in the case that X is a Kähler surface and Σ is a holomorphic curve; see [Kr] for a statement of a conjecture modeled on the 2-dimensional result of Seshadri [Se], also proved in [Bi2]. A weakened form of this result (for the orbifold metric) is proved in [KrM], and while this is adequate for the applications, the conjecture as stated in [Kr] remains attractive.

Acknowledgements. The first author wishes to thank Simon Donaldson for his help and unerring advice. Both authors are grateful for the generous hospitality of the Mathematisches Forschungsinstitut at Oberwolfach, where much of this work was carried out.

2. A framework for gauge theory

This section contains more precise statements of some of the main definitions and results concerning the moduli spaces. The proofs are contained for the most part in section 5, after some preliminary work in sections 3 and 4. One exception is the dimension formula (1.6), which is postponed until section 6. The remaining results, not discussed in this section, relate to the weak compactness theorem and are dealt with in sections 7 and 8, at the end of the paper.

(i) The moduli spaces.

Let X be a smooth, closed, oriented 4-manifold and Σ a closed embedded surface. For simplicity we shall suppose Σ to be orientable and oriented, though the following constructions can be modified for the non-orientable case. We shall also suppose that both X and Σ are connected. Let N be a closed tubular neighbourhood of Σ, diffeomorphic to the unit disk bundle of the normal bundle, and let Y be the boundary of N, which acquires the structure of a circle bundle over Σ via this diffeomorphism. Let $i\eta$ be a connection 1-form for the circle bundle; so η is an S^1-invariant 1-form on Y which coincides with the 1-form $d\theta$ on each circle fibre. Here we write (r, θ) for polar coordinates in some local trivialization of the disk bundle, and we choose these so that $dr \wedge d\theta$ fixes the correct orientation for the normal plane. By radial projection we extend η to $N \backslash \Sigma$.

The matrix-valued 1-form given on $X \backslash \Sigma$ by the expression

$$i \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} \eta$$

has the asymptotic behaviour of (1.4), but is not globally defined. To make an SU(2) connection on $X \backslash \Sigma$ which has this form near Σ, begin with an SU(2) bundle $\bar{E}$ on X and choose a C^∞ decomposition of $\bar{E}$ on N as $\bar{E}|_N = \bar{L} \oplus \bar{L}^*$, compatible with the hermitian metric. (We shall tend to use the overbar for objects defined over all of X rather than just over $X \backslash \Sigma$.) Although $\bar{E}$ is trivial on N, we need not suppose that $\bar{L}$ is: there are two topological invariants in this situation, which we write

$$k = c_2(\bar{E})[X]$$
$$l = -c_1(\bar{L})[\Sigma].$$

The minus sign is for later convenience. Choose any smooth SU(2) connection $\bar{A}^0$ on $\bar{E}$ which respects the decomposition over N, so

$$\bar{A}^0|_N = \begin{pmatrix} b & 0 \\ 0 & -b \end{pmatrix},$$

where b is a smooth connection in $\bar{L}$. Finally, choose a number α in the range $0 < \alpha < \frac{1}{2}$, and define a connection A^α on $E = \bar{E}_{X \backslash \Sigma}$ by

$$A^\alpha = \bar{A}^0 + i\beta(r) \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} \eta. \tag{2.1}$$

Here β is a smooth cut-off function equal to 1 in a neighbourhood of 0 and equal to 0 for $r \geq \frac{1}{2}$. This mixture of local and global notation is best explained by saying that the second term is an element of $\Omega^1_{N\setminus\Sigma}(\mathfrak{g}_E)$ expressed in terms of a trivialization compatible with the decomposition, and that this 1-form is extended by zero to all of $X\setminus\Sigma$. Note that the curvature of A^α extends to a smooth 2-form with values in $\mathfrak{g}_{\bar{E}}$ on the whole of X. This is because the 2-form $i\,d\eta$ is smooth on N: it is the pull-back to N of the curvature form of the circle bundle Y, which can be regarded as a smooth 2-form on the base Σ.

The expression (2.1) defines a connection over $X\setminus\Sigma$ whose holonomy around small linking circles is asymptotically equal to (1.5). We now define an affine space of connections modeled on A^α by choosing some p bigger than 2 and setting

$$\mathcal{A}^{\alpha,p} = \{\, A^\alpha + a \mid a, \nabla_{A^\alpha} a \in L^p(X\setminus\Sigma)\,\}. \tag{2.2}$$

Similarly we define a gauge group

$$\mathcal{G}^p = \{\, g \in \mathrm{Aut}(E) \mid \nabla_{A^\alpha} g, \nabla^2_{A^\alpha} g \in L^p(X\setminus\Sigma)\,\}. \tag{2.3}$$

The L^p space is defined using the measure inherited from any smooth measure on X. The notation is not meant to suggest that a and g are smooth on $X\setminus\Sigma$; in local trivializations away from Σ their matrix entries will be in L^p_1 and L^p_2 respectively. We do not include mention of α in the notation for the gauge group because we shall see in section 3 that $\mathcal{G}^p$ is independent of the holonomy parameter. Sometimes we shall write $\mathcal{A}^{\alpha,p}_{k,l}$ when the instanton and monopole numbers need to be mentioned; more frequently we shall write just $\mathcal{A}^\alpha$, or even $\mathcal{A}$, if the context makes the parameters either clear or unimportant.

Proposition 2.4. (*i*) *The spaces $\mathcal{A}^{\alpha,p}$ and $\mathcal{G}^p$ are independent of the choice of $\bar{A}^0$ and the choice of the connection 1-form η.*

(*ii*) *The space $\mathcal{G}^p$ is a Banach Lie group which is independent of α, and it acts smoothly on $\mathcal{A}^{\alpha,p}$. The stabilizer of A is $\{\pm 1\}$ or S^1 according as A is irreducible or reducible respectively.*

(*iii*) *There exists a continuous function $p(\alpha)$, with $p(\alpha) > 2$ for $\alpha \in (0, \frac{1}{2})$, such that when $2 < p < p(\alpha)$ the quotient $\mathcal{B}^{\alpha,p}$ is a Banach manifold except at points $[A]$ corresponding to reducible connections.*

Remark. The definition of l used the orientation of the surface Σ twice. First, the orientation of the unit circle in the normal bundle was used in determining the holonomy; secondly, the orientation was used to calculate the first Chern class of the line bundle. The result is that, up to gauge equivalence, the space $\mathcal{A}^{\alpha,p}_{k,l}$ is independent of the chosen orientation. The monopole number does not change sign when the orientation of Σ is reversed.

Although the domain of definition of the connections in $\mathcal{A}^{\alpha,p}$ is strictly speaking $X\setminus\Sigma$, we shall sometimes refer to these twisted connections as living "over (X, Σ)", or sometimes even "over X" when this lack of precision is not likely to cause confusion. Sometimes we shall talk of a twisted connection over (X, Σ) as

being carried by a bundle pair (E, L), and we will mean that (E, L) was the pair of bundles used in the definition of the space $\mathcal{A}$.

We now introduce the anti-self-duality equations. Pick a smooth Riemannian metric $\bar{g}$ on X and let $\mathcal{F}^p$ denote $L^p(X\backslash\Sigma, \Lambda^+(\mathfrak{g}_E))$, the Banach space of self-dual 2-forms in $L^p(X\backslash\Sigma)$ with values in $\mathfrak{g}_E$. The gauge group acts on $\mathcal{F}^p$ with kernel $\{\pm 1\}$, so over the open set of irreducible connections $(\mathcal{B}^{\alpha,p})^* \subset \mathcal{B}^{\alpha,p}$ we can form the Banach vector bundle $\mathcal{V}^p$ with fibre $\mathcal{F}^p$. The multiplication theorem (3.8) shows that $A \mapsto F_A^+$ defines a smooth map from $\mathcal{A}^{\alpha,p}$ to $\mathcal{F}^p$. We define

$$M^\alpha_{k,l} = \{\, A \in \mathcal{A}^{\alpha,p} \mid F_A^+ = 0 \,\}/\mathcal{G}^p.$$

Thus $M^\alpha_{k,l}$ is a subspace of $\mathcal{B}^{\alpha,p}$. We write $(M^\alpha_{k,l})^*$ for its intersection with the space of irreducible connections $(B^{\alpha,p})^*$, so $(M^\alpha_{k,l})^*$ is the zero set of a smooth section Φ of $\mathcal{V}^p$. For sufficiently small p we have all the usual Fredholm theory for these moduli spaces:

Proposition 2.5. *There exists a continuous function $p(\alpha)$, with $p(\alpha) > 2$ for $\alpha \in (0, \frac{1}{2})$, such that for p in the range $2 < p < p(\alpha)$ the following results hold:*

(i) the moduli space $M^\alpha_{k,l}$ is independent of p;

(ii) in local trivializations of $\mathcal{V}^p$, the section Φ which cuts out $(M^\alpha_{k,l})^*$ is Fredholm;

(iii) the index of Φ is

$$8k + 4l - 3(b^+ - b^1 + 1) - (2g - 2).$$

The third proposition of this section is a generic metrics theorem, modeled on that of [FU]. As usual, we note that the anti-self-duality equations can be defined using metrics which are, say, of class C^r for some r; such metrics are parametrized by an open subset of a Banach space, and the term "generic" below refers to the complement of a first-category subset of this space.

Proposition 2.6. *Fix an α in the range $0 < \alpha < \frac{1}{2}$, and consider the moduli space $M^\alpha_{k,l}$.*

(i) For a generic choice of Riemannian metric $\bar{g}$ on X, the section Φ cuts out the moduli space $(M^\alpha_{k,l})^*$ transversely, except perhaps at flat connections.

(ii) If b^+ is positive, then for a generic Riemannian metric the moduli space contains no reducible solutions, except perhaps for flat ones.

To summarize, we have described moduli spaces which depend on two topological numbers k and l, and a real parameter α, which determines the holonomy. If b^+ is positive and a moduli space contains no flat connections, then for a generic choice of metric it is a smooth manifold of the dimension given by the formula (1.6). We refer to k and l as the instanton and monopole numbers of the solution.

(ii) The holonomy as a parameter.

As the holonomy α varies, the moduli spaces M^α sweep out a space of one higher dimension. To set this up correctly, let I be a compact sub-interval of the interval $(0, \frac{1}{2})$, and introduce the space

$$\hat{\mathcal{B}}^p = \bigcup_{\alpha \in I} \mathcal{B}^{\alpha,p} \times \{\alpha\}.$$

Proposition 2.7. *Let $p(\alpha)$ be the function from* (2.4) *and let p_0 be its minimum value on the interval I. Then, for $p < p_0$, the space $\hat{\mathcal{B}}^p$ has the structure of a smooth Banach manifold away from the reducibles , and the map $\hat{\mathcal{B}}^p \to I$ is a smooth submersion.*

In a similar way we introduce the moduli space $\hat{M}_{k,l}$ as the union of the $M^\alpha_{k,l}$ over I. This space is a subset of $\hat{\mathcal{B}}^p$, and its irreducible part is the zero set of a section $\hat{\Phi}$ of a vector bundle over the Banach manifold. Summarizing the analogues of Propositions 2.5 and 2.6 for this situation, we state:

Proposition 2.8. *Suppose b^+ is at least 2 and $p < p_0$. Then, except at flat connections and for a generic Riemannian metric, the space $\hat{M}_{k,l}$ consists only of irreducible solutions and is a smooth manifold cut out transversely by $\hat{\Phi}$. The map to I is smooth and the fibres are the moduli spaces $M^\alpha_{k,l}$.*

Since our moduli spaces are independent of p once p is sufficiently small, we can take the union over increasing intervals I to obtain a moduli space $\hat{M}_{k,l}$ over the whole interval $(0, \frac{1}{2})$. Note that b^+ must be one higher than in the previous proposition, in order to avoid reducible connections in the 1-parameter family. We shall call this $\hat{M}_{k,l}$ the *extended* moduli space.

(iii) Cone-like metrics.

We have been using a smooth metric on X to define the moduli spaces, but this is not the only possibility. We can take a metric which, near to the surface Σ, is modeled on

$$ds^2 = du^2 + dv^2 + dr^2 + (\frac{1}{\nu^2}) r^2 d\theta^2,$$

where u and v are coordinates on Σ and ν is a real parameter, not less than 1. To obtain a global metric on N of this shape we just replace $du^2 + dv^2$ by the pull-back of any smooth metric on Σ and we replace $d\theta$ by the 1-form η. We then patch the metric to a smooth one on the complement of N to extend it to the rest of X. The resulting metric has a cone-angle of $2\pi/\nu$ in the normal planes to Σ. When $\nu = 1$, this metric is smooth on X, and when ν is an integer greater than 1 the metric is an orbifold metric: locally there is a ν-fold branched cover on which the metric is smooth. (Moduli spaces of the usual kind over 4-manifolds with cone-like metrics of this sort were previously studied in [W].) We

write g^ν for a typical metric of this form, and $M^\alpha_{k,l}(g^\nu)$ for the corresponding moduli space.

Everything we have stated above for the smooth metric carries over to metrics of this form. We emphasize that the spaces $\mathcal{A}$, $\mathcal{G}$ and $\mathcal{B}$ have not been changed; the local coordinates r and θ are used just as they were before to define the model connection A^α, and the L^p norms are equivalent to the old ones since g^ν differs boundedly from a smooth metric. Most importantly, the covariant derivative ∇_{A^α} used to define $\mathcal{A}$ and $\mathcal{G}$ does not involve the Levi-Civita connection of the cone-like metric but only the smooth structure of X. The new metric only enters in defining the anti-self-duality condition on $X\backslash\Sigma$.

Remark. We will have more to say in section 3(iv) on the subject of comparing the norms defined by the different Levi-Civita derivatives. The choice we have made here is not the only way to make the theory go.

Proposition 2.9. *For any $\nu \geq 1$, the statements of Propositions* 2.5*,* 2.6 *and* 2.8 *continue to hold with metrics of the form g^ν in place of smooth metrics.*

The only additional word of explanation needed here is to say what is meant now by a generic metric of the form g^ν, since the geometry of the model metric we have described is rather special. We use the simplest route which allows the desired conclusion by fixing a particular model metric and considering C^r perturbations of this which are compactly supported in some fixed subset of $X\backslash N$. When we say that a property holds for a generic metric, we mean that it holds for a generic perturbation of this form.

(iv) The extension problem and SO(3) bundles.

In the introduction, we made reference to what we might call the *extension* problem. To put this in its sharpest form, we can ask whether every anti-self-dual connection A with finite action on $X\backslash\Sigma$ is gauge-equivalent to a connection in one of the spaces $\mathcal{A}^\alpha$ we have defined. This problem is chiefly an analytic one, and as such can be studied locally in coordinate patches near Σ; but there is also a global, topological side to the question, which, though it is elementary, can cause confusion. It is this topological aspect which we wish to discuss now.

Suppose A is given: a connection in an SU(2) bundle E over $X\backslash\Sigma$. To keep the analytic difficulties at bay, let us assume of our connection that it is reducible on $N\backslash\Sigma$ and that its curvature 2-form extends smoothly across Σ. So A has the same sort of regularity near Σ as our model connections A^α. From this it is not hard to deduce (or alternatively the reader can add this to the hypotheses) that each point of Σ has a neighbourhood $U \subset X$ such that the connection is gauge-equivalent on $U\backslash\Sigma$ to a connection matrix of the form

$$\begin{pmatrix} b_U & 0 \\ 0 & -b_U \end{pmatrix} + i\beta(r) \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} \eta, \tag{2.10}$$

where b_U is a smooth 1-form on U and α is a constant in the interval $[0, \frac{1}{2}]$, (compare (2.1) above). We shall suppose that α lies in the interior of the interval; its value is uniquely determined by the gauge-equivalence class of the connection. The nub of the topological difficulty is that the definition of $\mathcal{A}^\alpha$ began with bundles $\bar{E}$ and $\bar{L}$ defined on all of X, not just on $X \backslash \Sigma$, and these bundles were used in the definition of k and l. We have only a bundle on the open manifold, and this has no characteristic classes. We want to see how k and l can be recovered.

Suppose V is a nearby open set and that we have a similar trivialization there, with b_V a smooth 1-form on V. Let g be the transition function on $U \cap V$ which relates the two trivializations. Since the connection matrices lie in the Lie algebra of the subgroup S^1 in both trivializations, g must take values in the normalizer of S^1, which is $S^1 \cup \varepsilon S^1$, where ε is a representative of the non-trivial element of the Weyl group. However, since in both trivializations the holonomy around small circles is close to

$$h(\alpha) = \exp 2\pi i \begin{pmatrix} -\alpha & 0 \\ 0 & \alpha \end{pmatrix}, \tag{2.11}$$

the odd component εS^1 is ruled out (ε changes the sign of α). It follows that we can write

$$g = \begin{pmatrix} \psi & 0 \\ 0 & \psi^{-1} \end{pmatrix},$$

and the circle-valued function ψ satisfies $d\psi = \psi(b_U - b_V)$. Since b_U and b_V are smooth, it follows easily that ψ extends across Σ and is also smooth on $U \cap V$.

Covering the whole of N with such neighbourhoods, we obtain a system of S^1-valued transition functions, which define a reducible bundle $\bar{L} \oplus \bar{L}^{-1}$. Outside Σ, this bundle is canonically identified with E, and the connection matrices b_U in the local trivializations satisfy the correct relations to define a smooth connection in this bundle. This achieves our aim of realizing A as arising from one of our model connections, and shows how k and l can be recovered.

We have an application for this construction. Suppose A and E are as above and suppose in addition that the homology class of Σ is a multiple of 2. This means that on $X \backslash \Sigma$ there is a flat, real line bundle ξ whose holonomy on the small circles linking Σ is -1. Consider the connection $A' = A \otimes \xi$ on the tensor product bundle $E' = E \otimes \xi$ over $X \backslash \Sigma$. Certainly this is still reducible on N, and its curvature is a smooth 2-form as before. It follows, as above, that A' arises as a model connection, and has associated a holonomy parameter $\alpha' \in (0, \frac{1}{2})$ and an instanton and monopole number k' and l'. The relation between the new parameters and old is not quite obvious:

Lemma 2.12. *In the situation above, the parameters of E' are related to those of E by*

$$\begin{aligned} \alpha' &= \tfrac{1}{2} - \alpha \\ k' &= k + l - \tfrac{1}{4}\Sigma \cdot \Sigma \\ l' &= -l + \tfrac{1}{2}\Sigma \cdot \Sigma. \end{aligned}$$

The formula for α' is easy to verify: on small circles, the holonomy of A is asymptotic to $h(\alpha)$ (2.11), so the holonomy of A' approximates $-h(\alpha)$, which is conjugate to $h(\frac{1}{2}-\alpha)$ in SU(2). The formulae for k' and l' can be extracted by analysing the construction of the extension as outlined above. Some care is needed to get this right, and we prefer to postpone the proof until after our discussion of the Chern-Weil formula in section 5(ii), at which point the formulae can be proved by an argument which is less prone to error. Lemma 2.12 will be used in our proof of the index formula (1.6).

Notice that the induced connections on the Lie algebra bundles $\mathfrak{g}_E$ and $\mathfrak{g}_{E'}$ are isomorphic. It follows straight away that the construction $A \mapsto A \otimes \xi$ gives a bijection between the spaces of connections $\mathcal{A}^{\alpha}_{k,l}$ and $\mathcal{A}^{\alpha'}_{k',l'}$, and therefore between the two moduli spaces

$$\otimes\xi \,:\, M^{\alpha}_{k,l} \to M^{\alpha'}_{k',l'}. \tag{2.13}$$

If we extend our theory to include SO(3) bundles with non-zero Stiefel-Whitney class w_2, we can generalize the correspondence (2.13), dropping the constraint that Σ should be a multiple of 2. We finish now by explaining this. The discussion of SO(3) bundles will not be used elsewhere in this paper.

So let $\bar{E}$ be an oriented $\mathbb{R}^3$ bundle on X and let there be given a reduction of $\bar{E}$ to SO(2) on N. By a reduction to SO(2), we mean strictly that a section of the associated S^2 bundle is given. This means that $\bar{E}$ is decomposed as $\mathbb{R} \oplus \bar{K}$ where $\bar{K}$ is an SO(2) bundle and both summands *are given orientations.* In this situation, define

$$\begin{aligned} k &= -\tfrac{1}{4} p_1(\bar{E}) \\ l &= -\tfrac{1}{2} e(\bar{K})[\Sigma], \end{aligned}$$

taking values in $\frac{1}{4}\mathbb{Z}$ and $\frac{1}{2}\mathbb{Z}$ respectively. For $\alpha \in (0, \frac{1}{2})$, construct a model connection A^{α} on E in the form

$$A^{\alpha} = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} + \beta(r) \begin{pmatrix} 0 & & \\ & 0 & 2\alpha \\ & -2\alpha & 0 \end{pmatrix} \eta, \tag{2.14}$$

where B is a smooth connection in K and the connection matrix is written in a trivialization which is compatible with the orientations. Writing w for the Stiefel-Whitney class of E, we obtain moduli spaces

$$M^{\alpha}_{k,l,w}.$$

The characteristic classes are constrained by

$$\begin{aligned} -4k &= w^2 \pmod 4 \\ 2l &= w[\Sigma] \pmod 2. \end{aligned} \tag{2.15}$$

Consider now the extension problem in this situation. We are given an SO(3) bundle E only on $X \backslash \Sigma$, and A is a connection which is compatible with a reduction $\mathbb{R} \oplus K$ near Σ and has smooth curvature. From this data it is *not*

possible to recover α, k and l uniquely. One first needs an orientation of K. Once an orientation is chosen, there is an essentially unique gauge in which the connection matrix looks like (2.14), and the construction goes through as in the SU(2) case, giving rise to an extension $(\bar{E}, \bar{K})$ and associated invariants k, l.

