Commun. Math. Phys. 280, 285–313 (2008)
Digital Object Identifier (DOI) 10.1007/s00220-008-0466-9

Communications in
**Mathematical
Physics**

Moduli Spaces of Self-Dual Connections over Asymptotically Locally Flat Gravitational Instantons

Gábor Etesi[1,2], **Marcos Jardim**[2]

[1] Department of Geometry, Mathematical Institute, Faculty of Science, Budapest University of Technology and Economics, Egry J. u. 1, H ép., 1111 Budapest, Hungary. E-mail: etesi@math.bme.hu; etesi@ime.unicamp.br
[2] Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, C.P. 6065, 13083-859, Campinas, SP, Brazil. E-mail: jardim@ime.unicamp.br

Received: 3 November 2006 / Accepted: 22 October 2007
Published online: 28 March 2008 –

Abstract: We investigate Yang–Mills instanton theory over four dimensional asymptotically locally flat (ALF) geometries, including gravitational instantons of this type, by exploiting the existence of a natural smooth compactification of these spaces introduced by Hausel–Hunsicker–Mazzeo. First referring to the codimension 2 singularity removal theorem of Sibner–Sibner and Råde we prove that given a smooth, finite energy, self-dual SU(2) connection over a complete ALF space, its energy is congruent to a Chern–Simons invariant of the boundary three-manifold if the connection satisfies a certain holonomy condition at infinity and its curvature decays rapidly. Then we introduce framed moduli spaces of self-dual connections over Ricci flat ALF spaces. We prove that the moduli space of smooth, irreducible, rapidly decaying self-dual connections obeying the holonomy condition with fixed finite energy and prescribed asymptotic behaviour on a fixed bundle is a finite dimensional manifold. We calculate its dimension by a variant of the Gromov–Lawson relative index theorem. As an application, we study Yang–Mills instantons over the flat $\mathbb{R}^3 \times S^1$, the multi-Taub–NUT family, and the Riemannian Schwarzschild space.

1. Introduction

By a *gravitational instanton* one usually means a connected, four dimensional complete hyper-Kähler Riemannian manifold. In particular, these spaces have $\mathrm{SU}(2) \cong \mathrm{Sp}(1)$ holonomy; consequently, they are Ricci flat, and hence solutions of the Riemannian Einstein's vacuum equation.

The only compact, four dimensional hyper-Kähler spaces are, up to (universal) covering, diffeomorphic to the flat torus T^4 or a $K3$ surface.

The next natural step would be to understand non-compact gravitational instantons. Compactness in this case should be replaced by the condition that the metric be complete and decay to the flat metric at infinity somehow such that the Pontryagin number of the manifold be finite.

Such open hyper-Kähler examples can be constructed as follows. Consider a connected, orientable compact four-manifold $\overline{M}$ with connected boundary $\partial\overline{M}$ which is a smooth three-manifold. Then the open manifold $M := \overline{M} \setminus \partial\overline{M}$ has a decomposition $M = K \cup W$, where K is a compact subset and $W \cong \partial\overline{M} \times \mathbb{R}^+$ is an open annulus or neck. Parameterize the half-line $\mathbb{R}^+$ by r. Assume $\partial\overline{M}$ is fibered over a base manifold B with fibers F and the complete hyper-Kähler metric g asymptotically and locally looks like $g \sim \mathrm{d}r^2 + r^2 g_B + g_F$. In other words the base B of the fibration blows up locally in a Euclidean way as $r \to \infty$, while the volume of the fiber remains finite. By the curvature decay, g_F must be flat, hence F is a connected, compact, orientable, flat manifold. On induction of the dimension of F, we can introduce several cases of increasing transcendentality, using the terminology of Cherkis and Kapustin [7]:

(i) (M, g) is ALE (asymptotically locally Euclidean) if $\dim F = 0$;
(ii) (M, g) is ALF (asymptotically locally flat) if $\dim F = 1$, in this case necessarily $F \cong S^1$ must hold;
(iii) (M, g) is ALG (this abbreviation by induction) if $\dim F = 2$, in this case $F \cong T^2$;
(iv) (M, g) is ALH if $\dim F = 3$, in this case F is diffeomorphic to one of the six flat orientable three-manifolds.

Due to their relevance in quantum gravity or recently rather in low-energy supersymmetric solutions of string theory and, last but not least, their mathematical beauty, there has been some effort to classify these spaces over the past decades. Trivial examples for any class is provided by the space $\mathbb{R}^{4-\dim F} \times F$ with its flat product metric.

The first two non-trivial, infinite families were discovered by Gibbons and Hawking in 1976 [19] in a rather explicit form. One of these families are the A_k ALE or multi-Eguchi–Hanson spaces. In 1989, Kronheimer gave a full classification of ALE spaces [26] constructing them as minimal resolutions of $\mathbb{C}^2/\Gamma$, where $\Gamma \subset \mathrm{SU}(2)$ is a finite subgroup, i.e. Γ is either a cyclic group A_k, $k \geq 0$, dihedral group D_k with $k > 0$, or one of the exceptional groups E_l with $l = 6, 7, 8$.

The other infinite family of Gibbons and Hawking is the A_k ALF or multi-Taub–NUT family. Recently another D_k ALF family has been constructed by Cherkis and Kapustin [10] and in a more explicit form by Cherkis and Hitchin [8].

Motivated by string theoretical considerations, Cherkis and Kapustin have suggested a classification scheme for ALF spaces as well as for ALG and ALH [7] although they relax the above asymptotical behaviour of the metric in these later two cases in order to obtain a natural classification. They claim that the A_k and D_k families with $k \geq 0$ exhaust the ALF geometry (in this enumeration D_0 is the Atiyah–Hitchin manifold). For the ALG case if we suppose that these spaces arise by deformations of elliptic fibrations with only one singular fiber, it is conjectured that the possibilities are D_k with $0 \leq k \leq 5$ (cf. [9]) and E_l with $l = 6, 7, 8$. An example for a non-trivial ALH space is the minimal resolution of $(\mathbb{R} \times T^3)/\mathbb{Z}_2$. The trouble is that these spaces are more transcendental as $\dim F$ increases, hence their constructions, involving twistor theory, Nahm transform, etc. are less straightforward and explicit.

To conclude this brief survey, we remark that the restrictive hyper-Kähler assumption on the metric, which appeared to be relevant in the more recent string theoretical investigations, excludes some examples considered as gravitational instantons in the early eighties. An important non-compact example which satisfies the ALF condition is for instance the Riemannian Schwarzschild space, which is Ricci flat but not hyper-Kähler [22]. For a more complete list of such "old" examples cf. [14].

From Donaldson theory we learned that the moduli spaces of SU(2) instantons over compact four-manifolds encompass a lot of information about the original manifold, hence understanding SU(2) instantons over gravitational instantons also might be helpful in their classification. On the compact examples T^4 and the $K3$'s, Yang–Mills instantons can be studied via the usual methods, especially the celebrated Hitchin–Kobayashi correspondence. The full construction of SU(2) instantons in the hyper-Kähler ALE case was carried out in important papers by Nakajima [29] and Kronheimer–Nakajima in 1990 [27]. However, the knowledge regarding moduli spaces of instantons over non-trivial ALF spaces is rather limited, even in the hyper-Kähler ALF case, due to analytical difficulties. One has only sporadic examples of explicit solutions (cf. e.g. [17,18]). Also very little is known about instanton theory over the Riemannian Schwarzschild space [6,16]. The only well studied case is the flat $\mathbb{R}^3 \times S^1$ space; instantons over this space, also known as *calorons*, have been extensively studied in the literature, cf. [4,5,30]. Close to nothing is known about instantons over non-trivial ALG and ALH geometries.

Studying Yang–Mills instanton moduli spaces over ALF spaces is certainly interesting not only because understanding the reducible solutions already leads to an encouraging topological classification result in the hyper-Kähler case [15], but also due to their physical significance. In this paper we set the foundations for a general theory of Yang–Mills instantons over ALF spaces in the broad sense adopted in the eighties, i.e. including not hyper-Kähler examples, too.

In Sect. 2 we exploit the existence of a natural smooth compactification X of an ALF space introduced by Hausel–Hunsicker–Mazzeo [21]. Working over this compact space, the asymptotical behaviour of any finite energy connection over an ALF space can be analyzed by the codimension 2 singularity removal theorem of Sibner–Sibner [33] and Råde [32]. This guarantees the existence of a locally flat connection ∇_Γ with fixed constant holonomy in infinity to which the finite energy connection converges. First we prove in Sect. 2 that the energy of a smooth, self-dual SU(2) connection of finite energy which satisfies a certain holonomy condition (cf. condition (11) here) and has rapid curvature decay (in the sense of condition (16) in the paper), is congruent to the Chern–Simons invariant $\tau_N(\Gamma_\infty)$ of the boundary N of the ALF space (Theorem 2.2 here). If the holonomy condition holds then ∇_Γ is in fact flat and Γ_∞ is a fixed smooth gauge for the limiting flat connection restricted to the boundary. The relevant holonomy condition can be replaced by a simple topological criterion on the infinity of the ALF space, leading to a more explicit form of this theorem (cf. Theorem 2.3).

Then in Sect. 3 we introduce framed instanton moduli spaces $\mathcal{M}(e, \Gamma)$ of smooth, irreducible, rapidly decaying self-dual SU(2) connections, obeying the holonomy condition, with fixed energy $e < \infty$ and asymptotical behaviour described by the flat connection ∇_Γ on a fixed bundle. Referring to a variant of the Gromov–Lawson relative index theorem [20] (cf. Theorem 3.1 here) we will be able to demonstrate that a framed moduli space over a Ricci flat ALF space is either empty or forms a smooth manifold of dimension

$$\dim \mathcal{M}(e, \Gamma) = 8\,(e + \tau_N(\Theta_\infty) - \tau_N(\Gamma_\infty)) - 3b^-(X),$$

where Θ_∞ is the restriction to N of the trivial flat connection ∇_Θ in some smooth gauge and $b^-(X)$ is the rank of the negative definite part of the intersection form of the Hausel–Hunsicker–Mazzeo compactification (cf. Theorem 3.2).

In Sect. 4 we apply our results on three classical examples, obtaining several novel facts regarding instantons over them.

First, we prove in Theorem 4.1 that any smooth, finite energy caloron over $\mathbb{R}^3 \times S^1$ automatically satisfies our holonomy condition, has integer energy $e \in \mathbb{N}$ if it decays rapidly and that the dimension of the moduli space in this case is $8e$, in agreement with [4]. These moduli spaces are non-empty for all positive integer e [5].

For the canonically oriented multi-Taub–NUT spaces, we show that the dimension of the framed moduli of smooth, irreducible, rapidly decaying anti-self-dual connections satisfying the holonomy condition is divisible by 8 (cf. Theorem 4.2). Known explicit solutions [18] show that at least a few of these moduli spaces are actually non-empty.

Finally, we consider the Riemannian Schwarzschild case, and prove in Theorem 4.3 that all smooth finite energy instantons obey the holonomy condition, have integer energy e if they decay rapidly and the dimension is $8e-3$. Moreover, this moduli space is surely non-empty at least for $e = 1$. We also enumerate the remarkably few known explicit solutions [6,16], and observe that these admit deformations.

Section 5 is an Appendix containing the proof of the relative index theorem used in the paper, Theorem 3.1.

2. The Spectrum of the Yang–Mills Functional

In this section we prove that the spectrum of the Yang–Mills functional evaluated on self-dual connections satisfying a certain analytical and a topological condition over a complete ALF manifold is "quantized" by the Chern–Simons invariants of the boundary. First, let us carefully define the notion of ALF space used in this paper, and describe its useful topological compactification, first used in [21].

Let (M, g) be a connected, oriented Riemannian four-manifold. This space is called an *asymptotically locally flat (ALF) space* if the following holds. There is a compact subset $K \subset M$ such that $M \setminus K = W$ and $W \cong N \times \mathbb{R}^+$, with N being a connected, compact, oriented three-manifold without boundary admitting a smooth S^1-fibration

$$\pi : N \xrightarrow{F} B_\infty \tag{1}$$

whose base space is a compact Riemann surface B_∞. For the smooth, complete Riemannian metric g there exists a diffeomorphism $\phi : N \times \mathbb{R}^+ \to W$ such that

$$\phi^*(g|_W) = \mathrm{d}r^2 + r^2(\pi^* g_{B_\infty})' + h'_F, \tag{2}$$

where g_{B_∞} is a smooth metric on B_∞, h_F is a symmetric 2-tensor on N which restricts to a metric along the fibers $F \cong S^1$ and $(\pi^* g_{B_\infty})'$ as well as h'_F are some finite, bounded, smooth extensions of $\pi^* g_{B_\infty}$ and h_F over W, respectively. That is, we require $(\pi^* g_{B_\infty})'(r) \sim O(1)$ and $h'_F(r) \sim O(1)$ and the extensions for $r < \infty$ preserve the properties of the original fields. Furthermore, we impose that the curvature R_g of g decays like

$$|\phi^*(R_g|_W)| \sim O(r^{-3}). \tag{3}$$

Here R_g is regarded as a map $R_g : C^\infty(\Lambda^2 M) \to C^\infty(\Lambda^2 M)$ and its pointwise norm is calculated accordingly in an orthonormal frame. Hence the Pontryagin number of our ALF spaces is finite.

Examples of such metrics are the natural metric on $\mathbb{R}^3 \times S^1$ which is in particular flat; the multi-Taub–NUT family [19] which is moreover hyper-Kähler or the Riemannian Schwarzschild space [22] which is in addition Ricci flat only. For a more detailed description of these spaces, cf. Sect. 4.

We construct the compactification X of M simply by shrinking all fibers of N into points as $r \to \infty$ like in [21]. We put an orientation onto X induced by the orientation of the original M. The space X is then a connected, oriented, smooth four-manifold without boundary. One clearly obtains a decomposition $X \cong M \cup B_\infty$, and consequently we can think of B_∞ as a smoothly embedded submanifold of X of codimension 2. For example, for $\mathbb{R}^3 \times S^1$ one finds $X \cong S^4$ and B_∞ is an embedded S^2 [15]; for the multi-Taub–NUT space with the orientation induced by one of the complex structures, X is the connected sum of s copies of $\overline{\mathbb{C}P}^2$'s (s refers to the number of NUTs) and B_∞ is homeomorphic to S^2 providing a generator of the second cohomology of X; in case of the Riemannian Schwarzschild geometry, $X \cong S^2 \times S^2$ and B_∞ is again S^2, also providing a generator for the second cohomology (cf. [15,21]).

