

## TOPOLOGY, GEOMETRY AND PHYSICS: BACKGROUND FOR THE WITTEN CONJECTURE I.

GREGORY L. NABER

Communicated by Izu Vaisman

---

**Abstract.** The profound, beautiful and, at times, rather mysterious symbiosis between mathematics and physics has always been a source of wonder, but, in the past twenty years, the intensity of the mutual interaction between these two has become nothing short of startling. Our objective here is to provide an introduction, in terms as elementary as possible, to one small aspect of this relationship. Toward this end we shall tell a story. Although we make no attempt to relate it chronologically, the story can be said to begin with the efforts of Yang and Mills to construct a nonabelian generalization of classical electromagnetic theory, and to culminate in a remarkable conjecture of Witten concerning the Donaldson invariants of a smooth four-manifold.

---

### 1. Instantons and four-Manifolds

The central characters in our story are all “classical gauge theories” and we will eventually introduce them in some generality (Section 3), but we would like to begin by getting to know a few of them personally. For this we first recall the construction of the **quaternionic Hopf bundle**

$$Sp(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{H}\mathbb{P}^1. \tag{1.1}$$

Here  $Sp(1)$  is the Lie group of unit quaternions (those  $g \in \mathbb{H}$  satisfying  $|g| = 1$ ). As a manifold it is diffeomorphic to  $\mathbb{S}^3$ , but it is also isomorphic to the Lie group  $SU(2)$  of  $2 \times 2$  complex matrices  $U$  that are unitary ( $U^{-1} = \bar{U}^T$ ) and satisfy  $\det U = 1$ . Indeed, every such  $U$  can be written in the form  $U = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$ , where  $\alpha, \beta \in \mathbb{C}$  satisfy  $|\alpha|^2 + |\beta|^2 = 1$  and the map

$$\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \longrightarrow \alpha + \beta j = \alpha^1 + \alpha^2 i + (\beta^1 + \beta^2 i) j \\ = \alpha^1 + \alpha^2 i + \beta^1 j + \beta^3 k \tag{1.2}$$

is an isomorphism. We will allow ourselves the luxury of adopting whichever view of this Lie group is most convenient in any given situation. We will usually think of  $\mathbb{S}^7$  as  $\{p = (q^1, q^2) \in \mathbb{H}^2; |q^1|^2 + |q^2|^2 = 1\}$ . Then we can define a smooth right action  $\sigma : \mathbb{S}^7 \times Sp(1) \rightarrow \mathbb{S}^7$  of  $Sp(1)$  on  $\mathbb{S}^7$  by

$$\sigma(p, g) = p \cdot g = (q^1, q^2) \cdot g = (q^1 g, q^2 g).$$

This action is clearly free and the orbits  $p \cdot Sp(1)$  are submanifolds of  $\mathbb{S}^7$  diffeomorphic to  $\mathbb{S}^3$ . The orbit space  $\mathbb{S}^7/Sp(1)$  is, by definition, the **quaternionic projective line**  $\mathbb{H}\mathbb{P}^1$ . We will denote by  $\pi : \mathbb{S}^7 \rightarrow \mathbb{H}\mathbb{P}^1$  the projection map which carries  $p = (q^1, q^2)$  onto its orbit  $\pi(p) = \pi(q^1, q^2) = [q^1, q^2]$ . Obviously,  $\pi(p \cdot g) = \pi(p)$  for all  $p \in \mathbb{S}^7$  and  $g \in Sp(1)$ .  $\mathbb{H}\mathbb{P}^1$  is given the quotient topology determined by  $\pi$  and the differentiable structure determined by the atlas consisting of the two charts  $(U_k, \varphi_k)$ ,  $k = 1, 2$ , defined as follows:

$$\begin{aligned} U_k &= \left\{ x = [q^1, q^2] \in \mathbb{H}\mathbb{P}^1; q^k \neq 0 \right\}, \quad k = 1, 2 \\ \varphi_k : U_k &\longrightarrow \mathbb{H} \cong \mathbb{R}^4, \quad k = 1, 2 \\ \varphi_1(x) &= \varphi_1([q^1, q^2]) = q^2 (q^1)^{-1} \\ \varphi_2(x) &= \varphi_2([q^1, q^2]) = q^1 (q^2)^{-1}. \end{aligned} \tag{1.3}$$