The choice of orientation in K is important: if we chose the opposite orientation, we would not simply recover the same $(\bar{E}, \bar{K})$ with the orientation of $\bar{K}$ reversed; the complication stems from the convention that the holonomy parameter should lie in $(0, \frac{1}{2})$. Changing the orientation of K gives rise to a different extension, with invariants α', k' and l' as in (2.12). The Stiefel Whitney class w' is different also: we have

$$w' = w + \text{P.D.}[\Sigma],$$

as can be deduced from the formula for l' in (2.12) and the constraint (2.15).

The correct generalization of (2.13) therefore is that there is a natural identification of the two moduli spaces

$$M^{\alpha}_{k,l,w} \cong M^{\alpha'}_{k',l',w'}. \tag{2.16}$$

If we wish to think of the SO(3) connections in one of our spaces $\mathcal{A}$ as arising from smooth connections in a smooth bundle by adding a singular term, then there is an ambiguity in identifying the smooth bundle: for each singular connection, there is a choice of two.

As a last word on this subject, we can note that to distinguish between α and $\frac{1}{2} - \alpha$ it is not quite an orientation of K which is being used; all that is really needed is an isomorphism between the orientation bundle of K and the orientation bundle of the normal to Σ: it is because we are assuming a standard orientation for the latter that it did not enter in the discussion above. This shows how the theory can be extended to include the case that Σ is not orientable. In general, the topological data should consist of a principle SO(3) bundle $P \to X$, a reduction of $P|_N$ to an O(2) bundle Q, and an isomorphism χ between the $(\mathbb{Z}/2)$-bundle $Q/\,\mathrm{SO}(2)$ and the orientation bundle of Σ. Under these circumstances, the Euler class of Q can be evaluated on Σ as a signed integer $-2l$ and the holonomy parameter $\alpha \in (0, \frac{1}{2})$ can be interpreted unambiguously.

3. Function spaces and multiplication theorems

We have defined the space of connections $\mathcal{A}^{\alpha}$ as the space of all $A^{\alpha} + a$, where the covariant derivative of a is in L^p. This is a covariant version of the Sobolev space L^p_1. On a compact manifold, the Sobolev norms defined using a covariant derivative are independent of the choice of connection used, at least up to equivalence. But for our twisted connections, which are singular along Σ, a little care is called for. In this section we will examine these Sobolev norms.

(i) Weighted Sobolev spaces.

For p in the range $1 < p < \infty$, let L^p_k denote the usual Sobolev space of functions f on X with k derivatives in L^p. Sometimes we prefer to regard the same L^p_k as a space of functions on $X \backslash \Sigma$; there is no essential difference. We shall introduce also the following weighted Sobolev spaces. Extend the function r (originally defined on N) to a smooth function on $X \backslash \Sigma$, taking values greater than or equal to 1 on $X \backslash N$. Define W^p_k as the completion of the space of compactly supported smooth functions on $X \backslash \Sigma$ in the norm

$$\|f\|_{W^p_k} = \|\frac{1}{r^k} f\|_p + \|\frac{1}{r^{k-1}} \nabla f\|_p + \cdots + \|\nabla^k f\|_p.$$

We begin with two elementary lemmas concerning these norms.

Lemma 3.1. *If f is in $L^p_{k,\mathrm{loc}}$ on $X \backslash \Sigma$ and if the integrals which define the W^p_k norm converge, then f is in the space W^p_k.*

Proof. The lemma asserts that we can find a sequence f_n in C^∞_0 which are Cauchy in W^p_k and converge to f. The matter of smoothness is quite standard, so we just seek compactly supported f_n. It is a routine matter to verify that the simplest strategy is successful: let $\beta(r)$ be the standard cut-off function supported in N and equal to 1 near Σ, and define f_n as $(1 - \beta(nr))f$. □

Lemma 3.2. *A function f is in W^p_k if and only if $(1/r^i)f$ is in L^p_{k-i} for $i = 0, \ldots, k$.*

Proof. The forward direction is immediate; for the converse, use the previous lemma. □

Next we carry over the usual embedding theorems to the weighted spaces. Recall that the Sobolev embedding theorems are governed by the conformal weights of the function spaces L^p_k, which in dimension 4 are defined as

$$w(p, k) = k - 4/p.$$

Lemma 3.3. *If $k \geq l$ and $w(p, k) \geq w(q, l)$ then there is an inclusion $W^p_k \hookrightarrow W^q_l$. If both inequalities are strict, then the inclusion is compact.*

Proof. Use Lemma 3.2 to reduce to the ordinary embedding theorem $L^p_{k-i} \hookrightarrow L^q_{l-i}$, for $i \leq l$. □

The multiplication theorems also carry over, for various combinations of spaces:

Lemma 3.4. *The multiplication map $(f, g) \mapsto fg$ is continuous as a map from $W^p_k \times L^q_l$ or $L^p_k \times W^q_l$ to W^r_m provided that k and l are not less than m and one of the following conditions holds:*

(i) $w(p, k) < 0$, $w(q, l) < 0$ and $w(p, k) + w(q, l) \geq w(r, m)$;

(*ii*) $w(p,k) \geq 0$, $w(q,l) \leq 0$ *and* $w(q,l) > w(r,m)$;
(*iii*) $w(p,k) > 0$, $w(q,l) > 0$ *and* $w(p,k), w(q,l) \geq w(r,m)$.

In each of these cases, if $k > m$ *and the final inequality is strict, then the map* $f \mapsto fg$*, for fixed* g*, is compact.*

Proof. The same strategy applies as was successful with Lemma 3.3: these reduce to the multiplication theorems for the unweighted Sobolev spaces. □

Since $W^p_k \subset L^p_k$, we also have multiplication theorems of the shape $W^p_k \times W^q_l \to W^r_m$ and $W^p_k \times W^q_l \to L^r_m$, as well as a compact inclusion of W^p_k in C^0 when the weight is positive.

(ii) Holonomy and weighted norms.

To take a closer look at the weighted norms, consider a local situation. As a model for a neighbourhood of a patch of Σ in X take the space $D^2 \times D^2$ containing the surface $D^2 \times \{0\}$. Write Z for the complement of the patch of surface,

$$Z = D^2 \times (D^2 \backslash \{0\}),$$

and equip Z with the flat product metric. Over Z, let K be the trivial line bundle $\mathbb{C} \times Z$ and let ∇_α be the connection

$$\nabla_\alpha = \nabla + i\alpha\, d\theta,$$

where (r, θ) are polar coordinates on $D^2 \backslash \{0\}$. We shall suppose that α is not an integer, so the holonomy is not 1. For compactly-supported sections of K we can consider two families of norms. On the one hand, using the canonical trivialization of K, we can regard sections of K as functions and compute the W^p_k norms. On the other hand, we can use the covariant derivative to define the covariant Sobolev norms:

$$\sum_0^k \|\nabla^i_\alpha s\|_p.$$

To define these higher derivatives, we need also a connection on the tangent bundle. We choose here the Levi-Civita connection for the flat metric on the patch, or just the product connection, (but see section (iv) later). We use $L^p_{k,\alpha}$ to denote these last norms. An important observation for us is that the two families of norms are equivalent. The main point is contained in the following lemma. We use the notation $c(\alpha)$ for the absolute value of the difference between α and the nearest integer. So c is a non-negative saw-tooth function.

Lemma 3.5. *For compactly supported sections* s *of* K *we have an inequality*

$$\|\frac{1}{r}s\|_p \leq \text{const.}\|\nabla_\alpha s\|_p,$$

with a constant of the form $C/c(\alpha)$*, where* C *is independent of* α *and* p*.*

Proof. Let S be the unit circle in one of the normal planes to D^2. The differential operator $\partial_\alpha = (\partial/\partial\theta) + i\alpha$ on S, has no kernel, since α is not an integer. Being skew-adjoint, it is invertible, and the inverse is given by convolution with the Green's function

$$G(\theta) = (1 - e^{-2\pi i\alpha})^{-1} e^{i\alpha\theta}, \qquad (0 < \theta < 2\pi).$$

Since G is bounded, the C^0 norm of f is controlled by the L^1 norm of $\partial_\alpha f$; and so the weaker inequality also holds:

$$\int |f|^p \, d\theta \leq (\text{const.})^p \int |\partial_\alpha f|^p \, d\theta.$$

The L^∞ norm of G will serve as the constant here, and this is bounded by an expression of the form $C/c(\alpha)$. This shows that, in comparing the integrals which define the two norms in the lemma, the contribution from S in the first integral dominates the same contribution in the second. Since Z is a union of similar circles, and since the two norms scale in the same way, the lemma follows. □

Note that although the constant depends on α, we will have a uniform bound as long as there is a lower bound on the difference between α and the nearest integer. We will make use of a sharper version of this lemma in section 4.

The lemma says that the $L^p_{1,\alpha}$ norm dominates the W^p_1 norm. On the other hand, since ∇_α differs from ∇ by a term of size $1/r$, the reverse inequality is clear, so the two norms are equivalent. Repeating this argument k times proves:

Lemma 3.6. *The $L^p_{k,\alpha}$ and W^p_k norms are equivalent on sections of K.* □

(iii) Function spaces for gauge theory.

Let A^α be the model twisted connection over (X, Σ) constructed in section 2. Fix a number p greater than 2, and write $\mathcal{A}^\alpha$ for the space of connections $\mathcal{A}^{\alpha,p}$. We shall write L^p_{k,A^α} for the covariant Sobolev norms on bundle-valued functions or forms. Recall then that $\mathcal{A}^\alpha$ is the space of all $A^\alpha + a$ where a is in L^p_{1,A^α}. We give $\mathcal{A}^\alpha$ the topology of an affine space over this Banach space.

In a neighbourhood of some patch of Σ, choose a smooth trivialization of $\bar{L}$ and so obtain a trivialization of $\bar{E} = \bar{L} \oplus \bar{L}^*$. We shall call a trivialization obtained in this way a *diagonal* trivialization. Note that the chosen trivialization of $\bar{L}$ is supposed to be defined (and smooth) on Σ too, not just on the complement. The Lie algebra bundle, restricted to N, decomposes as $\mathbb{R} \oplus \bar{L}^2$; the two summands are respectively the diagonal and off-diagonal parts of the two-by-two matrix in our trivialization. The connection ∇_{A^α} restricts as the trivial connection ∇ on the $\mathbb{R}$ summand and a connection close to $\nabla_{2\alpha}$ on the L^2 summand (in fact, differing from $\nabla_{2\alpha}$ by something smooth). From Lemma 3.6 we therefore deduce:

Proposition 3.7. *A section s of $\mathfrak{g}_E$ or of $\mathrm{End}(E)$ which is in $L^p_{k,\mathrm{loc}}$ is in L^p_{k,A^α} if and only if, in diagonal local trivializations near Σ, the diagonal components*

of s are in L^p_k and the off-diagonal components are in W^p_k. In particular, $\mathcal{A}^\alpha$ consists of connections $A^\alpha + a$ where the diagonal and off-diagonal components of a are in L^p_1 and W^p_1 respectively, while the gauge transformations $g \in \mathcal{G}$ have their diagonal and off-diagonal parts in L^p_2 and W^p_2. □

Combining the first part of this proposition with the multiplication theorems of (3.4), we obtain multiplication theorems for the spaces $L^p_{k,\alpha}(\mathrm{End}\, E)$. They are proved by considering the various components separately. Thus, when multiplying two sections of $\mathrm{End}(E)$, the product of an off-diagonal term and a diagonal term must be an off-diagonal term etc., so we must check that (3.4) provides a continuous multiplication on the corresponding spaces.

Lemma 3.8. *On sections of* $\mathrm{End}(E)$*, the multiplication map $(s_1, s_2) \mapsto s_1 s_2$ is continuous in any of the following norms:*

$$L^p_{2,A^\alpha} \times L^p_{2,A^\alpha} \to L^p_{2,A^\alpha}$$
$$L^p_{2,A^\alpha} \times L^p_{1,A^\alpha} \to L^p_{1,A^\alpha}$$
$$L^p_{1,A^\alpha} \times L^p_{1,A^\alpha} \to L^p.$$

In the last two of these three cases, if s_2 is fixed, the map $s_1 \mapsto s_1 s_2$ is compact. □

We can now prove parts (i) and (ii) of Proposition 2.4. For part (i), note that changing $\bar{A}^0$ or η only changes A^α by the addition of a smooth term, so the equivalence class of the norm L^p_{k,A^α} is unaffected. (We already used this fact tacitly in (3.7).) For part (ii), the continuity of the multiplication map on $\mathcal{G}$ follows from the first case of (3.8); the fact that $\mathcal{G}$ is a Lie group with Lie algebra $L^p_{2,A^\alpha}(\mathfrak{g}_E)$ is standard given this multiplication and the inclusion of L^p_{2,A^α} in C^0. The second case of (3.8) shows that $\mathcal{G}$ acts smoothly on $\mathcal{A}^\alpha$. The last sentence of part (ii) is standard, and really defines irreducibility for us. In passing, note that for any $A \in \mathcal{A}^\alpha$ the difference between ∇_A and ∇_{A^α} (regarded as an operator from L^p_{k,A^α} to L^p_{k-1,A^α} for $k = 1$ or 2) is a compact operator. This follows from the rider to the proposition. The same statement holds for the associated operators d_A and d_A^+.

The third part of Proposition 2.4 requires a slice theorem for the action of $\mathcal{G}$ on $\mathcal{A}^\alpha$, and this will have to wait until we have developed the elliptic theory.

As in section 2(ii), we can now regard α as a parameter in the definition of A^α. One corollary of (3.7) is that the equivalence class of the norm L^p_{k,A^α} is independent of α. So (as we mentioned without proof in section 2) the gauge group $\mathcal{G}$ is independent of α, and the space $\mathcal{A}^\alpha$ has the form $A^\alpha + \Omega$, where Ω is a Banach space which is independent of α also. The extended space of connections

$$\hat{\mathcal{A}} = \bigcup_\alpha \big((A^\alpha + \Omega) \times \{\alpha\}\big)$$

can therefore be topologized as the product $\Omega \times (0, \frac{1}{2})$, and the space $\hat{\mathcal{B}}$ is the quotient of $\hat{\mathcal{A}}$ by an action of $\mathcal{G}$. (Note however that the product structure on $\hat{\mathcal{A}}$ is not natural, and the action of $\mathcal{G}$ is different on each fibre.)

The difference between the operators ∇_{A^α} for two values of the holonomy, say α and α', is a multiplication operator by a form whose size is of the order $(\alpha - \alpha')(1/r)$ pointwise on $X \backslash \Sigma$. It therefore follows from (3.5) that the operators ∇_A, d_A and d_A^+ vary continuously in operator norm as A runs over $\hat{\mathcal{A}}$.

(iv) Holonomy on a cone-like metric.

Let us go back to the local model space Z from subsection (ii) above, but now let us equip Z with the cone-like metric

$$g^\nu = du^2 + dv^2 + dr^2 + \big(\frac{1}{\nu^2}\big) r^2 d\theta^2.$$

Let K be the same line bundle as before, and let $L^p_{k,\alpha}$ continue to denote the Sobolev spaces defined using the covariant derivative ∇_α and the Levi-Civita connection of the original flat metric. Let $L^p_{k,\alpha}(g^\nu)$ temporarily denote the covariant Sobolev spaces defined using the Levi-Civita connection of the metric g^ν coupled to ∇_α. When considering such norms on spaces of forms $\Omega^i_Z(K)$, we shall use the covariant derivative also on Λ^i.

The norms $L^p_{k,\alpha}$ and $L^p_{k,\alpha}(g^\nu)$ are not always equivalent, since the Levi-Civita connection of g^ν has holonomy itself, and this may cancel the holonomy in K on some components. Whenever this phenomenon occurs, it will eventually be the metric-independent norm $L^p_{k,\alpha}$ with which we shall want to work. Here we find conditions which ensure the equivalence of the two families. Recall the saw-tooth function $c(\alpha)$ from section (ii) above.

Lemma 3.9. *The norms $L^p_{k,\alpha}$ and $L^p_{k,\alpha}(g^\nu)$ are equivalent on $\Omega^i_Z(K)$ provided that $c(\alpha)$ exceeds $(i + k - 1)/\nu$.*

Proof. The holonomy of the Levi-Civita connection of g^ν acts on the complexified cotangent bundle $T^*_{\mathbb{C}} Z$ with eigenvalues 1 (twice) and $\exp(\pm 2\pi i/\nu)$. For the action on $(T^*Z)^{\otimes j} \otimes \Lambda^i$, the eigenvalues are $\exp(\pm 2\pi i \beta)$ for various β in the range $|\beta| \leq (i + j)/\nu$. If $c(\alpha)$ exceeds the largest such β then the holonomy on all components of $E^{i,j} = (T^*Z)^{\otimes j} \otimes \Lambda^i_Z(K)$ will be non-trivial. Lemma 3.5 then implies that the $L^p_{1,\alpha}(g^\nu)$ norm is equivalent to the W^p_1 norm on sections of $E^{i,j}$ in the standard local trivialization. Applying this argument $k - 1$ times proves that, under the hypothesis of the lemma, the $L^p_{k,\alpha}(g^\nu)$ norm is equivalent to the W^p_k norm on $\Omega^i(K)$. □

4. Elliptic theory on S^4

The basic properties of the anti-self-dual moduli spaces are usually deduced from properties of the elliptic operator $d_A^* + d_A^+$ acting on 1-forms. One of the first difficulties in developing the theory in the context of twisted connections is the apparent possibility that this operator is not Fredholm as a map from $L^p_{1,A}$ to L^p, at least for smooth Riemannian metrics on the base, and that the kernel of d_A^* does not provide a complement to the image of d_A. It seems likely that the usual elliptic results can be recovered if the operator d^* is replaced by a suitable weighted adjoint of the operator d, with a weight appropriate to the spaces introduced in the previous section. The authors succeeded in carrying this through, with a direct analysis of the equations in the case $p = 2$, as far as establishing the Fredholm nature of the linear operator, but the necessary L^p theory for $p > 2$ appeared too daunting.

While not doubting that a direct approach could be successful, we have followed a different and rather indirect route. There are two ingredients. The first is an idea taken from [W], that we can use the constructions of [DS] to compare an orbifold metric with a smooth one on X. The second idea is to use a sharp version of Lemma 3.5 to estimate the difference between the operators as the holonomy is varied. In this section we deal with the local theory by tackling the analysis on S^4.

(i) Estimates on the orbifold.

Let S^4 be the unit sphere in $\mathbb{R}^5$, let Π^3 be the 3-plane in $\mathbb{R}^5$ spanned by the last three basis vectors, and let Σ be the 2-sphere in which Π^3 meets S^4. There is a 1-parameter family of 4-dimensional half-planes with boundary Π^3 in $\mathbb{R}^5$, uniformly parametrized by an angle φ running from 0 to 2π. The form $d\varphi$ is a 1-form on $S^4 \backslash \Sigma$ whose integral on small loops linking Σ is 1. The loci of constant φ are 3-dimensional hemi-spheres in S^4, analogous to the lines of longitude on the 2-sphere.

Let the cyclic group $\mathbb{Z}_\nu$ of order ν act linearly on S^4 so that the generator fixes Σ and increases the angle φ by $2\pi/\nu$. Let $\check{S}^4$ be the quotient space and let θ be the coordinate $\nu\varphi$, which is single valued mod 2π on $\check{S}^4$. The integral of $d\theta$ on small circles linking Σ in $\check{S}^4$ is 1 again. On $\check{S}^4 \backslash \Sigma$ we put the Riemannian metric $\check{g}$ inherited from the round metric on S^4. Topologically, $\check{S}^4$ is a sphere, but the metric has an orbifold singularity along Σ.

On S^4, take the trivial complex line bundle $\mathbb{C} \times S^4$ with the product connection. Lift the action of $\mathbb{Z}_\nu$ to this line bundle by making the generator act as $e^{2\pi i a/\nu}$ on the fibre. The number a should be an integer in the range $1 < a < \nu$. Taking the quotient by $\mathbb{Z}_\nu$, we obtain a line bundle K on $\check{S}^4 \backslash \Sigma$ with a connection whose holonomy is non-trivial: in some global trivialization of K, the connection 1-form is $i\alpha_0\, d\theta$, where

$$\alpha_0 = a/\nu.$$

We shall write ∇_{α_0} for this covariant derivative, and we let d_{α_0} and $d^+_{\alpha_0}$ stand for the usual operators in the complex

$$\Omega^0(K) \xrightarrow{d_{\alpha_0}} \Omega^1(K) \xrightarrow{d^+_{\alpha_0}} \Omega^+(K).$$

The formal adjoint of d_{α_0} for the metric $\check{g}$ will be written $d^*_{\alpha_0}$.

Until section (iii) below, when considering the covariant derivatives on $\Omega^i(K)$ we shall have in mind that the Levi-Civita connection of the orbifold metric $\check{g}$ is used. We shall write $\check{\nabla}_{\alpha_0}$ for the Levi-Civita connection coupled to ∇_{α_0}. With this said, let T be the kernel of $d^*_{\alpha_0}$ acting on the completion of $\Omega^1(K)$ in the norm

$$\|a\|_T = \|\check{\nabla}_{\alpha_0} a\|_p,$$

and consider the operator

$$d^+_{\alpha_0} : T \to L^p(\check{S}^4, \Lambda^+(K)).$$

This operator is invertible since it can be regarded as just the $\mathbb{Z}_\nu$-invariant part of the corresponding operator with trivial coefficients on the smooth S^4 upstairs (the latter operator is invertible since S^4 has no first or second cohomology). Let Q_{α_0} denote the inverse. We need to estimate the operator norm of Q_{α_0}.