Let $M_R := M\backslash(N \times (R, \infty))$ be the truncated manifold with boundary $\partial M_R \cong N \times \{R\}$. Taking into account that $W \cong N \times (R, \infty)$, a normal neighbourhood V_R of B_∞ in X has a model like $V_R \cong N \times (R, \infty]/\sim$, where $\sim$ means that $N \times \{\infty\}$ is pinched into B_∞. We obtain $W = V_R\backslash V_\infty$, with $V_\infty \cong B_\infty$. By introducing the parameter $\varepsilon := R^{-1}$ we have another model V_ε provided by the fibration

$$V_\varepsilon \xrightarrow{B^2_\varepsilon} B_\infty$$

whose fibers are two-balls of radius ε. In this second picture we have the identification $V_0 \cong B_\infty$, so that the end W looks like

$$V^*_\varepsilon := V_\varepsilon\backslash V_0. \tag{4}$$

Choosing a local coordinate patch $U \subset B_\infty$, then locally $V_\varepsilon|_U \cong U \times B^2_\varepsilon$ and $V^*_\varepsilon|_U \cong U \times (B^2_\varepsilon\backslash\{0\})$. We introduce local coordinates (u, v) on U and polar coordinates (ρ, τ) along the discs B^2_ε with $0 \leq \rho < \varepsilon$ and $0 \leq \tau < 2\pi$. Note that in fact $\rho = r^{-1}$ is a global coordinate over the whole $V_\varepsilon \cong V_R$. For simplicity we denote $V_\varepsilon|_U$ as U_ε and will call the set

$$U^*_\varepsilon := U_\varepsilon\backslash U_0 \tag{5}$$

an *elementary neighbourhood*. Clearly, their union covers the end W. In this ε-picture we will use the notation $\partial M_\varepsilon \cong N \times \{\varepsilon\}$ for the boundary of the truncated manifold $M_\varepsilon = M\backslash(N \times (0, \varepsilon))$, and by a slight abuse of notation we will also think of the end sometimes as $W \cong \partial M_\varepsilon \times (0, \varepsilon)$.

We do not expect the complete ALF metric g to extend over this compactification, even conformally. However the ALF property (2) implies that we can suppose the existence of a smooth positive function $f \sim O(r^{-2})$ on M such that the rescaled metric $\tilde{g} := f^2 g$ extends smoothly as a tensor field over X (i.e., a smooth Riemannian metric degenerated along the singularity set B_∞). In the vicinity of the singularity we find $\tilde{g}|_{V_\varepsilon} = \mathrm{d}\rho^2 + \rho^2(\pi^* g_{B_\infty})'' + \rho^4 h''_F$ via (2), consequently we can choose the coordinate system (u, v, ρ, τ) on U_ε such that $\{\mathrm{d}u, \mathrm{d}v, \mathrm{d}\rho, \mathrm{d}\tau\}$ forms an oriented frame on $T^*U^*_\varepsilon$ and with some bounded, finite function φ, the metric looks like $\tilde{g}|_{U^*_\varepsilon} = \mathrm{d}\rho^2 + \rho^2\varphi(u, v, \rho)(\mathrm{d}u^2 + \mathrm{d}v^2) + \rho^4(\mathrm{d}\tau^2 + \ldots)$. Consequently we find that

$$\mathrm{Vol}_{\tilde{g}}(V_\varepsilon) \sim O(\varepsilon^5). \tag{6}$$

We will also need a smooth regularization of $\tilde{g}$. Taking a monotonously increasing smooth function f_ε supported in $V_{2\varepsilon}$ and equal to 1 on V_ε such that $|\mathrm{d}f_\varepsilon| \sim O(\varepsilon^{-1})$, as well as picking up a smooth metric h on X, we can regularize $\tilde{g}$ by introducing the smooth metric

$$\tilde{g}_\varepsilon := (1 - f_\varepsilon)\tilde{g} + f_\varepsilon h \tag{7}$$

over X. It is clear that $\tilde{g}_0$ and $\tilde{g}$ agree on M.

Let F be an SU(2) vector bundle over X endowed with a fixed connection ∇_Γ and an invariant fiberwise scalar product. Using the rescaled-degenerated metric $\tilde{g}$, define Sobolev spaces $L^p_{j,\Gamma}(\Lambda^* X \otimes F)$ with $1 < p \leq \infty$ and $j = 0, 1, \ldots$ as the completion of $C^\infty_0(\Lambda^* X \otimes F)$, smooth sections compactly supported in $M \subset X$, with respect to the norm

$$\|\omega\|_{L^p_{j,\Gamma}(X)} := \left(\lim_{\varepsilon \to 0} \sum_{k=0}^{j} \|\nabla^k_\Gamma \omega\|^p_{L^p(M_\varepsilon, \tilde{g}|_{M_\varepsilon})} \right)^{\frac{1}{p}}, \tag{8}$$

where

$$\|\nabla^k_\Gamma \omega\|^p_{L^p(M_\varepsilon, \tilde{g}|_{M_\varepsilon})} = \int\limits_{M_\varepsilon} |\nabla^k_\Gamma \omega|^p *_{\tilde{g}} 1.$$

Throughout this paper, Sobolev norms of this kind will be used unless otherwise stated. We will write simply L^p for $L^p_{0,\Gamma}$. Notice that every 2-form with finite L^2-norm over (M, g) will also belong to this Sobolev space, by conformal invariance and completeness.

Next, we collect some useful facts regarding the Chern–Simons functional. Let E be a smooth SU(2) bundle over M. Since topological G-bundles over an open four-manifold are classified by $H^2(M, \pi_1(G))$, note that E is necessarily trivial. Put a smooth SU(2) connection ∇_B onto E. Consider the boundary ∂M_ε of the truncated manifold. The restricted bundle $E|_{\partial M_\varepsilon}$ is also trivial. Therefore any restricted SU(2) connection $\nabla_B|_{\partial M_\varepsilon} := \nabla_{B_\varepsilon}$ over $E|_{\partial M_\varepsilon}$ can be identified with a smooth $\mathfrak{su}(2)$-valued 1-form B_ε. The *Chern–Simons functional* is then defined to be

$$\tau_{\partial M_\varepsilon}(B_\varepsilon) := -\frac{1}{8\pi^2} \int\limits_{\partial M_\varepsilon} \operatorname{tr}\left(\mathrm{d}B_\varepsilon \wedge B_\varepsilon + \frac{2}{3} B_\varepsilon \wedge B_\varepsilon \wedge B_\varepsilon \right).$$

This expression is gauge invariant up to an integer. Moreover, the representation space

$$\chi(\partial M_\varepsilon) := \mathrm{Hom}(\pi_1(\partial M_\varepsilon), \mathrm{SU}(2))/\mathrm{SU}(2)$$

is called the *character variety* of $\partial M_\varepsilon \cong N$ and parameterizes the gauge equivalence classes of smooth flat SU(2) connections over N.

Lemma 2.1. *Fix an $0 < \rho < \varepsilon$ and let $\nabla_{A_\rho} = \mathrm{d} + A_\rho$ and $\nabla_{B_\rho} = \mathrm{d} + B_\rho$ be two smooth SU(2) connections in a fixed smooth gauge on the trivial SU(2) bundle $E|_{\partial M_\rho}$. Then there is a constant $c_1 = c_1(B_\rho) > 0$, depending on ρ only through B_ρ, such that*

$$|\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(B_\rho)| \leq c_1 \|A_\rho - B_\rho\|_{L^2(\partial M_\rho)}, \tag{9}$$

that is, the Chern–Simons functional is continuous in the L^2 norm.

Moreover, for each ρ, $\tau_{\partial M_\rho}(A_\rho)$ is constant on the path connected components of the character variety $\chi(\partial M_\rho)$.

Proof. The first observation follows from the identity

$$\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(B_\rho) =$$
$$-\frac{1}{8\pi^2} \int\limits_{\partial M_\rho} \operatorname{tr}\left((F_{A_\rho} + F_{B_\rho}) \wedge (A_\rho - B_\rho) - \frac{1}{3}(A_\rho - B_\rho) \wedge (A_\rho - B_\rho) \wedge (A_\rho - B_\rho) \right),$$

which implies that there is a constant $c_0 = c_0(\rho, B_\rho)$ such that

$$|\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(B_\rho)| \le c_0 \|A_\rho - B_\rho\|_{L^{\frac{3}{2}}_{1,B_\rho}(\partial M_\rho)},$$

that is, the Chern–Simons functional is continuous in the $L^{\frac{3}{2}}_{1,B_\rho}$ norm. Then applying the Sobolev embedding $L^{\frac{3}{2}}_{1,B_\rho} \subset L^2$ over a compact three-manifold, we find a similar inequality with a constant $c_1 = c_1(\rho, B_\rho)$. The metric locally looks like $\tilde{g}|_{\partial M_\rho \cap U^*_\varepsilon} = \rho^2\varphi(\mathrm{d}u^2 + \mathrm{d}v^2) + \rho^4(\mathrm{d}\tau^2 + 2h_{\tau,u}\mathrm{d}\tau\mathrm{d}u + 2h_{\tau,v}\mathrm{d}\tau\mathrm{d}v + h_{u,v}\mathrm{d}u^2 + h_{v,v}\mathrm{d}v^2 + 2h_{u,v}\mathrm{d}u\mathrm{d}v)$ with φ and $h_{\tau,u}$, etc. being bounded functions of (u, v, ρ) and (u, v, ρ, τ) respectively, hence the metric coefficients are bounded functions of ρ; consequently we can suppose that c_1 does not depend explicitly on ρ.

Concerning the second part, assume ∇_{A_ρ} and ∇_{B_ρ} are two smooth, flat connections belonging to the same path connected component of $\chi(\partial M_\rho)$. Then there is a continuous path $\nabla_{A^t_\rho}$ with $t \in [0, 1]$ of flat connections connecting the given flat connections. Out of this we construct a connection ∇_A on $\partial M_\rho \times [0, 1]$ given by $A := A^t_\rho + 0 \cdot \mathrm{d}t$. Clearly, this connection is flat, i.e., $F_A = 0$. The Chern–Simons theorem [11] implies that

$$\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(B_\rho) = -\frac{1}{8\pi^2} \int\limits_{\partial M_\rho \times [0,1]} \mathrm{tr}(F_A \wedge F_A) = 0,$$

concluding the proof. □

The last ingredient in our discussion is the fundamental theorem of Sibner–Sibner [33] and Råde [32] which allows us to study the asymptotic behaviour of finite energy connections over an ALF space. Consider a smooth (trivial) SU(2) vector bundle E over the ALF space (M, g). Let ∇_A be a Sobolev connection on E with finite energy, i.e. $F_A \in L^2(\Lambda^2 M \otimes \mathrm{End}E)$. Taking into account completeness of the ALF metric, we have $|F_A|(r) \to 0$ almost everywhere as $r \to \infty$. Thus we have a connection defined on X away from a smooth, codimension 2 submanifold $B_\infty \subset X$ and satisfies $\|F_A\|_{L^2(X)} < \infty$.

Consider a neighbourhood $B_\infty \subset V_\varepsilon$ and write V^*_ε to describe the end W as in (4). Let ∇_Γ be an SU(2) connection on $E|_{V^*_\varepsilon}$ which is locally flat and smooth. The restricted bundle $E|_{V^*_\varepsilon}$ is trivial, hence we can choose some global gauge such that $\nabla_A|_{V^*_\varepsilon} = \mathrm{d} + A_{V^*_\varepsilon}$ and $\nabla_\Gamma = \mathrm{d} + \Gamma_{V^*_\varepsilon}$; we assume with some $j = 0, 1, \dots$ that $A_{V^*_\varepsilon} \in L^2_{j+1,\Gamma,loc}$. Taking into account that for the elementary neighbourhood $\pi_1(U^*_\varepsilon) \cong \mathbb{Z}$, generated by a τ-circle, it is clear that locally on $U^*_\varepsilon \subset V^*_\varepsilon$ we can choose a more specific gauge $\nabla_\Gamma|_{U^*_\varepsilon} = \mathrm{d} + \Gamma_m$ with a constant $m \in [0, 1)$ such that [33]

$$\Gamma_m = \begin{pmatrix} \mathbf{i}m & 0 \\ 0 & -\mathbf{i}m \end{pmatrix} \mathrm{d}\tau. \tag{10}$$

Here m represents the *local holonomy* of the locally flat connection around the punctured discs of the space U^*_ε, see [33]. It is invariant under gauge transformations modulo an integer. For later use we impose two conditions on this local holonomy. The embedding $i : U^*_\varepsilon \subset V^*_\varepsilon$ induces a group homomorphism $i_* : \pi_1(U^*_\varepsilon) \to \pi_1(V^*_\varepsilon)$. It may happen that this homomorphism has non-trivial kernel. Let l be a loop in U^*_ε such that $[l]$ generates $\pi_1(U^*_\varepsilon) \cong \mathbb{Z}$.

Definition 2.1. *A locally flat connection* ∇_Γ *on* $E|_{V_\varepsilon^*}$ *is said to satisfy the* **weak holonomy condition** *if for all* $U_\varepsilon^* \subset V_\varepsilon^*$ *the restricted connection* $\nabla_\Gamma|_{U_\varepsilon^*} = \mathrm{d} + \Gamma_m$ *has trivial local holonomy whenever* l *is contractible in* V_ε^**, i.e.*

$$[l] \in \operatorname{Ker} i_* \implies m = 0. \tag{11}$$

Additionally, ∇_Γ *is said to satisfy the* **strong holonomy condition** *if the local holonomy of any restriction* $\nabla_\Gamma|_{U_\varepsilon^*} = \mathrm{d} + \Gamma_m$ *vanishes, i.e.*

$$m = 0. \tag{12}$$

Clearly, ∇_Γ is *globally* a smooth, flat connection on $E|_{V_\varepsilon^*}$ if and only if the weak holonomy condition holds. Moreover, the strong holonomy condition implies the weak one. We are now in a position to recall the following fundamental regularity result [32,33].

Theorem 2.1. (Sibner–Sibner, 1992 and Råde, 1994) *There exist a constant* $\varepsilon > 0$ *and a flat* $SU(2)$ *connection* $\nabla_\Gamma|_{U_\varepsilon^*}$ *on* $E|_{U_\varepsilon^*}$ *with a constant holonomy* $m \in [0, 1)$ *such that on* $E|_{U_\varepsilon^*}$ *one can find a gauge* $\nabla_A|_{U_\varepsilon^*} = \mathrm{d} + A_{U_\varepsilon^*}$ *and* $\nabla_\Gamma|_{U_\varepsilon^*} = \mathrm{d} + \Gamma_m$ *with* $A_{U_\varepsilon^*} - \Gamma_m \in L^2_{1,\Gamma}(U_\varepsilon^*)$ *such that the estimate*

$$\|A_{U_\varepsilon^*} - \Gamma_m\|_{L^2_{1,\Gamma}(U_\varepsilon^*)} \leq c_2 \|F_A\|_{L^2(U_\varepsilon)}$$

holds with a constant $c_2 = c_2(\tilde{g}|_{U_\varepsilon}) > 0$ *depending only on the metric.*

This theorem shows that any finite energy connection is always asymptotic to a flat connection at least locally. It is therefore convenient to say that the finite energy connection ∇_A *satisfies the weak* or *the strong holonomy condition* if its associated asymptotic locally flat connection ∇_Γ, in the sense of Theorem 2.1, satisfies the corresponding condition in the sense of Definition 2.1. We will be using this terminology.