Clearly,  $\varphi_1$  and  $\varphi_2$  map onto  $\mathbb{H}$ ,  $\varphi_1^{-1}(q) = [1, q]$ ,  $\varphi_2^{-1}(q) = [q, 1]$  and, on  $\varphi_1(U_1 \cap U_2) = \varphi_2(U_1 \cap U_2) = \mathbb{H} - \{0\}$ ,

$$\varphi_1 \circ \varphi_2^{-1}(q) = \varphi_2 \circ \varphi_1^{-1}(q) = q^{-1}. \tag{1.4}$$

It follows from this that we have, indeed, defined a differentiable structure. One can further verify that, relative to this structure,  $\pi : \mathbb{S}^7 \rightarrow \mathbb{H}\mathbb{P}^1$  is smooth and each of the maps

$$\begin{aligned} \Psi_k : \pi^{-1}(U_k) &\longrightarrow U_k \times Sp(1), \quad k = 1, 2 \\ \Psi_k(p) &= \Psi_k(q^1, q^2) = \left( \pi(p), q^k / |q^k| \right) \end{aligned} \tag{1.5}$$

is a diffeomorphism that is equivariant with respect to the given right action of  $Sp(1)$  on  $\mathbb{S}^7$  and the natural right action of  $Sp(1)$  on  $U_k \times Sp(1)$  (this means that, writing  $\Psi_k(p) = (\pi(p), \psi_k(p))$ , where  $\psi_k(p) = q^k / |q^k|$ , we have  $\Psi_k(p \cdot g) = (\pi(p), \psi_k(p)g) = (\pi(p), \psi_k(p)) \cdot g$ ). To summarize, what we have just shown is that  $\mathbb{S}^7$  is a smooth principal  $Sp(1)$  (i.e.,  $SU(2)$ )-bundle over  $\mathbb{H}\mathbb{P}^1$  with

local trivializations given by (1.5). We note in passing that one can replace the quaternions  $\mathbb{H}$  by the complex numbers  $\mathbb{C}$  everywhere above and thereby construct the **complex Hopf bundle**

$$U(1) \hookrightarrow \mathbb{S}^3 \longrightarrow \mathbb{C}\mathbb{P}^1. \tag{1.6}$$

There are, of course, also the obvious higher dimensional generalizations  $Sp(1) \hookrightarrow \mathbb{S}^{4n-1} \rightarrow \mathbb{H}\mathbb{P}^{n-1}$  and  $U(1) \hookrightarrow \mathbb{S}^{2n-1} \rightarrow \mathbb{C}\mathbb{P}^{n-1}$  for  $n > 2$ .

**Remark.** Let  $\varphi_S$  and  $\varphi_N$  be the stereographic projection maps from the north and south poles of  $\mathbb{S}^4$ , respectively, and let  $\varphi_1$  and  $\varphi_2$  be as in (1.3). If  $\bar{\varphi}_1$  denotes the map  $\bar{\varphi}_1(x) = \overline{\varphi_1(x)}$ , then both  $\varphi_S^{-1} \circ \varphi_2$  and  $\varphi_N^{-1} \circ \bar{\varphi}_1$  are diffeomorphisms of  $\mathbb{H}\mathbb{P}^1$  minus a point to  $\mathbb{S}^4$  minus a point. On the intersection of their domains they agree and so determine a global diffeomorphism of  $\mathbb{H}\mathbb{P}^1$  onto  $\mathbb{S}^4$ . Composing with  $\pi : \mathbb{S}^7 \rightarrow \mathbb{H}\mathbb{P}^1$  gives a principal bundle

$$Sp(1) \hookrightarrow \mathbb{S}^7 \longrightarrow \mathbb{S}^4$$

which is also often referred to as the quaternionic Hopf bundle. Some caution is advised, however, since reversing the roles of  $\varphi_1$  and  $\varphi_2$  above gives another identification of  $\mathbb{H}\mathbb{P}^1$  with  $\mathbb{S}^4$ , but the corresponding  $Sp(1)$ -bundle over  $\mathbb{S}^4$  is not equivalent to the one we just described. This is most readily shown by computing their Chern numbers which turn out to be 1 in the first case and  $-1$  in the second (we will briefly review these calculations shortly).