Lemma 4.1. *The operator norm of Q_{α_0} on the spaces above is bounded by $M_p\sqrt{2}$, where M_p is a constant which can be taken to be independent ν and α and tends to 1 as p tends to 2.*

Proof. It is sufficient to estimate the norm of the operator Q upstairs on S^4, for this will bound the norm of Q_{α_0} for all ν and α_0. Further, it suffices to estimate the norm in the case $p = 2$, because the estimate for nearby p follows by interpolation (cf. [DS]). On the kernel of d^* on $\Omega^1(S^4)$, the Weitzenböck formula [FU] reads

$$\nabla^*\nabla a = 2d^*d^+a - \mathrm{Ricci}(a),$$

and since the Ricci curvature is positive on S^4 we obtain

$$\|\nabla a\|_2^2 \leq 2\|d^+a\|_2^2$$

by integration by parts. This gives the desired estimate. □

There is a second operator whose norm we need to estimate before leaving the simple orbifold. Let R^+ be the operator

$$\begin{aligned} R^+ : T &\to L^p(\check{S}^4, \Lambda^+(K)) \\ a &\to (a \wedge d\theta)^+, \end{aligned}$$

and let R^- be defined similarly. The fact that these are bounded comes from (3.5). Let $c(\alpha)$ again denote the saw-tooth function, and suppose that $1/\nu$ is less than $c(\alpha_0)$.

Lemma 4.2. *If $1/\nu$ is less than $c(\alpha)$, the operator norms of R^+ and R^- are bounded by $N_p/(\sqrt{2}(c(\alpha)-1/\nu))$, where N_p is a constant which tends to 1 as p tends to 2.*

Proof. Again, it is enough to treat the case $p=2$. The number $c(\alpha)$ is equal to the absolute value of smallest eigenvalue of the operator ∂_α of (3.5) acting on functions on the unit circle S^1. So, since $|d\theta|=1$ on S^1, we have

$$\|\partial_\alpha f \otimes d\theta\|^2 \geq c(\alpha)^2 \|f\, d\theta\|^2,$$

for the L^2 norms. If we replace functions f by 1-forms, we have to take account of the fact that the Levi-Civita connection for $\breve{g}$ has holonomy on some components of Λ^1; as in the proof of (3.9), this shifts α by $\pm 1/\nu$. So we have

$$\|\partial_\alpha a \otimes d\theta\|^2 \geq (c(\alpha)-1/\nu)^2 \|a\wedge d\theta\|^2,$$

This estimate is independent of the size of the circle; so by integrating over the circles which make up $\breve{S}^4$ we obtain

$$\|\breve{\nabla}_\alpha a\| \geq (c(\alpha)-1/\nu)\|a\wedge d\theta\|.$$

The extra factor of $\sqrt{2}$ arises because $R^+(a)$ and $R^-(a)$ are orthogonal 2-forms of equal length whose sum is $a\wedge d\theta$, and the self-dual and anti-self-dual parts of a decomposable 2-form have the same magnitude. □

(ii) Changing the conformal structure.

We now bring in the device used in [DS]. Let g be some other metric on $\breve{S}^4\backslash\Sigma$, different from $\breve{g}$. (We have in mind later that g will be a metric with a more open cone-angle than $\breve{g}$, but we keep the discussion general for now.) Only the conformal class of g is material.

We continue to use Ω^+ and Ω^- to denote the self-dual and anti-self-dual spaces for the metric $\breve{g}$, and we write P^+ and P^- for the projections. The anti-self-dual subspace for the metric g can be represented as the graph of a unique linear map

$$\mu : \Lambda^- \to \Lambda^+$$

defined by a bundle map whose pointwise operator norm is everywhere less than 1 [D2]. We shall suppose that there is a uniform upper bound $|\mu|$ which is *strictly* less than 1. The map μ encodes the conformal structure of g (see [D2]), so we can dispense with g and think only of μ as given. The anti-self-dual subspace for the new conformal structure is the kernel of $P^+ - \mu P^-$. We consider the operator

$$d^+_{\alpha_0,\mu} = d^+_{\alpha_0} - \mu\circ d^-_{\alpha_0} : T \to \Omega^+(K).$$

If we identify the self-dual space for μ with Ω^+ by the projection, then this operator is the anti-self-duality operator d^+ for the new conformal structure. The norms on T and $\Omega^+(K)$ are still defined using the metric $\breve{g}$.

Write S for the operator $d^-_{\alpha_0} \circ Q_{\alpha_0} : \Omega^+(K) \to \Omega^-(K)$. The next lemma is from [DS].

Lemma 4.3. *The operator norm of S on L^p is bounded by a constant C_p which tends to 1 as p approaches 2. This constant can be taken to be independent of ν and α_0.*

Proof. For the operators with trivial coefficients on S^4, this is (2.14) from [DS]. Take the $\mathbb{Z}_\nu$-invariant part to obtain this version of the result. □

From now on we shall suppose that p is sufficiently close to 2 to ensure that $C_p|\mu|$ is strictly less than 1, where C_p is the constant in the lemma above.

Lemma 4.4. *If $C_p|\mu|$ is less than 1 then the operator $d^+_{\alpha_0,\mu}$ is invertible. The operator norm of its inverse $Q_{\alpha_0,\mu}$ is bounded above by $M_p\sqrt{2}(1 - C_p|\mu|)^{-1}$, where M_p is the constant from Lemma 4.1.*

Proof. (This follows [DS]). The number $C_p|\mu|$ bounds the operator norm of

$$\mu \circ S : \Omega^+(K) \to \Omega^+(K)$$

in L^p. So if this number is less than 1 then $(1 - \mu S)$ is invertible by a power series:

$$(1 - \mu S)^{-1} = 1 + \mu S + (\mu S)^2 + \cdots.$$

An estimate for the norm of this inverse is $(1 - C_p|\mu|)^{-1}$. Combining this with Lemma 4.1 we obtain the bound $M_p\sqrt{2}(1-C_p|\mu|)^{-1}$ for the norm of the operator $Q_{\alpha_0}(1 - \mu S)^{-1}$, which is the inverse of $d^+_{\alpha_0,\mu}$. □

Now we get to the main point, which is to change the holonomy angle away from the orbifold value a/ν. Let $\alpha = \alpha_0 + \gamma$, with γ considered small. A connection ∇_α with this holonomy is obtained from ∇_{α_0} by adding the connection 1-form $i\gamma\, d\theta$. The corresponding anti-self-duality operator for the metric $\breve{g}$ is

$$d^+_\alpha = d^+_{\alpha_0} + i\gamma R^+ : T \to \Omega^+(K),$$

while for the conformal structure μ we have

$$d^+_{\alpha,\mu} = d^+_{\alpha_0,\mu} + i\gamma(R^+ - \mu R^-).$$

For small γ we can regard this operator as a perturbation of the operator $d^+_{\alpha_0,\mu}$, which we already know to be invertible with inverse $Q_{\alpha_0,\mu}$. The power series for the inverse of the perturbed operator will converge provided that the operator norm of

$$i\gamma(R^+ - \mu R^-) \circ Q_{\alpha_0,\mu}$$

is less than 1. The operator norm of $R^+ - \mu R^-$ is bounded by $(1 + |\mu|)$ times the constant in Lemma 4.2 as long as $c(\alpha_0)$ exceeds $1/\nu$; for the other factor $Q_{\alpha_0,\mu}$ we have the estimate (4.4). Putting together these estimates, we deduce:

Lemma 4.5. *The operator $d^+_{\alpha,\mu}$, with $\alpha = \alpha_0 + \gamma$ is invertible provided that $c(\alpha_0)$ exceeds $1/\nu$ and*

$$|\gamma| \frac{M_p N_p}{(c(\alpha_0) - 1/\nu)} \frac{(1+|\mu|)}{(1 - C_p|\mu|)} < 1.$$

□

(iii) Application to the smooth metric.

As a particular case, we now take g to be a metric on $\check{S}^4$ which makes this topological sphere look like a round sphere of unit radius. This g just stretches the θ directions by a factor of ν compared to $\check{g}$. Conformally, the metric $\check{g}$ is the same as the metric on $\mathbb{R}^4 \backslash \mathbb{R}^2$ given by

$$du^2 + dv^2 + dr^2 + \left(\frac{r^2}{\nu^2}\right) d\theta^2.$$

The new conformal structure $[g]$ will just lose the factor of ν^2 in the denominator; so the conformal structure is that of the flat metric on $\mathbb{R}^4$ or the round one on S^4. For the metric displayed above, an orthonormal basis of 1-forms at any point is given by

$$e_1 = du, \quad e_2 = dv, \quad e_3 = dr, \quad e_4 = \left(\frac{r}{\nu}\right) d\theta,$$

and the map μ which corresponds to $[g]$ is easily calculated in these terms; we have

$$\mu(e_1 e_2 - e_3 e_4) = -\left(\frac{\nu - 1}{\nu + 1}\right)(e_1 e_2 + e_3 e_4)$$

because the form

$$(e_1 e_2 - e_3 e_4) - \left(\frac{\nu - 1}{\nu + 1}\right)(e_1 e_2 + e_3 e_4) = \left(\frac{2}{\nu + 1}\right)(e_1 e_2 - \nu\, e_3 e_4)$$

is anti-self-dual for g. There are similar expressions for the other basis vectors of Ω^-, and the norm of μ is therefore

$$|\mu| = \frac{\nu - 1}{\nu + 1}.$$

Substituting this value into the estimate in Lemma 4.5, we find that the differential operator is invertible as long as

$$|\gamma| < \left(\frac{c(\alpha_0) - 1/\nu}{M_p N_p}\right) \frac{1 - \frac{1}{2}(C_p - 1)\nu}{\nu}.$$

Since the three unknown constants approach 1 as p approaches 2, the right hand side is eventually greater than $0.9(c(\alpha_0) - 1/\nu)/\nu$ when p tends to 2 with ν and α_0 fixed. This proves:

Proposition 4.6. *There is a positive number $\eta(\nu)$ such that for all p with $|p-2| < \eta(\nu)$ the operator $d^+_{\alpha,\mu}$ is invertible for α in the range*

$$|\alpha - \alpha_0| < 0.9 \left(\frac{c(\alpha_0) - 1/\nu}{\nu} \right).$$

□

We now shift our point of view. The space $\check{S}^4$ depends on an integer ν, but the geometry of this space equipped with the metric g is independent of ν: it looks like the round sphere. So let S now be a fixed round sphere containing a standard S^2, and let us identify each of the spaces $(\check{S}^4, g)_\nu$ isometrically with S. We have a line bundle K on S which we regard as the trivial bundle equipped with the connection $i\alpha d\theta$. The space of forms T on $\check{S}^4$ becomes identified with a space of forms on S which depends on ν and on a. Let us denote this space by $T_{a,\nu}$.

The round metric on S defines a Levi-Civita connection which, coupled to the connection on K, defines an $L^p_{1,\alpha}$ norm on $\Omega^1_S(K)$ and an $L^p_{2,\alpha}$ norm on $\Omega^0(K)$. According to (3.9), these two norms are equivalent to those defined by the orbifold derivative $\check{\nabla}_\alpha$ provided that $c(\alpha) > 1/\nu$.

We now let d^+_α stand for the usual anti-self-duality operator for the round metric; this operator is equivalent to the operator $d^+_{\alpha,\mu}$ considered in the proposition above. We have as usual the complex

$$\Omega^0_S(K) \xrightarrow{d_\alpha} \Omega^1_S(K) \xrightarrow{d^+_\alpha} \Omega^+_S(K) \tag{4.7}$$

and we can translate Proposition 4.6 as follows:

Proposition 4.8. *Consider the complex* (4.7) *for the round metric on the sphere, and let the three spaces be completed in the topologies $L^p_{2,\alpha}$, $L^p_{1,\alpha}$ and L^p. Suppose there exist integers a and ν such that, with $\alpha_0 = a/\nu$, we have*

$$\begin{aligned} c(\alpha) &> 1/\nu \\ c(\alpha_0) &> 1/\nu \\ |\alpha - \alpha_0| &< 0.9 \left(\frac{c(\alpha_0) - 1/\nu}{\nu} \right). \end{aligned}$$

Suppose also that $|p-2| < \eta(\nu)$. Then the operator d^+_α has a bounded right inverse Q mapping to the space $T_{a,\nu}$, which is a closed complement to the image of d_α.

□

We now see that for each rational number a/ν in the interval $(0,1)$ there is an open neighbourhood of acceptable values for the holonomy α. To tie things up, we observe that these open intervals cover the whole of $(0,1)$:

Lemma 4.9. *For each $\alpha \in (0,1)$ there exists a rational number $\alpha_0 = a/\nu$ such that the three inequalities displayed in* (4.8) *are satisfied.*

Proof. Take $a = 3$ and consider the sequence of values $\alpha_0 = 3/\nu$, for $\nu = 6, 7, \ldots$. For $\alpha \leq \frac{1}{2}$, we have $c(\alpha) = \alpha$ and the conditions are satisfied as long as

$$|\alpha - (3/\nu)| < 1.8/\nu^2.$$

Since the distance between $3/\nu$ and $3/(\nu+1)$ is $3/\nu(\nu+1)$, which is less than $1.8(1/\nu^2 + 1/(\nu+1)^2)$, the adjacent intervals overlap, so covering the whole of the bottom half of the interval $(0, 1)$. The top half is covered similarly by taking $a = \nu - 3$. □

Remark. There is nothing too delicate in this. We can replace 0.9 by any positive constant bigger than $\frac{1}{2}$ in (4.8) without spoiling the proof of (4.9); it is only necessary to replace the choice $a = 3$ in the proof by a larger integer. If we are prepared to sacrifice some explicitness in approximating reals by rationals then any non-zero constant ε can replace 0.9. The main thing is to approximate a real α with a rational a/ν with

$$\big|\alpha - \frac{a}{\nu}| < \frac{\varepsilon(a-1)}{\nu^2}.$$

For this problem, with α fixed, we can replace $(a-1)/\nu$ by a constant, and so consider the problem as one of approximating α by a/ν with $|\alpha - a/\nu| < \delta/\nu$, for some small δ. The existence of such an approximation is standard: it is the statement that the multiples of any irrational number, taken mod 1, are dense in the unit interval.

We now return to considering an SU(2) bundle $E = L \oplus L^{-1}$ on $S \backslash S^2$, where L is the trivial line bundle carrying the connection with holonomy parameter α. The Lie algebra bundle $\mathfrak{g}_E$ splits as $\mathbb{R} \oplus L^2$, where L^2 has holonomy 2α. We now take α in the range $(0, \frac{1}{2})$, and take L^2 as the K of the Proposition (4.8) (so there is a factor of 2 between the present α and that above). Let A^α denote the SU(2) connection, and let L^p_{k,A^α} be the covariant Sobolev norms defined in section 3. Consider the usual complex

$$\Omega^0(\mathfrak{g}_E) \xrightarrow{d_{A^\alpha}} \Omega^1(\mathfrak{g}_E) \xrightarrow{d^+_{A^\alpha}} \Omega^+(\mathfrak{g}_E) \tag{4.10}$$

This is the sum of two complexes corresponding to the decomposition of $\mathfrak{g}_E$. The $\mathbb{R}$ summand poses no problems: the kernel of the formal adjoint d^* for the round metric provides a complement to the image of d on which d^+ is invertible. For the L^2 summand we use the results above. Putting things together we obtain:

Proposition 4.11. *Let the spaces of forms in* (4.10) *be completed in the norms* L^p_{2,A^α}, L^p_{1,A^α} *and* L^p *respectively. Then there exists a continuous positive function* $\eta(\alpha)$ *on the interval* $(0, \frac{1}{2})$ *such that for all* p *with* $|p-2| < \eta(\alpha)$ *the operator* $d^+_{A^\alpha}$ *for the round metric has a right inverse* Q *mapping to a closed complement* T *of* $\mathrm{Im}(d_{A^\alpha})$. *The space* T *depends on* α, *but for any* α *there is an open neighbourhood for which the same* T *will serve, and in this region the operator norm*

of Q is uniformly bounded. Both Q and T are independent of p in the admissible range. □

The phrase "independent of p" is meant with the understanding that the L^p spaces are contained one in another. This rider to the proposition implies a little elliptic regularity for the operator $d^+_{A^\alpha}$ on T. So, for example, suppose $p \geq q$ are two numbers close to 2; then we have:

Lemma 4.12. *Let $a \in \Omega^1(\mathfrak{g}_E)$ be of class L^q_{1,A^α}, with $d^+_{A^\alpha} a$ in L^p. Then there is a $u \in L^q_{2,A^\alpha}$ with $a - d_{A^\alpha} u \in L^p_{1,A^\alpha}$.* □

The construction of the complement T seems artificial, and it is worth spelling it out again. Consider the decomposition of $\mathfrak{g}_E$ into diagonal and off-diagonal parts. Let d^1 be the usual d^* operator for the round metric, acting on the diagonal summand; and let d^2 be the formal adjoint of $d_{2\alpha_0}$ with respect to the orbifold metric, acting on the off-diagonal summand. Let d' be the operator

$$d' = d^1 \oplus d^2 : \Omega^1(\mathfrak{g}_E) \to \Omega^0(\mathfrak{g}_E) \tag{4.13}$$

Our slice T in the complex (4.10) is defined as the kernel of d', in which $2\alpha_0$ is a rational approximation to 2α satisfying the conditions of (4.8). Thus, on the off-diagonal part, it is the kernel of an operator defined using a different metric and a connection with a different holonomy. The second of these two features is probably an unnecessary artifact of our line of argument, but the need to avoid the usual Coulomb condition seems to be real.

(iv) Intermediate cone-angles.

Although we have pursued the argument only for the smooth metric on the sphere, we could instead have stretched out the metric $\breve{g}$ by a smaller factor, and so obtained the same results for any cone-angle less than 2π. Let $\tilde{\nu} > 1$ be some fixed number, and let ν large integer as before, greater than $\tilde{\nu}$. Form the same orbifold $\breve{S}^4$ with an orbifold metric of cone-angle $2\pi/\nu$. Let $\tilde{g}$ be the metric obtained by dilating the θ directions by a factor of $\nu/\tilde{\nu}$, so $\tilde{g}$ has cone-angle $\tilde{\nu}$. The new conformal structure is related to $\breve{g}$ by a map $\tilde{\mu} : \Lambda^- \to \Lambda^+$ which has norm

$$|\tilde{\mu}| = \left(\frac{\nu - \tilde{\nu}}{\nu + \tilde{\nu}} \right).$$

This is smaller than the norm of the μ in the previous section, so the estimates become more favourable. Let d' be defined again as in (4.13), with d^1 still being the formal adjoint with respect to the *round* metric (neither $\tilde{g}$ nor $\breve{g}$). Then we have

Proposition 4.14. *The statements of Propositions 4.11 and 4.12 continue to hold when the operator $d^+_{A^\alpha}$ is defined using the cone-like metric $\tilde{g}$ in place of the round metric on the sphere. In Proposition 4.11, the transverse slice T is still provided by the kernel of d', as in (4.13).*

Proof. When $\mathfrak{g}_E$ is decomposed again as $\mathbb{R} \oplus L^2$, the result for the off-diagonal part L^2 follows as above. The diagonal summand $\mathbb{R}$ is now a little different, since we are using the slice defined by the formal adjoint for the round metric while the anti-self-duality condition is defined with respect to the metric $\tilde{g}$. The required Fredholm property here follows from another application of the idea we have already used from [DS]: we regard the conformal structure $\tilde{g}$ as being obtained by deforming the round metric. (This is the same application as was made by Wang in [W]). □

(v) An alternative approach : stronger norms.

If we are satisfied with developing a theory which is valid only for α in a specified, compact subinterval, and if we are content with cone-like metrics, then we can simplify the constructions of section (iii) considerably: the standard L^2 adjoint d^* will define a slice, and we can use the stronger Sobolev norms L^2_k, for fixed, large k. We outline the argument in this section.

We take a large integer ν and go back to the orbifold $\check{S}^4 = S^4/\mathbb{Z}_\nu$ with the metric $\check{g}$. We define K again, with its connection ∇_{α_0} for $\alpha_0 = a/\nu$, as in section (i) above. Again, we insist that $0 < a < \nu$, so the holonomy is non-trivial. We define Sobolev norms $\check{L}^p_{k,\alpha_0}$ for $1 < p < \infty$, now using the connection $\check{\nabla}_{\alpha_0}$ coupled Levi-Civita connection *of the orbifold metric* $\check{g}$ (so these are the norms we called $L^p_k(g^\nu)$ in section 3(iv)). These norms are equivalent to the usual Sobolev norms on the branched cover, restricted to the subspace which is invariant under the weight a action of $\mathbb{Z}_\nu$. The differential operators are also just invariant parts of those on S^4. This immediately implies:

Lemma 4.15. *The operator*

$$\mathcal{D}_{\alpha_0} = d^*_{\alpha_0} \oplus d^+_{\alpha_0} : \Omega^1(K) \to \Omega^0(K) \oplus \Omega^+(K),$$

considered as acting from $\check{L}^p_{k,\alpha_0}$ *to* $\check{L}^p_{k-1,\alpha_0}$*, has an inverse* Q *whose operator norm is bounded by a constant* $N_{k,p}$ *which depends on* k *and* p *but is independent of* ν *and* α_0*.*

(Note that although the operator $\mathcal{D}$ is not surjective for trivial coefficients on S^4, it is surjective on the invariant part of weight a for $a \neq 0$. So there is no need to qualify Lemma 4.8 with a statement of a Fredholm alternative.)