This estimate can be globalized over the whole end V_ε^* as follows. Consider a finite covering $B_\infty = \cup_\alpha U_\alpha$ and denote the corresponding punctured sets as $U_{\varepsilon,\alpha}^* \subset V_\varepsilon^*$. These sets also give rise to a finite covering of V_ε^*. It is clear that the weak condition (11) is independent of the index α, since by Theorem 2.1, m is constant over all V_ε^*. Imposing (11), the local gauges Γ_m on $U_{\varepsilon,\alpha}^*$ extend smoothly over the whole $E|_{V_\varepsilon^*}$. That is, there is a smooth flat gauge $\nabla_\Gamma = \mathrm{d} + \Gamma_{V_\varepsilon^*}$ over $E|_{V_\varepsilon^*}$ such that $\Gamma_{V_\varepsilon^*}|_{U_{\varepsilon,\alpha}^*} = \gamma_\alpha^{-1}\Gamma_m\gamma_\alpha + \gamma_\alpha^{-1}\mathrm{d}\gamma_\alpha$ with smooth gauge transformations $\gamma_\alpha : U_{\varepsilon,\alpha}^* \to \mathrm{SU}(2)$. This gauge is unique only up to an arbitrary smooth gauge transformation. Since this construction deals with the topology of the boundary (1) only, we can assume that these gauge transformations are independent of the (global) radial coordinate $0 < \rho < \varepsilon$. Then we write this global gauge as

$$\nabla_A|_{V_\varepsilon^*} = \mathrm{d} + A_{V_\varepsilon^*}, \qquad \nabla_\Gamma = \mathrm{d} + \Gamma_{V_\varepsilon^*}. \tag{13}$$

A comparison with the local gauges in Theorem 2.1 shows that

$$(A_{V_\varepsilon^*} - \Gamma_{V_\varepsilon^*})|_{U_{\varepsilon,\alpha}^*} = \gamma_\alpha^{-1}(A_{U_{\varepsilon,\alpha}^*} - \Gamma_m)\gamma_\alpha. \tag{14}$$

Applying Theorem 2.1 in all coordinate patches and summing up over them we come up with

$$\|A_{V_\varepsilon^*} - \Gamma_{V_\varepsilon^*}\|_{L^2_{1,\Gamma}(V_\varepsilon^*)} \leq c_3 \|F_A\|_{L^2(V_\varepsilon)}, \tag{15}$$

with some constant $c_3 = c_3(\tilde{g}|_{V_\varepsilon}, \gamma_\alpha, \mathrm{d}\gamma_\alpha) > 0$. This is the globalized version of Theorem 2.1.

Let A_ε and Γ_ε be the restrictions of the connection 1-forms in the global gauge (13) to the boundary $E|_{\partial M_\varepsilon}$. The whole construction shows that Γ_ε is smooth and is independent of ε, consequently $\lim\limits_{\varepsilon\to 0}\Gamma_\varepsilon$ exists and is smooth. Note again that smoothness follows if and only if the weak holonomy condition (11) is satisfied. In this way $\lim\limits_{\varepsilon\to 0}\tau_{\partial M_\varepsilon}(\Gamma_\varepsilon) = \tau_N(\Gamma_0)$ also exists and gives rise to a *Chern–Simons invariant* of the boundary.

Assume that our finite energy connection ∇_A is smooth; then in the global gauge (13) $A_{V^*_\varepsilon}$ is also smooth, hence $\tau_{\partial M_\varepsilon}(A_\varepsilon)$ also exists for $\varepsilon > 0$.

Next we analyze the behaviour $\tau_{\partial M_\varepsilon}(A_\varepsilon)$ as ε tends to zero. First notice that the local flat gauge Γ_m in (10) does not have radial component, consequently $A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon} = A_\varepsilon - \Gamma_\varepsilon + A_r$, where A_r is the radial component of $A_{V^*_\varepsilon}$. Dividing the square of (9) by $\varepsilon > 0$ and then integrating it we obtain, making use of (15), that

$$\begin{aligned}\frac{1}{\varepsilon}\int_0^\varepsilon |\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(\Gamma_\rho)|^2\mathrm{d}\rho \le \frac{c_1^2}{\varepsilon}\int_0^\varepsilon \|A_\rho - \Gamma_\rho\|^2_{L^2(\partial M_\rho)}\mathrm{d}\rho \le\\ \le \frac{c_1^2}{\varepsilon}\int_0^\varepsilon \left(\|A_\rho - \Gamma_\rho\|^2_{L^2(\partial M_\rho)} + \|A_r\|^2_{L^2(\partial M_\rho)}\right)\mathrm{d}\rho\\ = \frac{c_1^2}{\varepsilon}\|A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon}\|^2_{L^2(V^*_\varepsilon)} \le \frac{(c_1c_3)^2}{\varepsilon}\|F_A\|^2_{L^2(V_\varepsilon)}.\end{aligned}$$

Finite energy and completeness implies that $\|F_A\|_{L^2(V_\varepsilon)}$ vanishes as ε tends to zero. However for our purposes we need a stronger decay assumption.

Definition 2.2. *The finite energy* $SU(2)$ *connection* ∇_A *on the bundle* E *over* M **decays rapidly** *if its curvature satisfies*

$$\lim_{\varepsilon\to 0}\frac{\|F_A\|_{L^2(V_\varepsilon)}}{\sqrt{\varepsilon}} = \lim_{R\to\infty}\sqrt{R}\,\|F_A\|_{L^2(V_R, g|_{V_R})} = 0 \qquad (16)$$

along the end of the ALF space.

Consequently for a rapidly decaying connection we obtain

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^\varepsilon |\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(\Gamma_\rho)|^2\mathrm{d}\rho = 0,$$

which is equivalent to

$$\lim_{\varepsilon\to 0}\tau_{\partial M_\varepsilon}(A_\varepsilon) = \tau_N(\Gamma_0). \qquad (17)$$

We are finally ready to state an energy identity for self-dual connections. Let ∇_A be a smooth, self-dual, finite energy connection on the trivial SU(2) bundle E over an ALF space (M, g):

$$F_A = *F_A, \quad \|F_A\|^2_{L^2(M,g)} < \infty.$$

Assume it satisfies the weak holonomy condition (11). In this case we can fix a gauge (13) along V^*_ε and both $A_{V^*_\varepsilon}$ and $\Gamma_{V^*_\varepsilon}$ are smooth. Restrict ∇_A onto $E|_{M_\varepsilon}$ with $\varepsilon > 0$. Exploiting self-duality, an application of the Chern–Simons theorem [11] along the boundary shows that

$$\|F_A\|^2_{L^2(M_\varepsilon, g|_{M_\varepsilon})} \equiv \tau_{\partial M_\varepsilon}(A_\varepsilon) \mod \mathbb{Z}.$$

Moreover if the connection decays rapidly in the sense of (16), then the right-hand side has a limit (17); therefore we have arrived at the following theorem:

Theorem 2.2. *Let* (M, g) *be an ALF space with an end* $W \cong N \times \mathbb{R}^+$. *Let* E *be an* $SU(2)$ *vector bundle over* M, *necessarily trivial, with a smooth, finite energy, self-dual connection* ∇_A. *If it satisfies the weak holonomy condition* (11) *and decays rapidly in the sense of* (16), *then there exists a smooth flat* $SU(2)$ *connection* ∇_Γ *on* $E|_W$ *and a smooth flat gauge* $\nabla_\Gamma = \mathrm{d} + \Gamma_W$, *unique up to a smooth gauge transformation, such that* $\lim\limits_{r\to\infty} \Gamma_W|_{N\times\{r\}} = \Gamma_\infty$ *exists, is smooth and*

$$\|F_A\|^2_{L^2(M,g)} \equiv \tau_N(\Gamma_\infty) \mod \mathbb{Z}.$$

That is, the energy is congruent to a Chern–Simons invariant of the boundary.

Remark. It is clear that the above result depends only on the conformal class of the metric. One finds a similar energy identity for manifolds with conformally cylindrical ends [28] (including ALE spaces, in accordance with the energies of explicit instanton solutions of [17] and [18]) and for manifolds conformally of the form $\mathbb{C} \times \Sigma$ as in [35]. We expect that the validity of identities of this kind is more general.

Taking into account the second part of Lemma 2.1 and the fact that the character variety of a compact three-manifold has finitely many connected components, we conclude that *the energy spectrum of smooth, finite energy, self-dual connections over ALF spaces which satisfy the weak holonomy condition and decay rapidly is discrete*.

For irreducible instantons, imposing rapid decay is necessary for having discrete energies. For instance, in principle the energy formula [30, Eq. 2.32] provides a continuous energy spectrum for calorons and calorons of fractional energy are known to exist (cf. e.g. [12]). But slowly decaying reducible instantons still can have discrete spectrum; this is the case e.g. over the Schwarzschild space, cf. Sect. 4.

Alternatively, instead of the rapid decay condition (16), one could also impose the possibly weaker but less natural condition that the gauge invariant limit

$$\lim_{\varepsilon\to 0} |\tau_{\partial M_\varepsilon}(A_\varepsilon) - \tau_{\partial M_\varepsilon}(\Gamma_\varepsilon)| = \mu$$

exists. Then the identity of Theorem 2.2 would become:

$$\|F_A\|^2_{L^2(M,g)} \equiv \tau_N(\Gamma_\infty) + \mu \mod \mathbb{Z}.$$

By analogy with the energy formula for calorons [30, Eq. 2.32], we believe that the extra term μ is related to the *overall magnetic charge* of an instanton while the modified energy formula would represent the decomposition of the energy into “electric” (i.e., Chern–Simons) and “magnetic” (i.e., proportional to the μ-term) contributions.

In general, proving the existence of limits for the Chern–Simons functional assuming only the finiteness of the energy of the connection ∇_A is a very hard analytical problem, cf. [28,35]. Therefore our rapid decay condition is a simple and natural condition which allows us to explicitly compute the limit of the Chern–Simons functional in our situation.

Another example illustrates that the weak holonomy condition is also essential in Theorem 2.2. Consider $\mathbb{R}^4$, equipped with the Taub–NUT metric. This geometry admits a smooth L^2 harmonic 2-form which can be identified with the curvature F_B of a self-dual, rapidly decaying U(1) connection ∇_B as in [17]; hence $\nabla_B \oplus \nabla_B^{-1}$ is a smooth, self-dual, rapidly decaying, reducible SU(2) connection. We know that $H^2(\mathbb{R}^4, \mathbb{Z}) = 0$, hence ∇_B lives on a trivial line bundle, consequently it can be rescaled by an arbitrary constant like $B \mapsto cB$ without destroying its self-duality and finite energy. But the smooth, self-dual family ∇_{cB} has continuous energy proportional to c^2. This strange phenomenon also appears over the multi-Taub–NUT spaces, although they are no more topologically trivial, cf. Sect. 4 for more details.

From our holonomy viewpoint, this anomaly can be understood as follows. Let $i : U^*_\varepsilon \subset W$ be an elementary neighbourhood as in (5) with the induced map $i_* : \pi_1(U^*_\varepsilon) \to \pi_1(W)$. On the one hand we have $\pi_1(U^*_\varepsilon) \cong \mathbb{Z}$ as usual. On the other hand for the Taub–NUT space the asymptotical topology is $W \cong S^3 \times \mathbb{R}^+$, hence $\pi_1(W) \cong 1$, consequently i_* has a non-trivial kernel. However for a generic c the connection $\nabla_{cB}|_{U^*_\varepsilon}$ has non-trivial local holonomy $m \neq 0$, hence it does not obey the weak holonomy condition (11); therefore Theorem 2.2 fails in this case.

The flat $\mathbb{R}^3 \times S^1$ space has contrary behaviour to the multi-Taub–NUT geometries. In this case we find $W \cong S^2 \times S^1 \times \mathbb{R}^+$ for the end, consequently $\pi_1(W) \cong \mathbb{Z}$ and the map i_* is an obvious isomorphism. Hence the weak holonomy condition is always obeyed. The character variety of the boundary is $\chi(S^2 \times S^1) \cong [0, 1)$, hence connected. Referring to the second part of Lemma 2.1 we conclude then that the energy of any smooth, self-dual, rapidly decaying connection over the flat $\mathbb{R}^3 \times S^1$ must be a non-negative integer in accordance with the known explicit solutions [5]. The case of the Schwarzschild space is similar, cf. Sect. 4.

These observations lead us to a more transparent form of Theorem 2.2 by replacing the weak holonomy condition with a simple, sufficient topological criterion, which amounts to a straightforward re-formulation of Definition 2.1.

Theorem 2.3. *Let (M, g) be an ALF space with an end $W \cong N \times \mathbb{R}^+$ as before and, referring to the fibration (1), assume N is an arbitrary circle bundle over $B_\infty \not\cong S^2, \mathbb{R}P^2$, or is a trivial circle bundle over S^2 or $\mathbb{R}P^2$.*

Then if E is the (trivial) $SU(2)$ vector bundle over M with a smooth, finite energy connection ∇_A, then it satisfies the weak holonomy condition (11).

Moreover if ∇_A is self-dual and decays rapidly as in (16), then its energy is congruent to one of the Chern–Simons invariants of the boundary N.

In addition if the character variety $\chi(N)$ is connected, then the energy of any smooth, self-dual, rapidly decaying connection must be a non-negative integer.

Proof. Consider an elementary neighbourhood $i : U^*_\varepsilon \subset W$ as in (5) and the induced map $i_* : \pi_1(U^*_\varepsilon) \to \pi_1(W)$. If $\operatorname{Ker} i_* = \{0\}$, then ∇_A obeys (11). However one sees that $\pi_1(U^*_\varepsilon) \cong \pi_1(F)$ and $\pi_1(W) \cong \pi_1(N)$, hence the map i_* fits well into the homotopy exact sequence

$$\ldots \longrightarrow \pi_2(B_\infty) \longrightarrow \pi_1(F) \xrightarrow{i_*} \pi_1(N) \longrightarrow \pi_1(B_\infty) \longrightarrow \ldots$$

of the fibration (1). This segment shows that $\operatorname{Ker} i_* \neq \{0\}$ if and only if N is either a non-trivial circle bundle over S^2 that is, $N \cong S^3/\mathbb{Z}_s$ a lens space of type $L(s, 1)$, or a non-trivial circle bundle over $\mathbb{R}P^2$.

The last part is clear via the second part of Lemma 2.1. $\square$

Finally we investigate the strong holonomy condition (12). As an important corollary of our construction we find

Theorem 2.4. (Sibner–Sibner, 1992; Råde, 1994) *Let (M, g) be an ALF space with an end W as before and E be the trivial $SU(2)$ vector bundle over M with a smooth, finite energy connection ∇_A and associated locally flat connection ∇_Γ on $E|_W$. If and only if the strong holonomy condition (12) is satisfied then both ∇_A, as well as ∇_Γ as a flat connection, extend smoothly over the whole X, the Hausel–Hunsicker–Mazzeo compactification of (M, g). That is, there exist bundles $\tilde{E}$ and $\tilde{E}_0 \cong X \times \mathbb{C}^2$ over X such that $\tilde{E}|_M \cong E$ and the connection ∇_A extends smoothly over $\tilde{E}$; in the same fashion $\tilde{E}_0|_W \cong E|_W$ and ∇_Γ extends smoothly as a flat connection over $\tilde{E}_0$.*

Proof. The restriction of the embedding $M \subset X$ gives $U_\varepsilon^* \subset U_\varepsilon$ and this later space is contractible. Consequently if $i : U_\varepsilon^* \subset X$ is the embedding, then for the induced map $i_* : \pi_1(U_\varepsilon^*) \to \pi_1(X)$ we always have $\operatorname{Ker} i_* = \pi_1(U_\varepsilon^*)$, hence the connections ∇_A and ∇_Γ extend smoothly over X via Theorem 2.1 if and only if the strong holonomy condition (12) holds. In particular the extension of ∇_Γ is a flat connection. $\square$

Remark. If a finite energy self-dual connection satisfies the strong holonomy condition, then its energy is integer via Theorem 2.4 regardless of its curvature decay. Consequently these instantons again have discrete energy spectrum. We may then ask ourselves about the relationship between the strong holonomy condition on the one hand and the weak holonomy condition imposed together with the rapid decay condition on the other hand.