Now we focus our attention on a fixed point  $p = (q^1, q^2) \in \mathbb{S}^7 \subseteq \mathbb{H}^2$ . The orbit of our  $Sp(1)$ -action containing  $p$  (i.e., the fiber of  $\pi : \mathbb{S}^7 \rightarrow \mathbb{H}\mathbb{P}^1$  above  $\pi(p)$ ) is a copy of  $\mathbb{S}^3$ . The subset of the tangent space  $T_p(\mathbb{S}^7)$  to  $\mathbb{S}^7$  at  $p$  consisting of vectors tangent to this fiber is called the **vertical space** at  $p$  and denoted  $\text{Vert}_p(\mathbb{S}^7)$ . It is a three-dimensional linear subspace of  $T_p(\mathbb{S}^7)$  which, in turn, can be identified with a linear subspace of  $T_p(\mathbb{H}^2) \cong T_p(\mathbb{R}^8) \cong \mathbb{R}^8$ .

Thus, relative to the usual Euclidean inner product on  $\mathbb{R}^8$ ,  $\text{Vert}_p(\mathbb{S}^7)$  has an orthogonal complement and we shall call the intersection of this orthogonal complement with  $T_p(\mathbb{S}^7)$  the **horizontal space** at  $p$  and denote it  $\text{Hor}_p(\mathbb{S}^7)$ . Thus, at each  $p \in \mathbb{S}^7$  we have a natural decomposition

$$T_p(\mathbb{S}^7) \cong \text{Vert}_p(\mathbb{S}^7) \oplus \text{Hor}_p(\mathbb{S}^7).$$

If one fixes a  $g \in Sp(1)$  and explicitly computes the derivative of the diffeomorphism  $\sigma_g : \mathbb{S}^7 \rightarrow \mathbb{S}^7$ ,  $\sigma_g(p) = \sigma(p, g) = p \cdot g$ , it is easy to see that

$$(\sigma_g)_{*p}(\text{Hor}_p(\mathbb{S}^7)) = \text{Hor}_{p \cdot g}(\mathbb{S}^7)$$

and so the distribution  $\{\text{Hor}_p(\mathbb{S}^7); p \in \mathbb{S}^7\}$  of four-dimensional subspaces satisfies the conditions required of a connection on the Hopf bundle. This is called the **natural connection** on  $Sp(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{H}\mathbb{P}^1$ . Now, any connection arises from a connection one-form, i.e., a Lie algebra-valued one-form on the principal bundle space whose kernel at each point is the horizontal space at that point. Identifying the Lie algebra of  $Sp(1)$  with the pure imaginary quaternions  $\text{Im}\mathbb{H}$  and defining an  $\text{Im}\mathbb{H}$ -valued one-form on  $\mathbb{H}^2$  by

$$\tilde{\omega} = \text{Im}(\bar{q}^1 dq^1 + \bar{q}^2 dq^2)$$

it is not difficult to see that the connection one-form  $\omega$  for the natural connection on the Hopf bundle is the restriction of  $\tilde{\omega}$  to  $\mathbb{S}^7$ , i.e.,

$$\omega = \iota^* \tilde{\omega} \tag{1.7}$$

where  $\iota : \mathbb{S}^7 \hookrightarrow \mathbb{H}^2$  is the inclusion map (the proof is on page 295 of [34]).

In the physics literature it is more common to describe connections (there called **gauge potentials**) locally on the base manifold by pulling back the connection one-form by sections corresponding to some trivializing cover of the bundle. For the trivializations  $(U_k, \Psi_k)$ ,  $k = 1, 2$ , of the Hopf bundle given by (1.5) each  $U_k$  covers all but one point of  $\mathbb{H}\mathbb{P}^1$  and it follows that the connection  $\omega$  is uniquely determined by either one of the corresponding pullbacks. For example, it is easy to verify that the inverse of  $\Psi_1 : \pi^{-1}(U_1) \rightarrow U_1 \times Sp(1)$  is given by  $\Psi_1^{-1}([q^1, q^2], g) = (|q^1|g, q^2(q^1/|q^1|)^{-1}g) \in \mathbb{S}^7 \subseteq \mathbb{H}^2$  and so the associated section  $s_1 : U_1 \rightarrow \pi^{-1}(U_1)$  is

$$s_1(x) = s_1([q^1, q^2]) = \Psi_1^{-1}([q^1, q^2], 1) = (|q^1|, q^2(q^1/|q^1|)^{-1}).$$