Lemma 4.16. *Let* $\alpha \in (0,1)$ *and an integer* k *be given. Then for all* $\nu > k/c(\alpha)$ *the multiplication operator on* $\Omega^1(K)$ *defined by*

$$a \mapsto a \wedge d\theta$$

is bounded as a map from $\check{L}^p_{k,\alpha}$ *to* $\check{L}^p_{k-1,\alpha}$*, and the bound is of the form* $M_{k,p}/(c(\alpha) - k/\nu)$*, where* $M_{k,p}$ *is independent of* ν *and* α*, and* c *is the saw-tooth function defined previously.*

Proof. The pointwise norm of $d\theta$ is $1/r$, so this lemma follows from (3.9) in the case $i = 1$. □

Once more, we set $\alpha = \alpha_0 + \gamma$, and we consider the operator $\mathcal{D}_\alpha$ as a perturbation of $\mathcal{D}_{\alpha_0}$. We can write

$$\mathcal{D}_\alpha = \mathcal{D}_{\alpha_0} + i\gamma P,$$

where P is a multiplication operator which satisfies bounds of the same order as the multiplication operator in Lemma 4.9 up to some overall algebraic factor. Adjusting the constant $M_{k,p}$ accordingly, we see that $\mathcal{D}_\alpha$ is going to be invertible as long as

$$|\gamma| < \frac{(c(\alpha_0) - k/\nu)}{N_{k,p} M_{k,p}}.$$

We mention also that we can apply the same argument to the formal adjoint $\mathcal{D}^*_\alpha$. So we have an interval surrounding the value $\alpha_0 = a/\nu$ for which the theory works. The radius of this interval exceeds $1/\nu$ provided that

$$c(\alpha_0) > (N_{k,p} M_{k,p} + k)/\nu,$$

and we can make the right-hand side less than any given ε by taking ν to be large. In that case, the union of the intervals centred on $1/\nu$, $2/\nu$, ... covers the subinterval of $(0, 1)$ defined by the condition $c(\alpha) \geq \varepsilon$: i.e., the subinterval $[\varepsilon, 1 - \varepsilon]$.

Finally, we switch back to the case of the SU(2) bundle $L \oplus L^{-1}$, where L has holonomy $\alpha \in (0, \frac{1}{2})$, and we consider the operator $\mathcal{D}_{A^\alpha}$ for the connection A^α on $\mathfrak{g}_E$. We write $\check{L}^p_{k,A^\alpha}$ for the Sobolev norms obtained using the orbifold Levi-Civita connection.

Proposition 4.17. *Given any compact subinterval $I \subset (0, \frac{1}{2})$ and any p and k_0, there exists a ν_0 such that for all $\alpha \in I$, all $\nu \geq \nu_0$ and all $k \leq k_0$, the elliptic theory for $\mathcal{D}_{A^\alpha}$ "works" on the Sobolev spaces $\check{L}^p_{k,A^\alpha}(\mathfrak{g}_E)$ over the orbifold $\check{S}^4 = S^4/\mathbb{Z}_\nu$. That is, both $\mathcal{D}_{A^\alpha}$ and its formal adjoint are Fredholm, acting from $\check{L}^p_k$ to $\check{L}^2_{k-1}$, the kernel of $\mathcal{D}_{A^\alpha}$ is trivial, its cokernel consists of the constant diagonal matrices in $\Omega^0(\mathfrak{g}_E)$, and the Fredholm alternative holds.* □

From this proposition, elliptic regularity follows for both $\mathcal{D}_{A^\alpha}$ and its adjoint. For example, if a is in $L^2(\Omega^1(\mathfrak{g}_E))$ on $\check{S}^4$ and satisfies $\mathcal{D}_{A^\alpha} a = \eta$ formally, with $\eta \in \check{L}^2_{j,A^\alpha}$, then a is in $\check{L}^2_{j+1,A^\alpha}$. This is because the proposition gives us some solution $\tilde{a} \in \check{L}^2_{j+1,A^\alpha}$ to the equation $\mathcal{D}_{A^\alpha} \tilde{a} = \eta$, and it follows that $a = \tilde{a}_s$. If a is only defined on an open subset $U \subset \check{S}^4$ then we multiply a by a cut-off function ψ of class $\check{L}^2_{k+1}$ to deduce the statement of local elliptic regularity by a bootstrapping argument. The remaining point worth noting is that (4.17) also tells us that the second-order operator $\mathcal{D}^*_{A^\alpha} \mathcal{D}_{A^\alpha}$ is Fredholm with kernel and cokernel equal to the constants, and this operator also satisfies the usual statement of elliptic regularity for formal L^2 solutions.

(vi) Hodge theory for the de Rham complex.

The argument of the previous subsection extends to other operators. Again take the bundle $\mathfrak{g}_E$ with the connection A^α over $(\check{S}^4, \check{g})$, and let $\Omega'(\mathfrak{g}_E)$ denote the complement in $\Omega^*(\mathfrak{g}_E)$ of the constant forms in Ω^0 and Ω^4. The operator

$$d^*_{A^\alpha} + d_{A^\alpha} : \Omega'(\mathfrak{g}_E) \to \Omega'(\mathfrak{g}_E)$$

in the topology $\check{L}^p_{1,A^\alpha} \to L^p$ is invertible in the case that the holonomy is a/ν; this comes from upstairs again. Just as in 4.17, if an interval I is given, we can argue that for large ν the operator will be invertible for all α in I.

The value of ν affects the operator d^* through the metric $\check{g}$, but the de Rham complex itself (and the L^p_1 norm, up to equivalence, once ν is large) do not see the metric. So, using the metric-dependent d^* as a prop, we deduce:

Proposition 4.18. *Let $\omega \in \Omega^i(\mathfrak{g}_E)$ be of class L^q, and suppose $d_{A^\alpha}\omega$ is in L^p in the distributional sense, with $1 < p, q < \infty$. Let the exponents be such as to ensure that $L^p_1 \subset L^q$. Then there exists a $\chi \in \Omega^{i-1}(\mathfrak{g}_E)$ of class L^q_{1,A^α} such that $\omega - d\chi \in L^p_{1,A^\alpha}$.* □

5. Global theory

We now have all the machinery in place to prove the propositions of section 2, with the exception of the dimension formula, which we postpone to section 6. The local elliptic theory of section 4 allows us to develop the global theory in more than one way. This reflects the fact that, if we take the standard elliptic complex

$$\Omega^0(\mathfrak{g}_E) \xrightarrow{d_{A^\alpha}} \Omega^1(\mathfrak{g}_E) \xrightarrow{d^+_{A^\alpha}} \Omega^+(\mathfrak{g}_E)$$

for the model twisted connection A^α on (S^4, S^2), then there is more than one way in which we can complete the spaces to obtain a Fredholm complex of Banach spaces.

First, if d^+ is defined with respect to the round conformal structure, then the three spaces can be completed in the topologies L^p_{2,A^α}, L^p_{1,A^α} and L^p respectively, for some p depending on α; this is the content of (4.11). If we take p a little bigger than 2 then the functions in L^p_{2,A^α} are continuous, so the non-linear aspects of the gauge theory can be developed and the results transferred to an arbitrary pair (X, Σ). This is the framework outlined in section 2; it is the most natural and general of these settings, and its details will occupy us for parts (i)–(iii) of this section.

Next, if we put a cone-like metric $\check{g}$ on (S^4, S^2), with cone-angle $2\pi/\nu$ for some arbitrary real ν greater than 1, and if d^+ is the corresponding operator, then the complex is again Fredholm, with the *same* completions of the spaces as above. This is the content of (4.14), and it allows us to develop the theory for a pair (X, Σ) equipped with a cone-like metric, as promised in Proposition 2.9.

Since the necessary modifications to the argument are very small, we will not go through the details.

Finally, if ν is an integer, then we can complete the spaces in the topologies $\check{L}^2_{k+1,A^\alpha}$, $\check{L}^2_{k,A^\alpha}$ and $\check{L}^2_{k-1,A^\alpha}$, for some $k \geq 2$, as in section 4(v). This gives rise to a more elementary theory. However, the holonomy parameter α is then restricted to lie in some compact sub-interval $I \subset (0, \frac{1}{2})$, which depends on ν. We shall sketch some of the necessary construction in part (iv) below, and we shall see that the moduli spaces which arise are not different from those which are obtained using the other completion.

(i) Moduli spaces.

We consider the set-up of section 2: X is a smooth, closed 4-manifold with an embedded surface Σ and a smooth metric $\bar{g}$; on $X \backslash \Sigma$ we have an SU(2) bundle E with a model twisted connection A^α having holonomy parameter $\alpha \in (0, \frac{1}{2})$; and we define $\mathcal{A}^\alpha$ and $\mathcal{G}$ as before, fixing an exponent p a little bigger than 2. As in (4.8), there is a rational number $\alpha_0 = a/\nu$ such that $|\alpha - \alpha_0|$ satisfies the necessary bounds, and we shall suppose p is chosen so that $p - 2 < \eta(\nu)$. Let q be the conjugate exponent; this is a little less than 2, and $|q - 2| < \eta(\nu)$ also.

Since the elliptic theory given by (4.11) does not deal with the usual operator d^*_A, we develop the global theory on the basis of having a parametrix for the operator d^+. In this we follow [DS], and the exposition below is modeled closely on section 4 of that paper, to which the reader can turn for more details at several points.

Let A be a connection in $\mathcal{A}^\alpha$ and consider the operators

$$T\mathcal{G} \xrightarrow{d_A} T\mathcal{A}^\alpha \xrightarrow{d^+_A} \mathcal{F}, \tag{5.1}$$

where the three spaces are the completions of $\Omega^0(\mathfrak{g}_E)$, $\Omega^1(\mathfrak{g}_E)$ and $\Omega^+(\mathfrak{g}_E)$ in the L^p_{2,A^α}, L^p_{1,A^α} and L^p topologies respectively. (These operators do not form a complex if A is not anti-self-dual.) The results of section 3 imply that the norms $L^p_{k,A}$ for $k = 1$, 2 are equivalent to the L^p_{k,A^α} norms, since $A^\alpha - A$ is in L^p_{1,A^α} and the multiplication theorems (3.8) apply. The same multiplication theorems show that the operators are bounded. It also follows immediately that the image of d_A is closed; its kernel is the space of covariant constant sections, which is either 0 or 1 dimensional.

Lemma 5.2. *The operator d^+_A has a right parametrix P, i.e. an operator $P : \mathcal{F} \to T\mathcal{A}^\alpha$ such that $d^+_A P = 1 + K$, with K compact.*

Proof. (See [DS], p. 210.) Cover X (not $X \backslash \Sigma$) with open sets U_i. We can arrange that these are of two types: the first type meet Σ and are such that the pairs $(U_i, U_i \cap \Sigma)$ are diffeomorphic to a standard ball-pair inside (S^4, S^2), with the metric being nearly Euclidean; the second type are disjoint from Σ, and the metric on these is also to be close to the Euclidean one. On each patch U_i,

choose a trivialization of the bundle so as to express the connection as

$$d_A = d_{A_0} + a$$

If U_i is of the first type then A_0 is to be the model α-twisted connection on (S^4, S^2); otherwise, A_0 is to be the trivial connection. In either case, a is in L^p_{1,A_0}.

Next construct local inversion operators Q_i for the operators $d^+_{A_0}$. Let $\{U_i'\}$ be a slight shrinking of the cover $\{U_i\}$, and let β_i be a cut-off function, equal to 1 on U_i' and supported in U_i. After identifying U_i with a ball in S^4, define an operator Q_i acting on sections of $\Lambda^+(\mathfrak{g}_E)$ over U_i by

$$Q_i(\omega) = Q(\beta_i \omega).$$

Here Q is the operator provided by (4.11) if U_i is of the first type; otherwise it is the standard operator for trivial coefficients on S^4. In either case, Q_i maps L^p to L^p_{1,A_0} continuously, and we have

$$d^+_{A_0} Q_i(\omega) = \omega \qquad \text{on } U_i'.$$

(We should note that the proof of (4.11) shows that the proposition remains valid for metrics which are close to the round one on S^4.)

Let $\{\gamma_i\}$ be a smooth partition of unity subordinate to $\{U_i'\}$ and define

$$P(\omega) = \sum_i \gamma_i Q_i(\omega|_{U_i}).$$

Then, as in [DS], we compute:

$$\begin{aligned} d^+_A P(\omega) &= \sum_i \gamma_i d^+_A Q_i(\omega|_{U_i}) \\ &\qquad + \sum_i (\nabla \gamma_i) Q_i(\omega|_{U_i}) \\ &= \sum_i \gamma_i \omega|_{U_i'} + \sum_i K_i(\omega) \\ &= \omega + \sum_i K_i(\omega) \end{aligned}$$

where $K_i(\omega) = (\gamma_i a^+ + \nabla\gamma_i) Q_i(\omega|_{U_i})$. Here a^+ denotes the operator $b \mapsto (a \wedge b)^+$.

The coefficient $R = (\gamma_i a^+ + \nabla \gamma_i)$ is in L^p_{1,A^α}, so the corresponding multiplication map

$$R : L^p_{1,A^\alpha} \to L^p$$

is compact by (3.8). This shows that each K_i is compact and proves the lemma, with $K = \sum_i K_i$. □

Corollary 5.3. *The image of d^+_A has finite codimension.* □

Let $H \subset \mathcal{F}$ denote the space of coupled self-dual harmonic forms of class L^p, the kernel of d_A on $\mathcal{F}$, and let $H^\perp$ be its annihilator in $\mathcal{F}$ under the L^2 inner product.

Lemma 5.4. *The image of d_A^+ is equal to $H^\perp$ and the cokernel is isomorphic to H. The operator has a bounded right inverse $Q_A : H^\perp \to T\mathcal{A}^\alpha$.*

Proof. The image is the annihilator of the space of coupled self-dual harmonic forms of class L^q (the dual exponent). So the first sentence follows if we show that the L^q harmonic forms are all in L^p. Again, a model for this regularity result is in [DS], on p.211. In local charts the equation is

$$(d_{A^\alpha} + a)\omega = 0,$$

with $a \in L^p_{1,A^\alpha}$. The multiplication theorems give $a\omega$, and hence $d_{A^\alpha}\omega$ in $L^{4/3}$. So by the Hodge theory (4.18) there is a $\chi \in L^q_{1,A^\alpha}$ with

$$\omega + d\chi \in L^{4/3}_1 \subset L^2$$

over a smaller open set. Since ω is self-dual, $d^-_{A^\alpha}\chi$ is in L^2; so using (4.12) we may alter χ by an exact 1-form so that $\chi \in L^2_{1,A^\alpha}$. This shows that $\omega \in L^2$. Repeat the whole thing to get $\omega \in L^p$ at the next iteration.

Given the parametrix (5.2), the construction of the one-sided inverse is standard; see [DS] for the model. □

Lemma 5.5 ([DS], Lemma 4.14). *For each A in $\mathcal{A}^\alpha$, the image of d_A has a closed complement $T_A \subset T\mathcal{A}^\alpha$.*

Proof. Lacking the usual d_A^*, we exploit the operator Q_A as in [DS]. Let $S = \mathrm{Im}(Q_A)$, and consider the intersection and sum:

$$S \cap \mathrm{Im}\, d_A$$
$$S + \mathrm{Im}\, d_A$$

The claim is that these have finite dimension and finite codimension respectively in $T\mathcal{A}^\alpha$. If this is established, the result follows, since a finite-dimensional modification of S will give the required transversal. We show first that $\mathrm{Im}\, d_A$ is closed. This follows if we show that, on any complement C to the finite-dimensional space of covariant constant sections, the norm on $T\mathcal{G}$ is equivalent to the norm $\|d_A u\|_{L^p_{1,A}}$. All that is at stake is to show that, if u_i is a sequence in $T\mathcal{G}$ which converges weakly to zero while $d_A u_i$ converges strongly to zero in $L^p_{1,A}$ norm, then u_i converges to zero strongly in L^p. Using Lemma 3.5, the hypothesis implies that $d_{\bar{A}^\circ} u_i$ is bounded in the ordinary Sobolev space L^p_1 defined by the smooth connection, and the existence of a strongly convergent L^p subsequence is then standard.

For the intersection, let a_i be a sequence in $S \cap \mathrm{Im}\, d_A$, so

$$a_i = d_A u_i = Q_A \psi_i,$$

and suppose $\|a_i\|_{L^p_{1,A}} = 1$. Since the inclusion of $L^p_{2,A}$ in C^0 is compact, we may suppose the u_i converge in C^0. Then $d^+_A a_i = [F^+_A, u_i] = \psi_i$ is L^p convergent, so $a_i = Q_A(\psi_i)$ converges in the $L^p_{1,A}$ topology of $T\mathcal{A}^\alpha$.

Now we turn to the sum. First note that $\mathrm{Ker}(d^+_A)$ is a transversal to S. In the case that A is anti-self-dual, the composite $d^+_A d_A$ is zero, so the image of d_A is contained in $\mathrm{Ker}(d^+_A)$ and the codimension of $S + \mathrm{Im}\, d_A$ is equal to the dimension of the cohomology $\mathrm{Ker}(d^+_A)/\mathrm{Im}(d_A)$. For general A, this description is not available as it stands. Following [DS] we introduce the operator

$$\delta_A = d_A - Q_A F^+_A$$

Here F^+ is thought of as an algebraic operator, the composite $d^+ d$, as above. The operator δ_A is a compact perturbation of d_A, and it is set up just so that the composite $d^+_A \delta_A$ is zero.

To complete the proof of the lemma, we need to show that $\mathrm{Ker}(d^+_A)/\mathrm{Im}(\delta_A)$ is finite dimensional. This means that given a sequence $\{a_i\}$ in $\mathrm{Ker}(d^+_A)$ with $\|a_i\|_{L^p_{1,A}} = 1$, we must find a sequence $\{u_i\}$ such that $\{a_i - \delta_A u_i\}$ has a convergent subsequence. As with the proof of (5.2), the u_i are constructed by a patching construction, covering X by open sets and applying (4.12) locally. A model can be found in [DS]. □

Proposition 5.6. *The orbit space $\mathcal{B}^\alpha = \mathcal{A}^\alpha/\mathcal{G}$ is a Hausdorff space. The space of irreducible connections $(\mathcal{B}^\alpha)^*$ is a Banach manifold, with a local chart at $[A]$ being provided by the projection map from $A + T_A$ to $\mathcal{B}^\alpha$.*

Proof. Entirely standard, now that we have the slice. □

Now let $M \subset \mathcal{B}^\alpha$ be the moduli space of anti-self-dual connections, and let $[A] \in M$. If $[A]$ is an irreducible connection then a neighbourhood of $[A]$ in M is modeled on the zero set of a smooth map (see section 3) $\Phi : T_A \to \mathcal{F}$. The derivative of Φ is

$$d^+_A : T_A \to \mathcal{F},$$

and the construction of T_A as a finite-dimensional modification of the image of Q_A shows that d^+_A is Fredholm on these spaces. It follows as usual that the neighbourhood is homeomorphic to the zero-set of a smooth map between finite-dimensional spaces,

$$\varphi : H^1_A \to H^2_A,$$

where the H^i_A are the cohomology groups of (5.1) (which is a complex when A is anti-self-dual). We also have the familiar description of the neighbourhoods in the case that A is reducible: the map φ is equivariant, and the model is $\varphi^{-1}(0)/S^1$.

Next we turn to the matter of showing that the moduli space M is independent of p, as asserted in part (i) of Proposition 2.5. Suppose $2 < p' < p < 2 + \eta$, and let $\mathcal{A}'$, $\mathcal{G}'$ and M' be the objects corresponding to p'. It is an elementary matter that if two connections in $\mathcal{A}^\alpha$ are gauge equivalent by a $g \in \mathcal{G}'$, then g

is actually in $\mathcal{G}$ (i.e., has the regularity of L^p_{2,A^α}). So there is a natural injective map from M to M'. To show surjectivity, let $A' \in \mathcal{A}'$ be an anti-self-dual connection and let us seek a gauge-equivalent connection in $\mathcal{A}^\alpha$. First, since $\mathcal{A}^\alpha$ is dense in $\mathcal{A}'$, we can find a nearby connection B of class L^p_{1,A^α}. There is a gauge transformation which carries B into $T_{A'}$; or thinking of this another way, there is a connection A, gauge-equivalent to A', with $B \in T_A$. Write $B = A + a$. Since T_A is constructed as a finite-dimensional modification of the image of Q_A, we have

$$a = Q_A d^+ a + s,$$

where s is regular (say of class L^p_{1,A^α}). On the other hand, we have

$$d^+ a = F_B^+ + (a \wedge a)^+,$$

and using the multiplication theorems gives

$$a \wedge a \in L^{p^*}$$

where

$$p^* = \frac{2p'}{4 - p'} > p' + (p' - 2).$$

Since Q_A maps L^p to L^p_{1,A^α}, this now shows that a is in L^p_{1,A^α} (if $p^* > p$) or in $L^{p^*}_{1,A^\alpha}$ otherwise, in any case giving an improvement on p'. By repeated application of the argument we eventually reach p, showing that $A = B - a$ is in $\mathcal{A}^\alpha$.