3. The Moduli Space

In this section we are going to prove that the moduli spaces of framed SU(2) instantons over ALF manifolds form smooth, finite dimensional manifolds, whenever non-empty. The argument will go along the by now familiar lines consisting of three steps: (i) Compute the dimension of the space of infinitesimal deformations of an irreducible, rapidly decaying self-dual connection, satisfying the weak holonomy condition, using a variant of the Gromov–Lawson relative index theorem [20] and a vanishing theorem; (ii) Use the Banach space inverse and implicit function theorems to integrate the infinitesimal deformations and obtain a local moduli space; (iii) Show that local moduli spaces give local coordinates on the global moduli space and that this global space is a Hausdorff manifold. We will carry out the calculations in detail for step (i) while just sketch (ii) and (iii) and refer the reader to the classical paper [2].

Let (M, g) be an ALF space with a single end W as in Sect. 2. Consider a trivial SU(2) bundle E over M with a smooth, irreducible, self-dual, finite energy connection ∇_A on it. By smoothness we mean that the connection 1-form is smooth in any smooth trivialization of E. In addition suppose ∇_A satisfies the weak holonomy condition (11) as well as decays rapidly in the sense of (16). Then by Theorem 2.2 its energy is determined by a Chern–Simons invariant. We will assume that this energy $e := \|F_A\|^2_{L^2(M,g)}$ is *fixed*.

Consider the associated flat connection ∇_Γ with holonomy $m \in [0, 1)$ as in Theorem 2.2. Extend ∇_Γ over the whole E and continue to denote it by ∇_Γ. Take the smooth gauge (13) on the neck. Since E is trivial, we can extend this gauge smoothly over the whole M and can write $\nabla_A = \mathrm{d} + A$ and $\nabla_\Gamma = \mathrm{d} + \Gamma$ for some smooth connection 1-forms A and Γ well defined over the whole M. We also fix this gauge once and for all in our forthcoming calculations. In particular the asymptotics of ∇_A is also *fixed* and is given by Γ. The connection ∇_Γ, the usual Killing form on $\mathrm{End}E$ and the rescaled metric $\tilde{g}$ are used to construct Sobolev spaces over various subsets of X with respect to the norm (8). Both the energy e and the asymptotics Γ are preserved under gauge transformations which tend to the identity with vanishing first derivatives everywhere in infinity. We suppose $\mathrm{Aut}E \subset \mathrm{End}E$ and define the gauge group to be the completion

$$\mathcal{G}_E := \overline{\{\gamma - 1 \in C_0^\infty(\mathrm{End}E) \mid \|\gamma - 1\|_{L^2_{j+2,\Gamma}(M)} < \infty, \gamma \in C^\infty(\mathrm{Aut}E) \text{ a.e.}\}},$$

and the gauge equivalence class of ∇_A under $\mathcal{G}_E$ is denoted by $[\nabla_A]$. Then we are seeking the virtual dimension of the framed moduli space $\mathcal{M}(e, \Gamma)$ of all such connections given up to these specified gauge transformations.

Consider the usual deformation complex

$$L^2_{j+2,\Gamma}(\Lambda^0 M \otimes \mathrm{End}E) \stackrel{\nabla_A}{\longrightarrow} L^2_{j+1,\Gamma}(\Lambda^1 M \otimes \mathrm{End}E) \stackrel{\nabla_A^-}{\longrightarrow} L^2_{j,\Gamma}(\Lambda^- M \otimes \mathrm{End}E),$$

where ∇_A^- refers to the induced connection composed with the projection onto the anti-self-dual side. Our first step is to check that the Betti numbers h^0, h^1, h^- of this complex, given by $h^0 = \dim H^0(\mathrm{End}E)$, etc., are finite. We therefore introduce an elliptic operator

$$\delta_A^* : L^2_{j+1,\Gamma}(\Lambda^1 M \otimes \mathrm{End}E) \longrightarrow L^2_{j,\Gamma}((\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End}E), \qquad (18)$$

the so-called *deformation operator* $\delta_A^* := \nabla_A^* \oplus \nabla_A^-$, which is a conformally invariant first order elliptic operator over $(M, \tilde{g})$, hence (M, g). Here ∇_A^* is the formal L^2 adjoint of ∇_A. We will demonstrate that δ_A^* is Fredholm, so it follows that $h^1 = \dim \mathrm{Ker}\, \delta_A^*$ and $h^0 + h^- = \dim \mathrm{Coker}\, \delta_A^*$ are finite.

Pick up the trivial flat SU(2) connection ∇_Θ on E; it satisfies the strong holonomy condition (12), hence it extends smoothly over X to an operator $\nabla_{\tilde{\Theta}}$ by Theorem 2.4. Using the regularized metric $\tilde{g}_\varepsilon$ of (7) it gives rise to an induced elliptic operator over the compact space $(X, \tilde{g}_\varepsilon)$ as

$$\delta^*_{\varepsilon,\tilde{\Theta}} : C_0^\infty(\Lambda^1 X \otimes \mathrm{End}\tilde{E}_0) \longrightarrow C_0^\infty((\Lambda^0 X \oplus \Lambda^- X) \otimes \mathrm{End}\tilde{E}_0).$$

Consequently $\delta^*_{\varepsilon,\tilde{\Theta}}$ hence its restrictions $\delta^*_{\varepsilon,\tilde{\Theta}}|_M$ and $\delta^*_{\varepsilon,\tilde{\Theta}}|_W$ are Fredholm with respect to any Sobolev completion. We construct a particular completion as follows. The smooth extended connection ∇_Γ on E can be extended further over X to a connection $\nabla_{\tilde{\Gamma}}$ such that it gives rise to an SU(2) Sobolev connection on the trivial bundle $\tilde{E}_0$. Consider a completion like

$$\delta^*_{\varepsilon,\tilde{\Theta}} : L^2_{j+1,\tilde{\Gamma}}(\Lambda^1 X \otimes \mathrm{End}\tilde{E}_0) \longrightarrow L^2_{j,\tilde{\Gamma}}((\Lambda^0 X \oplus \Lambda^- X) \otimes \mathrm{End}\tilde{E}_0). \qquad (19)$$

The operators in (18) and (19) give rise to restrictions. The self-dual connection yields

$$\delta_A^*|_W : L^2_{j+1,\tilde{\Gamma}}(\Lambda^1 W \otimes \mathrm{End}E|_W) \longrightarrow L^2_{j,\tilde{\Gamma}}((\Lambda^0 W \oplus \Lambda^- W) \otimes \mathrm{End}E|_W),$$

while the trivial connection gives

$$\delta^*_{\varepsilon,\tilde{\Theta}}|_W : L^2_{j+1,\tilde{\Gamma}}(\Lambda^1 W \otimes \mathrm{End}\tilde{E}_0|_W) \longrightarrow L^2_{j,\tilde{\Gamma}}((\Lambda^0 W \oplus \Lambda^- W) \otimes \mathrm{End}\tilde{E}_0|_W).$$

Notice that these operators actually act on isomorphic Sobolev spaces, consequently comparing them makes sense. Working in these Sobolev spaces we claim that

Lemma 3.1. *The deformation operator δ^*_A of* (18) *with $j = 0, 1$ satisfies the operator norm inequality*

$$\|(\delta^*_A - \delta^*_{\varepsilon,\tilde{\Theta}})|_{V^*_\varepsilon}\| \leq 3 \cdot 2^j \left(c_3 \|F_A\|_{L^2(V_\varepsilon)} + c_4 m\varepsilon^5 \right), \tag{20}$$

where $m \in [0, 1)$ is the holonomy of ∇_A and $c_4 = c_4(\tilde{g}|_{V_\varepsilon}, \gamma_\alpha, \mathrm{d}\gamma_\alpha) > 0$ is a constant depending only on the metric and the gauge transformations used in (14). *Consequently δ^*_A is a Fredholom operator over (M, g).*

Proof. Consider the restriction of the operators constructed above to the neck $W \cong V^*_\varepsilon$ and calculate the operator norm of their difference with an $a \in L^2_{j+1,\Gamma}(\Lambda^1 M \otimes \mathrm{End} E)$ as follows:

$$\|(\delta^*_A - \delta^*_{\varepsilon,\tilde{\Theta}})|_{V^*_\varepsilon}\| = \sup_{a \neq 0} \frac{\|(\delta^*_A - \delta^*_{\varepsilon,\tilde{\Theta}})a\|_{L^2_{j,\Gamma}(V^*_\varepsilon)}}{\|a\|_{L^2_{j+1,\Gamma}(V^*_\varepsilon)}}.$$

By assumption ∇_A satisfies the weak holonomy condition (11), hence the connection ∇_Γ is flat on $E|_{V^*_\varepsilon}$ and it determines a deformation operator $\delta^*_{\tilde{\Gamma}}|_{V^*_\varepsilon}$ over $(V^*_\varepsilon, \tilde{g}|_{V^*_\varepsilon})$. We use the triangle inequality:

$$\|(\delta^*_A - \delta^*_{\varepsilon,\tilde{\Theta}})|_{V^*_\varepsilon}\| \leq \|(\delta^*_A - \delta^*_{\tilde{\Gamma}})|_{V^*_\varepsilon}\| + \|(\delta^*_{\tilde{\Gamma}} - \delta^*_{\varepsilon,\tilde{\Theta}})|_{V^*_\varepsilon}\|.$$

Referring to the global gauge (13) and the metric $\tilde{g}$ we have $\delta^*_A a = (\delta + \mathrm{d}^-)a + A^* a + (A \wedge a + a \wedge A)^-$, and the same for $\delta^*_{\tilde{\Gamma}}$, consequently

$$(\delta^*_A - \delta^*_{\tilde{\Gamma}})|_{V^*_\varepsilon} = (A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon})^* + ((A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon}) \wedge \cdot + \cdot \wedge (A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon}))^-.$$

Taking $j = 0, 1$ and combining this with (15) we find an estimate for the first term like

$$\|(\delta^*_A - \delta^*_{\tilde{\Gamma}})|_{V^*_\varepsilon}\| \leq 3 \cdot 2^j \|A_{V^*_\varepsilon} - \Gamma_{V^*_\varepsilon}\|_{L^2_{j,\Gamma}(V^*_\varepsilon)} \leq 3 \cdot 2^j c_3 \|F_A\|_{L^2(V_\varepsilon)}.$$

Regarding the second term, the trivial flat connection ∇_Θ on $E|_{V^*_\varepsilon}$ satisfies the strong holonomy condition (12). In the gauge (13) we use, we may suppose simply $\Theta|_{V^*_\varepsilon} = 0$, consequently neither data from $\nabla_{\tilde{\Theta}}$ nor the perturbed metric $\tilde{g}_\varepsilon$ influence this term. Now take a partition of the end into elementary neighbourhoods (5) and use the associated simple constant gauges (10) then

$$\|(\delta^*_{\tilde{\Gamma}} - \delta^*_{\varepsilon,\tilde{\Theta}})|_{V^*_\varepsilon}\| \leq 3 \cdot 2^j \sum_\alpha \|\gamma_\alpha^{-1} \Gamma_m \gamma_\alpha\|_{L^2_{j,\Gamma}(U^*_{\varepsilon,\alpha})} \leq 3 \cdot 2^j c_4 m\varepsilon^5$$

with some constant $c_4 = c_4(\tilde{g}|_{V^*_\varepsilon}, \gamma_\alpha, \ldots, \nabla^j_\Gamma \gamma_\alpha) > 0$ via (6). Putting these together we get (20).

Taking into account that the right-hand side of (20) is arbitrarily small we conclude that $\delta^*_A|_{V^*_\varepsilon}$ is Fredholm because so is $\delta^*_{\varepsilon,\tilde{\Theta}}|_{V^*_\varepsilon}$ and Fredholmness is an open property. Clearly, $\delta^*_A|_{M \setminus V^*_\varepsilon}$ is also Fredholm, because it is an elliptic operator over a compact manifold. Therefore glueing the parametrices of these operators together, one constructs a parametrix for δ^*_A over the whole M (see [24] for an analogous construction). This shows that δ^*_A is Fredholm over the whole $(M, \tilde{g})$ hence (M, g) as claimed. □

We assert that $T_{[\nabla_A]}\mathcal{M}(e,\Gamma) \cong \text{Ker}\ \delta_A^*$, consequently $h^1 = \dim \mathcal{M}(e,\Gamma)$. Indeed, the anti-self-dual part of the curvature of a perturbed connection ∇_{A+a} is given by $F^-_{A+a} = \nabla^-_A a + (a \wedge a)^-$ with the perturbation $a \in L^2_{j+1,\Gamma}(\Lambda^1 M \otimes \text{End}E)$. Therefore if $a \in \text{Ker}\ \delta^*_A$, then both self-duality and the energy of ∇_{A+a} are preserved *infinitesimally* i.e., in first order. The perturbation vanishes everywhere in infinity hence the asymptotics given by Γ is also unchanged; in particular ∇_{A+a} continues to obey the weak holonomy condition.

Incidentally, we note however that *locally*, hence also *globally*, some care is needed when one perturbs a connection in our moduli spaces. For a perturbation with $a \in L^2_{j+1,\Gamma}(\Lambda^1 M \otimes \text{End}E)$ the asymptotics and in particular the weak holonomy condition are obeyed as we have seen. But concerning the energy, by repeating the calculation of Sect. 2 again, we obtain

$$\begin{aligned}
&\frac{1}{\varepsilon}\int\limits_0^\varepsilon |\tau_{\partial M_\rho}(A_\rho + a_\rho) - \tau_{\partial M_\rho}(\Gamma_\rho)|^2 \mathrm{d}\rho \\
&\leq \frac{1}{\varepsilon}\int\limits_0^\varepsilon \left(|\tau_{\partial M_\rho}(A_\rho + a_\rho) - \tau_{\partial M_\rho}(A_\rho)| + |\tau_{\partial M_\rho}(A_\rho) - \tau_{\partial M_\rho}(\Gamma_\rho)|\right)^2 \mathrm{d}\rho \\
&\leq \left(\frac{c_1}{\sqrt{\varepsilon}}\|a\|_{L^2(V_\varepsilon)} + \frac{c_1 c_3}{\sqrt{\varepsilon}}\|F_A\|_{L^2(V_\varepsilon)}\right)^2 .
\end{aligned}$$

If the last line tends to zero as $\varepsilon \to 0$, then the perturbed connection has the same limit as in (17). Consequently, we find that the energy is also unchanged. Since the original connection decays rapidly in the sense of (16) if the perturbations also decay rapidly, i.e., $\lim\limits_{\varepsilon\to 0} \varepsilon^{-\frac{1}{2}}\|a\|_{L^2(V_\varepsilon)} = 0$, then the energy is preseved by local (i.e. small but finite) perturbations as well.