Since  $U_1$  is also the domain of the standard chart  $(U_1, \varphi_1)$  on  $\mathbb{H}\mathbb{P}^1$  we can write the pullback  $s_1^* \omega$  in terms of these coordinates on  $\mathbb{H}\mathbb{P}^1$ . More precisely, we pull  $s_1^* \omega$  back to  $\mathbb{H} \cong \mathbb{R}^4$  by  $\varphi_1^{-1}$ . These calculations are carried out in detail on pages 256-258 of [34] and yield

$$\mathcal{A}_1 = (s_1 \circ \varphi_1^{-1})^* \omega = \text{Im}\left(\frac{\bar{q}}{1 + |q|^2} dq\right) \tag{1.8}$$

which we now simply regard as an  $\text{Im}\mathbb{H}$ -valued one-form on  $\mathbb{H}$ . Oddly enough, this one-form on  $\mathbb{R}^4$  first appeared in the physics literature [4] where it was initially referred to as a **pseudoparticle**. We will have more to say about this shortly.

Thus far we know only one connection on the Hopf bundle (the “natural” one) and we would now like to produce some more. Recall that an **automorphism** of

$Sp(1) \hookrightarrow \mathbb{S}^7 \xrightarrow{\pi} \mathbb{H}\mathbb{P}^1$  is a diffeomorphism  $f : \mathbb{S}^7 \rightarrow \mathbb{S}^7$  of  $\mathbb{S}^7$  onto itself that respects the group action  $(f(p \cdot g) = f(p) \cdot g)$  and that each such automorphism induces a diffeomorphism  $f_{\mathbb{H}\mathbb{P}^1} : \mathbb{H}\mathbb{P}^1 \rightarrow \mathbb{H}\mathbb{P}^1$  of  $\mathbb{H}\mathbb{P}^1$  onto itself by  $\pi \circ f = f_{\mathbb{H}\mathbb{P}^1} \circ \pi$ . If  $f_{\mathbb{H}\mathbb{P}^1}$  happens to be the identity on  $\mathbb{H}\mathbb{P}^1$ , then  $f$  is called a **(global) gauge transformation**. Now, if  $f$  is any automorphism and  $\omega$  is any connection one-form, then the pullback  $f^*\omega$  is also a connection one-form. If  $f$  is a gauge transformation, then the connections  $\omega$  and  $f^*\omega$  are said to be **gauge equivalent**.

**Remark.** *The motivation here is as follows. In physics, a (local) section  $s : U \rightarrow \pi^{-1}(U)$  of a principal bundle  $G \hookrightarrow P \xrightarrow{\pi} X$  is called a choice of **gauge** on  $U$  and is regarded as a selection, at each  $x \in U$ , of a frame (coordinate system) in some internal space. The **gauge principle** asserts that the laws of physics should be invariant under an arbitrary change of gauge and, more particularly, that quantities with the same set of gauge representations are to be regarded as physically equivalent. For example, if  $\omega$  is a connection and  $f$  is a gauge transformation, then, for any section  $s$ ,  $f \circ s$  is also a section and  $s^*(f^*\omega) = (f \circ s)^*\omega$ . Thus,  $\omega$  and  $f^*\omega$  have the same set of gauge potentials and so are taken to be “equivalent”. For future reference we note that a local gauge transformation on  $U$  can be identified with a map  $g : U \rightarrow G$  which gives rise to a new section/gauge  $s^g : U \rightarrow \pi^{-1}(U)$  defined by  $s^g(x) = s(x) \cdot g(x)$  and that if  $\omega$  is any connection,  $\mathcal{A} = s^*\omega$  and  $\mathcal{A}^g = (s^g)^*\omega$ , then*

$$\mathcal{A}^g = g^{-1}\mathcal{A}g + g^{-1}dg.$$

*Although not entirely obvious, one can show (pages 297-303 of [34]) that, by judiciously choosing automorphisms of the Hopf bundle by which to pull back the natural connection, one can produce a new connection  $\omega_{\lambda,n}$  for each  $(\lambda, n) \in (0, \infty) \times \mathbb{H}$  that is uniquely determined by the gauge potential*

$$\mathcal{A}_{\lambda,n} = \text{Im} \left( \frac{\bar{q} - \bar{n}}{\lambda^2 + |q - n|^2} dq \right) \quad (1.9)$$