So the natural map from M to M' is a continuous bijection. The gauge transformation in the argument above can be made a continuous function of a in small patches, so the inverse map is continuous too. □

The extended moduli spaces $\hat{M}$ now give us no problem. Given a compact subinterval $I \subset (0, \frac{1}{2})$ we can find a p less than the minimum value of $2 + \eta(\alpha)$ and construct the space $\hat{\mathcal{B}}$ over I. On the irreducible set $\hat{\mathcal{B}}^*$, local charts are obtained using the transversals $T_A \times \mathbb{R} \subset T\hat{\mathcal{A}}^\alpha$. Inside $\hat{\mathcal{B}}$ is the moduli space, modeled on the zero-set of a map

$$\hat{\Phi} : T_A \oplus \mathbb{R} \to \mathcal{F}.$$

Since the moduli space is independent of p, we can take the union over increasing intervals I to obtain $\hat{M}$.

(ii) The Chern-Weil formula.

This is a convenient moment to take up the Chern-Weil formula mentioned in the introduction (1.7), as we will need part of the discussion in the section to follow, when we talk about reducible solutions.

Let k and l again denote the instanton and monopole numbers of a model twisted connection A^α.

Proposition 5.7. *For all $A \in \mathcal{A}^\alpha$ we have the formula*

$$\frac{1}{8\pi^2}\int_{X\setminus\Sigma} \mathrm{tr}(F_A \wedge F_A) = k + 2\alpha l - \alpha^2 \Sigma\cdot\Sigma.$$

Proof. We begin by proving the formula for the model A^α in the simple case that A^α is globally reducible. So we suppose that $\bar{E} = \bar{L} \oplus \bar{L}^*$ globally, that $\bar{b}$ is a smooth connection on L and that A^α is reducible as

$$A^\alpha = \begin{pmatrix} b_\alpha & 0 \\ 0 & -b_\alpha \end{pmatrix},$$

where

$$b_\alpha = \bar{b} + i\alpha\beta(r)\eta.$$

Here, as in section 2, $\beta(r)$ is a cut-off function and $i\eta$ is a connection 1-form on the normal circle bundle to Σ. As we have already mentioned, the closed 2-form $d(i\beta(r)\eta)$ extends smoothly across Σ, since near Σ where $\beta = 1$ it is the pull back from Σ of the curvature form of the connection $i\eta$. We see from this description that the form

$$\frac{i}{2\pi} d(i\beta(r)\eta)$$

integrates on Σ to give $\Sigma\cdot\Sigma$, the degree of the normal bundle. Since the second cohomology of the neighbourhood is 1-dimensional, it follows that this closed 2-form represents the Poincaré dual of Σ (see [BT] for this construction of the Thom class). In de Rham cohomology we therefore have

$$\begin{aligned} \frac{i}{2\pi}[db_\alpha] &= \frac{i}{2\pi}[d\bar{b}] + \alpha[\Sigma] \\ &= c_1(\bar{L}) + \alpha[\Sigma]. \end{aligned}$$

Denoting the form on the left-hand side by ω, we now calculate the left hand side in (5.7) as

$$\begin{aligned} -\langle \omega\wedge\omega, X\rangle &= -c_1(\bar{L})^2 - 2\alpha c_1(\bar{L})[\Sigma] - \alpha^2[\Sigma]^2 \\ &= k + 2\alpha l - \alpha^2 \Sigma\cdot\Sigma, \end{aligned}$$

since $c_2(\bar{E}) = -c_1(\bar{L})^2$.

Although this calculation is global, it has an interpretation locally on N. Let $Y_\varepsilon \subset N$ be the 3-manifold circle bundle over Σ given by $r = \varepsilon$, and consider the Chern-Simons integral (see eg. [DK])

$$\tau_\varepsilon(A^\alpha) = \frac{1}{8\pi^2}\int_{Y_\varepsilon} \mathrm{tr}(dA\wedge A + \frac{2}{3} A\wedge A\wedge A).$$

The integral depends only on the homotopy class of the trivialization of the bundle on Y_ε with respect to which the connection matrix A is computed. But on Y_ε there is a distinguished trivialization, the one which extends to N, so

τ_ε can be defined as a real number. If X_ε denotes the complement of the ε-neighbourhood of Σ (a manifold with boundary Y_ε), then the theory of the Chern-Simons invariant gives us

$$\frac{1}{8\pi^2}\int_{X_\varepsilon} \mathrm{tr}(F_A \wedge F_A) = k + \tau_\varepsilon,$$

so the calculation above for the reducible solution tells us that

$$\lim_{\varepsilon \to 0} \tau_\varepsilon = 2\alpha l - \alpha^2 \Sigma\cdot\Sigma.$$

This local statement can now be applied elsewhere, so, by applying the above argument in reverse, we can conclude that the Chern-Weil formula (5.7) holds whenever A is a connection which is smooth and reducible near to Σ. Since such connections are dense in $\mathcal{A}^\alpha$ and the curvature integral is a continuous function of A in the L^p_{1,A^α} topology, the result follows. □

The Chern-Weil formula has a simple corollary:

Corollary 5.8. *If X is simply-connected and $\Sigma\cdot\Sigma$ is square-free, then the fundamental group of $X\backslash\Sigma$ has no non-trivial representations in* SU(2).

Proof. Write n for the self-intersection number. Suppose there were a non-trivial representation, and let A be the corresponding flat connection. Consider the restriction of A to Y_ε and the holonomy around the circle fibre. Since the conjugacy class of the circle generates the fundamental group of $X\backslash\Sigma$, the holonomy cannot be 1. Nor can it be -1 since the class of Σ cannot be a multiple of 2. So the holonomy is

$$\exp 2\pi i \begin{pmatrix} -\alpha & 0 \\ 0 & \alpha \end{pmatrix},$$

with $\alpha \in (0, \frac{1}{2})$. Since the circle's class is central in $\pi_1(Y_\varepsilon)$ and its nth power is a product of commutators, the representation must be reducible on Y_ε and α must be of the form a/n [T2]. We can regard A now as one of our model connections corresponding to some pair (k, l). The formula for the cohomology class of the form ω in the proof of the previous proposition gives

$$\langle \omega, \Sigma \rangle = -l + \alpha \Sigma\cdot\Sigma,$$

and since $\omega = 0$ (the bundle being flat) we see that $l = a$. Since the action is zero, the Chern-Weil formula gives

$$\begin{aligned} k &= -2\alpha l + \alpha^2 \Sigma\cdot\Sigma \\ &= -a^2/n. \end{aligned}$$

So a^2/n is an integer, which is impossible if n is square-free and $a < n$. □

Some of the arithmetic in this proof is worth extracting as a separate proposition, for future reference.

Proposition 5.9. *Let the self-intersection number of Σ be $n \neq 0$ and let A be a flat α-twisted connection. Then the holonomy parameter α is of the form a/n and the instanton and monopole numbers are given by*

$$l = a$$
$$k = -a^2/n$$

If on the other hand $\Sigma \cdot \Sigma$ is zero, then k and l are zero also. □

Remark. The result (5.8) is also true for SO(3) representations. The authors know of no example of an embedded surface satisfying the hypotheses of the theorem for which the fundamental group of the complement is non-trivial.

As another application of the Chern-Weil formula, we can tie up a loose end from section 2(iv):

Proof of Lemma 2.12. The original connection A and its half-twisted companion $A' = A \otimes \xi$ have the same curvature, so from the Chern-Weil formula it follows that the invariants (α', k', l') of A' are related to the invariants of A by

$$k + 2\alpha l - \alpha^2 \Sigma \cdot \Sigma = k' + 2\alpha' l' - (\alpha')^2 \Sigma \cdot \Sigma.$$

We have already noted the relation $\alpha' = \frac{1}{2} - \alpha$. As for k' and l', since the construction can be made continuously as α varies in $(0, \frac{1}{2})$, they must be independent of α. We can therefore substitute $\alpha' = \frac{1}{2} - \alpha$ in the formula above and then equate coefficients of α and the constant term. This gives

$$2l = -2l' + \Sigma \cdot \Sigma$$

and

$$k = k' + l' - \tfrac{1}{4} \Sigma \cdot \Sigma,$$

which are equivalent to the formulae in the lemma. □

(iii) Transversality.

We now establish the transversality results (2.6), sketching an appropriate variant of the usual proof [FU]. Details for this version of the argument are in [DK]. Let U be an open domain in X whose closure is disjoint from Σ. Let $\bar{g}$ be a fixed, smooth metric on X and let $\mathcal{C}$ be the space of all conformal classes of C^r metrics which differ from $\bar{g}$ only in the open set U. The integer r is chosen to be larger (by 2) than the degree of differentiability occuring in any of the Sobolev spaces we use for the gauge theory. The space $\mathcal{C}$ can be parametrized by an open subset of the Banach space $C^r(U, \mathrm{Hom}(\Lambda^-, \Lambda^+))$. Build now the *parametrized moduli space*

$$\mathcal{M}^{\alpha}_{k,l} \subset \mathcal{B}^{\alpha} \times \mathcal{C}$$

whose fibre over the conformal structure $[g]$ is the moduli space of anti-self-dual solutions $M^{\alpha}_{k,l}$ with respect to g. We also write $\mathcal{M}^*$ for the space of irreducible, non-flat solutions. If we show that $\mathcal{M}^*$ is a Banach manifold, then it will follow

that the fibre M^* is smooth for a generic g, by the Sard-Smale theorem, since the projection to $\mathcal{C}$ is Fredholm.

If $\mathcal{M}$ fails to be smooth at a point $([A],[g])$ then the usual line of argument shows that there must be a harmonic form $h \in H^2_A$ whose image, regarded as a bundle map from Λ^+ to $\mathfrak{g}_E$, is pointwise orthogonal to the image of the curvature on the open set U. From this it follows that A is locally reducible on U, and so must either be flat or globally reducible by a unique continuation argument [DK].

This completes the proof of part (i) of (2.6). Part (ii), concerning the reducible solutions also follows the standard proof [D2], once one has understood a necessary condition for the existence of a reducible solution in the presence of the twist. This condition can be read from the proof of (5.7). If a line bundle $\bar{L}$ on X admits an anti-self-dual α-twisted connection, then there must be a smooth, harmonic, anti-self-dual 2-form ω whose cohomology class represents

$$c_1(\bar{L}) + \alpha[\Sigma].$$

In particular, the translate of the integer lattice

$$H^2(X,\mathbb{Z}) + \alpha[\Sigma] \subset H^2(X,\mathbb{R})$$

must meet the harmonic space $\mathcal{H}^- \subset H^2$. The usual argument shows that, if b^+ is positive, this will not happen for a generic conformal class in $\mathcal{C}$, unless $\alpha[\Sigma]$ is an integer class, in which case the two sets intersect at 0. In this case, there may be a flat α-twisted connection on L. Note, however, that there is still a global topological obstruction to the existence of such a connection; it is necessary that the homology class of Σ be divisible.

Given these results, the additional arguments needed to deal with the extended moduli spaces are entirely formal. We introduce the parametrized extended moduli space $\hat{\mathcal{M}}$ as the zero set of a map

$$\hat{\mathcal{B}} \times \mathcal{C} \to \mathcal{F}.$$

That the derivative of this map is surjective at all points of $\hat{\mathcal{M}}$ follows from the corresponding statement for $\mathcal{M}$. So the parametrized extended moduli space is a Banach manifold, except at reducible or flat solutions. The extended moduli space $\hat{M}$ is a fibre of the Fredholm map $\hat{\mathcal{M}} \to \mathcal{C}$, and the Sard-Smale theorem tells us that the generic fibre is smooth. All this should be done first for a fixed compact subinterval $I \subset (0, \frac{1}{2})$; take the union over a countable increasing sequence of intervals to deduce the smoothness of $\hat{M}^*$ for a generic metric. Note that we do *not* assert that the projection map $\hat{M}^* \to (0, \frac{1}{2})$ will be a submersion: even though $\hat{M}^*$ will be smooth generically, we expect the holonomy parameter α to have critical points as a function on $\hat{M}^*$.

In order that $\hat{M}$ contains a reducible solution (other than a flat one) it is necessary that $\mathcal{H}^-$ should meet a 1-parameter family of translates of the integer lattice somewhere other than at 0. The spaces $\mathcal{H}^- \subset H^2$ meeting this condition are a subset K of the Grassmanian of codimension $(b^+ - 1)$ (a countable union

of submanifolds). If this codimension is positive, then for a generic metric there will be no solution. Indeed, the set of bad conformal classes in $\mathcal{C}$ has the same codimension, because the map

$$\begin{aligned} \mathcal{C} &\to \mathrm{Gr}_{b^-}(H^2) \\ [g] &\mapsto \mathcal{H}^-(g) \end{aligned}$$

is transverse to K (see [D2,DK]).

We have now completed the proofs of the transversality results (2.6) and (2.8). Bringing in the result of (5.7) we can state that, if X is simply connected, b^+ is at least 2 and the self-intersection number of Σ is square-free, then for a generic metric the moduli spaces $\hat{M}_{k,l}$ are all smooth, and are cut out transversely by the equations.

(iv) Stronger norms on the orbifold.

Let k be an integer not less than 2, let I be a compact sub-interval of $(0, \frac{1}{2})$, and let ν be a large integer, greater than the lower bound ν_0 which appears in Proposition 4.17. Put an orbifold metric g^ν on X with a cone-angle $2\pi/\nu$ along Σ, let A^α be a model α-twisted connection, with $\alpha \in I$, and consider the space of connections

$$\check{\mathcal{A}}^\alpha = A^\alpha + \check{L}^2_{k,A^\alpha}(\mathfrak{g}_E),$$

the gauge group

$$\check{\mathcal{G}} = \mathcal{G} \cap \check{L}^2_{k+1,A^\alpha}(\mathrm{End}(E))$$

and the space of curvature tensors

$$\check{\mathcal{F}} = \check{L}^2_{k-1,A^\alpha}(\mathfrak{g}_E).$$

All the higher Sobolev norms here are defined using the Levi-Civita derivative of g^ν, as in section 4(v). For $A \in \check{\mathcal{A}}^\alpha$, we consider the operators

$$T\check{\mathcal{G}} \xrightarrow{d_A} T\check{\mathcal{A}}^\alpha \xrightarrow{d_A^+} \check{\mathcal{F}},$$

and the first-order elliptic operator $\mathcal{D}_A$, acting in the topologies $\check{L}^2_{k,A^\alpha} \to \check{L}^2_{k-1,A^\alpha}$.

Starting with (4.17) we can prove that the operator $\mathcal{D}_A$ and its formal adjoint are Fredholm, and that the usual statements of elliptic regularity and the Fredholm alternative hold. The image of d_A is closed, and a transversal is provided by the slice

$$T = \mathrm{Ker}\, d_A^*.$$

From this it follows as usual that the quotient space $\check{\mathcal{B}}^\alpha = \check{\mathcal{A}}^\alpha/\mathcal{G}$ is a Banach manifold except at the reducible connections, with local charts provided by the slices $A+T$. The moduli space of anti-self-dual connections is finite-dimensional and modeled in the usual way in terms of the cohomology groups of the deformation complex. All this can follow the exposition of [DK] without much alteration,

and transversality results need no modification of the proof given above. We also have the extended moduli spaces $\hat{M}$ in this set-up, but only over the compact interval I: we cannot take an increasing sequence of intervals without taking a decreasing sequence of cone-angles, and changing the cone-angle changes the moduli space. To summarize these statements, we give a proposition.

Proposition 5.10. *For $\nu > \nu_0$, depending on I, the extended moduli spaces $\hat{M}$ over the interval I, using a metric g^ν with cone-angle $2\pi/\nu$, are finite-dimensional spaces. The local model at an irreducible connection $[A]$ is the zero-set of a smooth map*

$$\varphi : H^1_A \oplus \mathbb{R} \to H^2_A,$$

where the H^i_A are the cohomology groups of the deformation complex above. If X is simply-connected, b^+ is at least 2 and the self-intersection number of Σ is square-free, then the extended moduli space is cut out transversely by the equations and contains no reducible or flat solutions for a generic choice of metric g^ν. □

Of course, we do not have to use the stronger Sobolev spaces to construct the moduli spaces. As we said in the introduction to this section, we can stay with the 'L^p_1' framwork, rather than use the orbifold norms. The advantage is that, following the line of 4(iv), we can also construct moduli spaces for arbitrary irrational cone-parameter $\tilde{\nu}$, and arbitrary holonomy $\alpha \in (0, \frac{1}{2})$. For integer values of $\tilde{\nu}$ and a restricted range of α, we have a choice between the two Banach space frameworks. This leaves us with the business of showing that, in such cases, the two moduli spaces coincide.

In the case that ν is an integer, let us temporarily write M for the moduli space as described in parts (i)–(iii) of this section, and $\check{M}$ for the moduli space described in this part, using the orbifold norms. The norms defining $\check{M}$ are stronger, so there is a continuous map from $\check{M}$ to M. What is needed is a regularity result, to show that an anti-self-dual connection in M is gauge equivalent to one in $\check{M}$.

So let $A^\alpha + a$ represent an element of M, so that $\nabla_{A^\alpha} a \in L^p$, with ∇ the Levi Civita derivative for the round metric. By the embedding theorem (3.3), a is in L^q for some $q > 4$. If we go over to the orbifold metric and the corresponding derivative $\check{\nabla}$, it may be that the covariant derivative $\check{\nabla}_{A^\alpha} a$ is not in L^p, so a is not in $\check{L}^p_{1,A^\alpha}$. But we can still say that a is in L^q and that the exterior derivative $d_{A^\alpha} a$ is in L^p. We wish to find a gauge equivalent connection $A^\alpha + a'$ with a' in $\check{L}^p_{1,A^\alpha}$. The proof that we can achieve this is the usual combination of local elliptic theory and a patching argument. The local input is provided by the Hodge theory for the de Rham complex on the orbifold (see section 4(vi)), which tells us that if a is a 1-form in L^q whose exterior derivative is in L^p, then there is a 0-form χ of class $\check{L}^q_1$ such that $a - d\chi$ is in $\check{L}^p_1$.

Once one has a' in $\check{L}^p_{1,A^\alpha}$, it is a routine matter to show that the connection is equivalent to one in any of the higher Sobolev spaces on the orbifold, using the regularity results of section 4(v), just as one does in the usual theory.

Remark. The reason that some work had to be done at this point is that the L^p_1 norms on 1-forms are not the same in the orbifold setting as for the round metric, because the Levi-Civita connection is involved. It is in fact only on the diagonal components of the connection matrix that these norms differ, since on the off-diagonal components both are equivalent to the W^p_1 norm. In [W], based on the construction in [DS], gauge theory is developed for cone-like metrics of the sort we are considering, using connection matrices a in $L^{4+\epsilon}$ whose exterior derivatives (rather than full covariant derivatives) are in $L^{2+\delta}$. Such a space of connections has the advantage of being unaltered by the change from a smooth to a cone-like metric, and we have exploited this in the argument given above. These spaces would presumably provide another framework in which the present theory could be developed.

6. The index formula

(i) Setting up the index problem.

In this section we take up the dimension formula (2.5)(iii). So let A be an anti-self-dual twisted connection in one of our moduli spaces $M^\alpha_{k,l} = M^\alpha_{k,l}(X, \Sigma)$, associated to a pair (X, Σ). The quantity we wish to calculate is the alternating sum

$$\dim H^1_A - \dim H^0_A - \dim H^2_A \tag{6.1}$$

of the dimensions of the cohomology groups of the complex (4.10), and we wish to obtain the answer

$$8k + 4l - 3(b^+ - b^1 + 1) - (2g - 2). \tag{6.2}$$

In the usual development of the theory for anti-self-dual moduli spaces on a closed manifold, one introduces the formal adjoint d^* of the first operator d in the complex, and one interprets (6.2) as the index of the operator $\mathcal{D}_A = d^*_A + d^+_A$ acting on $\Omega^1(\mathfrak{g}_E)$. The advantage gained is that the operator $\mathcal{D}_A$ is defined whether or not A is anti-self-dual, whereas the cohomology groups in (6.1) are not; this allows one the flexibility needed to calculate the index by exploiting deformation-invariance and excision arguments.

In our theory, at least for the set-up with the smooth rather than the orbifold metric, the operator d^* is not available globally on X. The same point arose in [DS], and we borrow from there a device to avoid the problem. Let A be any connection in $\mathcal{A}^\alpha$ and let Q_A be the right-inverse to d^+_A constructed in (5.4). (The operator d^+_A is for the smooth metric, for the moment, and we use the L^p spaces). As in the proof of (5.5), we introduce the operator

$$\delta_A : T\mathcal{G} \to T\mathcal{A}^\alpha$$

defined by $\delta_A = d_A - Q_A F^+_A$. So the sequence

$$T\mathcal{G} \xrightarrow{\delta_A} T\mathcal{A}^\alpha \xrightarrow{d^+_A} \mathcal{F} \tag{6.3}$$

is a complex. Let $H^i_{A,Q}$ be the cohomology groups; they depend on Q.

Proposition 6.4. *The image of δ_A is closed and the cohomology groups $H^i_{A,Q}$ are finite-dimensional.*

Proof. The argument of [DS] (Proposition 4.11) applies; the difficult case is the finite-dimensionality of H^1, which we sketched in the proof of (5.5). □

We now define a quantity i as minus the Euler characteristic of the complex (6.3); we write

$$i_{k,l}(X, \Sigma) = \dim H^1_{A,Q} - \dim H^0_{A,Q} - \dim H^2_{A,Q}.$$

This integer is independent of the choice of Q first of all. Indeed, if we have two different choices we may interpolate linearly between them and so have a family of complexes of the form (6.3), with δ_A varying continuously in operator norm; in such a situation the Euler characteristic is constant in the family.

