S-duality in Abelian gauge theory revisited

Gábor Etesi

Department of Geometry, Mathematical Institute, Faculty of Science,

Budapest University of Technology and Economics,

Egry J. u. 1, H ép., H-1111 Budapest, Hungary *

Ákos Nagy

Budapest University of Technology and Economics †

February 22, 2020

Abstract

Definition of the partition function of U(1) gauge theory is extended to a class of four-manifolds containing all compact spaces and the asymptotically locally flat (ALF) ones including the multi-Taub–NUT sopaces. The partition function is calculated via zeta-function regularization and heat kernel techniques with special attention to its modular properties.

In the compact case, compared with the purely topological result of Witten, we find a non-trivial curvature correction to the modular weights of the partition function. But S-duality can be restored by adding gravitational counter terms to the Lagrangian in the usual way.

In the ALF case however we encounter non-trivial difficulties stemming from original non-compact ALF phenomena. Fortunately our careful definition of the partition function makes it possible to circumnavigate them and conclude that the partition function has the same modular properties as in the compact case.

AMS Classification: Primary: 81T13; Secondary: 81Q30, 57M50, 11F37, 35K08

Keywords: *S-duality; Partition function; L^2 cohomology; Zeta-function regularization; Heat kernel*

1 Introduction

The long standing conjecture asserts that quantum gauge theory has a symmetry exchanging strong and weak coupling as well as electric and magnetic fields. This conjecture originated with the work of Montonen and Olive [16] from 1977 who proposed a symmetry in quantum gauge theory with the above properties and also interchanging the gauge group G with its dual group $G^\vee$. It was soon realized however that this duality is more likely to hold in an $N = 4$ supersymmetrized theory [17].

*e-mail: `etesi@math.bme.hu`

†e-mail: `nagyak@math.bme.hu`

The original Montonen–Olive conjecture proposed a $\mathbb{Z}_2$ symmetry exchanging the electric and magnetic charges however in $N = 4$ theory it is naturally extended to an $\mathrm{SL}(2, \mathbb{Z})$ symmetry acting on the complex coupling constant

$$\tau := \frac{\theta}{2\pi} + \frac{4\pi}{e^2}\mathbf{i} \ \in \mathbb{C}^+ \tag{1}$$

combining the gauge coupling e and the θ parameter of the $N = 4$ theory. In this framework the conjecture can be formulated as follows [25].

We say that a not necessarily holomorphic function $f : \mathbb{C}^+ \to \mathbb{C}$ on the upper half-plane is an *unrestricted modular form of weight* (α, β) (conventionally supposed to be integers) if with respect to $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$ it transforms as

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^\alpha (c\overline{\tau} + d)^\beta f(\tau).$$

If $\alpha \neq 0$ or $\beta \neq 0$ we also say sometimes that a *modular anomaly* is present in f. Since $\mathrm{SL}(2, \mathbb{Z})$ is generated by $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ modular properties are sufficient to be checked under the simple transformations $f(\tau) \mapsto f(\tau + 1)$ and $f(\tau) \mapsto f(-1/\tau)$. Modular forms play an important role in classical number theory [15, 21].

The *electric-magnetic duality conjecture* in its simplest form asserts that over a four-manifold (M, g) the partition function of a (twisted) $N = 4$ supersymmetric quantum gauge theory with simply-laced gauge group G is modular in the sense that it satisfies

$$\begin{cases} Z(M, g, G, \tau + 1) = Z(M, g, G, \tau) & (\text{“level 1 property”}) \\ Z(M, g, G, -1/\tau) = Z(M, g, G^\vee, \tau) & (\text{“}S\text{-duality property”}) \end{cases}$$

where $G^\vee$ is the Goddard–Nuyts–Olive or Langlands dual group to G. The first symmetry is a classical one while the second is expected to reflect the true quantum nature of gauge theories in the sense that it connects two theories on the quantum level which are classically different. In general G is not isomorphic to its dual however for example if $G \cong \mathrm{U}(1)$ then it is. Moreover if $\theta = 0$ this duality reduces to the original Montonen–Olive conjecture.

We do not attempt here to survey the long, diverse and colourful history of the conjecture and its variants rather refer to the introductions of [25, 14]. We just mention that for instance it led to highly non-trivial predictions about the number of L^2 harmonic forms on complete manifolds due to Sen [20] (cf. also [10, 12, 19]) and the latest chapter of the story relates the electric-magnetic duality conjecture with the geometric Langlands program of algebraic geometry due to Kapustin and Witten [14], see also [9].

In this work, motivated by papers of Witten [26, 27] we focus attention to the modular properties of the partition function

$$Z(M, g, \tau) := Z(M, g, \mathrm{U}(1), \tau)$$

of (supersymmetric) Abelian or $\mathrm{U}(1)$ or Maxwell gauge theory over various four-manifolds because in this case the calculations can be carried out almost rigorously.

The paper is organized as follows. In Sect. 2 we offer an extended definition of the partition function of Abelian gauge theory (cf. Eq. (11) here) such that it gives back Witten's [26] in case of compact four-manifolds moreover the definition continues to make sense for a certain class of non-compact geometries, the so-called *asymptotically locally flat* (ALF) spaces including the flat $\mathbb{R}^3 \times S^1$, the multi-Taub–NUT spaces and the Riemannian Schwarzschild and Kerr solutions, etc. The subtle point of this extended definition is imposing a natural boundary condition at infinity on finite action classical solutions of Maxwell theory in order to save the modular properties of the partition function. This boundary condition is the so-called *strong holonomy condition* and is used to rule out certain classical solutions having non-integer energy. It has also appeared already in the approach to SU(2) instanton moduli spaces over ALF geometries [4, 7].

Then in Sect. 3 we repeat Witten's calculation [26] of $Z(M,g,\tau)$ in the compact case. After summing up over a discrete set leading to ϑ-functions the partition function is given by a formal infinite dimensional integral and can be calculated quite rigorously via ζ-function regularization. On the way we formulate a natural "Fubini principle" for these formal integrals which converts them into successive integrals. We come up with an expression involving the zero values of various ζ-functions and their derivatives. The zero values of these ζ-functions can be calculated by the aid of heat kernel techniques. With the help of these tools we find that

$$\begin{cases} Z(M,g,\tau+2) = Z(M,g,\tau) \\ Z(M,g,-1/\tau) = (-\mathbf{i})^{\frac{1}{2}\sigma(M)}\,\tau^{\alpha}\,\overline{\tau}^{\beta}\,Z(M,g,\tau) \end{cases} \tag{2}$$

i.e., up to a factor it is a level 2 modular form but with non-integer weights (cf. Eq. (20) here)

$$\alpha = \frac{1}{4}\left(\chi(M) + \sigma(M) + \text{curvature corrections}\right)$$

and

$$\beta = \frac{1}{4}\left(\chi(M) - \sigma(M) + \text{curvature corrections}\right)$$

yielding that the modular weights are not purely topological as claimed in [26, 27]. As a consequence S-duality fails in its simplest form but it can be saved by adding usual gravitational c-number terms to the naive Lagrangian of Maxwell theory on a curved background [26].

Before proceeding to the non-compact calculation we make a digression in Sect. 4 and clarify the role played by the strong holonomy condition. We will see that without it (i.e., simply taking the definition of the partition function from the compact case) the partition function for instance over the multi-Taub–NUT spaces would seriously fail to be modular as a consequence of the presence of classical solutions with continuous energy spectrum.

Finally as a novelty in Sect. 5 we repeat the calculation over ALF spaces. S-duality over non-compact geometries is less known (cf. the case of ALE spaces in [25]). However before obtaining some results we have to overcome another technical difficulty caused by non-compactness. Namely if one wishes to calculate the formal integrals by ζ-function regularization again then first one has to face the fact that the spectra of various differential operators are continuous hence the existence of their ζ-functions is not straightforward. Consequently we take a truncation of the original manifold and impose Dirichlet boundary condition on the boundary. This boundary condition is compatible with the finite action assumption on connections. After defining everything correctly in this framework, we let the boundary go to infinity. We will find that the modular weights converge in this limit and give back a formula very similar to the compact case above. The only difference is that the various Betti numbers which enter the modular weights are mixtures of true L^2 Betti numbers and limits of Dirichlet–Betti numbers which are remnants of the boundary condition (see Eq. (26) here).

Consequently the partition function transforms akin to (2) again however its modular anomaly cannot be cancelled by adding c-number terms. However up to a mild topological condition on the infinity of an ALF space (cf. Eq. (27) here) all these Betti numbers can be converted into L^2 ones again and the cancellation of the modular anomaly goes as in the compact case hence S-duality can be saved. It is interesting that while the Riemannian Schwarzschild and Kerr not, the multi-Taub–NUT family satisfies this topological condition.

We clarify at this point that throughout this paper questions related to the contributions of various determinants to the partition function will be suppressed.

We close this introduction by making a comment about the way of separating rigorous mathematical steps from intuitive ones in the text. To help the reader we decided to formulate every rigorous steps in the form of a lemma (with proof) or with a clear reference to the literature while the other considerations just appear continuously in the text.

Acknowledgement. The authors are grateful to Gyula Lakos and Szilárd Szabó for the stimulating discussions. The first author was partially supported by OTKA grant No. NK81203 (Hungary).

2 Preliminary calculations

In this section we formulate our problem as precisely as possible. Let (M,g) be a connected, oriented, complete Riemannian four-manifold without boundary. M can be either compact or non-compact. In the non-compact case we require (M,g) to have infinite volume.

Let L be a smooth complex line bundle over M. The isomorphism classes of these bundles are classified by the elements of the group $H^2(M;\mathbb{Z})$ via their first Chern class $c_1(L)\in H^2(M;\mathbb{Z})$. Putting a Hermitian structure onto L we can pick a $\mathrm{U}(1)$-connection ∇ with curvature F_∇. Under $\mathfrak{u}(1)\cong\mathbf{i}\mathbb{R}$ the curvature is an $\mathbf{i}\mathbb{R}$-valued 2-form closed by Bianchi identity hence $[\frac{F_\nabla}{2\pi\mathbf{i}}]\in H^2(M;\mathbb{R})$ is a de Rham cohomology class taking its values in the integer lattice. This cohomology class (the Chern–Weil class of L) is independent of the connection but it is a slightly weaker invariant of L namely it characterizes it up to a flat line bundle only. Note that flat line bundles over M satisfy $c_1(L)\in\mathrm{Tor}(H^2(M;\mathbb{Z}))$.