*on  $\mathbb{H}$  (thus,  $\mathcal{A}_1 = \mathcal{A}_{1,0}$ ). For reasons that we will discuss shortly,  $\mathcal{A}_{\lambda,n}$  is called the **BPST-instanton** with **center**  $n$  and **spread**  $\lambda$ . Although all of these differ from the natural connection by an automorphism we will see that distinct pairs  $(\lambda, n)$  give rise to connections that are not gauge equivalent.*

Any connection  $\omega$  has a curvature  $\Omega$  that can be calculated from the Cartan Structure Equation  $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$  and is uniquely determined by a family of pullbacks  $\mathcal{F} = s^*\Omega$ , called **gauge field strengths**, by sections corresponding to some trivializing cover. For the connection  $\omega_{\lambda,n}$  on the Hopf bundle the curvature  $\Omega_{\lambda,n}$  is uniquely determined by the single gauge field strength

$\mathcal{F}_{\lambda,n} = d\mathcal{A}_{\lambda,n} + \frac{1}{2}[\mathcal{A}_{\lambda,n}, \mathcal{A}_{\lambda,n}]$ . A rather tedious, but routine calculation (pages 284-289 of [34]) gives

$$\begin{aligned} \mathcal{F}_{\lambda,n} &= \frac{\lambda^2}{(\lambda^2 + |q-n|^2)^2} d\bar{q} \wedge dq \\ &= \frac{2\lambda^2}{(\lambda^2 + |q-n|^2)^2} \left( (dx^1 \wedge dx^2 - dx^3 \wedge dx^4)\mathbf{i} \right. \\ &\quad \left. + (dx^1 \wedge dx^3 + dx^2 \wedge dx^4)\mathbf{j} + (dx^1 \wedge dx^4 - dx^2 \wedge dx^3)\mathbf{k} \right) \end{aligned} \quad (1.10)$$

where  $x^1, x^2, x^3$  and  $x^4$  are standard coordinates on  $\mathbb{R}^4$ .

The  $\text{Im}\mathbb{H}$ -valued two-forms  $\mathcal{F}_{\lambda,n}$  on  $\mathbb{H} (\cong \mathbb{R}^4)$  have a number of crucial properties. If “ $*$ ” denotes the Hodge star operator on  $\mathbb{R}^4$  arising from the usual orientation and inner product and if we extend this operator to  $\text{Im}\mathbb{H}$ -valued forms componentwise, then each  $\mathcal{F}_{\lambda,n}$  is **anti-self-dual (ASD)** in the sense that

$$*\mathcal{F}_{\lambda,n} = -\mathcal{F}_{\lambda,n} \quad (1.11)$$

(page 333 of [34]). The Hodge star also gives a (pointwise) inner product on the spaces  $\Omega^p(\mathbb{R}^4)$ ,  $0 \leq p \leq 4$ , of real-valued  $p$ -forms on  $\mathbb{R}^4$  (if  $\mu$  and  $\nu$  are in  $\Omega^p(\mathbb{R}^4)$ , then  $*\nu$  is in  $\Omega^{4-p}(\mathbb{R}^4)$  so  $\mu \wedge *\nu$  is in  $\Omega^4(\mathbb{R}^4)$  and so is a multiple of the metric volume form **vol** on  $\mathbb{R}^4$  and one defines  $\langle \mu, \nu \rangle$  by  $\mu \wedge *\nu = \langle \mu, \nu \rangle \mathbf{vol}$ ). Combined with the Killing form on the Lie algebra this will give a pointwise inner product on any space of Lie algebra-valued forms on  $\mathbb{R}^4$ .