The integer i is also independent of the choice of A in $\mathcal{A}^\alpha$. Given a family of connections A_t, we would like to construct a continuous family of right-inverses Q_{A_t} and hence a continuous family of operators δ_{A_t} to deduce the invariance of the Euler characteristic. This is not quite possible, since the dimension of H^2_A need not be constant in the family. This point can be overcome by the usual procedure of stabilization. We replace the space $T\mathcal{A}^\alpha$ with $T\mathcal{A}^\alpha \oplus \mathbb{R}^N$ for some large N, and replace d^+_A with $d^+_A \oplus \psi$; the operator ψ and the integer N are chosen so as to make the sum surjective for all A_t in the family. Using a right-inverse then for the modified operators, one deduces the constancy of i quite formally; see [DS]. This argument only depends on the fact that the operators d_A and d^+_A are varying continuously in operator norm as A varies, and on the existence of the parametrix for d^+_A.

There are two other parameters which we can deform without affecting i. First, we can vary the holonomy parameter α and consider a family $A_t \in \mathcal{A}^{\alpha_t}$, varying continuously in the total space $\hat{\mathcal{A}}$. Our basic observation, Lemma 3.5, says that the operators d_{A_t} and $d^+_{A_t}$ vary continuously in operator norm as the holonomy changes, so the index i is independent of α (recall that the spaces $T\mathcal{A}^\alpha$ etc. are independent of the holonomy). Second, we can vary the cone-angle of the metric; recall from 4(iv) that the Fredholm theory continues to work with the same L^p spaces for metrics $g^{\tilde{\nu}}$ with cone-angle $2\pi/\tilde{\nu}$ for arbitrary $\tilde{\nu} \geq 1$. As the metric varies in such a family, the operators d^+_A vary. As in section 4(iv), these d^+ operators are equivalent to a family

$$d^+_A - \tilde{\mu} d^-_A$$

for a varying $\tilde{\mu}$ representing the change in conformal structure. As the cone-angle varies, $\tilde{\mu}$ varies continuously in $C^0(X \backslash \Sigma)$, so these operators vary continuously in operator norm. Again we deduce that the index is invariant.

Exploiting these last two properties, we can reduce the calculation to the case where the cone-angle is $2\pi/\nu$ for some integer ν, and the holonomy parameter α is a multiple a/ν. In this setting the analysis is of orbifold type; that is, the

manifold is locally a quotient and the operators are the usual ones (without holonomy after a gauge transformation) on the local covering spaces. This gives two advantages. The first advantage is a technical one. We can introduce the formal adjoint d_A^* of the operator d_A and the usual first-order elliptic operator $\mathcal{D}_A$; the index i is the index of this operator, as usual. We can also suppose that A is smooth in the orbifold sense (i.e. is smooth on the local coverings), and we can consider the operators acting on smooth orbifold sections, for we have the full range of elliptic regularity. The second advantage is more on the formal side: for orbifolds, one knows how to calculate the index, at least when the orbifold is globally a quotient, for one only need examine the invariant part of the cohomology of the complex on the smooth cover.

We will prove the index formula using excision to establish the general shape of the answer and then examining some special cases, global quotients by finite groups, to obtain the correct numerical coefficients.

(ii) Excision: the shape of the index formula.

Let us temporarily introduce the notation $i_k(X)$ for the index of the usual deformation complex (4.10) in the absence of any surface with twist, so [DK]

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

The result to be proved can then be written

$$i_{k,l}(X, \Sigma) - i_k(X) = 4l - (2g - 2). \tag{6.5}$$

The first point is that the difference on the left necessarily depends only on l and the geometry of the tubular neighbourhood N of Σ. That is, if $(X_i, \Sigma_i, k_i, l_i)$ are two such situations $(i = 1, 2)$, then we have:

Lemma 6.6. *The two differences $i_{k_i,l_i}(X_i, \Sigma_i) - i_{k_i}(X_i)$ are the same provided that $l_1 = l_2$ and that the surfaces Σ_1, Σ_2 have diffeomorphic neighbourhoods (which will be the case as long as they have the same genus and self-intersection number).*

The proof is by excision, as we now outline (see [DK] for a more detailed exposition). Suppose a 4-manifold X is written as the union of two open sets U and V, and let E be a vector bundle over X carrying a connection A. Suppose that $X' = U' \cup V'$ is another such manifold with a bundle E' and connection A', and consider the difference $i - i'$ between the indices of the corresponding operators $\mathcal{D}_A$ and $\mathcal{D}_{A'}$ in the two settings. The statement of the excision principle (due to Atiyah and Singer) in this particular setting is that, if there is an isometry between V and V' which lifts to a bundle isomorphism ψ intertwining the connections A and A', then the difference $i - i'$ depends only on the data over U and U'. (This data comprises the sets U, U' themselves, the bundles $E|_U$ and $E'|_{U'}$ with their connections, and the restriction of ψ to $U \cap V$.) A proof which stays within the framework of differential operators is given in [DK], and examination

of the argument shows that neither the statement of the result nor the proof need any modification if X and X' are orbifolds rather than manifolds. In view of the remarks in the previous subsection, the excision principle can therefore be applied to the calculation of $i_{k,l}(X, \Sigma)$.

Proof of Lemma 6.6. Given a pair (X, Σ) with an orbifold metric g^ν, write X as $U \cup V$ with V the complement of the closed tubular neighbourhood $N \supset \Sigma$ and U a neighbourhood of N. Let E be a bundle carrying a twisted connection A with holonomy parameter $\alpha = a/\nu$, regarded as an orbifold bundle, and with invariants (k, l). Let X' be the same 4-manifold, but equipped with a metric which is smooth near Σ, and let E' be a bundle with $c_2 = k$, equipped with a smooth $SU(2)$ connection A'. To be quite specific, we can refer to the construction of section 2(i), and for E' we can take the bundle we there called $\bar{E}$ and for A' we take a smooth connection which is reducible on N (such as the connection $\bar{A}^0$ in 2(i)). Then for A we can take the model connection A^α of (2.1). The construction of A^α is by a standard modification within U, so the excision principle applies. □

We have dwelt on the excision principle because we have other applications for it. We consider next a procedure for increasing the monopole number l by 1. Over the pair (S^4, S^2), let E_1 be a bundle carrying an α-twisted connection A_1, and choose these with invariants $l = 1$ and $k = 0$. Suppose that A_1 is flat near a point $p \in S^2$, and choose a local trivialization there in which A_1 coincides with the model connection A^α. Identify $S^4 \backslash \{p\}$ with $\mathbb{R}^4$, and apply a dilation to obtain a twisted connection over $(\mathbb{R}^4, \mathbb{R}^2)$ whose curvature is supported in a small ball B around the origin and which coincides with a model A^α outside the small ball. Now let (X, Σ) be any other pair, carrying a twisted connection A, living on a bundle E. Let U be a neighbourhood of a point $q \in \Sigma$ and suppose that A is flat in U, so that in some trivialization the connection is the same model A^α. Now construct a new bundle E' with connection A' by removing a small ball around q and replacing it with the standard ball about 0 in $\mathbb{R}^4$; the bundle E' is built from E and E_1, which are identified using trivializations on the edge of the ball in which the connections are standard. The result is to change the topological type, increasing l by one while leaving k unchanged (we will return to this construction in section 7(iv)).

Applying the excision principle to this situation shows that, when l is increased by one, the index changes by a quantity which is independent of (X, Σ) and (k, l). It follows that l enters linearly with a constant coefficient in the dimension formula, which therefore takes the form

$$i_{k,l}(X, \Sigma) = i_k(X) + Cl + D(g, (\Sigma \cdot \Sigma)),$$

where C is a constant and D is an unknown function of the genus and self-intersection number.

Next consider the following operation. Again, we suppose we have a pair (X, Σ) carrying a twisted connection A which is flat in some neighbourhood U of a point $q \in \Sigma$. Take a small ball Z inside U and disjoint from Σ. Inside Z, let

T be a torus in some standard embedding. Modify the surface Σ so as to make a new surface Σ' by joining Σ to T with a standard pipe contained in U. Using the flat structure, we modify the connection A in a standard way, to obtain a connection A' twisted along the new surface Σ', without affecting k or l. The effect is to increase the genus g by one, and by the excision principle, we deduce that the genus g enters linearly in the formula, with a constant coefficient.

As a final application of excision, we consider the self-intersection number. Let Z be a small ball near to Σ, as in the previous paragraph, and modify X by taking a connected sum with $\mathbb{C}\mathrm{P}^2$, making the sum in the small ball. The new manifold X' has different topology, and the formula for $i_k(X)$ shows that the index drops by 3. Now consider modifying Σ by joining it to by a pipe to a standard 2-sphere S inside $\mathbb{C}\mathrm{P}^2$, with $S\cdot S = 1$. The genus is unchanged and the self-intersection number of the surface increases by 1. As in the previous paragraph, we deduce that $\Sigma\cdot\Sigma$ enters in the index formula with a constant coefficient.

(iii) Calculation of some examples.

From the results of the previous subsection we now know that the index formula must have the shape

$$\begin{aligned} i_{k,l}(X,\Sigma) &= i_k(X) + m_1 l + m_2 g + m_3(\Sigma\cdot\Sigma) + m_4 \\ &= 8k - 3(b^+ - b_1 + 1) + m_1 l + m_2 g + m_3(\Sigma\cdot\Sigma) + m_4. \end{aligned}$$

We shall obtain the unknown coefficients m_i from some examples.

As a first example, take X to be S^4 and take Σ to be a surface of genus g (with a standard embedding). On the complement of Σ there is a flat line-bundle L with holonomy α; we take $E = L \oplus L^{-1}$, with $\alpha = 1/4$. The invariants k and l are zero in this example, by (5.9). The bundle of Lie algebras $\mathfrak{g}_E$ decomposes as

$$\mathfrak{g}_E = \mathbb{R} \oplus \xi \oplus \xi,$$

where $\mathbb{R}$ is the line bundle with trivial connection, and ξ is a real line bundle with a flat connection having holonomy -1. Take an orbifold metric g^ν on (S^4, Σ), with $\nu = 2$, and let $\tilde{X}$ be the branched double cover of S^4, branched along Σ, which has a smooth metric . If we lift $\mathfrak{g}_E$ to $\tilde{X}$ then we get a trivial $\mathbb{R}^3$-bundle with a flat connection which extends over Σ; the covering transformation is acting trivially on one $\mathbb{R}$ summand and non-trivially on the other two. Let $i(\tilde{X})$ denote the index of the operator $\mathcal{D}$ on $\tilde{X}$ acting with trivial real coefficients, and let $i(\tilde{X}) = i_+ + i_-$ denote the decomposition according to the action of the covering transformation. Then we have

$$i_{0,0}(S^4, \Sigma) = i_+ + 2i_-.$$

We now need to compute the invariants of $\tilde{X}$. The Euler number is given by

$$\begin{aligned} e(\tilde{X}) &= 2e(S^4) - e(\Sigma) \\ &= 4 + (2g - 2), \end{aligned}$$

while the signature $\sigma(\tilde{X})$ is zero, because there is an orientation-reversing diffeomorphism. So we have

$$\begin{aligned} i(\tilde{X}) &= -(b^+(\tilde{X}) - b_1(\tilde{X}) + 1) \\ &= -\tfrac{1}{2}(\sigma + e) \\ &= -2 - (g-1). \end{aligned}$$

The invariant part i_+ is just the index of $\mathcal{D}$ on S^4 with trivial coefficients $\mathbb{R}$, which is -1; so i_- is equal to $-1-(g-1)$. This gives the answer

$$i_{0,0}(S^4, \Sigma) = -3 - (2g-2),$$

and we deduce that, in the index formula above, the coefficient of g is -2 and the constant term m_4 is 2.

The most interesting coefficient in the index formula is the coefficient m_1 of the monopole number l. Curiously, we can argue that m_1 is equal to 4 just on the grounds of internal consistency. Let (X, Σ) be any pair with the property that the homology class of Σ is even, so there exists on $X \backslash \Sigma$ a flat real line bundle ξ with holonomy -1 around Σ (see section 2(iv)). This is the case, for example, for the standard S^2 in S^4. Let $A \in \mathcal{A}^\alpha$ be a twisted connection, let E be the bundle which carries it, and let (k, l) be the invariants. Consider the bundle $E' = E \otimes \xi$ with the tensor product connection A'. The associated Lie algebra bundles $\mathfrak{g}_E$ and $\mathfrak{g}_{E'}$ are isomorphic, so the two indices are the same. However, the topological invariants (k', l') for E' are different, as is shown by the formulae in Lemma 2.12. Equating the indices we obtain

$$8k + m_1 l = 8k' + m_1 l',$$

and on substituting the expressions from (2.12) this becomes

$$(2m_1 - 8)l + (8 \cdot \tfrac{1}{4} - \tfrac{1}{2} m_1) \Sigma \cdot \Sigma = 0.$$

Since this formula must be an identity, the coefficient m_1 must be 4.

We now know that the formula for the index has the shape

$$i_{k,l}(X, \Sigma) = 8k + 4l - 3(b^+ - b_1 + 1) - (2g-2) + m_3(\Sigma \cdot \Sigma),$$

and all that remains is to show that the coefficient m_3 is zero. For this purpose it will be enough if we calculate one case in which $\Sigma \cdot \Sigma$ is non-zero, to verify that our answer is consistent with the vanishing of m_3. Take X to be $\mathbb{C}\mathrm{P}^2$ and let Σ be a smooth algebraic curve of degree $2p$. Take $\alpha = 1/4$. If p is even then there exists on $X \backslash \Sigma$ a flat complex line bundle L with holonomy parameter $1/4$; for E we take the flat bundle $L \oplus L^{-1}$. The invariants k and l in this situation are not zero: from (5.9) we have

$$\begin{aligned} l &= \alpha \Sigma \cdot \Sigma \\ &= p^2, \end{aligned}$$

and

$$k = -\frac{1}{4} p^2$$

The genus of Σ is given by the adjunction formula, and comes out as

$$(2g-2) = 4p^2 - 6p.$$

If the coefficient m_3 is zero, then the formula predicts that the index is

$$\begin{aligned} 8k + 4l - 3(b^+ - b_1 + 1) - (2g-2) &= -2p^2 + 4p^2 - 6 - 4p^2 + 6p \\ &= -2p^2 + 6p - 6. \end{aligned}$$

We shall verify this answer. Put a cone-like metric g^ν on $\mathbb{C}\mathrm{P}^2$ with $\nu = 2$, so that the metric on the branched double cover $\tilde{X}$ is smooth. Then, as with our example of the sphere above, the index we wish to calculate can be expressed as $i_+ + 2i_-$, where

$$i(\tilde{X}) = i_+ + i_-$$

is the decomposition of the ordinary index on $\tilde{X}$ (with trivial real coefficients) into the parts corresponding to the ± 1 eigenspaces of the action of the covering transformation. The relevant invariants of $\tilde{X}$ are written down in [DK], from which we extract the fact that $b_1(\tilde{X}) = 0$ and

$$b^+(\tilde{X}) = p^2 - 3p + 3.$$

So the index with trivial coefficients is

$$\begin{aligned} i(\tilde{X}) &= -(b^+ - b^1 + 1) \\ &= -p^2 + 3p - 4. \end{aligned}$$

The invariant part i_+ is the index on $\mathbb{C}\mathrm{P}^2$, which is -2, so

$$\begin{aligned} i_+ &= -2 \\ i_- &= -p^2 + 3p - 2. \end{aligned}$$

Finally then, the index for the original deformation complex is

$$i_+ + 2i_- = -2p^2 + 6p - 6,$$

which is the answer we said we wanted above. This completes the proof of the index formula. □

Remark. There are other ways to prove the formula. For example, having used excision to reduce to the case of an orbifold which is globally a quotient, one can apply the general G-index theorem of [AS]. Another possibility is to use excision to reduce to the case of a Kähler orbifold, where the cohomology groups can be identified with sheaf cohomology groups associated with a bundle with parabolic structure, as outlined in [Kr]; this line will be taken up again in [KrM], where we shall use excision to compare not the indices but the moduli spaces themselves when holonomy is introduced.

In most lines of calculation, the sign of l is a worry, for it is easy to mistake $-l$ for l or $-\alpha$ for α. The proof we have given above, essentially deducing part of the index formula from the Chern-Weil formula via (2.12), provides a useful cross-check of the consistency of our sign conventions.

7. Regularity and compactness

Our next goal is a gauge fixing theorem, modeled after the basic result for connections on the ball due to Uhlenbeck. We wish to use a gauge fixing result to deduce a weak compactness result for our moduli spaces:

Proposition 7.1. *Let A_n be a sequence of twisted connections in the extended moduli space $\hat{M}_{k,l}$ over (X, Σ). Suppose that the holonomy parameters α_n for these connections converge to $\alpha \in (0, \frac{1}{2})$. Then there exists a sub-sequence, which we continue to call A_n, and gauge transformations $g_n \in \mathcal{G}$ such that the connections $g_n(A_n)$ converge, off a finite set of points $\{x_i\} \subset X$, to a connection A. The solution A extends across the finite set and defines a point in a moduli space $M^{\alpha}_{k',l'}$.*

Further, we can assign to each point x_i a pair of integers (k_i, l_i), with $l_i = 0$ unless $x_i \in \Sigma$, such that the curvature densities converge

$$|F_{A_n}|^2 \to |F_A|^2 + 8\pi^2 \sum (k_i + 2\alpha l_i)\delta_{x_i}$$

in the sense of measures. In this case also,

$$k = k' + \sum k_i$$
$$l = l' + \sum l_i.$$

This proposition is valid either for the L^p-based moduli spaces with the smooth metric, or for the moduli spaces of Proposition 5.10 with the orbifold metric, provided that α is in the compact interval I. It remains valid also if each A_n is anti-self-dual with respect to a different metric in a sequence which converges to some limit metric g.

A few words of explanation are needed here. Given a sequence of anti-self-dual connections of the sort described in the proposition, we can restrict our attention to $X \backslash \Sigma$, where the connections will be smooth and of bounded action, and by a simple application of Uhlenbeck's theorem (covering $X \backslash \Sigma$ by a countable collection of open sets) we can prove that there is a subsequence which converges on compact subsets of $X \backslash \Sigma$, except perhaps at a finite set of points, to some anti-self-dual connection A. Such an argument applies quite generally on a non-compact manifold, but it is not enough for our purposes. First of all, this line tells us nothing much about the limit A, except that it has finite action, and we would need to show that A had sufficient regularity along Σ to belong to our moduli spaces. Second, we would not have obtained a uniform understanding of how the A_n were approaching A near Σ.

The proposition above is giving us stronger convergence. It asserts that if $U \subset X$ is an open coordinate chart meeting Σ but not containing one of the points x_i then, in some gauge, the restriction of the A_n to U converge in the topology which defines $\mathcal{A}^{\alpha}$; that is, we can write

$$A_n = A^{\alpha_n} + a_n, \quad \text{on } U$$

where the A^{α_n} are the fixed model connections and the a_n converge strongly in L^p_{1,A^α} to a limit a (or converge in $\check{L}^2_{k-1,A}$ if we are using the orbifold moduli spaces of section 5(iv)).

To prove such a result, what is needed is a gauge fixing result for the connections $A^{\alpha_n} + a_n \in \mathcal{A}$ in a chart such as U; we need to know that if the curvature is small then we can find a gauge in which a_n is small in L^p_{1,A^α}. This is the nature of the gauge fixing theorem of the next section. Once it has been set up in this way, and given the local elliptic theory of section 4, the proof turns out to be a straightforward adaptation of the usual procedures. Note that this is a rather different, and more elementary, sort of theorem than the type of gauge-fixing discussed in [SS].

As the proposition says, the results can be developed both for the L^p framework and smooth metrics on X and for the orbifold framework, with the stronger norms. We shall treat only the first case in any detail, since the arguments are not much different in the technically easier orbifold case. The exposition follows the treatment in [DK] very closely.

(i) Gauge fixing.

Let B^4 be the standard 4-ball, with B^2 as a standard embedded surface. We identify (B^4, B^2) with the upper hemi-spheres in (S^4, S^2); here S^4 is a 4-sphere with the standard round metric $\bar{g}$. On B^4, the metric $\bar{g}$ is conformal to the flat one, by a bounded conformal factor, and defines Sobolev norms equivalent to those of the flat metric. Let $\alpha \in (0, \frac{1}{2})$ be given, and let A^α be the standard flat SU(2) connection on $S^4 \backslash S^2$ (and by restriction on $B^4 \backslash B^2$) with holonomy parameter α. This we regard as a twisted connection on the trivial bundle E over the manifold. Let $\alpha_0 = a/\nu$ be a nearby rational number, so that $|\alpha - \alpha_0|$ satisfies the bound of Proposition 4.8. As always, we suppose that $2 < p < 2 + \eta(\nu)$.