It is therefore convenient to introduce weighted Sobolev spaces with weight $\delta = \frac{1}{2}$ and to say that a and ∇_A *decay rapidly* if $a \in L^2_{\frac{1}{2},j+1,\Gamma}(\Lambda^1 M \otimes \text{End}E)$ and $F_A \in L^2_{\frac{1}{2},j,\Gamma}(\Lambda^2 M \otimes \text{End}E)$, respectively. These are gauge invariant conditions and for ∇_A with $j = 0$ it is equivalent to (16). In this framework the rough estimate

$$\begin{aligned}
\|F_{A+a}\|_{L^2_{\frac{1}{2},j,\Gamma}(M)} &= \|F_A + \nabla_A a + a \wedge a\|_{L^2_{\frac{1}{2},j,\Gamma}(M)} \\
&\leq \|F_A\|_{L^2_{\frac{1}{2},j,\Gamma}(M)} + \|\nabla_A a\|_{L^2_{\frac{1}{2},j,\Gamma}(M)} + c_5\|a\|^2_{L^2_{\frac{1}{2},j+1,\Gamma}(M)}
\end{aligned}$$

with some constant $c_5 = c_5(\tilde{g}) > 0$ implies that ∇_{A+a} also decays rapidly.

Let $\mathcal{A}_E$ denote the affine space of rapidly decaying $\mathrm{SU}(2)$ connections on E as well as $\mathcal{F}^-_E$ the vector space of the anti-self-dual parts of their curvatures. Then take a complex of punctured spaces, the global version of the deformation complex above:

$$(\mathcal{G}_E, 1) \xrightarrow{f_A} (\mathcal{A}_E, \nabla_A) \xrightarrow{\varrho^-_A} (\mathcal{F}^-_E, 0).$$

We have $(\mathcal{A}_E, \nabla_A) \cong L^2_{\frac{1}{2},j+1,\Gamma}(\Lambda^1 M \otimes \text{End}E)$ and $(\mathcal{F}^-_E, 0) \cong L^2_{\frac{1}{2},j,\Gamma}(\Lambda^- M \otimes \text{End}E)$. The global gauge fixing map is defined as $f_A(\gamma) := \gamma^{-1}\nabla_A\gamma - \nabla_A$, while $\varrho^-_A(a) :=$

$F^{-}_{A+a} = \nabla^{-}_{A} a + (a \wedge a)^{-}$. If ∇_A is irreducible then Ker $f_A \cong 1$. One easily shows that if both $\gamma - 1$ and a are pointwise small, then $f_A(\gamma) = \nabla_A(\log \gamma)$ and its formal L^2 adjoint satisfying

$$(f_A(\gamma), a)_{L^2(\mathcal{A}_E, \nabla_A)} = (\gamma, f^*_A(a))_{L^2(\mathcal{G}_E, 1)}$$

looks like $f^*_A(a) = \exp(\nabla^*_A a)$ if

$$(\gamma, \beta)_{L^2(\mathcal{G}_E,1)} := -\int_M \mathrm{tr}(\log \gamma \log \beta) *_{\tilde{g}} 1,$$

defined in a neighbourhood of $1 \in \mathcal{G}_E$. Hence we have a model for $O_{[\nabla_A]} \subset \mathcal{M}(e, \Gamma)$, the vicinity of $[\nabla_A]$, as follows:

$$O_{[\nabla_A]} \cong \mathrm{Ker}(f^*_A \times \varrho^{-}_{A}) \subset (\mathcal{A}_E, \nabla_A).$$

The derivative of $f^*_A \times \varrho^{-}_A$ at a is $\delta^*_A + (a \wedge \cdot + \cdot \wedge a)^{-}$, which is a Fredholm operator, hence $O_{[\nabla_A]}$ is smooth and finite dimensional. These local models match together and prove that the moduli space is indeed a smooth manifold of dimension h^1.

We return to the calculation of h^1. We will calculate the index of δ^*_A in (18) by referring to a relative index theorem. This provides us with the alternating sum $-h^0 + h^1 - h^-$ and then we show that $h^0 = h^- = 0$ via a vanishing theorem.

First we proceed to the calculation of the index of δ^*_A. This will be carried out by a variant of the Gromov–Lawson relative index theorem [20], which we will now explain. First, let us introduce some notation. For any elliptic Fredholm operator P, let $\mathrm{Index}_a P$ denote its analytical index, i.e., $\mathrm{Index}_a P = \dim \mathrm{Ker}\, P - \dim \mathrm{Coker}\, P$. If this P is defined over a compact manifold, $\mathrm{Index}_a P$ is given by a topological formula as in the Atiyah–Singer index theorem, which we denote by $\mathrm{Index}_t P$, the topological index of P. The following theorem will be proved in the Appendix (cf. [23]):

Theorem 3.1. *Let (M, g) be a complete Riemannian manifold, and let X be some smooth compactification of M. Let also $D_1 : L^2_{j+1,\Gamma_1}(F_1) \to L^2_{j,\Gamma'_1}(F'_1)$ and $D_0 : L^2_{j+1,\Gamma_0}(F_0) \to L^2_{j,\Gamma'_0}(F'_0)$ be two first order, elliptic Fredholm operators defined on complex vector bundles F_1, F'_1 and F_0, F'_0 over M with fixed Sobolev connections ∇_{Γ_1}, $\nabla_{\Gamma'_1}$ and ∇_{Γ_0}, $\nabla_{\Gamma'_0}$, respectively. Assume that given $\kappa > 0$, there is a compact subset $K \subset M$ such that, for $W = M \backslash K$, the following hold:*

- (i) *There are bundle isomorphisms $\phi : F_1|_W \cong F_0|_W$ and $\phi' : F'_1|_W \cong F'_0|_W$;*
- (ii) *The operators asymptotically agree, that is, in some operator norm $\|(D_1 - D_0)|_W\| < \kappa$.*

If arbitrary elliptic extensions $\tilde{D}_1$ and $\tilde{D}_0$ of D_1 and D_0 to X exist, then we have

$$\mathrm{Index}_a D_1 - \mathrm{Index}_a D_0 = \mathrm{Index}_t \tilde{D}_1 - \mathrm{Index}_t \tilde{D}_0$$

for the difference of the analytical indices.

In the case at hand, $F_1 = F_0 = \Lambda^1 M \otimes \mathrm{End} E \otimes \mathbb{C}$ and $F'_1 = F'_0 = (\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End} E \otimes \mathbb{C}$; then the complexification of δ^*_A, which by Lemma 3.1 is a Fredholm operator, plays the role of D_1 and via (20) it asymptotically agrees with the complexified $\delta^*_{\varepsilon,\tilde{\Theta}}|_M$, also Fredholm by construction, which replaces D_0.

To find the operators which correspond to $\tilde{D}_1$ and $\tilde{D}_0$, we proceed as follows. Remember that in general ∇_A does not extend over X (cf. Theorem 2.4). Let $\tilde{E}$ be the unique vector bundle over X constructed as follows. Since $\tilde{E}|_{X\setminus V_{2\varepsilon}} \cong E|_{M\setminus V^*_{2\varepsilon}}$ and $\tilde{E}|_{V_\varepsilon} \cong V_\varepsilon \times \mathbb{C}^2$, this bundle is uniquely determined by the glued connection $\nabla_{\tilde{A}} = (1-f_\varepsilon)\nabla_A + f_\varepsilon \nabla_{\tilde{\Theta}}$ on $\tilde{E}$, where f_ε is taken from (7). We can construct the associated operator $\delta^*_{\varepsilon,\tilde{A}}$ over $\mathrm{End}\tilde{E}$ with respect to the metric (7) whose complexification will correspond to $\tilde{D}_1$; the operator $\tilde{D}_0$ is given by the complexification of the operator $\delta^*_{\varepsilon,\tilde{\Theta}}$ on the trivial bundle $\mathrm{End}\tilde{E}_0$, which we have already constructed.

The right-hand side of the relative index formula in Theorem 3.1 is given by

$$\mathrm{Index}_t(\delta^*_{\varepsilon,\tilde{A}}) - \mathrm{Index}_t(\delta^*_{\varepsilon,\tilde{\Theta}}) = 8k - 3(1-b^1(X)+b^-(X)) + 3(1-b^1(X)+b^-(X)) = 8k,$$

where $k = \|F_{\tilde{A}}\|^2_{L^2(X)}$ is the second Chern number of the extended bundle $\tilde{E}$. Notice that this number might be different from the energy $e = \|F_A\|^2_{L^2(M,g)}$ of the original connection. We only know a priori that $k \leq e$. However we claim that

Lemma 3.2. *Using the notation of Theorem 2.2 and, in the same fashion, if $\nabla_\Theta = \mathrm{d}+\Theta_W$ on $E|_W$, then letting $\Theta_\infty := \lim\limits_{r\to\infty} \Theta_W|_{N\times\{r\}}$, in any smooth gauge*

$$k = e + \tau_N(\Theta_\infty) - \tau_N(\Gamma_\infty) \tag{21}$$

holds. Notice that this expression is gauge invariant.

Remark. Of course in any practical application it is worth taking the gauge in which simply $\tau_N(\Theta_\infty) = 0$.

Proof. Using the gauge (13) for instance and applying the Chern–Simons theorem for the restricted energies to M_ε, we find

$$(k-e)|_{M_\varepsilon} = \tau_{\partial M_\varepsilon}(\tilde{A}_\varepsilon) - \tau_{\partial M_\varepsilon}(A_\varepsilon) + 2\|F^-_{\tilde{A}}\|^2_{L^2(M_\varepsilon)}.$$

Therefore, since $\|F^-_{\tilde{A}}\|_{L^2(M_\varepsilon)} = \|F^-_{\tilde{A}}\|_{L^2(V_{2\varepsilon})}$ and the Chern–Simons invariants converge as in (17) by the rapid decay assumption, taking the limit one obtains

$$\begin{aligned} k - e &= \tau_N(\Theta_0) - \tau_N(\Gamma_0) + 2\lim_{\varepsilon\to 0}\|F^-_{\tilde{A}}\|^2_{L^2(V_{2\varepsilon})} \\ &= \tau_N(\Theta_\infty) - \tau_N(\Gamma_\infty) + 2\lim_{R\to\infty}\|F^-_{\tilde{A}}\|^2_{L^2(V_{2R},g|_{V_{2R}})}. \end{aligned}$$

Consequently we have to demonstrate that $\lim\limits_{\varepsilon\to 0}\|F^-_{\tilde{A}}\|_{L^2(V_{2\varepsilon})} = 0$. There is a decomposition of the glued curvature like $F_{\tilde{A}} = \Phi + \varphi$ with $\Phi := (1-f_\varepsilon)F_A + f_\varepsilon F_\Theta$, and a perturbation term as follows:

$$\varphi := -\mathrm{d}f_\varepsilon \wedge (A_{V^*_{2\varepsilon}} - \Theta_{V^*_{2\varepsilon}}) - f_\varepsilon(1-f_\varepsilon)(A_{V^*_{2\varepsilon}} - \Theta_{V^*_{2\varepsilon}}) \wedge (A_{V^*_{2\varepsilon}} - \Theta_{V^*_{2\varepsilon}}).$$

This shows that $F^-_{\tilde{A}} = \varphi^-$, consequently it is compactly supported in $V_{2\varepsilon}\setminus V_\varepsilon$. Moreover there is a constant $c_6 = c_6(\mathrm{d}f_\varepsilon, \tilde{g}|_{V_{2\varepsilon}}) > 0$, independent of ε, such that $\|\mathrm{d}f_\varepsilon\|_{L^2_{1,\Gamma}(V_{2\varepsilon})} \leq$

c_6; as well as $|f_\varepsilon(1-f_\varepsilon)| \leq \frac{1}{4}$ therefore, recalling the pattern of the proof of Lemma 3.1, we tame φ^- like

$$\|\varphi^-\|_{L^2(V_{2\varepsilon})} \leq \|\varphi\|_{L^2(V_{2\varepsilon})} \leq c_5c_6\|A_{V^*_{2\varepsilon}} - \Theta_{V^*_{2\varepsilon}}\|_{L^2_{1,\Gamma}(V^*_{2\varepsilon})} + \frac{c_5}{4}\|A_{V^*_{2\varepsilon}} - \Theta_{V^*_{2\varepsilon}}\|^2_{L^2_{1,\Gamma}(V^*_{2\varepsilon})} \leq$$
$$\leq c_5c_6\left(c_3\|F_A\|_{L^2(V_{2\varepsilon})} + c_4m(2\varepsilon)^5\right) + \frac{c_5}{4}\left(c_3\|F_A\|_{L^2(V_{2\varepsilon})} + c_4m(2\varepsilon)^5\right)^2.$$

However we know that this last line can be kept as small as one likes providing the result. □

Regarding the left-hand side of Theorem 3.1, on the one hand we already know that

$$\mathrm{Index}_a\delta^*_A = -h^0 + h^1 - h^- = h^1 = \dim\mathcal{M}(e,\Gamma)$$

by the promised vanishing theorem. On the other hand, since $\mathrm{End}E\otimes\mathbb{C}\cong M\times\mathbb{C}^3$, we find

$$\mathrm{Index}_a(\delta^*_{\varepsilon,\tilde{\Theta}}|_M) = -3\left(b^0_{L^2}(M,\tilde{g}_\varepsilon|_M) - b^1_{L^2}(M,\tilde{g}_\varepsilon|_M) + b^-_{L^2}(M,\tilde{g}_\varepsilon|_M)\right),$$

where $b^i_{L^2}(M,\tilde{g}_\varepsilon|_M)$ is the i^{th} L^2 Betti number and $b^-_{L^2}(M,\tilde{g}_\varepsilon|_M)$ is the dimension of the space of anti-self-dual finite energy 2-forms on the rescaled-regularized manifold $(M,\tilde{g}_\varepsilon|_M)$, i.e. this index is the truncated L^2 Euler characteristic of $(M,\tilde{g}_\varepsilon|_M)$. We wish to cast this subtle invariant into a more explicit form at the expense of imposing a further but natural assumption on the spaces we work with.

Lemma 3.3. *Let (M,g) be a Ricci flat ALF space, and let X be its compactification with induced orientation. Then one has* $\mathrm{Index}_a(\delta^*_{\varepsilon,\tilde{\Theta}}|_M) = -3b^-(X)$*, where $b^-(X)$ denotes the rank of the negative definite part of the topological intersection form of X.*

Proof. Exploiting the stability of the index against small perturbations as well as the conformal invariance of the operator $\delta^*_{\varepsilon,\tilde{\Theta}}|_M$, without changing the index we can replace the metric $\tilde{g}_\varepsilon$ with the original ALF metric g. Consequently we can write

$$\mathrm{Index}_a(\delta^*_{\varepsilon,\tilde{\Theta}}|_M) = -3\left(b^0_{L^2}(M,g) - b^1_{L^2}(M,g) + b^-_{L^2}(M,g)\right)$$

for the index we are seeking.