**Remark.** *The conventions we adopt are as follows: The Lie algebra  $\text{Im}\mathbb{H}$  is isomorphic to the Lie algebra  $\mathfrak{su}(2)$  of  $2 \times 2$  complex matrices  $A$  that are skew-Hermitian ( $\bar{A}^T = -A$ ) and tracefree ( $\text{tr}A = 0$ ). We take as a basis for  $\mathfrak{su}(2)$  the matrices  $T_j = -\frac{1}{2}\mathbf{i}\sigma_j$ , where  $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\sigma_2 = \begin{pmatrix} 0 & -\mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}$  and  $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$  are the Pauli spin matrices. Thus, every element of  $\mathfrak{su}(2)$  can be written in the form*

$$A = A^1 T_1 + A^2 T_2 + A^3 T_3 = -\frac{1}{2} \begin{pmatrix} A^3 \mathbf{i} & A^2 + A^1 \mathbf{i} \\ -A^2 + A^1 \mathbf{i} & -A^3 \mathbf{i} \end{pmatrix}. \quad (1.12)$$

*As the inner product associated with the Killing form of  $\mathfrak{su}(2)$  we take  $\langle A, B \rangle = -2 \text{tr}(AB)$  so that  $\{T_1, T_2, T_3\}$  is an orthonormal basis. The structure constants for this basis are given by  $[T_i, T_j] = \varepsilon_{ijk} T_k$  ( $\varepsilon_{ijk}$  is the Levi-Civita symbol with*

$\varepsilon_{123} = 1$ ). The structure constants for the basis  $\{\frac{1}{2}\mathbf{i}, \frac{1}{2}\mathbf{j}, \frac{1}{2}\mathbf{k}\}$  for  $\text{Im}\mathbb{H}$  are the same ( $[x, y] = xy - yx = 2\text{Im}(xy)$  on  $\text{Im}\mathbb{H}$ ) so an isomorphism of  $\text{Im}\mathbb{H}$  onto  $\mathfrak{su}(2)$  is  $a^1\mathbf{i} + a^2\mathbf{j} + a^3\mathbf{k} \rightarrow (2a^1)T_1 + (2a^2)T_2 + (2a^3)T_3$ . In particular, our Lie algebra inner product is four times the usual  $\mathbb{R}^3$  inner product  $\text{Im}\mathbb{H}$ .

Now, if  $\mathcal{F} = \mathcal{F}^1\mathbf{i} + \mathcal{F}^2\mathbf{j} + \mathcal{F}^3\mathbf{k}$  is an  $\text{Im}\mathbb{H}$ -valued form on  $\mathbb{R}^4$ , each component has a squared norm  $\|\mathcal{F}^i\|^2 = \langle \mathcal{F}^i, \mathcal{F}^i \rangle$  given by the Hodge dual as above. Our Lie algebra squared norm  $\|\mathcal{F}\|^2$  for  $\mathcal{F}$  is then taken to be four times the sum of these

$$\|\mathcal{F}\|^2 = 4 \sum_{i=1}^3 \|\mathcal{F}^i\|^2. \quad (1.13)$$

Writing  $\mathcal{F}$  as an  $\mathfrak{su}(2)$  matrix of complex-valued forms in the manner indicated above,  $*\mathcal{F}$  is computed entrywise. Defining the wedge product  $\mathcal{F} \wedge *\mathcal{F}$  to be the matrix product with entries multiplied by the ordinary wedge, a simple calculation shows that

$$-\text{tr}(\mathcal{F} \wedge *\mathcal{F}) = \frac{1}{2} \|\mathcal{F}\|^2 \text{vol}. \quad (1.14)$$

To compute  $\|\mathcal{F}_{\lambda,n}\|^2$  for the BPST gauge field strength given by (1.10) one observes that, for example,

$$\begin{aligned} & (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \wedge * (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \\ &= - (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \wedge (dx^1 \wedge dx^2 - dx^3 \wedge dx^4) \\ &= 2dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 = 2 \text{vol} \end{aligned}$$

and similarly for the rest so

$$\|\mathcal{F}_{\lambda,n}\|^2 = 4 \left( 3 \left( \frac{8\lambda^4}{(\lambda^2 + |q - n|^2)^4} \right) \right) = \frac{96\lambda^4}{(\lambda^2 + |q - n|^2)^4}. \quad (1.15)$$

Notice that  $\|\mathcal{F}_{\lambda,n}\|^2$  has a maximum value of  $96/\lambda^4$  at  $q = n$  and that, for a fixed  $n$ , its variation with  $\lambda$  (illustrated in the Figure 1) is such that the “total field strength”

$$\frac{1}{2} \int_{\mathbb{R}^4} \|\mathcal{F}_{\lambda,n}\|^2 \text{vol} = \int_{\mathbb{R}^4} \frac{48\lambda^4}{(\lambda^2 + |q - n|^2)^4} \text{vol} = 8\pi^2 \quad (1.16)$$

remains constant at  $8\pi^2$ .