Let

$$d' : \Omega^1(\mathfrak{g}_E) \to \Omega^0(\mathfrak{g}_E)$$

be the operator (4.13), and let $T = T_{\nu,a}$ be the kernel of d'. This T is the slice which appears in Proposition 4.11. By restriction, d' is an operator also over B^4. Let $\tilde{B}^4$ be any slightly smaller ball. The gauge fixing result is the following:

Proposition 7.2. *Let $A = A^\alpha + a$ be a twisted connection over (B^4, B^2), with $a \in L^p_{1,A^\alpha}$. There are constants κ_1, M and M', independent of A, such that if the curvature of A satisfies*

$$\|F_A\|^2_{L^2} < \kappa_1^2$$

then there is a gauge-equivalent connection $\tilde{A} = A^\alpha + \tilde{a}$ satisfying the following conditions on the smaller ball $\tilde{B}^4$:

(i) $d'\tilde{a} = 0$;

(ii) $\|\tilde{a}\|_{L^2_{1,A^\alpha}(\tilde{B}^4)} \leq M \|F_A\|_{L^2(B^4)}$;

(iii) $\|\tilde{a}\|_{L^p_{1,A^\alpha}(\tilde{B}^4)} \leq M' \|F_A\|_{L^p(B^4)}$.

The gauge transformation which achieves this is of class L^p_{2,A^α}. All the constants, and κ_1 in particular, depend on α.

We prove the result first in the case that $A = A^\alpha + a$ with a smooth and the off-diagonal part of a vanishing near B^2. (Such a are dense in $\mathcal{A}^\alpha$). The proof is by the method of continuity. Let $m_t : B^4 \to B^4$ be the map $x \mapsto tx$, and write $A_t = m_t^*(A)$. Since m_t respects A^α, we can write $A_t = A^\alpha + a_t$, and the a_t tend to zero in L^p_{1,A^α} as t goes to zero, while $A_1 = A$. The A_t satisfy the same L^2 curvature bound as A. As in [DK], we find it convenient to transfer these connections to the 4-sphere. So let $p : (S^4, S^2) \to (B^4, B^2)$ be the map which is the identity on the upper hemi-sphere and maps the lower to the upper. Approximate p by a smooth map p_ε, equal to p outside the ε-neighbourhood of the equatorial 3-sphere and with ∇p_ε uniformly bounded, and pull back A_t by this p_ε. We can arrange that $p_\varepsilon^*(A^\alpha) = A^\alpha$, since this is true of p, and so write the resulting connection on S^4 as $A^\alpha + r_\varepsilon^*(a_t)$. As in [DK], the squared norm of the curvature on S^4 will be below $2\kappa_1^2$ once ε is sufficiently small. (This step, as presented in [DK], uses the fact that F_A is in L^∞; this is the only reason we restrict our attention, for the moment, to the case of smooth a).

By the two devices of the previous paragraph, we come to consider a family of connections over (S^4, S^2), which we shall continue to call A_t. We shall prove that we can find gauge-equivalent connections $\tilde{A}_t = A^\alpha + \tilde{a}_t$ for which the two conditions of the proposition hold on S^4: that is, we find $g_t \in \mathcal{G}$ such that $g_t(A_t) = A^\alpha + \tilde{a}_t$ on S^4 with

(*i*) $d'\tilde{a}_t = 0$,

(*ii*) $\|\tilde{a}_t\|_{L^2_{1,A^\alpha}} \leq M\|F_A\|_{L^2}$,

(*iii*) $\|\tilde{a}_t\|_{L^p_{1,A^\alpha}} \leq M'\|F_A\|_{L^p}$.

Lemma 7.3. *Let $A = A^\alpha + a$ be a connection on S^4 with $d'a = 0$. There are constants κ_2, M and M', independent of a, such that if $\|a\|_{L^4} < \kappa_2$ then*

$$\|a\|_{L^2_{1,A^\alpha}} \leq M\|F_A\|_{L^2},$$
$$\|a\|_{L^p_{1,A^\alpha}} \leq M'\|F_A\|_{L^p}.$$

Proof. This is the usual re-arrangement argument. Since $d'a = 0$, the 1-form a is in the image of the operator Q given in (4.11). So we have

$$a = Q(F_A^+ - (a \wedge a)^+).$$

Since Q maps L^2 to L^2_{1,A^α}, we have

$$\|a\|_{L^2_{1,A^\alpha}} \leq C_1(\|F\|_2 + C_2\|a\|_4\|a\|_{L^2_{1,A^\alpha}}).$$

This gives the result for L^2, with $M = C_1/2$, once $\kappa_2 < 1/(2C_2)$. A simple bootstrapping argument gives the L^p estimate. □

We now go back to the family of connections $A^\alpha + a_t$ on S^4 and seek gauge transformations g_t as above. We show that the set of t for which g_t exists satisfying (i) and (ii) is open and closed, provided κ_1 is small. This is enough, since $a_0 = 0$ and $g_0 = 1$ will do here.

Openness. By (4.7), the image if d_{A^α} is a complement to $\operatorname{Ker} d'$ in the L^p_{1,A^α} topology. We want first to show that this is still true if we replace d_{A^α} by d_A, for some $A = A^\alpha + a \in \mathcal{A}$ on (S^4, S^2) with $\|a\|$ sufficiently small in L^2_{1,A^α}. Certainly, with a sufficiently small in this way, the operator

$$d_A : L^2_{2,A^\alpha} \to L^2_{1,A^\alpha}$$

has image which is a complement to $\operatorname{Ker} d'$ in the L^2_{1,A^α} topology; this is because the operator d_A is close to d_{A^α} in operator norm for this topology. This means that given $b \in L^p_{1,A^\alpha}(S^4 \backslash S^2, \Lambda^1(\mathfrak{g}_E))$ we can solve the equation for u and ω:

$$b - d_A u = Q\omega$$

with u in L^2_{2,A^α} and ω in L^2. However, we then have

$$\omega = d^+_{A^\alpha} \circ Q\omega = d^+_{A^\alpha} b - d^+ A^\alpha([a, b]),$$

and the multiplication theorems give $\omega \in L^p$. So u is in L^p_2, and we have proved that the image of d_A is a complement to $\operatorname{Ker} d'$ also in the L^p_{1,A^α} topology.

Now suppose that we have found a g_t as above for a particular $t = s$. We may as well suppose $g_s = 1$, and we write $a_s = a$ and $A_s = A^\alpha + a$. For nearby t, we write $A_t = A_s + b$ with b small in L^p_{1,A^α}. Since a is small in L^2_{1,A^α}, a straightforward application of the implicit function theorem, based on the linear result above, allows us to find a small gauge transformation achieving condition (i) above. By Lemma 7.3, the same gauge transformation also achieves (ii), a priori, provided κ_1 was small enough: this is because (ii) is assumed to hold for a_s and gives us a bound on the L^4 norm of a_s and of the nearby $a_t = a_s + b$.

Closedness. Suppose we have solved the problem above for a sequence of times t_i converging to t. So we have $A^\alpha + \tilde{a}_{t_i}$, gauge equivalent to $A^\alpha + a_{t_i}$, satisfying (i)–(iii). The curvatures of the connections A_{t_i} are converging in L^p to the curvature of A_t, so by condition (iii) above we have a uniform L^p_{1,A^α} bound on the $\tilde{a}_{t_i}$ and hence a subsequence converging weakly to some $\tilde{a}$ in L^p_{1,A^α}. We have $d'\tilde{a} = 0$, and it is a straightforward matter to show that $A^\alpha + \tilde{a}$ is gauge-equivalent to A_t.

This completes the proof of (7.2) under the additional hypothesis that a is smooth. To remove this hypothesis, approximate a by a sequence of smooth forms a_ε, find the gauge-equivalent forms $\tilde{a}_\varepsilon$ given by (7.2), and transfer these to the 4-sphere using a cut-off function supported in $\tilde{B}^4$ and equal to 1 on a slightly smaller ball. As in the proof of closedness above, we find a weak limit to solve the problem; see [DS], p.231 for a model.

(ii) Compactness and patching constructions.

The basic gauge-fixing result (7.1) gives a local convergence result for a sequence of anti-self-dual solutions:

Proposition 7.4. *Let $A_i = A^\alpha + a_i$ be a sequence of anti-self-dual twisted connections over (B^4, B^2). There is a constant κ, independent of the sequence, such that if*

$$\|F_{A_i}\| < \kappa,$$

then we can find a gauge-equivalent sequence $A^\alpha + \tilde{a}_i$ satisfying the estimate

$$\|\tilde{a}_i\|_{L^p_{1,A^\alpha}(\tilde{B}^4)} \leq M\|F\|_{L^2}$$

on a smaller ball $\tilde{B}^4$. Further, the $\tilde{a}_i$ have a subsequence converging strongly in L^p_{1,A^α}. This proposition remains valid also in the case that we have a sequence of connections $A_i = A^{\alpha_i} + a_i$ with different holonomy parameters α_i converging to $\alpha \in (0, \frac{1}{2})$.

Proof. This is quite standard, given (7.2). See [DS] Lemma 6.4 for the correct way to deduce the estimate by transferring to the 4-sphere. The estimate is valid also for $p' > p$, as long as $p' < 2 + \eta(\nu)$, and this is why one can extract a convergent subsequence. To prove the rider, we can assume that the α_i are all close to a/ν, satisfying the bound of (4.8); then nothing else in the proof needs modification. □

The proof, as usual, adapts to cover the case that the metric on the ball is not flat but is close to the flat metric; the main point is that Proposition 4.11 is valid for metrics close to the round one, as the proof clearly shows. It also follows that the proposition above remains true if the connections are anti-self-dual with respect to metrics which converge to some near-standard metric. This will be the basis for the proof of the last rider to Proposition 7.1, but we shall not make reference to this again.

To obtain convergence results on domains other than (B^4, B^2), we need to patch together gauge transformations. The crux of the procedure, in the ordinary case (without holonomy), is given in [DK], on page 159. The lemma there extends without change to cover our situation also, so we have the following result. Let (X, Σ) be a 4-manifold with an embedded surface, A^α a model twisted connection in a bundle E, and $\Omega \subset X$ an open domain.

Proposition 7.5 ([DK], Corollary 4.4.8.) *Suppose $A_n = A^\alpha + a_n$ is a sequence of anti-self-dual connections in a bundle E over Ω, with $a_n \in L^p_{1,A^\alpha}$. Suppose that for each point $x \in \Omega$ there is a neighbourhood B of x a subsequence $\{n'\}$ and gauge transformations $h_{n'}$ defined over B such that $h_{n'}(A^\alpha + a_{n'}) - A^\alpha$ converges in L^p_{1,A^α} over B. Then there is a single subsequence $\{n''\}$ and gauge transformations $g_{n''}$ defined over all of Ω, such that $g_{n''}(A^\alpha + a_{n''}) = A^\alpha + \tilde{a}_{n''}$ with $\tilde{a}_{n''}$ converging over all of Ω.* □

We also note that the following carries over to our case (see [DK], Proposition 4.4.10). As in [DK], we say that Ω is strongly simply connected if it is covered by balls $B_1, \ldots, B_m$ such that the intersections $B_r \cap (B_1 \cup \ldots B_{r-1})$ are connected.

Proposition 7.6. *Suppose that Ω is strongly simply connected and that the model connection A^α is flat on Ω. Let $A = A^\alpha + a$ be an anti-self-dual connection. For any interior domain Ω', there are constants κ_Ω, M and M', independent of A, such that if*

$$\|F_A\|_{L^2(\Omega)} < \kappa_\Omega$$

then A is gauge-equivalent to $A^\alpha + \tilde{a}$ over Ω' with

$$\|a\|_{L^4(\Omega')} \leq M \|F_A\|_{L^2(\Omega)}$$

and

$$\|a\|_{L^p_{1,A^\alpha}(\Omega')} \leq M' \|F_A\|_{L^2(\Omega)}.$$

□

Note that in both these propositions, the gauge fixing is done on balls of Ω, not $\Omega \backslash \Sigma$. It is Ω, not $\Omega \backslash \Sigma$, which needs to be strongly simply connected in Proposition 7.6.

(iii) Removability of singularities.

The third and last ingredient we shall need before proving the basic compactness theorem (7.1) is a version of Uhlenbeck's theorem on the removal of singularities. Again, let (B^4, B^2) be the standard ball-pair, carrying the model connection A^α. We shall say that a bundle-valued form f is in $L^p_{k,A^\alpha,\mathrm{loc}}$ on a domain $\Omega \subset B^4$ if every point $x \in \Omega$ has a neighbourhood on which f is in L^p_{k,A^α}. We spell this out to emphasize that the case $x \in B^2$ is not excluded.

Proposition 7.7. *Let $A = A^\alpha + a$ be a connection matrix with a in $L^p_{1,A^\alpha,\mathrm{loc}}$ on $B^4 \backslash \{0\}$. Suppose that A is anti-self-dual and*

$$\int_{B^4 \backslash \{0\}} |F_A|^2 < \infty.$$

Then there is a gauge transformation g in $L^p_{2,A^\alpha,\mathrm{loc}}$ on $B^4 \backslash \{0\}$, such that $g(A) = A^\alpha + \tilde{a}$ and $\tilde{a}$ extends across $\{0\}$ to define a $\mathfrak{g}_E$-valued 1-form of class L^p_{1,A^α}.

Proof. The proof from [DK] (also used in [DS]) applies. We recall a sketch of the argument. Cover $B^4 \backslash \{0\}$ with overlapping conformally-equivalent 4-dimensional annuli W_n diffeomorphic to $S^3 \times I$. On the W_n, the L^2 norm of the curvature goes to zero as n goes to infinity, so use (7.6) to find a gauge transformation g_n on W_n in which $g_n(A) = A^\alpha + a_n$, with $\|a_n\|_{L^4}$ approaching zero. In such a gauge, multiply a_n by a cut-off function ψ whose gradient is supported in W_n,

and obtain a connection $A_n = A^\alpha + \psi a_n$ which extends across $\{0\}$. The L^4 bound on a_n ensures that the anti-self-duality condition is not much damaged, and $F^+_{A_n}$ goes to zero. Restricting to a smaller ball if necessary, apply (7.2) to put A_n in a good gauge, so $g(A_n) = A^\alpha + \tilde{a}_n$, with $d'\tilde{a}_n = 0$ and $\|\tilde{a}_n\|_{L^2_{1,A^\alpha}}$ bounded. Then extract a subsequence converging weakly to an L^2_{1,A^α} connection $A^\alpha + \tilde{a}$. This limit is gauge-equivalent to $A^\alpha + a$ and therefore anti-self-dual; and $d'\tilde{a} = 0$. To complete the proof, mimic the proof of Lemma 6.4 of [DS] to establish the extra regularity of the L^2_{1,A^α} solution. □

(iv) Proof of the compactness theorem.

Given the local compactness result (7.4), the patching argument (7.5) and the removability of singularities theorem (7.7), the proof of the compactness theorem (7.1) is standard. The only point we need to dwell on is the statement about the action densities, for this involves the modified Chern-Weil formula.

Let (W, Σ) be the pair $(S^3 \times I, S^1 \times I)$, thought of as the complement of one ball-pair (B^4, B^2) inside a slightly larger one, and let $\mathcal{G}_W$ be the gauge group we have associated to this pair, for the model twisted connection A^α in the trivial SU(2) bundle $\mathbb{C} \oplus \mathbb{C}$. This group contains as a dense subgroup the set of smooth gauge transformations which are diagonal near Σ. On the other hand it is contained in the group of continuous gauge transformations which are diagonal on Σ. It follows easily that all three groups have the same component group, which we can identify as

$$\begin{aligned}[\,(S^3, S^1),\ (SU(2), U(1))\,] &= [\,(S^3, S^1),\ (S^3, S^1)\,] \\ &= \mathbb{Z} \oplus \mathbb{Z}.\end{aligned}$$

To each gauge transformation we can therefore assign a pair of integers (κ, λ). Now suppose that $(\bar{E}, \bar{L})$ is a bundle-pair over (X, Σ). Pick a point $x \in \Sigma$ and a neighbourhood pair (B^4, B^2). Over this neighbourhood, choose a diagonal trivialization of the bundle pair as $\mathbb{C} \oplus \mathbb{C}$. Now construct a new bundle $(\bar{E}', \bar{L}')$ over (X, Σ) by cutting out the neighbourhood and gluing it back in using a gauge transformation on the boundary belonging to the class (κ, λ) as above. Let (k', l') be the topological invariants of $(\bar{E}', \bar{L}')$. With one choice of signs in the isomorphism above, we have

$$\begin{aligned} k &= k' + \kappa \\ l &= l' + \lambda. \end{aligned}$$

In the proof of the removability of singularities theorem, we must choose a gauge transformation on an annulus such as W; the constraint on the gauge transformation is that the connection should be close to A^α in the chosen gauge, first in the L^4 topology and then, because the equations hold, in the L^p_{1,A^α} topology also. Comparing the chosen gauge on W with the unique gauge component which extends across (B^4, B^2), we obtain a pair of integers as above, say (κ, λ). We use these to define (k_i, l_i) in the statement of (7.1) in the case that

$x_i \in \Sigma$. The formula relating the topological numbers is then a consequence of the difference formula above.

The statement about the action densities is now a formal consequence of the convergence and the following result.

Lemma 7.8. *Let $A = A^\alpha + a$ be a twisted connection in the trivial bundle over (B^4, B^2), and suppose that A is flat over $B^4 \backslash \frac{1}{2}B^4$. Let g be the gauge transformation over the annulus $B^4 \backslash \frac{1}{2}B^4$ such that $g(A) = A^\alpha$ (this gauge transformation is unique up to a constant), and let (κ, λ) be the integers associated with g. Then we have*

$$\frac{1}{8\pi^2}\int_{B^4 \backslash B^2} \mathrm{tr}(F_A \wedge F_A) = \kappa + 2\alpha\lambda.$$

Proof. Using the gauge g, attach the connection to the sphere-pair (S^4, S^2) so that it is flat outside the ball. Then apply the Chern-Weil formula (5.7). □

8. A topological bound on bubbling off

(i) Admissible pairs on the sphere.

In the situation described by the compactness theorem (7.1), we associate a pair of integers (k_i, l_i) with each point x_i at which the curvature concentrates. These pairs are not unconstrained. First of all, the quantity $k_i + 2\alpha l_i$ must be positive, because of the statement about the action density. But this is not all. Roughly speaking, the change in the integers (k, l) in the weak limit must be accounted for by solutions which live on spheres. Let us say that a pair of integers (κ, λ) is an admissible pair for (S^4, S^2) if the moduli space $M^\alpha_{\kappa,\lambda}$ is non-empty for (S^4, S^2), given the round metric. This notion depends on α a priori. For consistency, let us also say that k is an admissible integer for S^4 if there is an ordinary instanton on S^4 with charge k. Of course, this is the same as saying that k is nonnegative, though our arguments could be arranged so as not to exploit this fact.

Proposition 8.1. *In the situation of Proposition 7.1, let x_i be a point of concentration. Suppose $x_i \in \Sigma$ and (k_i, l_i) the associated pair. Then there exist pairs $(\kappa_{i,m}, \lambda_{i,m})$, admissible for (S^4, S^2), and integers $\kappa_{i,n}$, admissible for S^4, such that*

$$k_i = \sum \kappa_{i,m} + \sum \kappa_{i,n}$$
$$l_i = \sum \lambda_{i,m}.$$

This proposition is not useful until we know which pairs are admissible on (S^4, S^2). We make the following conjecture:

Conjecture 8.2. *For any given α, a pair (κ, λ) is admissible for (S^4, S^2) if and only if*

$$\kappa \geq 0 \qquad \textit{and} \qquad \kappa + \lambda \geq 0.$$

Remark. These two inequalities are the two extremes of the inequality which comes from the Chern-Weil formula as α ranges over $(0, \frac{1}{2})$. Note that they are interchanged by the transformation

$$\begin{aligned} \kappa' &= \kappa + \lambda \\ \lambda' &= -\lambda \end{aligned}$$

which is effected by twisting the bundle by -1, as in Lemma 2.12; so proving the necessity of either inequality implies them both. The conjecture is analogous to a result due to Murray [Mu] on the admissible values for the magnetic charges of non-abelian magnetic monopoles.

Proposition 8.1 will be proved below. We have stated it above with the smooth metric in mind, but it is valid also for the orbifold metric, provided we make the obvious modification to the definition of admissible by considering solutions on the orbifold sphere. We cannot prove the conjecture (8.2), but we can prove a close approximation to it in the orbifold case. Let I_ε be the compact subinterval $I = [\varepsilon, \frac{1}{2} - \varepsilon]$ inside $(0, \frac{1}{2})$. Let $\nu = \nu_\varepsilon$ be any integer greater than the lower bound provided by Proposition 4.10, so the elliptic theory works on the orbifolds with cone-angle $2\pi/\nu$ on the Sobolev spaces L^2_k, say for $k \leq 3$.

Proposition 8.3. *With ε and ν as above, a necessary condition for (κ, λ) to be admissible on (S^4, S^2) with the orbifold metric g^ν, for some $\alpha \in I_\varepsilon$, is that the following two inequalities should hold:*

$$\kappa + 2\varepsilon\lambda \geq 0 \qquad \textit{and} \qquad \kappa + (1 - 2\varepsilon)\lambda \geq 0.$$

Proof of 8.3 (assuming 8.1). First of all, for a solution A on the orbifold sphere, the obstruction space H^2_A in the deformation complex vanishes identically. The proof is the standard one, exploiting a Weitzenböck formula and the positive curvature of the sphere, but since the proof involves integration by parts, some care is needed. Suppose the obstruction space is not zero, and let ω be a harmonic 2-form representing an element of H^2_A. Our regularity results in the orbifold framework tell us that ω is of class L^2_{1,A^α} at least, so we can find a sequence ω_n of smooth forms approaching ω in L^2_{1,A^α} norm, with each ω_n compactly supported in the complement of Σ. For these forms, the Weitzenböck formula [FU] can be integrated by parts to give

$$2\|d_A\omega_n\|^2 = \|\nabla_A\omega_n\|^2 + R\|\omega_n\|^2,$$

where R is a positive constant coming from the curvature of the sphere. Since both sides of the equation are continuous functions of ω_n in the L^2_{1,A^α} topology, we can take the limit as n goes to infinity to obtain

$$0 = \|\nabla_A \omega\|^2 + R\|\omega\|^2,$$

showing that ω is zero. Note that in the framework we have been using for the *smooth* metric, we would only have ω in L^p, and the argument would not go through. (This is the only obstruction to our proving (8.2) for the smooth metric).