With real constants e and θ the usual action of (Euclidean) electrodynamics extended with the so-called θ-term over (M,g) looks like

$$S(\nabla,e,\theta):=-\frac{1}{2e^2}\int\limits_M F_\nabla\wedge *F_\nabla+\frac{\mathbf{i}\theta}{16\pi^2}\int\limits_M F_\nabla\wedge F_\nabla. \tag{3}$$

The θ-term is a characteristic class hence its variation is identically zero consequently the Euler–Lagrange equations of this theory are just the usual vacuum Maxwell-equations

$$\mathrm{d}F_\nabla=0,\quad \delta F_\nabla=0. \tag{4}$$

Remark. If one considers the underlying quantum field theory then θ in (3), being a non-dynamical variable, remains well-defined at the full quantum level meanwhile the coupling constant e runs. Therefore, in order to keep its meaning at the full quantum level we extend (3), as usual, to an $N=4$ supersymmetric theory [17, 25]. This extension yields additional terms to (3) however their presence do not influence our forthcoming calculations therefore we shall not mention them explicitly in this paper.

Rather we introduce the complex coupling constant (1) taking its values on the upper half-plane and supplementing (3) re-write the relevant part of the action as

$$S(\nabla,\tau)=\frac{\mathbf{i}\pi}{2}\tau\left(\frac{1}{8\pi^2}\int\limits_M(F_\nabla\wedge *F_\nabla+F_\nabla\wedge F_\nabla)\right)+\frac{\mathbf{i}\pi}{2}(-\overline{\tau})\left(\frac{1}{8\pi^2}\int\limits_M(F_\nabla\wedge *F_\nabla-F_\nabla\wedge F_\nabla)\right). \tag{5}$$

For clarity we note that this expression is exactly the same as (3).

The orientation and the metric on M is used to form various Sobolev spaces. Fix a line bundle L over M and a connection ∇^0_L on it such that $\mathbf{i}F_{\nabla^0_L}\in L^2(M;\Lambda^2M)$. For any integer $l\geqq 2$ then set

$$\mathscr{A}(\nabla^0_L):=\{\nabla_L\mid\nabla_L:=\nabla^0_L+a \text{ with } \mathbf{i}a\in L^2_l(M;\Lambda^1M)\}.$$

This is the L^2_l Sobolev space of $\mathrm{U}(1)$ connections on L relative to ∇^0_L. Notice that this is a vector space (not an affine space) and if $L_0\cong M\times\mathbb{C}$ is the trivial line bundle and $\nabla^0_{L_0}=\mathrm{d}$ is the trivial flat connection in the trivial gauge on it then $\mathscr{A}(\nabla^0_{L_0})\cong L^2_l(M;\Lambda^1M)$. Furthermore write $\mathscr{U}(1)_L$ for the L^2_{l+1}-completion of the space of gauge transformations

$$\{\gamma-\mathrm{Id}_L\in C^\infty_0(M;\mathrm{End}L)\mid \|\gamma-\mathrm{Id}_L\|_{L^2_{l+1}(M)}<+\infty\,;\ \gamma\in C^\infty(M;\mathrm{Aut}L)\text{ a.e.}\}.$$

Under these assumptions the Sobolev multiplication theorem ensures us that $\mathbf{i}F_{\nabla_L}\in L^2(M;\Lambda^2M)$ that is, the curvature 2-form is always in L^2 over (M,g). The space $\mathscr{A}(\nabla^0_L)$ is acted upon by $\mathscr{U}(1)_L$ in the usual way; the orbit space $\mathscr{A}(\nabla^0_L)/\mathscr{U}(1)_L$ of gauge equivalence classes with its quotient topology is denoted by $\mathscr{B}(\nabla^0_L)$ as usual.

Next we record an obvious decomposition of the action. This decomposition plays a crucial role because it underlies the definition of the partition function. Suppose that $\nabla^0_L\in\mathscr{A}(\nabla^0_L)$ is a *finite action classical solution on* L i.e., it satisfies (4) and has finite action (3) or (5); moreover let $\nabla_L\in\mathscr{A}(\nabla^0_L)$ be another connection on the same line bundle L. Then it follows from (5) that

$$S(\nabla_L,\tau)=S(\nabla^0_L+a,\tau)=S(\nabla^0_L,\tau)+\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)} \tag{6}$$

where we write $\|\mathrm{d}a\|^2_{L^2(M)}=-\int_M\mathrm{d}a\wedge *\mathrm{d}a$ since a is pure imaginary. This decomposition is a straightforward calculation if M is compact while follows from the L^2 control on the curvature of ∇^0_L and the L^2_1 control on the perturbation a if M is non-compact.

Now we turn to clarify what shall we mean by a "partition function" in this paper. In our attempt to define it we have to be careful because we want to include the case of non-compact base spaces as well while we require that our definition should agree with Witten's [26, 27] in the compact case. The above decomposition of the action indicates that finding the partition function i.e., the integration over all finite-action connections should be carried out in two steps: (i) summation over a certain set of classical solutions ∇^0_L (this will coincide with summation over all line bundles L in the compact case) and then (ii) integration over the orbit space $\mathscr{B}(\nabla^0_{L_0})$ for the trivial flat connection $\nabla^0_{L_0}$ in Coulomb gauge on the trivial line bundle L_0.

We begin with the description of the orbit space. First we have a usual gauge fixing lemma.

Lemma 2.1. *Let (M,g) be as above i.e., a connected, oriented complete Riemannian four-manifold which is either closed (compact without boundary) or open (non-compact without boundary) with infinite volume. Consider the trivial line bundle $L_0\cong M\times\mathbb{C}$ on it.*

If $\nabla_{L_0}=\mathrm{d}+A$ with $A\in\mathscr{A}(\nabla^0_{L_0})$ is any $\mathrm{U}(1)$-connection on L_0 then there exists an L^2_{l+1} gauge transformation $\gamma:M\to\mathrm{U}(1)$ such that the resulting L^2_l connection $\nabla'_{L_0}=\mathrm{d}+A'$ satisfies the Coulomb gauge condition $\delta A'=0$.

Proof. Take a gauge transformation $\gamma = e^{\mathbf{i}f}$ with a function $f : M \to \mathbb{R}$. We want to solve the equation $0 = \delta A' = \delta(A + \mathbf{i}df)$ i.e., $\triangle_0 f = \mathbf{i}\delta A$. This equation is solvable if and only if the real-valued function $\mathbf{i}\delta A$ is orthogonal in L^2 to the cokernel of the scalar Laplacian $\triangle_0 = \delta d$ on (M, g). Note that $\triangle_0$ is formally self-adjoint hence in fact we need $\mathbf{i}\delta A \perp_{L^2} \ker\triangle_0$. However $\ker\triangle_0 \cong \mathbb{R}$ if M is closed hence the result follows from Stokes' theorem. While $\ker\triangle_0 \cong \{0\}$ if M is non-compact but complete and has infinite volume by a theorem of Yau [29] hence the result also follows. $\diamond$

The time has come to introduce L^2 cohomology groups on (M, g) in the usual way [12]: we say that a real k-form φ belongs to the k^{th} *reduced* L^2 *cohomology group* $\overline{H}^k_{L^2}(M)$ if and only if $\|\varphi\|_{L^2(M)} < +\infty$ as well as $d\varphi = 0$ and $\delta\varphi = 0$. Over a complete manifold we can equivalently say that φ has finite L^2 norm and is harmonic i.e., $\triangle_k\varphi = 0$ where $\triangle_k$ is the Laplacian on k-forms. If $\mathscr{H}^k_{L^2}(M)$ denotes the space of L^2 harmonic forms then by definition $\overline{H}^k_{L^2}(M) \cong \mathscr{H}^k_{L^2}(M)$ consequently in this paper we will write $\mathscr{H}^k_{L^2}(M)$ for these groups. In general we call a non-trivial L^2 harmonic form on a complete Riemannian manifold *non-topological* if either it is exact or not cohomologous to a compactly supported differential form. Roughly speaking the existence of non-topological L^2 harmonic forms are not predictable by topological means (cf. [19]). Note that in the compact case the L^2 groups reduce to ordinary de Rham cohomology groups $H^k(M;\mathbb{R})$ and of course there are no non-topological L^2 harmonic forms. Write $b^k_{L^2}(M) := \dim \mathscr{H}^k_{L^2}(M)$ for the corresponding L^2 Betti number (if finite). Of course in the compact case $b^k_{L^2}(M) = b^k(M)$ are just the ordinary finite Betti numbers.

Most of these groups admit natural interpretations in terms of our theory (3) as we will see shortly.

Lemma 2.2. *Let M be an arbitrary connected manifold and denote by $\widehat{M}$ the space of gauge equivalence classes of L^2 flat* U(1)*-connections on M (also called the* character variety *of M).*

(i) *There is an identification*

$$\widehat{M} \cong T^{b_1(M)} \times \mathrm{Tor}(H_1(M;\mathbb{Z}))$$

where T^k is a k-torus.

(ii) *If (M, g) is a Riemannian four-manifold as in the previous lemma then we have a more informative identification*

$$\widehat{M} \cong \frac{\mathscr{H}^1_{L^2}(M)}{2\pi\Lambda^1_{L^2}} \times \mathrm{Tor}(H_1(M;\mathbb{Z}))$$

where $\Lambda^1_{L^2}$ is a co-compact integer lattice in $\mathscr{H}^1_{L^2}(M)$.

Moreover keep in mind that $\mathscr{H}^1_{L^2}(M) \cong H^1(M;\mathbb{R})$ is ordinary de Rham cohomology if M is compact.

Remark. The lemma implies that if M is a connected, oriented, non-compact, complete four-manifold with infinite volume then $b^1_{L^2}(M) = b^1(M)$ hence there are no non-topological L^2 harmonic 1-forms.

Proof. The gauge equivalence classes of flat U(1)-connections are classified by the space

$$\widehat{M} \cong \mathrm{Hom}(\pi_1(M)\,;\ \mathrm{U}(1))/\mathrm{ad}\,\mathrm{U}(1) = \mathrm{Hom}(\pi_1(M)\,;\ \mathrm{U}(1))$$

since U(1) is Abelian. Also commutativity implies that the commutator group $[\pi_1(M), \pi_1(M)]$ lies in the kernel of any such homomorphism hence $\pi_1(M)$ can be replaced by

$$\pi_1(M)/\,[\pi_1(M), \pi_1(M)] \cong H_1(M;\mathbb{Z}).$$

Write $H_1(M;\mathbb{Z}) = \Lambda_1 \times \mathrm{Tor}(H_1(M;\mathbb{Z}))$ with $\Lambda_1 \cong \mathbb{Z}^{b_1(M)}$. Homomorphisms $\rho : \mathbb{Z} \to \mathrm{U}(1)$ are classified by their trace $\mathrm{tr}(\rho(1)) \in S^1 \subset \mathbb{C}$ on the generator $1 \in \mathbb{Z}$ hence the parameter space is S^1. The homomorphisms of the finite Abelian group $\rho : \mathrm{Tor}(H_1(M;\mathbb{Z})) \to \mathrm{U}(1)$ are certainly classified by its elements. This yields part (i) of the lemma.