Remember that this metric is complete and asymptotically looks like (2). This implies that (M,g) has infinite volume, hence a theorem of Yau [36] yields that $b^0_{L^2}(M,g)=0$. Moreover if we assume the curvature of (M,g) not only satisfies (3) but is furthermore Ricci flat, then a result of Dodziuk [13] shows that in addition $b^1_{L^2}(M,g)=0$. Concerning $b^-_{L^2}(M,g)$ we use the result of [15] (based on [21, Corollary 7]) to observe that any finite energy anti-self-dual 2-form over (M,g) extends smoothly as a (formally) anti-self-dual 2-form over $(X,\tilde{g})$ showing that $b^-_{L^2}(M,g)=b^-(X)$ as desired. □

Finally we prove the vanishing of the numbers h^0 and h^-. The proof is a combination of the standard method [2] and a Witten-type vanishing result ([31, Lemma 4.3]). In the adjoint of the elliptic complex (18) we find $\mathrm{Index}_a\delta_A = \dim\mathrm{Ker}\,\delta_A - \dim\mathrm{Coker}\,\delta_A = h^0 - h^1 + h^-$, hence proving $\mathrm{Ker}\,\delta_A = \{0\}$ is equivalent to $h^0 = h^- = 0$.

Lemma 3.4. *Assume (M, g) is an ALF space as defined in Sect. 2. If ∇_A is irreducible and $\psi \in L^2_{j,\Gamma}((\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End}E)$ satisfies $\delta_A \psi = 0$, then $\psi = 0$.*

Proof. Taking into account that $\mathrm{Ker}\,\delta_A = \mathrm{Ker}(\delta_A^* \delta_A)$ consists of smooth functions by elliptic regularity and is conformally invariant, we can use the usual Weitzenböck formula with respect to the original ALF metric g as follows:

$$\begin{aligned} \delta_A^* \delta_A \quad &= \quad \nabla_A^* \nabla_A + W_g^- + \frac{s_g}{3} : \quad C_0^\infty((\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End}E) \\ &\longrightarrow C_0^\infty((\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End}E). \end{aligned}$$

In this formula $\nabla_A : C_0^\infty((\Lambda^0 M \oplus \Lambda^- M) \otimes \mathrm{End}E) \to C_0^\infty((\Lambda^1 M \oplus \Lambda^3 M) \otimes \mathrm{End}E)$ is the induced connection while $W_g^- + \frac{s_g}{3}$ acts only on the $\Lambda^- M$ summand as a symmetric, linear, algebraic map. This implies that if $\delta_A \psi = 0$ such that $\psi \in L^2_{j,\Gamma}(\Lambda^0 M \otimes \mathrm{End}E)$ only or $W_g^- + \frac{s_g}{3} = 0$, then $\nabla_A \psi = 0$. If ∇_A is irreducible, then both $\Lambda^0 M \otimes \mathrm{End}E \otimes \mathbb{C} \cong S^0\Sigma^- \otimes \mathrm{End}E$ and $\Lambda^- M \otimes \mathrm{End}E \otimes \mathbb{C} \cong S^2\Sigma^- \otimes \mathrm{End}E$ are irreducible $\mathrm{SU}(2)^- \times \mathrm{SU}(2)$ bundles, hence $\psi = 0$ follows.

Concerning the generic case, then as before, we find $h^0 = 0$ by irreducibility, consequently we can assume $\psi \in L^2_{j,\Gamma}(\Lambda^- M \otimes \mathrm{End}E)$ only and $W_g^- + \frac{s_g}{3} \neq 0$. Since $\mathrm{Ker}\,\delta_A$ consists of smooth functions it follows that if $\delta_A \psi = 0$, then ψ vanishes everywhere at infinity.

Let $\langle \cdot\,, \cdot \rangle$ be a pointwise SU(2)-invariant scalar product on $\Lambda^- M \otimes$ EndE and set $|\psi| := \langle \psi, \psi \rangle^{\frac{1}{2}}$. Assume $\delta_A \psi = 0$ but $\psi \neq 0$. Then $\langle \nabla_A^* \nabla_A \psi, \psi \rangle = -\langle (W_g^- + \frac{s_g}{3})\psi, \psi \rangle$ by the Weitzenböck formula above. Combining this with the pointwise expression $\langle \nabla_A^* \nabla_A \psi, \psi \rangle = |\nabla_A \psi|^2 + \frac{1}{2} \triangle |\psi|^2$ and applying $\frac{1}{2}\triangle|\psi|^2 = |\psi| \triangle |\psi| + |\mathrm{d}|\psi||^2$ as well as Kato's inequality $|\mathrm{d}|\psi|| \leq |\nabla_A \psi|$, valid away from the zero set of ψ, we obtain

$$2\, \frac{|\mathrm{d}|\psi||^2}{|\psi|^2} \leq \left| W_g^- + \frac{s_g}{3} \right| - \frac{\triangle |\psi|}{|\psi|} = \left| W_g^- + \frac{s_g}{3} \right| - \triangle \log |\psi| - \frac{|\mathrm{d}|\psi||^2}{|\psi|^2},$$

that is,

$$3\,|\mathrm{d}\log|\psi||^2 + \triangle \log |\psi| \leq \left| W_g^- + \frac{s_g}{3} \right|.$$

Let $\lambda : \mathbb{R}^+ \to W \cong N \times \mathbb{R}^+$ be a naturally parameterized ray running toward infinity and let $f(r) := \log|\psi(\lambda(r))|$. Observe that f is negative in the vicinity of a zero of ψ or for large r's. The last inequality then asymptotically cuts down along λ to

$$3(f'(r))^2 - \frac{2}{r} f'(r) - f''(r) \leq \frac{c_7}{r^3},$$

using the expansion of the Laplacian for a metric like (2) and referring to the curvature decay (3) providing a constant $c_7 = c_7(g) \geq 0$. This inequality yields $-c_7 r^{-2} \leq (rf(r))''$. Integrating it we find $c_7 r^{-1} + a \leq (rf(r))' \leq rf'(r)$, showing $c_7 r^{-2} + ar^{-1} \leq f'(r)$ with some constant a, hence there is a constant $c_8 \geq 0$ such that $c_8 := -|\inf\limits_{r \in \mathbb{R}^+} f'(r)|$. Integrating $c_7 r^{-1} + a \leq (rf(r))'$ again we also obtain $c_7 (\log r) r^{-1} + a + br^{-1} \leq f(r) \leq 0$, for some real constants a, b.

Let $x_0 \in M$ be such that $\psi(x_0) \neq 0$; then by smoothness there is another point $x \in M$ with this property such that $|x_0| < |x|$. Integrating again the inequality $-c_7 r^{-2} \leq (rf(r))''$ twice from x_0 to x along the ray λ connecting them we finally get

$$|\psi(x_0)| \leq |\psi(x)| \exp\left((c_7|x_0|^{-1} + c_8|x_0|)(1 - |x_0||x|^{-1})\right).$$

Therefore either letting x be a zero of ψ along λ or, if no such point exists, taking the limit $|x| \to \infty$, we find that $\psi(x_0) = 0$ as desired. $\square$

Finally, putting all of our findings together, we have arrived at the following theorem:

Theorem 3.2. *Let (M, g) be an ALF space with an end $W \cong N \times \mathbb{R}^+$ as before. Assume furthermore that the metric is Ricci flat. Consider a rank 2 complex $SU(2)$ vector bundle E over M, necessarily trivial, and denote by $\mathcal{M}(e, \Gamma)$ the framed moduli space of smooth, irreducible, self-dual $SU(2)$ connections on E satisfying the weak holonomy condition (11) and decaying rapidly in the sense of (16) such that their energy $e < \infty$ is fixed and are asymptotic to a fixed smooth flat connection ∇_Γ on $E|_W$.*

Then $\mathcal{M}(e, \Gamma)$ is either empty or a manifold of dimension

$$\dim \mathcal{M}(e, \Gamma) = 8\,(e + \tau_N(\Theta_\infty) - \tau_N(\Gamma_\infty)) - 3b^-(X),$$

where ∇_Θ is the trivial flat connection on $E|_W$ and τ_N is the Chern–Simons functional of the boundary, while X is the Hausel–Hunsicker–Mazzeo compactification of M with induced orientation.

Remark. Of course, we get a dimension formula for anti-instantons by replacing $b^-(X)$ with $b^+(X)$. Notice that our moduli spaces contain framings, since we have a fixed flat connection and a gauge at infinity. The virtual dimension of the moduli space of unframed instantons is given by $\dim \mathcal{M}(e, \Gamma) - 3$, which is is the number of effective free parameters.

A dimension formula in the presence of a magnetic term μ mentioned in Sect. 2 is also easy to work out because in this case (21) is simply replaced with $k = e + \tau_N(\Theta_\infty) - \tau_N(\Gamma_\infty) - \mu$ and then this should be inserted into the dimension formula of Theorem 3.2.

Note also that our moduli spaces are naturally endowed with weighted L^2 metrics. An interesting problem is to investigate the properties of these metrics.

4. Case Studies

In this section we present some applications of Theorem 3.2. We will consider rapidly decaying instantons over the flat $\mathbb{R}^3 \times S^1$, the multi-Taub–NUT geometries and the Riemannian Schwarzschild space. We also have the aim to enumerate the known Yang–Mills instantons over non-trivial ALF geometries. However we acknowledge that our list is surely incomplete, cf. e.g. [1].

The flat space $\mathbb{R}^3 \times S^1$. This is the simplest ALF space, hence instanton (or also called caloron i.e., instanton at finite temperature) theory over this space is well-known (cf. [4,5]). We claim that

Theorem 4.1. *Take $M = \mathbb{R}^3 \times S^1$ with a fixed orientation and put the natural flat metric onto it. Let ∇_A be a smooth, self-dual, rapidly decaying $SU(2)$ connection on a fixed rank 2 complex vector bundle E over M. Then ∇_A satisfies the weak holonomy condition and has non-negative integer energy. Let $\mathcal{M}(e, \Gamma)$ denote the framed moduli space of these connections which are moreover irreducible as in Theorem 3.2. Then*

$$\dim \mathcal{M}(e, \Gamma) = 8e$$

and $\mathcal{M}(e, \Gamma)$ is not empty for all $e \in \mathbb{N}$.

Proof. Since the metric is flat, the conditions of Theorem 3.2 are satisfied. Furthermore, in the case at hand the asymptotical topology of the space is $W \cong S^2 \times S^1 \times \mathbb{R}^+$, consequently $N \cong S^2 \times S^1$, hence its character variety $\chi(S^2 \times S^1) \cong [0, 1)$ is connected. Theorem 2.3 therefore guarantees that any smooth, self-dual, rapidly decaying connection has non-negative integer energy. In the gauge in which $\tau_{S^2 \times S^1}(\Theta_\infty) = 0$, we also get $\tau_{S^2 \times S^1}(\Gamma_\infty) = 0$. Moreover we find that $X \cong S^4$ for the Hausel–Hunsicker–Mazzeo compactification [15] yielding $b^-(X) = b^+(X) = 0$; putting these data into the dimension formula in Theorem 3.2 we get the dimension as stated, in agreement with [4].

The moduli spaces $\mathcal{M}(e, \Gamma)$ are not empty for all $e \in \mathbb{N}$; explicit solutions with arbitrary energy were constructed via a modified ADHM construction in [4,5]. $\square$

The Multi-Taub–NUT (or A_k ALF, or ALF Gibbons–Hawking) spaces. The underlying manifold M_V topologically can be understood as follows. There is a circle action on M_V with s distinct fixed points $p_1, \dots, p_s \in M_V$, called NUTs. The quotient is $\mathbb{R}^3$ and we denote the images of the fixed points also by $p_1, \dots, p_s \in \mathbb{R}^3$. Then $U_V := M_V \setminus \{p_1, \dots, p_s\}$ is fibered over $Z_V := \mathbb{R}^3 \setminus \{p_1, \dots, p_s\}$ with S^1 fibers. The degree of this circle bundle around each point p_i is one.

The metric g_V on U_V looks like (cf. e.g. [14, p. 363])

$$\mathrm{d}s^2 = V(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2) + \frac{1}{V}(\mathrm{d}\tau + \alpha)^2,$$

where $\tau \in (0, 8\pi m]$ parameterizes the circles and $x = (x, y, z) \in \mathbb{R}^3$; the smooth function $V : Z_V \to \mathbb{R}$ and the 1-form $\alpha \in C^\infty(\Lambda^1 Z_V)$ are defined as follows:

$$V(x, \tau) = V(x) = 1 + \sum_{i=1}^{s} \frac{2m}{|x - p_i|}, \qquad \mathrm{d}\alpha = *_3 \mathrm{d}V.$$

Here $m > 0$ is a fixed constant and $*_3$ refers to the Hodge-operation with respect to the flat metric on $\mathbb{R}^3$. We can see that the metric is independent of τ, hence we have a Killing field on (M_V, g_V). This Killing field provides the above mentioned $U(1)$-action. Furthermore it is possible to show that, despite the apparent singularities in the NUTs, these metrics extend analytically over the whole M_V providing an ALF, hyper–Kähler manifold. We also notice that the Killing field makes it possible to write a particular Kähler-form in the hyper-Kähler family as $\omega = \mathrm{d}\beta$, where β is a 1-form of linear growth.

Then we can assert that

Theorem 4.2. *Let (M_V, g_V) be a multi-Taub–NUT space with s NUTs and orientation induced by any of the complex structures in the hyper–Kähler family. Consider the framed moduli space $\mathcal{M}(e, \Gamma)$ of smooth, irreducible, rapidly decaying anti-self-dual connections satisfying the weak holonomy condition on a fixed rank 2 complex $SU(2)$ vector bundle E as in Theorem 3.2. Then $\mathcal{M}(e, \Gamma)$ is either empty or a manifold of dimension*

$$\dim \mathcal{M}(e, \Gamma) = 8\left(e + \tau_{L(s,1)}(\Theta_\infty) - \tau_{L(s,1)}(\Gamma_\infty)\right),$$

where $L(s, 1)$ is the lens space representing the boundary of M_V. The moduli spaces are surely not empty for $\tau_{L(s,1)}(\Theta_\infty) = \tau_{L(s,1)}(\Gamma_\infty) = 0$ and $e = 1, \ldots, s$.

Proof. This space is non-flat, nevertheless its curvature satisfies the cubic curvature decay (3), hence it is an ALF space in our sense. Since it is moreover hyper–Kähler, the conditions of Theorem 3.2 are satisfied. However this time the asymptotic topology is $W \cong L(s, 1) \times \mathbb{R}^+$, therefore $N \cong L(s, 1)$ is a non-trivial circle bundle over S^2; consequently the weak holonomy condition (11) must be imposed. If the connection in addition decays rapidly as in (16) then its energy is determined by a Chern–Simons invariant via Theorem 2.2. The character variety of the boundary lens space, $\chi(L(s, 1))$ is also non-connected if $s > 1$ and each connected component has a non-trivial fractional Chern–Simons invariant which is calculable (cf., e.g. [3,25]). By the result in [15] the compactified space X with its induced orientation is isomorphic to the connected sum of s copies of $\overline{\mathbb{C}P}^2$'s, therefore $b^+(X) = 0$ and $b^-(X) = s$. Inserting these into the dimension formula of Theorem 3.2 for anti-self-dual connections we get the dimension.