With the vanishing theorem out of the way, we turn to the rest of the proof. Suppose the second inequality in the proposition is false and let $J \subset I_\varepsilon$ be the set of α for which there is a solution which violates this inequality. We shall show that J is the whole of I_ε by showing that it is both open and closed. The vanishing of H^2_A shows that for any κ and λ the map from the parametrized moduli spaces $\hat{M}_{\kappa,\lambda} \to I_\varepsilon$ is a submersion, because the fibres are cut out transversely. The image of $\hat{M}_{\kappa,\lambda}$ is therefore open, and this implies the openness of the set J.

To show closedness, let A_n be a sequence of solutions in $\hat{M}_{\kappa,\lambda}$, with (κ, λ) violating the inequality, and let $\alpha \in I_\varepsilon$ be the limit of their respective holonomies α_n. We need to show that α is in J. The sequence A_n has a weak limit on the orbifold sphere, and by combining Propositions 7.1 and 8.1, we see that there are pairs (κ_i, λ_i), admissible for the holonomy α, and integers κ_j admissible for S^4 such that

$$\kappa = \sum \kappa_i + \sum \kappa_j$$
$$\lambda = \sum \lambda_i.$$

Since the κ_j are non-negative, it follows from these equalities that at least one of the pairs (κ_i, λ_i) violates the second inequality, so α is in the bad set J.

The argument above establishes, in particular, that there is an admissible pair (κ, λ) violating the second inequality with holonomy α equal to the end-point $(\frac{1}{2} - \varepsilon)$ of the interval I_ε. This gives a contradiction, since the Chern-Weil formula for (S^4, S^2) gives the action of such a solution as $\kappa + (1 - 2\varepsilon)\lambda$, and this quantity must therefore be non-negative. The proof of the first inequlity is similar, but uses the other end-point of the interval. □

Corollary 8.4. *In the situation of Proposition 7.1, for an orbifold metric g^ν, the formal dimension of the moduli space $M^\alpha_{k',l'}$ is not greater than the formal dimension of $M^\alpha_{k,l}$, and the dimension is strictly smaller if any k_i or l_i is non-zero.*

Proof. Combining the statements of (7.1) and (8.1) again, we have

$$k - k' = \sum \kappa_{i,m} + \sum \kappa_{i,n}$$
$$l - l' = \sum \lambda_{i,m},$$

where $(\kappa_{i,m}, \lambda_{i,m})$ is admissible pairs and $\kappa_{i,n}$ are non-negative integers. Adding the two inequalities in (8.3) gives $2\kappa_{i,m} + \lambda_{i,m} \geq 0$, with equality only if $\kappa_{i,m} = \lambda_{i,m} = 0$. So $2(k - k') + (l - l')$ is greater than or equal to 0, and zero only if all $\kappa_{i,m}$ $\kappa_{i,n}$ and $\lambda_{i,m}$ are zero. The difference in the formal dimensions of the moduli spaces is four times this quantity, so the result follows. □

(ii) Proof of Proposition 8.1.

The formalities of our proof are modeled on an argument used by Taubes [T1] to study minimizing sequences for the Yang-Mills functional. The technical side of the argument is rather simpler here, since we are dealing with anti-self-dual solutions. We present the argument in the L^p setting, for the smooth metric. The proof adapts quite easily to the orbifold case.

The problem is essentially a local one, so we consider a sequence of solutions on the unit 4-ball B^4. To keep the exposition simple, we suppose at first that B^4 has the standard flat metric. The setup is the following: we have a sequence of anti-self-dual twisted connections A_n over (B^4, B^2); the holonomy parameter for A_n is α_n and we suppose these converge to $\alpha \in (0, \frac{1}{2})$. We suppose that the A_n are converging weakly in the sense of (7.1) to a solution A, and that curvature is concentrating at the centre of the ball and at no other place; so the action densities $f_n = |F(A_n)|^2 d\mathrm{Vol}$ are converging in the sense of measures

$$f_n \to f_A + \mu\delta_0,$$

where f_A is the action density of A,

$$\mu = 8\pi^2(k + 2\alpha l),$$

and k and l are the integers associated with the point of concentration by the construction of section 7(iv). We aim to show that (k, l) is a sum of admissible pairs (k_i, l_i) and admissible integers k_j.

The proof of the compactness theorem shows that μ is not less than the critical constant κ_1 which appears in the gauge-fixing result (7.2). Let $N(\mu)$ be the largest integer not exceeding μ/κ_1. We shall prove the proposition by induction on $N(\mu)$. Note in passing that the total action of any non-trivial solution on (S^4, S^2) cannot be less than κ_1, for otherwise we could construct a sequence of solutions contradicting the local compactness result by applying conformal transformations.

Let τ_i be a sequence of numbers less than 1 and approaching 0, and let W_i be the shell in B^4 bounded by the 3-spheres of radii τ_i and $\tau_i/2$. Write the limit connection A in some gauge as

$$A = A^\alpha + a$$

where A^α is the model twisted connection, and arrange that the sequence τ_i is decreasing sufficiently fast that the L^p_{1,A^α} norm of a restricted to the ball $B(\tau_i)$

of radius τ_i is less than C_1/i, for some constant C_1. The action density therefore satisfies

$$\int_{B(\tau_i)} f_A \leq C_2/i.$$

Since the A_n converge strongly to A on W_i in some gauge, we can find an integer n_i and a gauge representative on W_i

$$A_{n_i} = A^{\alpha_{n_i}} + a_{n_i}$$

such that the L^p_{1,A^α} norm of a_{n_i} on W_i satisfies a bound

$$\|a_{n_i}\|_{L^p_{1,A^\alpha}(W_i)} \leq C_3/i,$$

with a similar estimate for the L^4 norm. We can also arrange that the curvature density f_{n_i} is concentrated in a much smaller ball; specifically, if V_i denotes the complement of the ball of radius τ_i/i inside B^4, we can arrange that

$$\int_{V_i} f_{n_i} \leq C_4/i.$$

Let β_i be a standard cut-off function, with derivative supported in the shell W_i and equal to 0 outside the shell and 1 inside. Using the good gauge representative above, cut off the connection A_{n_i} so as to obtain a twisted connection on all of $(\mathbb{R}^4, \mathbb{R}^2)$:

$$\tilde{A}_{n_i} = A^{\alpha_{n_i}} + \beta_i a_{n_i}.$$

The new connection is not anti-self-dual since $F^+(A_{n_i})$ will be non-zero in the shell W_i. But because the L^4 norm of the connection matrix a_{n_i} was small, the L^2 norm of F^+ is bounded by some C_5/i, and we can continue to suppose that the action density $\tilde{f}_{n_i}$ of the connection $\tilde{A}_{n_i}$ satisfies the estimate above:

$$\int_{V_i} \tilde{f}_{n_i} \leq C_4/i.$$

Now we apply a conformal transformation c_i to $(\mathbb{R}^4, \mathbb{R}^2)$ so that the action density $c_i^*(\tilde{f}_{n_i})$ of the pulled-back connection satisfies:

- (a) the centre of mass of $c_i^*(\tilde{f}_{n_i})$ lies on the plane $\Sigma^\perp$;
- (b) the amount of action lying outside the unit ball is exactly $\kappa_1/2$, where κ_1 is the critical constant of the gauge-fixing theorem.

(Here we write Σ for our standard $\mathbb{R}^2$ on which we have the twist, and we write $\Sigma^\perp$ for the complementary plane through 0.) We achieve this in two steps: first we apply a translation t_i parallel to Σ to achieve (a); then we apply a dilation d_i centred at 0 to achieve (b). The conformal transformation c_i is the composite $d_i t_i$.

The reason for arranging that the integral of $\tilde{f}_{n_i}$ is small on the domain V_i is that it ensures that the size of the translation t_i is less than $C_6\tau_i/i$ and that the factor of the dilation d_i is greater than $C_7 i/\tau_i$. This implies that, for

the connection $c_i^*(\tilde{A}_{n_i})$, the self-dual curvature F^+ is supported near infinity, outside a ball of radius $C_8 i$.

Now identify $\mathbb{R}^4$ with the complement of the south pole (the point at infinity) in S^4 by stereographic projection, mapping the unit ball to the northern hemisphere. The twisted connections $c_i^*(\tilde{A}_{n_i})$ extend across the south pole since they are flat there.

To summarize the situation so far, we have transfered the solutions from the ball to the sphere by cutting off the connection in a good gauge, and then we have applied conformal transformations to spread out the curvature on the sphere. The topological invariants of the new connections are the same pair (k, l), and their total actions approach the limit μ. If we could show that the connections $c_i^*(\tilde{A}_{n_i})$ converged strongly to an anti-self-dual twisted connection $\tilde{A}$, then we would be home, for this would establish that (k, l) was an admissible pair. Unfortunately, the curvature may still be concentrating at points, so we need to be able to apply our inductive hypothesis to complete the proof.

First, by condition (b) above, we can apply the gauge-fixing theorem (7.2) to the connections on the southern hemisphere, and by the same argument as we used for the removability of singularities theorem, some subsequence will converge there to an anti-self-dual limit; the convergence is strong in L_1^p away from the point at infinity. On an open neighbourhood of the northern hemi-sphere, $c^*(\tilde{A}_{n_i})$ are solutions of the equations, and the argument of the compactness theorem (7.1) applies. From these two facts and the patching argument, we can conclude as in (7.1) that there is a solution $\tilde{A}$ on (S^4, S^2) with action density $\tilde{f}$, and a finite set of points x_r, $(r = 1, \dots p)$, in the closed northern hemisphere, such that the action densities $c_i^*(\tilde{f}_{n_i})$ converge to a limit

$$\tilde{f} + \sum \mu_r \delta_{x_r},$$

while, in some gauge, the connections $c^*(\tilde{A}_{n_i})$ are converging to $\tilde{A}$ except at the points x_r. Further, we have

$$\mu_r = k_r + 2\alpha l_r,$$

where (k_r, l_r) is the pair of integers associated with the point of concentration $x_r \in S^4$. As in (7.1) we have

$$k = k' + \sum k_r$$
$$l = l' + \sum l_r,$$

where (k', l') are the invariants of the limit $\tilde{A}$. Finally, since the limit of the total action is μ, we have

$$\mu = \int \tilde{f} + \sum_{r=1}^{p} \mu_r.$$

To complete the proof of (8.1), suppose first either that $\tilde{f}$ is non-zero or that $p \geq 2$. Under this assumption, since each μ_r must be at least as large as κ_1, and $\int \tilde{f}$ is also bounded below by κ_1 if it is non-zero, it follows that each μ_r is less

than μ, with a difference not less than κ_1. So the integer $N(\mu_r)$ is strictly smaller than $N(\mu)$, and we can complete the proof by applying the inductive hypothesis to the points of concentration x_r.

The only remaining possibility is that $p = 1$ and $\tilde{f} = 0$. In this case, all the curvature is concentrating at a single point x_1, with total mass μ. The main point now is that x_1 cannot be on Σ. Indeed, if x_1 were on Σ then by condition (a) above it would have to be the origin $0 \in \mathbb{R}^4$ (the north pole); but this would contradict condition (b), for not all the curvature can be concentrating in the ball of radius $\frac{1}{2}$. It follows now that l is zero, and so $\mu = 8\pi^2 k$ where k is an admissible integer for S^4.

This completes the proof for the standard metric on B^4. For the general case, we suppose A_n is anti-self-dual with respect to some metric g_n, and that the g_n converge to some general g as n goes to infinity. The restriction of g_{n_i} to the ball of radius τ_i is approximately standard for large i, and the rescaled connections $c_i^*(\tilde{A}_{n_i})$ on (S^4, S^2) are therefore anti-self-dual (except in the cut-off region) with respect to metrics $\tilde{g}_i$ which converge to the round metric as i goes to infinity. Proposition 7.1 can still be applied, and the proof by induction continues as before. □

Appendix

In this paper we have not so far seen an explicit, non-trivial solution of the anti-self-duality equations with the sort of singularity which defines the moduli spaces M^α. We shall remedy this short-coming here. The first explicit solution to have appeared in print, as far as the authors are aware, is an anti-self-dual connection belonging, in our notation, to the moduli space $M^{1/4}(S^4, S^2)$, for the standard round pair; this solution was described in [FHP], and was later generalized by the same authors to give solutions with arbitrary α.

We shall write down a solution on $T^2 \times D^2$, where T^2 is a torus and D^2 is a disk, both with their standard metrics. The singular surface Σ will be the torus $T^2 \times \{0\}$. We take polar coordinates (r, θ) on D^2 and standard orthogonal coordinates (u, v) on (the univesal cover of) the torus. The bundle $\bar{E}$ will be the trivial bundle $\mathbb{C} \oplus \mathbb{C}$ with the trivial decomposition into line bundles, so that $l = 0$. In this standard trivialization, we seek a solution which is in radial gauge (so there is no dr term in the connection matrix) and which is independent of θ, u and v. This means that it can be written

$$A = f(r) \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} dr + g(r) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + h(r) \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \tag{A1}$$

for some real functions $f(r)$, $g(r)$ and $h(r)$. The anti-self-duality equations now become

$$\begin{aligned} r^{-1}(df/dr) + 2gh &= 0 \\ (dg/dr) + 2r^{-1}fh &= 0 \\ (dh/dr) + 2r^{-1}fg &= 0. \end{aligned}$$

The equations take on a more familiar form if we substitute $r = e^{-t}$, and then put $F = \frac{1}{2} - f$, $G = e^{-t}g$ and $H = e^{-t}h$; we obtain a particular reduction of *Nahm's equations*:

$$\begin{aligned} dF/dt &= -2GH \\ dG/dt &= -2HF \\ dH/dt &= -2FG. \end{aligned}$$

There is essentially only one solution which has suitable behaviour as t goes to infinity (i.e. r goes to zero); this is given by

$$\begin{aligned} F &= c\coth(2ct) \\ G = H &= c\mathrm{cosech}(2ct). \end{aligned}$$

If we put $c = \frac{1}{2} + \alpha$ and return to the original coordinates, we obtain the solution in the form (A1) with

$$\begin{aligned} f &= 1/2 - (1/2 + \alpha)\left(\frac{1 + r^{2+4\alpha}}{1 - r^{2+4\alpha}}\right) \\ g = h &= \frac{(2\alpha + 1)}{r^{-2\alpha} - r^{2+2\alpha}}. \end{aligned}$$

The solution we have obtained blows up as r approaches 1 at the boundary of the disk; but our interest is in the behaviour as r goes to zero. Here we have

$$A = \alpha \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + (2\alpha + 1) r^{2\alpha} \begin{pmatrix} 0 & d\bar{w} \\ dw & 0 \end{pmatrix} + O(r^{1+4\alpha}),$$

where $dw = du + idv$. The size of the curvature near $r = 0$ has leading term

$$|F_A| \sim \text{const.} r^{-1+2\alpha}$$

(the constant is $2\alpha(1+2\alpha)$ for the solution here), which shows that the curvature is in L^p near Σ provided that $p < 2/(1 - 2\alpha)$.

The leading term in the curvature comes from its four components in the planes which mix directions parralel to and orthogonal to Σ, such as the $dr\,du$ plane. In the planes $du\,dv$ and $dr\,d\theta$ (the planes of T^2 and D^2) the curvature approaches zero as r goes to zero. In general, this can only be the case if l is zero, but nevertheless it seems plausible to suggest that, in any event, the curvature in the planes parallel to and orthogonal to Σ will remain bounded as we approach the singularity. (These two components are equal by the anti-self-duality condition.)

References

[AS] Atiyah, M.F., Singer, I.M.: The index of elliptic operators. III, Ann. of Math. **87** (1968), 546–604

[Bi1] Biquard, O.: Fibrés parabolique stables et connexions singulières plates, (preprint, 1990)

[Bi2] Biquard, O.: Prolongement d'un fibré holomorphe hermitien á courbure L^p sur une courbe ouverte, (preprint, 1991)

[BW] Boileau, M., Weber, C.: Le probléme de J. Milnor sur le nombre gordien des noeuds algébriques. L'Enseignement Mathématique **30** (1984), 173–222

[BT] Bott, R., Tu, L.W.: *Differential forms in algebraic topology,* Springer, New York (1982)

[D1] Donaldson, S.K.: An application of gauge theory to four dimensional topology, J. Differential Geometry **18** (1983), 279–315

[D2] Donaldson, S.K.: Connections, cohomology and the intersection forms of four-manifolds, J. Differential Geometry **24** (1986), 275–341

[D3] Donaldson, S.K.: Polynomial invariants for smooth four-manifolds, Topology **29** (1990), 257–315

[DK] Donaldson, S.K., Kronheimer, P.B.: *The geometry of four-manifolds,* Oxford University Press (1990)

[DS] Donaldson, S.K., Sullivan, D.P.: Quasiconformal four-manifolds, Acta Mathematica **163** (1990), 181–252

[FHP] Forgacs, P., Horvath, Z., Palla, M.: An exact fractionally charged self dual solution, Phys. Rev. Letters **46** (1981), 392

[F] Freedman M.H.: The topology of four-dimensional manifolds, J. Differential Geometry **17**, 357–454

[FS] Fintushel, R., Stern, R.: Pseudofree orbifolds, Annals of Math. **122** (1985), 335–64

[FU] Freed, D.S., Uhlenbeck, K.K.: *Instantons and four-manifolds,* M.S.R.I. Publications, Vol. 1, Springer, New York (1984)

[GH] Griffiths, P., Harris, J.: *Principles of algebraic geometry,* Wiley, New York (1978)

[HS] Hsiang, W.C., Szczarba, R.H.: On embedding surfaces in four-manifolds. In: Algebraic Topology. (Proc. Sympos. Pure Math., vol.22.) Amer. Math. Soc., Providence, RI, 1971, pp.97–103

[KeM] Kervaire, M., Milnor, J.: On 2-spheres in 4-manifolds, Proc. Nat. Acad. Science USA **47** (1961), 1651–1657

[Ko] Kotschick, D.: On manifolds homeomorphic to $\mathbb{C}\mathrm{P}^2\#8\overline{\mathbb{C}\mathrm{P}^2}$, Inventiones Mathematicae **95** (1989), 591–600

[Kr] Kronheimer, P.B.: Embedded surfaces in 4-manifolds. In: Proc. Intern. Congr. Mathematicians, Kyoto 1990, Springer, 1991, vol.I, pp.527–539

[KrM] Kronheimer, P.B., Mrowka, T.S.: Gauge theory for embedded surfaces: II, (in preparation)

[Ku] Kuga, K.: Representing homology classes of $S^2 \times S^2$, Topology **23** (1984), 133–137

[LW] Lee, R., Wilczyński, D.M.: Locally flat 2-spheres in simply-connected 4-manifolds, Comm. Math. Helv. **65** (1990), 388–412

[LP] Looijenga, E., Peters, C.: Torelli theorems for Kähler $K3$-surfaces, Compositio Math. **42** (1981), 145–186

[Ma] Matumoto, T.: On diffeomorphisms of a $K3$ surface, Hiroshima University Reprint (1985)
[Mi] Milnor, J.: *Singular Points of Complex Hypersurfaces,* (Ann. of Math. Studies, vol.61.) Princeton University Press (1968)
[MMR] Morgan, J.W., Mrowka, T., Ruberman, D.: (in preparation)
[Mu] Murray, M.K.: Monopoles and spectral curves for arbitrary Lie groups, Commun. Math. Phys. **90** (1983) 263–271
[R1] Rohlin, V.A.: New results in the theory of four-dimensional manifolds, Dokl. Akad. Nauk. S.S.S.R. **84** (1952), 221–224 (Russian)
[R2] Rohlin, V.A.: Two-dimensional submanifolds of four-dimensional manifolds. Funktional. annal. i Priložen. **5** (1971) 48–60 (Russian). [English transl.: Functional Anal. Appl. **5** (1971) 39–48]
[Ru] Rudolph, L.: Some knot theory of complex plane curves. L'Enseignement Mathématique **29** (1983), 185–208
[Se] Seshadri, C.S.: Moduli of vector bundles on curves with parabolic structures. Bull. Amer. Math. Soc. **83** (1977) 124–126
[SS] Sibner, L.M., Sibner, R.J.: Classification of singular Sobolev connections by their holonomy, Commun. Math. Phys. **144** (1992) 337–350
[T1] Taubes, C.H.: Path-connected Yang-Mills moduli spaces, J. Differential Geometry **19** (1984), 337–392
[T2] Taubes, C.H.: L^2 moduli spaces on manifolds with cylindrical ends, (preprint 1990)
[Tr] Tristram, A.G.: Some cobordism invariants for links, Math. Proc. Camb. Phil. Soc. **66** (1969) 251–264
[W] Wang, S.: *Gauge theory and involutions,* D.Phil thesis, Oxford (1990)

This version: 8 July 1992

This article was processed by the author using the TEX macro package from Springer-Verlag.