Regarding the second part, suppose that $\nabla^0_{L_0}$ is the trivial flat U(1)-connection on the trivial bundle $L_0 \cong M \times \mathbb{C}$ in the straightforward gauge i.e., when $\nabla^0_{L_0} = \mathrm{d}$. It follows from (6) that its perturbation $\nabla_{L_0} := \nabla^0_{L_0} + a$ with $a \in \mathscr{A}(\nabla^0_{L_0})$ is flat if and only if $\mathrm{d}a = 0$. Imposing the Coulomb gauge condition $\delta a = 0$ (cf. Lemma 2.1) apparently the gauge-inequivalent perturbations are parameterized by the group $\mathscr{H}^1_{L^2}(M)$. Indeed, if two flat connections $\mathrm{d} + a$ and $\mathrm{d} + b$ on L_0 are in Coulomb gauge then it easily follows from the proof of Lemma 2.1 that there is no gauge transformation between them of the form $\gamma = \mathrm{e}^{\mathbf{i}f}$ with a single-valued function $M \to \mathbb{R}$.

However if M is not simply-connected then there may still exists a gauge transformation such that $b = a + \gamma^{-1}\mathrm{d}\gamma$ with $\gamma = \mathrm{e}^{\mathbf{i}F}$ where F is a multi-valued real function on M (or a single-valued function on the universal cover $\widetilde{M}$). More precisely, if $\ell : [0,1] \to M$ is a smooth loop with base point $x_0 \in M$ then $\gamma(\ell(0)) = \gamma(\ell(1)) = \gamma(x_0)$ implies

$$F(x_0) = F(\ell(1)) = F(\ell(0)) + 2\pi n_\ell = F(x_0) + 2\pi n_\ell$$

with some integer n_ℓ. Consider a flat connection $\mathrm{d} + a$ with $a \in \mathscr{A}(\nabla^0_{L_0})$ and also pick a differentiable loop ℓ as before. Associated with this loop take a gauge transformation $\gamma_\ell = \mathrm{e}^{\mathbf{i}F_\ell}$ such that F_ℓ is supported in a tubular neighbourhood of ℓ and $F_\ell(\ell(t)) = F_\ell(x_0) + 2\pi t$. Then obviously $\gamma_\ell \in \mathscr{U}(1)_{L_0}$ and

$$a + \gamma_\ell^{-1}\mathrm{d}\gamma_\ell = a + 2\pi\mathbf{i}\,\mathrm{d}t$$

moreover the new connection depends only on the homotopy class $[\ell]$. Consider the lattice Λ' in $\mathscr{H}^1_{L^2}(M)$ generated by the above translations when $[\ell] \in \pi_1(M, x_0)$ runs over the homotopy classes. It follows that the truely gauge-inequivalent flat connections are given by the quotient $\mathscr{H}^1_{L^2}(M)/\Lambda'$. But comparing this quotient with part (i) of the lemma we conclude that Λ' is co-compact hence it must coincide with a full integer lattice of $\mathscr{H}^1_{L^2}(M)$. This gives a connected component of $\widehat{M}$ in the second picture.

Note that all the flat L^2 perturbations of the trivial flat connection are connections on the same trivial bundle L_0. The second integer cohomology has the straightforward decomposition

$$H^2(M;\mathbb{Z}) = \Lambda^2 \times \mathrm{Tor}(H^2(M;\mathbb{Z})) \tag{7}$$

into its free part $\Lambda^2 \cong \mathbb{Z}^{b^2(M)}$ and its torsion. We obtain that generic elements $(a, \alpha) \in \widehat{M}$ are flat connections of the shape $\nabla^0_L + a$ where ∇^0_L is a fixed flat connection on the non-trivial flat line bundle satisfying

$$c_1(L) = \alpha \in \mathrm{Tor}(H^2(M;\mathbb{Z})) \cong \mathrm{Tor}(H_1(M;\mathbb{Z})) \tag{8}$$

and $a \in \mathscr{H}^1_{L^2}(M)/2\pi\Lambda^1_{L^2}$ as above. This observation provides the description of $\widehat{M}$ in part (ii) of the lemma. $\diamond$

Therefore in light of the previous lemmata and (6) elements of $\mathscr{H}^1_{L^2}(M)$ can be interpreted as flat L^2 perturbations of a connection ∇^0_L on a given line bundle L over M.

Now we can provide an explicit description of our relevant orbit space as follows.

Lemma 2.3. *The orbit space of gauge inequivalent L^2_l Abelian connections ($l \geqq 2$) over the trivial line bundle L_0 with respect to the trivial flat connection $\nabla^0_{L_0}$ and the Coulomb gauge condition admits a decomposition*

$$\mathscr{B}(\nabla^0_{L_0}) \cong \frac{\mathscr{H}^1_{L^2}(M)}{2\pi\Lambda^1_{L^2}} \times \left((\mathscr{H}^1_{L^2}(M))^\perp \cap \overline{\ker\delta}\right), \quad \overline{\ker\delta} \subset L^2_l(M;\Lambda^1 M).$$

That is, it is a Hilbert space bundle over a finite dimensional torus representing the gauge inequivalent flat connections on L_0 (cf. part (ii) *of Lemma 2.2).*

We note again that if M is compact then L^2 objects reduce to ordinary de Rham cohomology.

Proof. Consider the trivial flat connection $\nabla^0_{L_0}$ on L_0 and write any connection in the form $\nabla_{L_0} = \nabla^0_{L_0} + a$ with $a \in \mathscr{A}(\nabla^0_{L_0})$ as usual. By the aid of Lemma 2.1 let us impose the Coulomb gauge condition $\delta a = 0$ on them.

An advantage of L^2 cohomology is that when $l \geqq 2$ the usual Hodge decomposition continues to hold [12] in the form

$$L^2_l(M;\Lambda^1 M) \cong \overline{\mathrm{im}\,\mathrm{d}} \oplus \overline{\mathrm{im}\,\delta} \oplus \mathscr{H}^1_{L^2}(M) \tag{9}$$

and obviously $\overline{\mathrm{im}\,\delta} \oplus \mathscr{H}^1_{L^2}(M) \subseteq \overline{\ker\delta}$. If $\gamma \in \mathscr{U}(1)_{L_0}$ is any gauge transformation on the trivial bundle then $\gamma^{-1}\mathrm{d}\gamma$ is closed moreover $\delta(a + \gamma^{-1}\mathrm{d}\gamma) = 0$ together with $\delta a = 0$ implies that $\gamma^{-1}\mathrm{d}\gamma$ is also co-closed and of course $\mathbf{i}\gamma^{-1}\mathrm{d}\gamma \in L^2(M;\Lambda^1 M)$. In other words

$$\mathbf{i}\gamma^{-1}\mathrm{d}\gamma \in \mathscr{H}^1_{L^2}(M).$$

Hence Hodge decomposition shows that $\mathscr{U}(1)_{L_0}$ acts trivially on $(\mathscr{H}^1_{L^2}(M))^\perp \cap \overline{\ker\delta}$. Consequently all elements here are gauge inequivalent while Lemma 2.2 shows that the finite dimensional torus $\mathscr{H}^1_{L^2}(M)/2\pi\Lambda^1_{L^2} \cong T^{b_1(M)}$ enumerates the gauge inequivalent flat connections. $\diamond$

Now we move on to clarify the set of classical solutions used in this paper. This will also provide us with a physical interpretation of the space $\mathscr{H}^2_{L^2}(M)$ as containing the curvatures of finite action classical solutions over a 4-space (M, g). First we introduce the class of four-manifolds considered in this paper. By definition this class contains all compact geometries moreover the so-called *asymptotically locally flat* (ALF) geometries. These latter sub-class contains non-compact but complete four dimensional Riemannian manifolds with a special asymptotical geometry. For the definition of an ALF space we refer to Sect. 5. Many important non-compact manifolds in mathematical physics are of ALF type. Examples are the multi-Taub–NUT spaces, the Riemannian Schwarzschild and Kerr geometries, etc. It follows from [12, Corollary 9] that an ALF space always has finite dimensional second L^2 cohomology.

Writing $H^2(M;\mathbb{R}) \subseteq \mathscr{H}^2_{L^2}(M)$ i.e., embedding compactly supported de Rham cohomology into L^2 cohomology by the unique harmonic representative in each cohomology class, we can suppose that $\Lambda^2 \subset \mathscr{H}^2_{L^2}(M)$ where Λ^2 is the integer lattice from (7).

Definition 2.1. *Let (M, g) be a compact or an ALF four-manifold. Consider its 2^{nd} L^2 cohomology. We define a co-compact lattice $\Lambda^2_{L^2} \subset \mathscr{H}^2_{L^2}(M)$ as follows. We say that $\omega \in \Lambda^2_{L^2}$ if and only if*

(i) *either $\omega \in \Lambda^2$ i.e., it is a harmonic representative of a usual compactly supported integer de Rham cohomology class;*

(ii) *or if $\omega \in \Lambda^2_{L^2}$ and $\omega \notin \Lambda^2$ then there exists an* U(1)*-connection ∇^0_L on a line bundle L with curvature $F_{\nabla^0_L}/2\pi\mathbf{i} = \omega$ and this connection satisfies the strong holonomy condition at infinity.*

For the definition of the strong holonomy condition we refer to [7, Definition 2.1]. Roughly speaking if a finite action U(1)-connection satisfies the strong holonomy condition then it has trivial holonomy at infinity hence in an appropriate L^2_1 norm it approaches the trivial flat U(1)-connection on the infinitely distant boundary of (M, g). This provides us that the action (3), or equivalently (5), of such a connection is an integer, cf. [7, Theorem 2.2]. Moreover it is clear that if (M, g) is compact then part (ii) of Definition 2.1 is vacuous.

Chern–Weil theory says that if $\omega \in \Lambda^2_{L^2}$ there exists a line bundle L with $c_1(L) \in \Lambda^2 \subseteq \Lambda^2_{L^2}$, the integer lattice in (7), and an U(1)-connection ∇^0_L on the bundle such that $\omega = F_{\nabla^0_L}/2\pi\mathbf{i}$. This connection is obviously a finite action classical solution to (4). Moreover it follows from part (ii) of Lemma 2.2 and especially from (8) that the remaining line bundles in (7) i.e., those with $c_1(L) \in \mathrm{Tor}(H^2(M;\mathbb{Z}))$ also carry classical solutions ∇^0_L to (4) namely vacuum solutions i.e., flat connections.

We conclude that all elements of the lattice $\Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M;\mathbb{Z}))$ give rise to classical solutions of the Abelian gauge theory (3). In terms of (5) or the projected fields $F^{\pm}_{\nabla^0_L} = \frac{1}{2}(F_{\nabla^0_L} \pm *F_{\nabla^0_L})$ all of them satisfy

$$\frac{1}{8\pi^2}\int\limits_M \left(F_{\nabla^0_L} \wedge *F_{\nabla^0_L} \pm F_{\nabla^0_L} \wedge F_{\nabla^0_L}\right) = \frac{1}{4\pi^2}\int\limits_M F^{\pm}_{\nabla^0_L} \wedge *F^{\pm}_{\nabla^0_L} \in \mathbb{Z}. \tag{10}$$

We emphasize again that the validity of the energy quantization (10) for the non-topological sector of $\Lambda^2_{L^2}$ follows from the boundary condition imposed on it in part (ii) of Definition 2.1.