Concerning non-emptiness, since lacking a general ADHM-like construction, we may use a conformal rescaling method [17,18]. Take the natural orthonormal frame

$$\xi^0 = \frac{1}{\sqrt{V}}(\mathrm{d}\tau + \alpha), \quad \xi^1 = \sqrt{V}\mathrm{d}x, \quad \xi^2 = \sqrt{V}\mathrm{d}y, \quad \xi^3 = \sqrt{V}\mathrm{d}z$$

over U_V and introduce the quaternion-valued 1-form $\boldsymbol{\xi} := \xi^0 + \xi^1\mathbf{i} + \xi^2\mathbf{j} + \xi^3\mathbf{k}$. Moreover pick up the non-negative function $f : U_V \to \mathbb{R}^+$ defined as

$$f(x) := \lambda_0 + \sum_{i=1}^{s} \frac{\lambda_i}{|x - p_i|}$$

with $\lambda_0, \lambda_1, \ldots, \lambda_s$ being real non-negative constants and also take the quaternion-valued 0-form

$$\mathbf{d}\log f := -V\frac{\partial \log f}{\partial \tau} + \frac{\partial \log f}{\partial x}\mathbf{i} + \frac{\partial \log f}{\partial y}\mathbf{j} + \frac{\partial \log f}{\partial z}\mathbf{k}$$

(notice that actually $\frac{\partial \log f}{\partial \tau} = 0$). Over $U_V \subset M_V$ we have a gauge induced by the above orthonormal frame on the positive spinor bundle $\Sigma^+|_{U_V}$. In this gauge consider 't Hooft-like SU(2) connections $\nabla^+_{\lambda_0,\ldots,\lambda_s}|_{U_V} := \mathrm{d} + A^+_{\lambda_0,\ldots,\lambda_s,U_V}$ on $\Sigma^+|_{U_V}$ with

$$A^+_{\lambda_0,\ldots,\lambda_s,U_V} := \operatorname{Im}\frac{(\mathbf{d}\log f)\,\boldsymbol{\xi}}{2\sqrt{V}}.$$

It was demonstrated in [18] that these connections, parameterized by $\lambda_0, \lambda_1, \ldots, \lambda_s$ up to an overall scaling, extend over M_V and provide smooth, rapidly decaying anti-self-dual connections on Σ^+. They are irreducible if $\lambda_0 > 0$, are non-gauge equivalent and have trivial holonomy at infinity; hence satisfy the weak holonomy condition (11) and in particular the corresponding Chern–Simons invariants vanish. Consequently their energies are always integers equal to $e = n$, where $0 \leq n \leq s$ is the number of non-zero λ_i's with $i = 1, \ldots, s$. Therefore moduli spaces consisting of anti-instantons of these energies cannot be empty. $\square$

Remark. It is reasonable to expect that the higher energy moduli spaces are also not empty. Furthermore, it is not clear whether fractionally charged, rapidly decaying irreducible instantons actually exist over multi-Taub–NUT or over more general ALF spaces at all.

Consider the family ∇_i^+ defined by the function $f_i(x) = \frac{\lambda_i}{|x-p_i|}$. These are unital energy solutions which are reducible to U(1) (in fact these are the only reducible points in the family, cf. [18]). These provide as many as $s = 1 + b^2(M_V)$ non-equivalent reducible, rapidly decaying anti-self-dual solutions and if $\nabla_i^+|_{U_V} = \mathrm{d} + A^+_{i,U_V}$, then the connection $\nabla^+|_{U_V} := \mathrm{d} + \sum\limits_i A^+_{i,U_V}$ decays rapidly, is reducible and non-topological; hence it admits arbitrary rescalings yielding the strange continuous energy solutions mentioned in Sect. 2. Of course these generically rescaled connections violate the weak holonomy condition.

The Riemannian (or Euclidean) Schwarzschild space. The underlying space is $M = S^2 \times \mathbb{R}^2$. We have a particularly nice form of the metric g on a dense open subset $(\mathbb{R}^2\backslash\{0\}) \times S^2 \subset M$ of the Riemannian Schwarzschild manifold. It is convenient to use polar coordinates (r, τ) on $\mathbb{R}^2\backslash\{0\}$ in the range $r \in (2m, \infty)$ and $\tau \in [0, 8\pi m)$, where $m > 0$ is a fixed constant. The metric then takes the form

$$\mathrm{d}s^2 = \left(1 - \frac{2m}{r}\right)\mathrm{d}\tau^2 + \left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2,$$

where $\mathrm{d}\Omega^2$ is the line element of the round sphere. In spherical coordinates $\theta \in (0, \pi)$ and $\varphi \in [0, 2\pi)$ it is $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$ on the open coordinate chart $(S^2\backslash(\{S\} \cup \{N\})) \subset S^2$. Consequently the above metric takes the following form on the open, dense coordinate chart $U := (\mathbb{R}^2\backslash\{0\}) \times S^2\backslash(\{S\} \cup \{N\}) \subset M$:

$$\mathrm{d}s^2 = \left(1 - \frac{2m}{r}\right)\mathrm{d}\tau^2 + \left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2).$$

The metric can be extended analytically to the whole M as a complete Ricci flat metric, however this time $W^\pm \neq 0$. Nevertheless we obtain

Theorem 4.3. *Let (M, g) be the Riemannian Schwarzschild manifold with a fixed orientation. Let ∇_A be a smooth, rapidly decaying, self-dual $SU(2)$ connection on a fixed rank 2 complex vector bundle E over M. Then ∇_A satisfies the weak holonomy condition and has non-negative integer energy. Let $\mathcal{M}(e, \Gamma)$ denote the framed moduli space of these connections which are moreover irreducible as in Theorem 3.2. Then it is either empty or a manifold of dimension*

$$\dim \mathcal{M}(e, \Gamma) = 8e - 3.$$

The moduli space with $e = 1$ is surely non-empty.

Proof. The metric is Ricci flat, moreover a direct calculation shows that both $W^{\pm}$ satisfy the decay (3) hence Theorem 3.2 applies in this situation as well. Furthermore, the asymptotical topology and the character variety of the space is again $W \cong S^2 \times S^1 \times \mathbb{R}^+$ and $\chi(S^2 \times S^1) \cong [0, 1)$, consequently the energy of a rapidly decaying instanton is integer as in Theorem 4.1. We can again set the gauge in which all Chern–Simons invariants vanish and find moreover $X \cong S^2 \times S^2$ yielding $b^-(X) = b^+(X) = 1$; substituting these data into the dimension formula we get the desired result.

Regarding non-emptiness, very little is known. The apparently different non-Abelian solutions found by Charap and Duff (cf. the 1-parameter family (I) in [6]) are in fact all gauge equivalent [34] and provide a single rapidly decaying self-dual connection which is the positive chirality spin connection. It looks like $\nabla^+|_U := \mathrm{d} + A^+_U$ on $\Sigma^+|_U$ with

$$A^+_U := \frac{1}{2}\sqrt{1-\frac{2m}{r}}\mathrm{d}\theta\,\mathbf{i} + \frac{1}{2}\sqrt{1-\frac{2m}{r}}\sin\theta\ \mathrm{d}\varphi\,\mathbf{j} + \frac{1}{2}\left(\cos\theta\mathrm{d}\varphi - \frac{m}{r^2}\mathrm{d}\tau\right)\mathbf{k}.$$

One can show that this connection extends smoothly over Σ^+ as an $\mathrm{SO}(3) \times \mathrm{U}(1)$ invariant, irreducible, self-dual connection of unit energy, centered around the 2-sphere in the origin. □

Remark. Due to its resistance against deformations over three decades, it has been conjectured that this positive chirality spin connection is the only unit energy instanton over the Schwarzschild space (cf. e.g. [34]). However we can see now that in fact it admits a 2 parameter deformation. It would be interesting to find these solutions explicitly as well as construct higher energy irreducible solutions.

In their paper Charap and Duff exhibit another family of $\mathrm{SO}(3) \times \mathrm{U}(1)$ invariant instantons ∇_n of energies $2n^2$ with $n \in \mathbb{Z}$ (cf. solutions of type (II) in [6]). However it was pointed out in [16] that these solutions are in fact reducible to $\mathrm{U}(1)$ and locally look like $\nabla_n|_{U^{\pm}} = \mathrm{d} + A_{n,U^{\pm}}$ with

$$A_{n,U^{\pm}} := \frac{n}{2}\left((\mp 1 + \cos\theta)\mathrm{d}\varphi - \frac{1}{r}\mathrm{d}\tau\right)\mathbf{k}$$

over the charts $U^{\pm}$ defined by removing the north or the south poles from S^2 respectively. They extend smoothly as *slowly decaying* reducible, self-dual connections over the bundles $L_n \oplus L_n^{-1}$, where L_n is a line bundle with $c_1(L) = n$. Hence they are topological in contrast to the above mentioned Abelian instantons over the multi-Taub–NUT space. This constrains them to have discrete energy spectrum despite their slow decay. It is known that these are the only reducible $\mathrm{SU}(2)$ instantons over the Riemannian Schwarzschild space [16,21].

5. Appendix

In this Appendix we shall prove Theorem 3.1; this proof is taken from [23], and follows closely the arguments of [20]. In the course of the proof we shall use the notation introduced in the bulk of the paper.

The right-hand side of the index formula in Theorem 3.1 is called the *relative topological index* of the operators D_0 and D_1:

$$\mathrm{Index}_t(D_1, D_0) := \mathrm{Index}_t\tilde{D}_1 - \mathrm{Index}_t\tilde{D}_0.$$

Notice that it can be computed in terms of the topology of the topological extensions of the bundles F_j and F'_j to X, using the Atiyah–Singer index theorem. Furthermore, as we will see below, this quantity does not depend on how the operators D_0 and D_1 are extended to $\tilde{D}_0$ and $\tilde{D}_1$ (see Lemma 5.1 below).

The first step in the proof of Theorem 3.1 is the construction of a new Fredholm operator D'_1 as follows. Let β_1 and β_2 be cut-off functions, respectively supported over K and $W = M\backslash K$, and define

$$D'_1 = \beta_1 D_1 \beta_1 + \beta_0 D_0 \beta_0.$$

Now, it is clear that $D'_1|_W$ coincides with $D_0|_W$. Furthermore, since $\|(D'_1 - D_1)|_W\| < \kappa$ with κ arbitrarily small, we know that $\mathrm{Index}_a D'_1 = \mathrm{Index}_a D_1$. Our strategy is to establish the index formula for the pair D'_1 and D_0. In order to simplify notation however, we will continue to denote by D_1 and D_0 a pair of elliptic Fredholm operators which *coincide* at infinity.

Now recall that if D is any Fredholm operator over M, there is a bounded, elliptic pseudo-differential operator Q, called the *parametrix* of D, such that $DQ = I - S$ and $QD = I - S'$, where S and S' are compact *smoothing operators*, and I is the identity operator. Note that neither Q nor S and S' are unique. In particular, there is a bounded operator G, called the *Green's operator* for D, satisfying $DG = I - H$ and $GD = I - H'$, where H and H' are finite rank *projection operators*; the image of H is $\mathrm{Ker}D$ and the image of H' is $\mathrm{Coker}D$.

Let $K^H(x, y)$ be the Schwartzian kernel of the operator H. Its local trace function is defined by $\mathrm{tr}[H](x) = K^H(x, x)$; moreover, these are C^∞ functions [20]. If D is Fredholm, its (analytical) index is given by

$$\mathrm{Index}_a D = \dim \mathrm{Ker}D - \dim \mathrm{Coker}D = \int_M \big(\mathrm{tr}[H] - \mathrm{tr}[H']\big) \tag{22}$$

as it is well-known; recall that compact operators have smooth, square integrable kernels. Furthermore, if M is a closed manifold, we have [20]

$$\mathrm{Index}_t D = \int_M \big(\mathrm{tr}[S] - \mathrm{tr}[S']\big)\,.$$

Let us now return to the situation set up above. Consider the parametrices and Green's operators ($j = 0, 1$),

$$\begin{cases} D_j Q_j = I - S_j \\ Q_j D_j = I - S'_j \end{cases} \quad \begin{cases} D_j G_j = I - H_j \\ G_j D_j = I - H'_j. \end{cases} \tag{23}$$

The strategy of proof is to express both sides of the index formula of Theorem 3.1 in terms of integrals, as in (22); for its left-hand side, the *relative analytical index*, we have

$$\begin{aligned} \mathrm{Index}_a(D_1, D_0) &:= \mathrm{Index}_a D_1 - \mathrm{Index}_a D_0 \\ &= \int_M \big(\mathrm{tr}[H_1] - \mathrm{tr}[H'_1]\big) - \int_M \big(\mathrm{tr}[H_0] - \mathrm{tr}[H'_0]\big)\,. \end{aligned} \tag{24}$$

For the relative topological index, we have the following

Lemma 5.1. *Under the hypothesis of Theorem 3.1, we have that*

$$\mathrm{Index}_t(D_1, D_0) = \int_M \left(\mathrm{tr}[S_1] - \mathrm{tr}[S_1']\right) - \int_M \left(\mathrm{tr}[S_0] - \mathrm{tr}[S_0']\right). \tag{25}$$

Proof. Denote by $\widetilde{W} \subset X$ the compactification of the end $W \subset M$. Extend D_j $(j = 0, 1)$ to operators $\tilde{D}_j$, both defined over the whole X. Then the parametrices $\tilde{Q}_j$ of $\tilde{D}_j$ are extensions of the parametrices Q_j of D_j, and the corresponding compact smoothing operators $\tilde{S}_j$ and $\tilde{S}_j'$ are extensions of S_j and S_j'.