Remark. In general on a given bundle these classical solutions are not unique because of two reasons. The first is if $b^2(M) < b^2_{L^2}(M)$ which can happen if M is not compact, cf. Sect. 4. The second is if $\mathscr{H}^1_{L^2}(M) \neq \{0\}$ which can happen if M is not simply connected, cf. Lemma 2.2.

Now we are in a position to carefully define the central object of our interest here. Consider a finite action classical solution ∇^0_L such that either its curvature represents an element in $\Lambda^2_{L^2}$ or it is a flat connection whose gauge class is in $\mathrm{Tor}(H^2(M;\mathbb{Z}))$. In this case we write simply $[\nabla^0_L] \in \Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M;\mathbb{Z}))$ and these solutions will be referred to as *allowed classical solutions.* Moreover let (M, g) be a compact or an ALF four-manifold. Our primary concern in this paper will be the calculation of the formal integral

$$Z(M, g, \tau) := \sum_{[\nabla^0_L] \in \left(\Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M;\mathbb{Z}))\right)} \frac{1}{\mathrm{Vol}(\mathscr{U}(1)_L)} \int\limits_{\nabla_L \in \mathscr{A}(\nabla^0_L)} \mathrm{e}^{-S(\nabla_L,\tau)} \mathrm{D}\nabla_L \tag{11}$$

or equivalently

$$Z(M, g, \tau) := \sum_{[\nabla^0_L] \in \left(\Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M;\mathbb{Z}))\right)} \int\limits_{[\nabla_L] \in \mathscr{B}(\nabla^0_L)} \mathrm{e}^{-S(\nabla_L,\tau)} \mathrm{D}[\nabla_L]$$

which gives rise to the relevant part of the (Euclidean) *partition function* of the supersymmetrized Abelian gauge theory (3) over (M, g). We note again that in the supersymmetric setting τ is well-defined at the full quantum level hence its appearance in $Z(M,g,\tau)$ is meaningful. Here $\mathrm{D}\nabla_L$ denotes the hypothetical (probably never definable) measure on the infinite dimensional vector space $\mathscr{A}(\nabla^0_L)$ and $\mathrm{D}[\nabla_L]$ is the induced one on the orbit space $\mathscr{B}(\nabla^0_L)$.

Notice that if M is compact the summation reduces to a summation over line bundles i.e., over $H^2(M;\mathbb{Z})$ in accord with [26, 27]. However if M is non-compact then the summation is taken over more finite action classical solutions than bundles. Moreover we will see in Sect. 4 that these allowed classical solutions are not the whole set of finite action classical solutions due to the strong holonomy condition from Definition 2.1.

Next we perform the straightforward summation in (11). Pick a bundle L and an allowed classical solution ∇^0_L on it. With respect to ∇^0_L write a connection $\nabla_L\in\mathscr{A}(\nabla^0_L)$ on the same bundle in the form ∇^0_L+a. Consider the associated decomposition (6) of the action. To be precise we regard ∇_L on $L\cong L\otimes L_0$ as a perturbation of ∇^0_L on L with a connection $\mathrm{d}+a$ on L_0. It then follows that the integral in (11) looks like

$$\int\limits_{[\nabla_L]\in\mathscr{B}(\nabla^0_L)}\mathrm{e}^{-S(\nabla_L,\tau)}\mathrm{D}[\nabla_L]=\mathrm{e}^{-S(\nabla^0_L,\tau)}\int\limits_{[a]\in\mathscr{B}(\nabla^0_L)}\mathrm{e}^{-\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)}}\mathrm{D}[a]$$

where $\mathrm{D}[a]=\mathrm{D}[\nabla_L]$ is the formal induced measure on $\mathscr{B}(\nabla^0_L)$.

As we have seen bundles with $c_1(L)\in\mathrm{Tor}(H^2(M;\mathbb{Z}))$ are flat i.e., the action (5) vanishes along them, consequently the summation in (11) over the torsion part simply gives a numerical factor in the partition function leaving us with a summation over the free lattice part as follows:

$$Z(M,g,\tau)=\left|\mathrm{Tor}(H^2(M;\mathbb{Z}))\right|\left(\sum_{[\nabla^0_L]\in\Lambda^2_{L^2}}\mathrm{e}^{-S(\nabla^0_L,\tau)}\right)\int\limits_{[a]\in\mathscr{B}(\nabla^0_{L_0})}\mathrm{e}^{-\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)}}\mathrm{D}[a]$$

where $\nabla^0_{L_0}$ is the trivial flat connection on the trivial line bundle L_0.

On the way we introduce our theta function. The left hand side of (10) provides us with an L^2 quadratic form q_M on $\Lambda^2_{L^2}\subset\mathscr{H}^2_{L^2}(M)$ which is indefinite according to the splitting $\Lambda^2_{L^2}=\Lambda^+_{L^2}\times\Lambda^-_{L^2}$ into (anti)self-dual parts. The second L^2 Betti number also splits like $b^2_{L^2}(M)=b^+_{L^2}(M)+b^-_{L^2}(M)$. Write $q_M=q^+_M\oplus q^-_M$ for the correspondig decomposition into definite parts. Note that if M is compact and simply connected then q_M is just the intersection form and $b^\pm_{L^2}(M)$ are just the usual signature decomposition of $b^2(M)$. Then $\vartheta_{q^\pm_M}:\mathbb{C}^+\to\mathbb{C}$ is defined by

$$\vartheta_{q^\pm_M}(\tau):=\sum_{n\,\in\underbrace{\mathbb{Z}\times\cdots\times\mathbb{Z}}_{b^\pm_{L^2}(M)}}\mathrm{e}^{\mathbf{i}\pi q^\pm_M(n,n)\tau}$$

and has the following properties taking into account the unimodularity of the intersection form (cf. e.g. [21, Sec. VII.6]): it is holomorphic on the upper half-plane moreover always satisfies the functional equations

$$\begin{cases}\vartheta_{q^\pm_M}(\tau+2)=\vartheta_{q^\pm_M}(\tau) & \text{(``level 2 property'')}\\ \vartheta_{q^\pm_M}(-1/\tau)=(\tau/\mathbf{i})^{\frac12 b^\pm_{L^2}(M)}\,\vartheta_{q^\pm_M}(\tau) & \text{(``modularity of weight }\tfrac12 b^\pm_{L^2}(M)\text{ property'')}\end{cases}\qquad(12)$$

where the square root is the principal value cut along the negative real axis.[1] By the aid of this function we proceed as follows. Making use of (10) the summation in (11) over the remaining

[1] In certain cases $\vartheta_{q^\pm_M}(\tau+1)=\vartheta_{q^\pm_M}(\tau)$ also holds; for example if M is a compact spin manifold.

lattice $\Lambda^2_{L^2}$ gives

$$Z(M,g,\tau) = \left|\mathrm{Tor}(H^2(M;\mathbb{Z}))\right| \vartheta_{q^+_M}(\tau)\vartheta_{q^-_M}(-\overline{\tau}) \int\limits_{\mathscr{B}(\nabla^0_{L_0})} \mathrm{e}^{-\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)}}\,\mathrm{D}[a]. \qquad (13)$$

Note that if $b^-_{L^2}(M) > 0$ then $Z(M,g,\tau)$ is not holomorphic in τ.

We procced further and ask ourselves how to perform the remaining integral in (13) by the aid of Lemma 2.3. It shows that integration in Coulomb gauge is to be taken over a finite dimensional torus and an infinite dimensional Hilbert space. The formal measure for the latter space is $\mathrm{D}[a] = (\det{}'\triangle_0)\mathrm{D}a$ where $\det{}'\triangle_0$ is the formal determinant of the scalar Laplacian without its zero eigenvalues and $\mathrm{D}a$ is some formal measure on the Hilbert space $L^2_l(M;\Lambda^1M)$ restricted to $\overline{\ker\delta} \subset L^2_l(M;\Lambda^1M)$. This determinant enters the story as the Faddeev–Popov determinant in Coulomb gauge in U(1) gauge theory. Of course $\det{}'\triangle_0$ is ill-defined; standard ζ-function regularization might be used to define it. These issues will be investigated in the forthcoming sections.

Referring to Lemma 2.3 since the action vanishes along the torus integrating over it we obtain

$$\int\limits_{\mathscr{B}(\nabla^0_{L_0})} \mathrm{e}^{-\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)}}\,\mathrm{D}[a] = \mathrm{Vol}\left(\frac{\mathscr{H}^1_{L^2}(M)}{2\pi\Lambda^1_{L^2}}\right)\det{}'\triangle_0\left(\int\limits_{(\mathscr{H}^1_{L^2}(M))^\perp\cap\overline{\ker\delta}} \mathrm{e}^{-\frac{\mathrm{Im}\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M)}}\,\mathrm{D}a\right). \qquad (14)$$

In the Coulomb gauge, provided $a \in \mathscr{A}(\nabla^0_{L_0})$ we can re-express the last term in (6) as

$$\|\mathrm{d}a\|^2_{L^2(M)} = (a\,,\,\triangle_1 a)_{L^2(M)}$$

where $\triangle_1 = \delta\mathrm{d} + \mathrm{d}\delta$ is the Laplacian acting on 1-forms and $(\cdot\,,\,\cdot)_{L^2(M)}$ is the L^2 scalar product on the space of 1-forms.

Consequently, collecting all of our findings sofar, we obtain that we have eventually cut down the original integral (11) to a yet highly non-trivial formal integral

$$\int\limits_{(\mathscr{H}^1_{L^2}(M))^\perp\cap\overline{\ker\delta}} \mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1 a\right)_{L^2(M)}}\,\mathrm{D}a, \qquad \overline{\ker\delta} \subset L^2_l(M;\Lambda^1M). \qquad (15)$$

Our aim in the forthcoming sections will be to calculate this integral over various manifolds.

*Remark***.** We make an important comment here related with the non-trivial moduli of flat connections. This comment is motivated by the difficulties we have to face when try to repeat this procedure in the non-compact case in Sect. 4 and 5. As we already noted, given a fixed line bundle L over M with a fixed allowed classical solution $[\nabla^0_L] \in \Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M;\mathbb{Z}))$ to (4) then $\nabla^1_L = \nabla^0_L + a$ is another solution if $a \in \mathscr{H}^1_{L^2}(M)$ since a is simply a flat perturbation of ∇^0_L. Therefore nothing prevents us to carry out the summation procedure again over $\Lambda^2_{L^2} \times \mathrm{Tor}(H^2(M,\mathbb{Z}))$ but this time starting with the decomposition (6) relative to ∇^1_L instead of ∇^0_L. Fortunately since $S(\nabla^0_L,\tau) = S(\nabla^1_L,\tau)$ we obtain the same result (13) and (14).

If M is compact then all classical solutions arise this way hence the summation is unambigous consequently the shape of (13) and (14) is well-defined. However we will see that for non-compact manifolds the situation is not so simple.

3 Compact spaces

In this section we calculate (15) in the compact case via ζ-function and heat kernel techniques. So throughout this section (M,g) denotes a connected, compact, oriented Riemannian four-manifold without boundary. In this situation all the L^2 cohomology groups appeared sofar reduce to ordinary de Rham cohomology groups.

The operator $c\triangle_k$ with $c>0$ real constant is a positive symmetric operator on the orthogonal complement $(\mathscr{H}^k(M))^\perp\subset L^2_l(M;\Lambda^kM)$ with $l\geqq 2$. By the finite dimensional analogue it is therefore natural to define the Gaussian-like integral to be

$$\int\limits_{(\mathscr{H}^k(M))^\perp}\mathrm{e}^{-(a\,,\,c\triangle_k a)_{L^2(M)}}\,\mathrm{D}a:=\pi^{\frac{1}{2}\mathrm{rk}'(c\triangle_k)}\left(\det{}'\,(c\triangle_k)\right)^{-\frac{1}{2}}$$

where the regularized rank and the determinant is yet to be defined somehow. The familiar way to do this is by making use of ζ-function regularization. Since the spectrum of the Laplacian over a compact manifold is non-negative real and discrete, one sets

$$\zeta_{\triangle_k}(s):=\sum_{\lambda\in\mathrm{Spec}\,\triangle_k-\{0\}}\lambda^{-s},\quad\text{with } s\in\mathbb{C}\text{ and }\mathrm{Re}\,s>0\text{ sufficiently large}$$

and observes that this function can be meromorphically continued over the whole complex plane (cf. e.g. [18, Theroem 5.2]) having no pole at $s=0\in\mathbb{C}$. A formal calculation then convinces us that the regularized rank and the determinant of the Laplacian should be

$$\mathrm{rk}'\,\triangle_k:=\zeta_{\triangle_k}(0),\quad \det{}'\triangle_k:=\mathrm{e}^{-\zeta'_{\triangle_k}(0)}$$

yielding $\mathrm{rk}'(c\triangle_k)=\zeta_{\triangle_k}(0)$ and $\det'(c\triangle_k)=c^{\zeta_{\triangle_k}(0)}\mathrm{e}^{-\zeta'_{\triangle_k}(0)}$. Hence

Definition 3.1. *Putting $c:=\mathrm{Im}\tau/8\pi$ in the formal integral above we set*

$$\int\limits_{(\mathscr{H}^k(M))^\perp}\mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_k a\right)_{L^2(M)}}\,\mathrm{D}a:=\pi^{\frac{1}{2}\zeta_{\triangle_k}(0)}\,\mathrm{e}^{\frac{1}{2}\zeta'_{\triangle_k}(0)}\left(\frac{\mathrm{Im}\,\tau}{8\pi}\right)^{-\frac{1}{2}\zeta_{\triangle_k}(0)}=\mathrm{e}^{\frac{1}{2}\zeta'_{\triangle_k}(0)}\left(\frac{\mathrm{Im}\,\tau}{8\pi^2}\right)^{-\frac{1}{2}\zeta_{\triangle_k}(0)}.$$

Therefore the restricted formal integral (15) taking place on the closed subspace $(\mathscr{H}^1(M))^\perp\cap\ker\delta$ only is defined to be

$$\int\limits_{(\mathscr{H}^1(M))^\perp\cap\ker\delta}\mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1 a\right)_{L^2(M)}}\,\mathrm{D}a:=\mathrm{e}^{\frac{1}{2}\zeta'_{\triangle_1|(\mathscr{H}^1(M))^\perp\cap\ker\delta}(0)}\left(\frac{\mathrm{Im}\,\tau}{8\pi^2}\right)^{-\frac{1}{2}\zeta_{\triangle_1|(\mathscr{H}^1(M))^\perp\cap\ker\delta}(0)}$$

where the restricted function $\zeta_{\triangle_1|(\mathscr{H}^1(M))^\perp\cap\ker\delta}$ is defined in the analogous way.

Since this restricted ζ-function is not easy to find we derive a sort of "Fubini principle" to obtain our integral successively from simpler ones.

Lemma 3.1. *With respect to Definition 3.1 we find*

$$\int\limits_{(\mathscr{H}^1(M))^\perp\cap\ker\delta}\mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1 a\right)_{L^2(M)}}\,\mathrm{D}a=\mathrm{e}^{\frac{1}{2}\zeta'_{\triangle_1}(0)-\zeta'_{\triangle_0}(0)}\left(\frac{\mathrm{Im}\,\tau}{8\pi^2}\right)^{\frac{1}{2}\left(\zeta_{\triangle_0}(0)-\zeta_{\triangle_1}(0)\right)}.\tag{16}$$

This is the value of the remaining integral (15) in the compact case.

Remark. Inserting (16) into (14) and then into (13) the calculation of the partition function (11) over compact spaces is now complete.[2]

Proof. In addition to the Hodge decomposition (9) we obviously know that $\mathrm{im}\,\delta\oplus\mathscr{H}^1(M)\subseteq\ker\delta$ and $\mathrm{im}\,\mathrm{d}\cap\ker\delta=\{0\}$ hence $L^2_l(M;\Lambda^1M)\cong\mathrm{im}\,\mathrm{d}\oplus\ker\delta$. Intersecting this with $(\mathscr{H}^1(M))^\perp$ and taking into account that $\mathrm{im}\,\mathrm{d}\cong(\mathscr{H}^0(M))^\perp$ we obtain the further decomposition

$$(\mathscr{H}^1(M))^\perp\cong(\mathscr{H}^0(M))^\perp\oplus\left((\mathscr{H}^1(M))^\perp\cap\ker\delta\right). \tag{17}$$

Applying (17) in the compact case we can write any element $\mathbf{i}a\in L^2_l(M;\Lambda^1M)$ uniquely in the form $a=\mathrm{d}f+\alpha$ with $\mathbf{i}f\in L^2_{l+1}(M;\Lambda^0M)$ a function and $\mathbf{i}\alpha\in L^2_l(M;\Lambda^1M)$ satisfying $\delta\alpha=0$. A simple calculation ensures us that if $\mathbf{i}f\in L^2_{l+2}(M;\Lambda^0M)$ then

$$(a\,,\,\triangle_1a)_{L^2(M)}=(\mathrm{d}f+\alpha\,,\,\triangle_1(\mathrm{d}f+\alpha))_{L^2(M)}=\left(f\,,\,\triangle_0^2f\right)_{L^2(M)}+(\alpha\,,\,\triangle_1\alpha)_{L^2(M)} \tag{18}$$

where $\triangle_0^2$ is the square of the scalar Laplacian on (M,g). Taking into account (17) and (18) we obtain that $(\mathrm{Spec}\,\triangle_1-\{0\})=(\mathrm{Spec}\,\triangle_0^2-\{0\})\sqcup(\mathrm{Spec}\,\triangle_1|_{(\mathscr{H}^1(M))^\perp\cap\ker\delta})$. This decomposition together with the proof of [18, Theorem 5.2] ensures us that

$$\zeta_{\triangle_1}=\zeta_{\triangle_0^2}+\zeta_{\triangle_1|(\mathscr{H}^1(M))^\perp\cap\ker\delta}$$

consequently Definition 3.1 yields the following Fubini-like successive formula

$$\int\limits_{(\mathscr{H}^1(M))^\perp}\mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1a\right)_{L^2(M)}}\,\mathrm{D}a=\left(\int\limits_{(\mathscr{H}^0(M))^\perp}\mathrm{e}^{-\left(f\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_0^2f\right)_{L^2(M)}}\,\mathrm{D}f\right)\left(\int\limits_{(\mathscr{H}^1(M))^\perp\cap\ker\delta}\mathrm{e}^{-\left(\alpha\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1\alpha\right)_{L^2(M)}}\,\mathrm{D}\alpha\right).$$

Taking into account that $\zeta_{\triangle_0^2}(s)=\zeta_{\triangle_0}(2s)$ hence $\mathrm{rk}'(c\triangle_0^2(0))=\zeta_{\triangle_0}(0)$ as well as $\det'(c\triangle_0^2)=c^{\zeta_{\triangle_0}(0)}\mathrm{e}^{-2\zeta'_{\triangle_0}(0)}$ we obtain the lemma. $\diamond$

The S-duality properties of the partition function are concentrated in its $\vartheta_{q^+M}(\tau)\vartheta_{q^-M}(-\overline{\tau})$ term and the exponent of $\mathrm{Im}\tau/8\pi^2$ in (16). Hence we shall focus our attention to this exponent. Over a compact four-manifold (M,g) without boundary it is well-known [18, Theorem 5.2] that

$$\zeta_{\triangle_k}(0)=-\dim\ker\triangle_k+\frac{1}{16\pi^2}\int\limits_M\mathrm{tr}(u_k^4)\mathrm{d}V \tag{19}$$

where the sections $u_k^p\in C^\infty(M;\mathrm{End}(\Lambda^kM))$ with $p=0,1,\dots$ appear [18, Chapter 3] in the coefficients of the short time asymptotic expansion of the heat kernel for the k-Laplacian

$$\sum_{\lambda\in\mathrm{Spec}\triangle_k}\mathrm{e}^{-\lambda t}\sim\frac{1}{(4\pi t)^2}\sum_{p=0}^{+\infty}\left(\int\limits_M\mathrm{tr}(u_k^p)\mathrm{d}V\right)t^{\frac{p}{2}}\quad\text{as } t\to0.$$

These functions are expressible with the curvature of (M,g) and one can demonstrate [11, p. 340] that

$$u_0^4=\frac{1}{360}\left(2|R|^2-2|r|^2+5s^2\right)$$

[2]Including the Faddeev–Popov determinant the full contribution of the determinants to the partition function (11) is $(\det'\triangle_1)^{-\frac12}(\det'\triangle_0)^2$ that is, $\mathrm{e}^{\frac12\zeta'_{\triangle_1}(0)-2\zeta'_{\triangle_0}(0)}$. This could be further analyzed however we skip this here.

and

$$\mathrm{tr}(u_1^4) = \frac{1}{360}\left(-22|R|^2 + 172|r|^2 - 40s^2\right)$$

where R is the Riemann, r is the Ricci and s is the scalar curvature of the metric on M. On substituting these into the exponent of $\mathrm{Im}\tau/8\pi^2$ in (16) we find that it takes the shape

$$\frac{1}{2}\left(\zeta_{\triangle_0}(0) - \zeta_{\triangle_1}(0)\right) = \frac{1}{2}\left(b^1(M) - b^0(M) + \frac{1}{\pi^2}\int\limits_M \left(\frac{1}{120}|R|^2 - \frac{87}{2880}|r|^2 + \frac{1}{128}s^2\right)\mathrm{d}V\right).$$

The time has come to write down the modular weights of the partition function (11). First note that

$$\mathrm{Im}\left(-\frac{1}{\tau}\right) = \frac{1}{\tau\overline{\tau}}\mathrm{Im}\,\tau$$

that is, it is modular of holomorphic and anti-holomorphic weights $(-1,-1)$ respectively. Secondly it follows from (12) that up to $\mathbf{i}$'s the modular weight of $\vartheta_{q^+_M}(\tau)$ and $\vartheta_{q^-_M}(-\overline{\tau})$ is $\frac{1}{2}b^\pm_{L^2}(M)$ hence referring to the shape of the partition function in (13) we find that its holomorphic and anti-holomorphic modular weights (α,β) from (2) are

$$\frac{1}{4}\left(\chi(M) \pm \sigma(M) - \frac{1}{\pi^2}\int\limits_M \left(\frac{1}{60}|R|^2 - \frac{87}{1440}|r|^2 + \frac{1}{64}s^2\right)\mathrm{d}V\right) \tag{20}$$

(to be precise the "+" is for holomorphic and the "−" is for anti-holomorphic). In this formula $\chi(M)\pm\sigma(M) = 2b^0(M) - 2b^1(M) + 2b^\pm(M)$ are the linear combinations of the Euler characteristic and the signature of the manifold.