As explained above, we can assume that the operators D_1 and D_0 coincide at infinity. This means that $D_1|_W \simeq D_0|_W$, hence also $\tilde{D}_1|_{\widetilde{W}} \simeq \tilde{D}_0|_{\widetilde{W}}$, and therefore

$$S_1|_W \simeq S_0|_W \ \text{ and } \ S_1'|_W \simeq S_0'|_W;$$
$$\tilde{S}_1|_{\widetilde{W}} \simeq \tilde{S}_0|_{\widetilde{W}} \ \text{ and } \ \tilde{S}_1'|_{\widetilde{W}} \simeq \tilde{S}_0'|_{\widetilde{W}}.$$

It follows that the operators $\tilde{S}_1 - \tilde{S}_0$ and $\tilde{S}_1' - \tilde{S}_0'$ are supported on $K = M \backslash W = X \backslash \widetilde{W}$ furthermore,

$$\tilde{S}_1 - \tilde{S}_0 = (S_1 - S_0)|_K \ \text{ and } \ \tilde{S}_1' - \tilde{S}_0' = (S_1' - S_0')|_K.$$

It follows that

$$\begin{aligned}
\mathrm{Index}_t(D_1, D_0) &= \mathrm{Index}_t \tilde{D}_1 - \mathrm{Index}_t \tilde{D}_0 \\
&= \int_X \left(\mathrm{tr}[\tilde{S}_1] - \mathrm{tr}[\tilde{S}_1']\right) - \int_X \left(\mathrm{tr}[\tilde{S}_0] - \mathrm{tr}[\tilde{S}_0']\right) \\
&= \int_X \left(\mathrm{tr}[\tilde{S}_1] - \mathrm{tr}[\tilde{S}_0]\right) - \int_X \left(\mathrm{tr}[\tilde{S}_1'] - \mathrm{tr}[\tilde{S}_0']\right) \\
&= \int_M (\mathrm{tr}[S_1] - \mathrm{tr}[S_0]) - \int_M \left(\mathrm{tr}[S_1'] - \mathrm{tr}[S_0']\right) \\
&= \int_M \left(\mathrm{tr}[S_1] - \mathrm{tr}[S_1']\right) - \int_M \left(\mathrm{tr}[S_0] - \mathrm{tr}[S_0']\right)
\end{aligned}$$

as desired. □

As we noted before, the proof of the lemma shows also that the definition of the relative topological index is independent of the choice of extensions $\widetilde{D}_0$ and $\widetilde{D}_1$.

Before we step into the proof of Theorem 3.1 itself, we must introduce some further notation. Let $f : [0, 1] \to [0, 1]$ be a smooth function such that $f = 1$ on $\left[0, \frac{1}{3}\right]$, $f = 0$ on $\left[\frac{2}{3}, 1\right]$ and $f' \approx -1$ on $\left[\frac{1}{3}, \frac{2}{3}\right]$. Pick up a point $x_0 \in M$ and let $d(x) = \mathrm{dist}(x, x_0)$. For each $m \in \mathbb{Z}^*$, consider the functions

$$f_m(x) = f\left(\frac{1}{m}\, \mathrm{e}^{-d(x)}\right).$$

Note that $\mathrm{supp}\, \mathrm{d} f_m^{\frac{1}{2}} \subset B_{\log\left(\frac{3}{4m}\right)} - B_{\log\left(\frac{3}{2m}\right)}$ and

$$\|\mathrm{d} f_m\|_{L^2} \leq \frac{c_9}{m}, \tag{26}$$

where $c_9 = \left(\int_X \mathrm{e}^{-d(x)}\right)^{\frac{1}{2}}$. Here, $B_r = \{x \in M \mid d(x) \leq r\}$, which is compact by the completeness of M.

Proof of Theorem 3.1. All we have to do is to show that the right-hand sides of (24) and (25) are equal.

In fact, let $U \subset V$ be small neighbourhoods of the diagonal within $M \times M$ and choose $\psi \in C^\infty(M \times M)$ supported on V and such that $\psi = 1$ on U. Let Q_j be the operator whose Schwartzian kernel is $K^{Q_j}(x, y) = \psi(x, y)K^{G_j}(x, y)$, where G_j is the Green's operator for D_j. Then Q_j is a parametrix for D_j for which the corresponding smoothing operators S_j and S'_j, as in (23), satisfy

$$\mathrm{tr}[S_j] = \mathrm{tr}[H_j] \text{ and } \mathrm{tr}[S'_j] = \mathrm{tr}[H'_j], \tag{27}$$

where H_j and H'_j are the finite rank projection operators associated with the Green's operator G_j, as in (23).

But it is not necessarily the case that the two parametrices Q_0 and Q_1 so obtained must coincide at W. In order to fix that, we will glue them with the common parametrix of $D_0|_W$ and $D_1|_W$, denoted Q (with corresponding smoothing operators S and S'), using the cut-off functions f_m defined above (assume that the base points are contained in the compact set K). More precisely, for a section s,

$$Q_j^{(m)}(s) = f_m^{\frac{1}{2}}\, Q_j\left(f_m^{\frac{1}{2}}\, s\right) + (1 - f_m)^{\frac{1}{2}} Q\left((1 - f_m)^{\frac{1}{2}} s\right);$$

clearly, for each m, the operators $Q_0^{(m)}$ and $Q_1^{(m)}$ coincide at W. For the respective smoothing operators, we get (see [20, Prop. 1.24])

$$\begin{cases} S_j^{(m)}(s) = f_m^{\frac{1}{2}}\, S_j\left(f_m^{\frac{1}{2}}\, s\right) + (1 - f_m)^{\frac{1}{2}} S\left((1 - f_m)^{\frac{1}{2}} s\right) + (Q_j\left(f_m^{\frac{1}{2}}\, s\right) - Q\left((1 - f_m)^{\frac{1}{2}} s\right) \mathrm{d} f_m^{\frac{1}{2}}, \\ S_j^{(m)\prime}(s) = f_m^{\frac{1}{2}}\, S'_j\left(f_m^{\frac{1}{2}}\, s\right) + (1 - f_m)^{\frac{1}{2}} S'\left((1 - f_m)^{\frac{1}{2}} s\right). \end{cases}$$

Therefore

$$\begin{aligned} \mathrm{tr}[S_j^{(m)}] - \mathrm{tr}[S_j^{(m)'}] &= f_m^{\frac{1}{2}}\left(\mathrm{tr}[S_j] - \mathrm{tr}[S'_j]\right) \\ &\quad + (1 - f_m)^{\frac{1}{2}}\left(\mathrm{tr}[S] - \mathrm{tr}[S']\right) + \mathrm{tr}\left[(Q_j - Q)\mathrm{d} f_m^{\frac{1}{2}}\right] \end{aligned}$$

and

$$\begin{aligned} &\mathrm{tr}[S_1^{(m)}] - \mathrm{tr}[S_1^{(m)'}] - \mathrm{tr}[S_0^{(m)}] + \mathrm{tr}[S_0^{(m)'}] \\ &\quad = f_m^{\frac{1}{2}}\left(\mathrm{tr}[S_1] - \mathrm{tr}[S'_1] - \mathrm{tr}[S_0] + \mathrm{tr}[S'_0]\right) + \mathrm{tr}\left[(Q_1 - Q)\mathrm{d} f_m^{\frac{1}{2}}\right] - \mathrm{tr}\left[(Q_0 - Q)\mathrm{d} f_m^{\frac{1}{2}}\right], \end{aligned}$$

so finally we obtain

$$\begin{aligned} &\mathrm{tr}[S_1^{(m)}] - \mathrm{tr}[S_1^{(m)'}] - \mathrm{tr}[S_0^{(m)}] + \mathrm{tr}[S_0^{(m)'}] \\ &\quad = f_m^{\frac{1}{2}}\left(\mathrm{tr}[S_1] - \mathrm{tr}[S'_1] - \mathrm{tr}[S_0] + \mathrm{tr}[S'_0]\right) + \mathrm{tr}\left[(Q_1 - Q_0)\mathrm{d} f_m^{\frac{1}{2}}\right]. \end{aligned} \tag{28}$$

We must now integrate both sides of (28) and take limits as $m \to \infty$. For m sufficiently large, $\operatorname{supp}(1 - f_m) \subset W$, hence the left hand side of the identity (28) equals the relative topological index $\operatorname{Index}_t(D_1, D_0)$, by Lemma 5.1. On the other hand, the first summand inside on the right-hand side of (28) equals $\operatorname{Index}_a(D_1, D_0)$ by (27) and (24). Thus, it is enough to show that the integral of the last two terms on the right-hand side of (28) vanishes as $m \to \infty$. Indeed, note that

$$\operatorname{tr}\left[(Q_1 - Q_0)\mathrm{d} f_m^{\frac{1}{2}}\right] = \mathrm{d} f_m^{\frac{1}{2}} \operatorname{tr}[Q_1 - Q_0],$$

hence, since $\operatorname{supp}(\mathrm{d} f_m) \subset W$ for sufficiently large m and using also (26), it follows that

$$\int\limits_W \operatorname{tr}\left[(Q_1 - Q_0)\mathrm{d} f_m^{\frac{1}{2}}\right] \leq \frac{c_9}{m} \int\limits_W \operatorname{tr}[G_1 - G_0] \to 0 \text{ as } m \to \infty \tag{29}$$

if the integral on the right-hand side of the above inequality is finite. Indeed, let $D = D_1|_W = D_0|_W$; from the parametrix equation, we have $D((G_1 - G_0)|_W) = (H_1 - H_0)|_W$. Observe that $\mathcal{H} = \operatorname{Ker}((H_1 - H_0)|_W)$ is a closed subspace of finite codimension in $L^2_{j+1,\Gamma_1}(F_1|_W)$. Moreover $\mathcal{H} \subseteq \operatorname{Ker} D$; thus, $(G_1 - G_0)|_W$ has finite dimensional range and hence it is of trace class, i.e. the integral on the right-hand side of inequality (29) does converge (see also [20, Lemma 4.28]). This concludes the proof. $\square$

Acknowledgement. The authors thank S. Cherkis, T. Hausel and D. Nógrádi for careful reading of the manuscript and making several important comments. We also thank the organizers of the 2004 March AIM-ARCC workshop on L^2 harmonic forms for bringing the team together. G.E. is supported by the CNPq post-doctoral grant No. 150854/2005-6 (Brazil) and the OTKA grants No. T43242 and No. T046365 (Hungary). M.J. is partially supported by the CNPq grant No. 300991/2004-5 and the FAPESP grant No. 2005/04558-0.

References

1. Aliev, A.N., Saclioglu, C.: Self-dual fields harbored by a Kerr–Taub-BOLT instanton. Phys. Lett. **B632**, 725–727 (2006)
2. Atiyah, M.F., Hitchin, N.J., Singer, I.M.: Self-duality in four dimensional Riemannian geometry. Proc. Roy. Soc. London **A362**, 425–461 (1978)
3. Auckly, D.R.: Topological methods to compute Chern–Simons invariants. Math. Proc. Camb. Phil. Soc. **115**, 229–251 (1994)
4. Bruckmann, F., van Baal, P.: Multi-caloron solutions. Nucl. Phys. **B645**, 105–133 (2002)
5. Bruckmann, F., Nógrádi, D., van Baal, P.: Higher charge calorons with non-trivial holonomy. Nucl. Phys. **B698**, 233–254 (2004)
6. Charap, J.M., Duff, M.J.: Space-time topology and a new class of Yang-Mills instanton. Phys. Lett. **B71**, 219–221 (1977)
7. Cherkis, S.A.: *Self-dual gravitational instantons*. Talk given at the AIM-ARCC workshop "L^2 cohomology in geometry and physics", Palo Alto, USA, March 16–21, 2004
8. Cherkis, S.A., Hitchin, N.J.: Gravitational instantons of type D_k. Commun. Math. Phys. **260**, 299–317 (2005)
9. Cherkis, S.A., Kapustin, A.: Hyper-Kähler metrics from periodic monopoles. Phys. Rev. **D65**, 084015 (2002)
10. Cherkis, S.A., Kapustin, A.: D_k gravitational instantons and Nahm equations. Adv. Theor. Math. Phys. **2**, 1287–1306 (1999)
11. Chern, S., Simons, J.: Characteristic forms and geometric invariants. Ann. Math. **99**, 48–69 (1974)
12. Derek, H.: Large scale and large period limits of symmetric calorons. J. Math. Phys. **48**, 082905 (2007)
13. Dodziuk, J.: Vanishing theorems for square-integrable harmonic forms. Proc. Indian Acad. Sci. Math. Sci. **90**, 21–27 (1981)

14. Eguchi, T., Gilkey, P.B., Hanson, A.J.: Gravity, gauge theories and differential geometry. Phys. Rep. **66**, 213–393 (1980)
15. Etesi, G.: The topology of asymptotically locally flat gravitational instantons. Phys. Lett. **B641**, 461–465 (2006)
16. Etesi, G., Hausel, T.: Geometric interpretation of Schwarzschild instantons. J. Geom. Phys. **37**, 126–136 (2001)
17. Etesi, G., Hausel, T.: Geometric construction of new Yang–Mills instantons over Taub–NUT space. Phys. Lett. **B514**, 189–199 (2001)
18. Etesi, G., Hausel, T.: On Yang–Mills instantons over multi-centered gravitational instantons. Commun. Math. Phys. **235**, 275–288 (2003)
19. Gibbons, G.W., Hawking, S.W.: Gravitational multi-instantons. Phys. Lett. **B78**, 430–432 (1976)
20. Gromov, M., Lawson, H.B. Jr.: Positive scalar curvature and the Dirac operator on complete Riemannian manifolds. Inst. Hautes Études Sci. Publ. Math. **58**, 83–196 (1983)
21. Hausel, T., Hunsicker, E., Mazzeo, R.: Hodge cohomology of gravitational instantons. Duke Math. J. **122**, 485–548 (2004)
22. Hawking, S.W.: Gravitational instantons. Phys. Lett. **A60**, 81–83 (1977)
23. Jardim, M.: *Nahm transform of doubly periodic instantons*. PhD Thesis, University of Oxford, 110 pp, http://arXiv.org/list/math.DG/9912028, 1999
24. Jardim, M.: Nahm transform and spectral curves for doubly-periodic instantons. Commun. Math. Phys. **225**, 639–668 (2002)
25. Kirk, P., Klassen, E.: Chern–Simons invariants and representation spaces of knot groups. Math. Ann. **287**, 343–367 (1990)
26. Kronheimer, P.B.: The construction of ALE spaces as hyper-Kähler quotients. J. Diff. Geom. **29**, 665–683 (1989)
27. Kronheimer, P.B., Nakajima, N.: Yang–Mills instantons on ALE gravitational instantons. Math. Ann. **288**, 263–307 (1990)
28. Morgan, J.W., Mrowka, T., Ruberman, D.: *The L^2 moduli space and a vanishing theorem for Donaldson polynomial invariants*. Monographs in geometry and topology, Volume II, Cambridge, MA: Int. Press, 1994
29. Nakajima, H.: Moduli spaces of anti-self-dual connections on ALE gravitational instantons. Invent. Math. **102**, 267–303 (1990)
30. Nye, T.M.W.: *The geometry of calorons*. PhD Thesis, University of Edinburgh, 147 pp, http://arXiv.org/list/hep-th/0311215, 2001
31. Parker, T., Taubes, C.H.: On Witten's proof of the positive energy theorem. Commun. Math. Phys. **84**, 223–238 (1982)
32. Råde, J.: Singular Yang–Mills fields. Local theory II. J. Reine Angew. Math. **456**, 197–219 (1994)
33. Sibner, L.M., Sibner, R.J.: Classification of singular Sobolev connections by their holonomy. Commun. Math. Phys. **144**, 337–350 (1992)
34. Tekin, B.: Yang-Mills solutions on Euclidean Schwarzschild space. Phys. Rev. **D65**, 084035 (2002)
35. Wehrheim, K.: Energy identity for anti-self-dual instantons on $\mathbb{C} \times \Sigma$. Math. Res. Lett. **13**, 161–166 (2006)
36. Yau, S-T.: Harmonic functions on complete Riemannian manifolds. Commun. Pure. Appl. Math. **28**, 201–228 (1975)

Communicated by G.W. Gibbons