Since (20) is not zero in general we conclude that in the naive theory (3) S-duality breaks down because its partition function (11) is not modular in τ. Moreover, comparing (20) with Witten's calculation [26, 27] whose result is simply the topological term $\frac{1}{4}(\chi(M)\pm\sigma(M))$ we find an analytic correction which vanishes only in the rare situation if (M,g) happens to be flat (there exist only 27 connected compact orientable flat four-manifolds and as many as 74 if the non-orientable ones are also included [13]).

Nevertheless this curvature correction does not destroy the main conclusion in [26, 27] namely that Abelian S-duality over a compact space can be restored within the framework of local quantum field theories by adding appropriate gravitational terms to the Lagrangian (3) in order to cancel the modular anomaly coming from (20). As it was observed by Witten [26, 27] it is quite remarkable that although the individual Betti numbers not, their combinations $\chi(M)\pm\sigma(M)$ are expressible as integrals of local curvature densities (cf. e.g. [1, p. 370-371]): the Gauss–Bonnet–Chern theorem gives

$$\chi(M) = \frac{1}{8\pi^2}\int\limits_M \mathrm{tr}(R\wedge *R) = \frac{1}{8\pi^2}\int\limits_M \left(|R|^2 - \left|r - \frac{s}{4}g\right|^2\right)\mathrm{d}V \tag{21}$$

for the Euler characteristic and the Hirzebruch signature theorem asserts that

$$\sigma(M) = -\frac{1}{24\pi^2}\int\limits_M \mathrm{tr}(R\wedge R) \tag{22}$$

holds for the signature. Hence because obviously the full weights (20) continue to be integrals of local densities, modular anomaly cancels by adding further gravitational terms to the Lagrangian (3) (called "c-numbers") however they are not of the form $a(\tau,\overline{\tau})\mathrm{tr}(R\wedge *R)+b(\tau,\overline{\tau})\mathrm{tr}(R\wedge R)$ as claimed in [26, 27]. Their shape can be read off from (20).

4 Significance of the strong holonomy condition

In this section we clarify the role of the strong holonomy condition imposed on connections in Definition 2.1. This condition excludes certain finite action classical solutions from the set of connections contributing to the partition function (11). It will turn out now that without this condition the partition function would have pathological behaviour. To show this we take the underlying space to be the multi-Taub–NUT spaces; these are quite important hyper-Kähler ALF spaces (see next section).

So let (M_V, g_V) be the 1-Taub–NUT space. This is a non-flat hyper-Kähler geometry on $\mathbb{R}^4\cong M_V$ and represents the $s=1$ member of the multi-Taub–NUT series ($s\in\mathbb{N}$ refers to the number of NUTs). For our purposes here we refer to [7] for a description of this space. First of all one can demonstrate that it is an ALF space and its unique (up to scale) non-topological L^2 harmonic 2-form arises as follows. Put an orientation onto M_V induced by any complex structure in the hyper-Kähler family. As it is well-known the 1-Taub–NUT space admits a non-trivial L^2 harmonic 2-form ω. This 2-form can be constructed as the exterior derivative of the metric dual of the Killing field generating an isometric action of U(1) on the 1-Taub–NUT space [10]. One can also obtain it by the conformal rescaling method when hunting for SU(2) anti-instantons [6]. Taking into account that $H^2(M_V;\mathbb{R})=\{0\}$ the condition $\mathrm{d}\omega=0$ is equivalent to the existence of an imaginary valued 1-form A such that $\mathrm{d}A=\mathbf{i}\omega$. In other words there exists an U(1)-connection $\nabla^1_{L_0}:=\mathrm{d}+A$ on the trivial bundle $L_0\cong M_V\times\mathbb{C}$ such that $\mathbf{i}F_{\nabla^1_{L_0}}\in L^2(M_V;\Lambda^2M_V)$ as well as $\mathrm{d}F_{\nabla^1_{L_0}}=0$ and $\delta F_{\nabla^1_{L_0}}=0$. That is, $\nabla^1_{L_0}$ is a finite action classical solution to the Maxwell equations (4) over the 1-Taub–NUT space (moreover it is anti-self-dual). As a consequence of the linearity of the Abelian gauge theory and that ω is topologically trivial, for all $c\in\mathbb{R}$ the rescaled connection $\nabla^c_{L_0}:=\mathrm{d}+cA$ is another gauge inequivalent solution on the same bundle with action proportional to c^2. In fact this 1-parameter family of solutions is anti-self-dual hence from (5) we obtain that

$$S(\nabla^c_{L_0},\tau)=\mathbf{i}\pi\overline{\tau}c^2,\quad c\in\mathbb{R}$$

in contrast to the quantized nature of (10). Now picking any L^2_l perturbation $a\in\mathscr{A}(\nabla^0_{L_0})$ as before we find via (6) that

$$S(\nabla^c_{L_0}+a,\tau)=\mathbf{i}\pi\overline{\tau}c^2+\frac{\mathrm{Im}\,\tau}{8\pi}\|\mathrm{d}a\|^2_{L^2(M_V)}.$$

Moreover note that if $c\neq 0$ then of course $\nabla^c_{L_0}\notin\mathscr{A}(\nabla^0_{L_0})$ (otherwise we would find $F_{\nabla^c_{L_0}}=0$ by Stokes' theorem) hence treating these solutions not as L^2_l perturbations of the trivial flat connection is correct even from the functional analytic viewpoint.

Suppose now that we want to calculate the partition function by simply mimicing the calculation in [26, 27] designed for the compact case i.e., integrating over connections on a given bundle and then summing over line bundles. Then, taking into account that $M_V\cong\mathbb{R}^4$ is simply connected hence there are no non-trivial flat connections and there is only one line bundle on it the partition function (11) is expected to look like

$$Z(M_V, g_V, \tau) = \left(\int\limits_{-\infty}^{+\infty} \mathrm{e}^{\mathbf{i}\pi(-\overline{\tau})c^2}\mathrm{d}c\right) \det{}'\triangle_0 \left(\int\limits_{(\mathscr{H}^1_{L^2}(M))^{\perp}\cap\overline{\ker\delta}} \mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1 a\right)_{L^2(M)}}\,\mathrm{D}a\right). \tag{23}$$

This formula replaces (13) and the last term is just the formal integral already appeared in (15). Since $\mathrm{d}c$ is *a fortiori* the Lebesgue measure on $\mathbb{R}$ and $\mathrm{Im}\,\tau > 0$ the Gaussian integral in the front converges and we plainly obtain

$$Z(M_V, g_V, \tau) = (\mathbf{i}/\overline{\tau})^{\frac{1}{2}}\,(\det{}'\triangle_0) \left(\int\limits_{(\mathscr{H}^1_{L^2}(M_V))^{\perp}\cap\,\overline{\ker\delta}} \mathrm{e}^{-\left(a\,,\,\frac{\mathrm{Im}\tau}{8\pi}\triangle_1 a\right)_{L^2(M_V)}}\,\mathrm{D}a\right).$$

However this formula is far from being able to satisfy something like (2) moreover the expected weights of τ and $\overline{\tau}$ do not look like in the compact case consequently modular anomaly cannot be cancelled with the same mechanism.

The situation gets even worse over the multi-Taub–NUT spaces (M_V, g_V) with $s > 1$ NUTs. In this case the aforementioned anti-self-dual L^2 solutions $\nabla^c_{L_0} = \mathrm{d} + cA$ for all $c \in \mathbb{R}$ exist on L_0 (cf. [7]). Hence if L is a generic line bundle over M_V carrying a classical solution ∇^0_L of the Maxwell equations then we obtain a similar 1-parameter family $\nabla^c_L := \nabla^0_L + cA$ on $L \cong L \otimes L_0$. Consequently in this case we would end up with more complicated ill expressions for the partition function.

Observe that we have run into this divergence problem not because of using some inappropriate regularization method for the infinite dimensional integral (15); rather the problem arose from the multi-Taub–NUT geometry itself in the sense that it possesses “too many” finite action classical solutions, enough to distort the partition function. We encounter the same difficulty over the Riemannian Schwarzschild and Kerr geometries, too (these are also ALF spaces, cf. [5]). These problematic solutions are truely non-compact phenomena in the sense that the L^2 harmonic 2-forms their curvatures represent are non-topological. This is the moment where the powerful nature of Definition 2.1 shows up: it rules out most of these pathological solutions but not all of them! We will see in Sect. 5 that if all of these non-topological problematic solutions are excluded then the modular properties would be destroyed again but in a different way.

In fact the remaining classical finite action solutions excluded by Definition 2.1 represent surface operators (cf. e.g. [22, 23]) attached to the infinitely distant surface $i : B_\infty \subset X$ (see Sect. 5) where X is a natural compactification of the ALF space [4, 7, 12]. That is, they provide us with observables of the form $\nabla^c_{\tilde{L}} \mapsto O_{B_\infty}(\nabla^c_{\tilde{L}}) := \exp(\int_{B_\infty} i^* F_{\nabla^c_{\tilde{L}}}) \in \mathbb{C}$ in our theory where $\nabla^c_{\tilde{L}}$ is a singular connection on the extended line bundle $\tilde{L}$ over X.

We also note that if the Gaussian integral in (23) could be somehow replaced by

$$\frac{1}{\mathbf{i}}\int\limits_{-\infty}^{+\infty} \mathrm{e}^{\mathbf{i}\pi(-\overline{\tau})c^2}\cot(\pi c)\mathrm{d}c := \lim_{\varepsilon\to 0}\frac{1}{\mathbf{i}}\int\limits_{-\infty}^{+\infty} \mathrm{e}^{\mathbf{i}\pi(-\overline{\tau})(c+\mathbf{i}\varepsilon)^2}\cot(\pi(c+\mathbf{i}\varepsilon))\mathrm{d}c = \vartheta_{q^-_{M_V}}(-\overline{\tau})$$

i.e., the restricted Feynman measure would turn out to be the singular measure $\frac{\cot(\pi c)}{\mathbf{i}}\mathrm{d}c$ then modular properties of $Z(M_V, g_V, \tau)$ are also recovered.

Accepting the definition of the partition function via (11) a second problem arises if one tries to calculate the infinite dimensional integral (15) over an ALF space akin to (16) i.e., by making use of a ζ-function regularization. This time one has to face the problem that in general the spectra of Laplacians on k-forms are not discrete hence the existence of their ζ-functions is not obvious. This obstacle will be resolved in the next section by a straightforward regularization method based on truncating the non-compact space and imposing Dirichlet boundary condition on the boundary.

5 Asymptotically locally flat spaces

In this section we repeat our calculations of (11) over ALF spaces. The calculation goes along the same lines as in the compact space with obvious technical modifications.

ALF spaces are also referred to sometimes as *gravitational instantons of ALF type* in the broad or narrow sense if in addition their metric is Ricci-flat or hyper-Kähler respectively. Ricci flat examples are the Riemannian Schwarzschild and Kerr manifolds while the flat space $\mathbb{R}^3 \times S^1$ and the multi-Taub–NUT spaces (also called A_k ALF or ALF Gibbons–Hawking spaces) and the D_k ALF spaces (including the Atiyah–Hitchin manifold as the D_0 case) provide hyper-Kähler examples. The multi-Taub–NUT spaces apparently are subject to recent investigations in mathematical physics [2, 3, 8, 28].

To begin with, we recall the definition of these spaces taken from [7] for instance. 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

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

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

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

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 $(\pi^* g_{B_\infty})'(\rho) \sim O(1)$ and $h'_F(\rho) \sim O(1)$ and the extensions for $\rho < +\infty$ preserve the properties of the original fields. Furthermore, we require the Riemann curvature R of g to decay like

$$|\phi^*(\nabla^k R|_W)| \sim O(\rho^{-3-k}), \quad k = 0, 1, 2, \dots \tag{24}$$

where R is regarded as a map $R : C^\infty(M; \Lambda^2 M) \to C^\infty(M; \Lambda^2 M)$ and its pointwise norm is calculated accordingly in an orthonormal frame. The definition of the metric shows that the volume of our spaces is infinite however from the curvature decay (24) follows that both $\chi_{L^2}(M)$ and $\sigma_{L^2}(M)$ defined by (21) and (22) respectively, remain finite.

Take any ALF space (M, g) as above. Fix a real number $\rho > 0$ and let $(\overline{M}_\rho, g|_{\overline{M}_\rho})$ be the truncated manifold i.e., $\overline{M}_\rho$ is a four-manifold with connected boundary $\partial\overline{M}_\rho$ such that the distance of any point $x \in \overline{M}_\rho$ satisfies $d(x_0, x) \leqq \rho$ with respect to an interior point $x_0 \in M_\rho$.

With $l \geqq 2$ define the spaces

$$LD^2_l(\overline{M}_\rho; \Lambda^k \overline{M}_\rho) := \{\varphi \in L^2_l(\overline{M}_\rho; \Lambda^k \overline{M}_\rho) \mid \varphi|_{\partial\overline{M}_\rho} = 0 \text{ and } \delta\varphi|_{\partial\overline{M}_\rho} = 0 \text{ a.e.}\}$$

and

$$\mathscr{H}^k_D(\overline{M}_\rho) := \{\varphi \in LD^2_l(\overline{M}_\rho;\Lambda^k\overline{M}_\rho)\mid \mathrm{d}\varphi = 0 \text{ and } \delta\varphi = 0\} \cong \ker(\triangle_k|_{\overline{M}_\rho}).$$

Also write $b^k_D(\overline{M}_\rho) := \dim \mathscr{H}^k_D(\overline{M}_\rho)$ for the corresponding Dirichlet–Betti numbers. Note that pushing $\partial\overline{M}_\rho$ toward infinity these numbers satisfy

$$b^k_D(\overline{M}_\rho) \leqq b^k_{L^2}(M) \tag{25}$$

where these latter numbers are the true L^2 Betti numbers of the original open manifold.

Now we are in a position to calculate the partition function over an ALF space as follows. First define $Z(M,g,\tau)$ by (11) including the remaining integral (15). To calculate this integral we observe the following things. Let us work over $(\overline{M}_\rho, g|_{\overline{M}_\rho})$. Referring to the Hodge decomposition theorem with respect to the Dirichlet boundary condition [24, Proposition 5.9.8] the decomposition (17) continues to hold providing us with the validity of (18). Secondly, under the Dirichlet condition the Laplacian has a discrete spectrum on LD^2_l consequently introducing $\zeta_{\triangle_k}$ also makes sense. This enables us to express the integral (15) over $(\overline{M}_\rho, g|_{\overline{M}_\rho})$ with (16). Then sending the boundary to infinity if the limit exists then the integral (15) is defined by this limit. In this way we also obtain an expression for the original partition function (11).

Now following [11] we quickly summarize the changes of the heat kernel formulæ in the situation of the Dirichlet boundary condition. The expression for the zero value of the ζ-function looks like

$$\zeta_{\triangle_k|_{\overline{M}_\rho}}(0) = -\dim\ker(\triangle_k|_{\overline{M}_\rho}) + \frac{1}{16\pi^2}\int\limits_{\overline{M}_\rho}\mathrm{tr}(u^4_k)\mathrm{d}V + \frac{1}{16\pi^2}\int\limits_{\partial\overline{M}_\rho}\mathrm{tr}(v^4_k)\mathrm{d}V|_{\partial\overline{M}_\rho}$$

and the integrals are again defined via the short time heat kernel expansion

$$\sum_{\lambda\in\mathrm{Spec}(\triangle_k|_{\overline{M}_\rho})}\mathrm{e}^{-\lambda t} \sim \frac{1}{(4\pi t)^2}\sum_{p=0}^{+\infty}\left(\int\limits_{\overline{M}_\rho}\mathrm{tr}(u^p_k)\mathrm{d}V + \int\limits_{\partial\overline{M}_\rho}\mathrm{tr}(v^p_k)\mathrm{d}V|_{\partial\overline{M}_\rho}\right)t^{\frac{p}{2}} \quad \text{as } t\to 0.$$

The curvature expressions on the bulk are similar to the compact case [11, Theorem 4.5.1]:

$$u^4_0 = \frac{1}{360}\left(2|R|^2 - 2|r|^2 + 5s^2 + 12\triangle_0 s\right)$$

and

$$\mathrm{tr}(u^4_1) = \frac{1}{360}\left(-22|R|^2 + 172|r|^2 - 40s^2 + 48\triangle_0 s\right).$$

However the contribution of the boundary is quite complicated. Consider a collar $U\subset\overline{M}_\rho$ of the boundary which looks like $U\cong\partial\overline{M}_\rho\times(-1,0]$. Pick an orthonormal frame field $(e_1,\dots,e_4)$ along U such that e_4 is orthogonal to the boundary. Then define the second fundamental form of the boundary by

$$\Pi_{ij} := g|_{\overline{M}_\rho}(\nabla_{e_i}e_j\,,\,e_4),\quad i,j=1,2,3$$

as well as let $\overline{\nabla}$ be the restriction of the Levi–Civita connection to the boundary. The boundary contributions are [11, Theorem 4.5.1]

$$\begin{aligned} v^4_0 \;=\; & \frac{1}{360}\Big(-138\nabla_{e_4}s + 140s\Pi_{ii} + 4R_{i4i4}\Pi_{jj} - 12R_{i4j4}\Pi_{ij} + 4R_{ijkj}\Pi_{ik} \\ & + 24\overline{\nabla}^2_{e_i,e_i}\Pi_{jj} + \frac{40}{21}\Pi_{ii}\Pi_{jj}\Pi_{kk} - \frac{88}{7}\Pi_{ij}\Pi_{ij}\Pi_{kk} + \frac{320}{21}\Pi_{ij}\Pi_{jk}\Pi_{ik}\Big)\end{aligned}$$

and

$$\begin{aligned}\mathrm{tr}(v_1^4) &= \frac{1}{360}\Big(- 192\nabla_{e_4}s + 200s\Pi_{ii} + 16R_{i4i4}\Pi_{jj} - 48R_{i4j4}\Pi_{ij} + 16R_{ijkj}\Pi_{ik} \\ &+ 96\overline{\nabla}^2_{e_i,e_i}\Pi_{jj} + \frac{160}{21}\Pi_{ii}\Pi_{jj}\Pi_{kk} - \frac{352}{7}\Pi_{ij}\Pi_{ij}\Pi_{kk} + \frac{1280}{21}\Pi_{ij}\Pi_{jk}\Pi_{ik}\Big).\end{aligned}$$

Now we demonstrate that all contributions from the boundary get vanish as we move the boundary toward infinity.

Lemma 5.1. *Let (M,g) be an ALF space carefully defined above and let $(\overline{M}_\rho, g|_{\overline{M}_\rho})$ be its truncation.*

Then

$$\lim_{\rho\to+\infty}\int\limits_{\overline{M}_\rho}(\triangle_0 s)\mathrm{d}V = 0 \quad \textit{and} \quad \lim_{\rho\to+\infty}\int\limits_{\partial\overline{M}_\rho}(\nabla_{e_4}s)\mathrm{d}V|_{\partial\overline{M}_\rho} = 0$$

moreover $\Pi_{ij}\sim O(\rho^{-1})$ for all $i,j=1,2,3$.

Proof. Referring to the definition of an ALF space we find that $\mathrm{Vol}(\partial\overline{M}_\rho)\sim O(\rho^2)$ and from the curvature decay (24) we get $|\nabla r|\sim O(\rho^{-4})$ and $|\nabla s|\sim O(\rho^{-4})$. Regarding the first integral, applying Stokes' theorem to convert it into an integral of ∇r along the boundary, the result follows. The second integral also decays in a straightforward way.

Finally, since asymptotically $\Pi_{ij}(x)\sim\Gamma^4_{ij}(x)$ for all $x\in\partial\overline{M}_\rho$, a simple calculation yields that over an ALF space Π also decays as claimed. $\diamond$

We proceed further and take the limit $\rho\to+\infty$ to recover the original ALF space (M,g). Applying the lemma and the curvature decay (24) we can see that all the terms involving $\triangle_0 s$, $\nabla_{e_4}s$ and Π do not contribute to the integrals. Therefore, since $\dim\ker(\triangle_k|_{\overline{M}_\rho}) = b^k_D(\overline{M}_\rho)$ we obtain over the original ALF space (M,g) that the exponent of $\mathrm{Im}\tau/8\pi^2$ in (16) converges and gives again

$$\frac{1}{2}\left(b^1_D(M) - b^0_D(M) + \frac{1}{\pi^2}\int\limits_M\left(\frac{1}{120}|R|^2 - \frac{87}{2880}|r|^2 + \frac{1}{128}s^2\right)\mathrm{d}V\right) < +\infty$$

as expected. Note that by (25) the limiting Dirichlet–Betti numbers $b^k_D(M)$ are well-defined.

From (13) we already know that the contributions of the ϑ-functions to the modular weights are $b^\pm_{L^2}(M)$ hence we eventually obtain that over an ALF space (M,g) if the partition function (11) exists[3] then its modular weights are equal to:

$$\frac{1}{4}\left(2b^0_D(M) - 2b^1_D(M) + 2b^\pm_{L^2}(M) - \frac{1}{\pi^2}\int\limits_M\left(\frac{1}{60}|R|^2 - \frac{87}{1440}|r|^2 + \frac{1}{64}s^2\right)\mathrm{d}V\right). \tag{26}$$

It is worth comparing this with the compact case (20). The maximum principle provides us that $b^0_D(\overline{M}_\rho)=0$ hence $b^0_D(M)=0$ too, moreover Yau's theorem [29] gives $b^0_{L^2}(M)=0$ since an ALF space is complete and has infinite volume. Additionally, since [24, Proposition 5.9.9]

$$b^k_D(\overline{M}_\rho) = \dim H^k(\overline{M}_\rho, \partial\overline{M}_\rho\,;\mathbb{R})$$

[3] We continue to abandon questions about the existence of the contribution of the determinants to the partition function in the limiting ALF case i.e., we do not check the limit of $\exp\left(\frac{1}{2}\zeta'_{\triangle_1|_{\overline{M}_\rho}}(0) - 2\zeta'_{\triangle_0|_{\overline{M}_\rho}}(0)\right)$ as $\rho\to+\infty$.

the part of the relative de Rham exact sequence

$$\{0\} \cong H^0(\overline{M}_\rho, \partial\overline{M}_\rho\,; \mathbb{R}) \longrightarrow H^0(\overline{M}_\rho\,; \mathbb{R}) \longrightarrow H^0(\partial\overline{M}_\rho\,; \mathbb{R})$$

$$\longrightarrow H^1(\overline{M}_\rho, \partial\overline{M}_\rho\,; \mathbb{R}) \longrightarrow H^1(\overline{M}_\rho\,; \mathbb{R}) \longrightarrow H^1(\partial\overline{M}_\rho\,; \mathbb{R})$$

shows that if $H^1(\partial\overline{M}_\rho\,; \mathbb{R}) = \{0\}$ then $b^1_D(M) = b^1(M)$ moreover $b^1(M) = b^1_{L^2}(M)$ by the remark of Lemma 2.2. Consequently for an ALF space (M, g) with $M = K \cup W$ and W representing its neck, if the mild topological condition

$$H^1(W; \mathbb{R}) = \{0\} \tag{27}$$

holds for its neck then

$$2b^0_D(M) - 2b^1_D(M) + 2b^{\pm}_{L^2}(M) = \chi_{L^2}(M) \pm \sigma_{L^2}(M)$$

that is, the modular weights can be written as integrals of the curvature like in the case of compact spaces. Consequently the modular anomaly cancels as before (cf. Sect. 3). For example, the multi-Taub–NUT family (whose members are hyper-Kähler ALF spaces) satisfies (27) while the Riemannian Schwarzschild and Kerr spaces (which are Ricci flat ALF spaces) not.

6 Conclusion

Let us summarize our approach to the partition function of Abelian gauge theory over ALF geometries.

We define the partition function via (11) which is a natural generalization of Witten's formula for compact manifolds. The crucial concept here is to impose the strong holonomy condition at infinity to attenuate the finite action classical solutions contributing to the partition function. After summation the resulting integral (15) is calculated by a regularization method based on truncating the original manifold and imposing the Dirichlet boundary condition on the boundary and then sending the boundary to infinity.

As a result the partition function can be calculated and up to a numerical factor it is a level 2 modular form like in the compact case i.e., it transforms as (2) with $b^{\pm}(M)$ replaced by $b^{\pm}_{L^2}(M)$. The (anti-)holomorphic modular weights are in general not zero hence S-duality fails in the naive theory (3). However, if in addition the simple condition (27) on the topology of the infinity of the ALF space holds then the modular anomaly can be removed by adding gravitational counter terms to the Lagrangian. Consequently S-duality can be restored within the realm of local quantum field theories as in the compact case.

This program can be carried out at least in the case of the multi-Taub–NUT geometries. For instance for the 1-Taub–NUT space (M_V, g_V) with its standard orientation the result is as follows. On $M_V \cong \mathbb{R}^4$ there are no non-trivial flat connections and regarding L^2 cohomology one knows that $b^0_{L^2}(M_V) = b^1_{L^2}(M_V) = 0$ moreover $b^2_{L^2}(M_V) = b^-_{L^2}(M_V) = 1$ yielding $\Lambda^2_{L^2} = \Lambda^-_{L^2} \cong \mathbb{Z}$ and $\chi_{L^2}(M_V) = 1$ and $\sigma_{L^2}(M_V) = -1$. Moreover for the hyper-Kähler metric $r = 0$ and $s = 0$ hence by (21) the exponent of $\mathrm{Im}\,\tau/8\pi^2$ is

$$\frac{1}{240\pi^2}\int\limits_{M_V}|R|^2\mathrm{d}V = \frac{1}{30}\cdot\frac{1}{8\pi^2}\int\limits_{M_V}\mathrm{tr}(R\wedge *R) = \frac{1}{30}\chi_{L^2}(M_V) = \frac{1}{30}.$$

Consequently

$$Z(M_V, g_V, \tau) = \frac{(\det{}'\triangle_0)^2}{(\det{}'\triangle_1)^{\frac{1}{2}}}\,\vartheta_{q^-_{M_V}}(-\overline{\tau})\left(\frac{\mathrm{Im}\,\tau}{8\pi^2}\right)^{\frac{1}{30}}.$$

The modular weights (26) are $(-\frac{1}{30}, \frac{7}{15})$. Since $H^1(W;\mathbb{R}) \cong H^1(S^3;\mathbb{R}) = \{0\}$ holds for the infinity of this space, condition (27) is satisfied therefore the modular anomaly can be removed by adding counter terms to the original naive action (3).

Finally we remark that in our opinion the topological condition (27) is an artifact and should be removed from the construction by calculating $\zeta_{\triangle_k}(0)$ directly on the non-compact manifold without truncating it and imposing any boundary condition.

References

[1] Besse, A.: *Einstein manifolds*, Springer, Berlin (1987);

[2] Cherkis, S.: *Moduli spaces of instantons on the Taub–NUT space*, Commun. Math. Phys. **290**, 719-736 (2009);

[3] Cherkis, S.: *Instantons on the Taub–NUT space*, preprint, 32 pp., arXiv: `0902.4724 [hep-th]` (2009);

[4] Etesi, G.: *The topology of asymptotically locally flat gravitational instantons*, Phys. Lett. **B641**, 461-465 (2006);

[5] Etesi, G., Hausel, T.: *Geometric interpretation of Schwarzschild instantons*, Journ. Geom. Phys. **37**, 126-136 (2001);

[6] Etesi, G., Hausel, T.: *Geometric construction of new Yang–Mills instantons over Taub–NUT space*, Phys. Lett. **B514**, 189-199 (2001);

[7] Etesi, G., Jardim, M.: *Moduli spaces of self-dual connections over asymptotically locally flat gravitational instantons*, Commun. Math. Phys. **280**, 285-313 (2008), Erratum: ibid. **288**, 799-800 (2009);

[8] Etesi, G., Szabó, Sz.: *Harmonic functions and instanton moduli spaces on the multi-Taub–NUT space*, Commun. Math. Phys. **301**, 175-214 (2011);

[9] Frenkel, E.: *Gauge theory and Langlands duality*, preprint, 32 pp., arXiv: `0906.2747 [math.RT]` (2009);

[10] Gibbons, G.W.: *The Sen conjecture for fundamental monopoles of distinct types*, Phys. Lett. **B382**, 53-59 (1996);

[11] Gilkey, P.B.: *Invariance theory, the heat equation, and the Atiyah–Singer index theorem*, CRC Press, Boca Raton, Florida (1995);

[12] Hausel, T., Hunsicker, E., Mazzeo, R.: *Hodge cohomology of gravitational instantons*, Duke Math. Journ. **122**, 485-548 (2004);

[13] Hillman, J.: *Four-manifolds, geometries and knots*, Geometry and Topology Monographs **5**, Geometry and Topology Publications, Warick (2002);

[14] Kapustin, A., Witten, E.: *Electric-magnetic duality and the geometric Langlands program*, preprint, 225 pp., arXiv: `hep-th/0604151` (2006);

[15] Knapp, A.W.: *Elliptic curves*, Princeton Univ. Press, Princeton (1992);

[16] Montonen, C., Olive, D.I.: *Magnetic monopoles as gauge particles?*, Phys. Lett. **B72**, 117-120 (1977);

[17] Olive, D.I., Witten, E.: *Supersymmetry algebras that include topological charges*, Phys. Lett. **B78**, 97-101 (1978);

[18] Rosenberg, S.: *The Laplacian on a Riemannian manifold*, Cambridge Univ. Press, Cambridge (1997);

[19] Segal, G., Selby, A.: *The cohomology of the space of magnetic monopoles*, Commun. Math. Phys. **177**, 775–787 (1996);

[20] Sen, A.: *Dyon-monopole bound states, self-dual harmonic forms on the multi-monopole moduli space, and* $\mathrm{SL}(2,\mathbb{Z})$ *invariance in string theory*, Phys. Lett. **B329**, 217-221 (1994);

[21] Serre, J.-P.: *A course in arithmetic*, GTM **7**, Springer-Verlag, New York (1973);

[22] Tan, M.-C.: *Surface operators in Abelian gauge theory*, Journ. High Energy Phys. 0905:104.2009 (2009);

[23] Tan, M.-C.: *Surface operators in* $N = 2$ *Abelian gauge theory*, Journ. High Energy Phys. 0909:047.2009 (2009);

[24] Taylor, M.E.: *Partial differential equations I*, AMS **115**, Springer, New York (1996);

[25] Vafa, C. Witten, E.: *Strong coupling test of S-duality*, Nucl. Phys. **B431**, 3-77 (1994);

[26] Witten, E.: *On S-duality in Abelian gauge theory*, Selecta Math. **1**, 383-410 (1995);

[27] Witten, E.: *Conformal field theory in four and six dimensions*, preprint, 16 pp., arXiv: `0712.0157 [math.RT]` (2007);

[28] Witten, E.: *Branes, instantons and Taub–NUT spaces*, Journ. High Energy Phys. 0906:067 (2009);

[29] Yau, S.-T.: *Harmonic functions on complete Riemannian manifolds*, Commun. Pure Appl. Math. **28**, 201-228 (1975).