Thus, the gauge potential  $\mathcal{A}_{\lambda,n}$  on  $\mathbb{R}^4$  has field strength that is “centered” at  $n$  in  $\mathbb{R}^4$  with a “spread” that is determined by  $\lambda$  (hence the terminology introduced

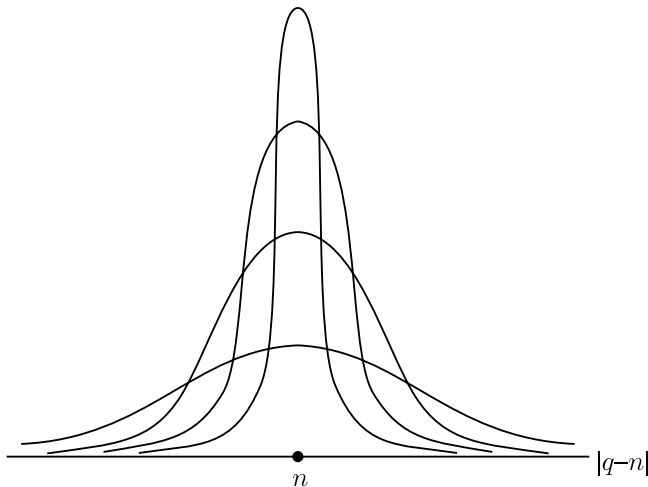


Figure 1. BPST Field Strengths

earlier), and a total field strength that is independent of  $\lambda$  and  $n$  (the reason for this is, as we will see shortly, deeper than it might seem).

Let us now temporarily suppress from our minds where the potentials  $\mathcal{A}_{\lambda,n}$  came from (i.e., the Hopf bundle) and regard them simply as Lie algebra-valued one-forms on  $\mathbb{R}^4$ . Any such Lie algebra-valued one-form  $\mathcal{A}$  on  $\mathbb{R}^4$  can be thought of as a gauge potential for a connection on the trivial  $Sp(1)$ - (or  $SU(2)$ -) bundle over  $\mathbb{R}^4$  and so has a gauge field strength  $\mathcal{F} = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}]$  on  $\mathbb{R}^4$ . We define the **Yang-Mills action**  $\mathcal{YM}(\mathcal{A})$  of  $\mathcal{A}$  by

$$\mathcal{YM}(\mathcal{A}) = \int_{\mathbb{R}^4} -\text{tr}(\mathcal{F} \wedge *\mathcal{F}) = \frac{1}{2} \int_{\mathbb{R}^4} \|\mathcal{F}\|^2 \text{vol}. \quad (1.17)$$

This integral might well be infinite, of course, but if it is not we will say that  $\mathcal{A}$  has **finite action** and think of  $\mathcal{YM}(\mathcal{A})$  as the total field strength of the gauge potential  $\mathcal{A}$ .

In an attempt to describe the isotopic spin of a nucleon, Yang and Mills [49] devised a non-abelian generalization of classical electromagnetic theory in which the electromagnetic potential was replaced by an  $\mathfrak{su}(2)$ -valued one-form  $\mathcal{A}$  on  $\mathbb{R}^4$  (actually, on Minkowski spacetime, but we will restrict our attention to the Euclidean version). The field strength for their potential was our  $\mathcal{F}$  and the “action” (i.e., Lagrangian) of the theory was  $\mathcal{YM}(\mathcal{A})$ . The field equations were the Euler-Lagrange equations for  $\mathcal{YM}(\mathcal{A})$  under variations of  $\mathcal{A}$  and it is not difficult to



show that these are

$$d^{\mathcal{A}} * \mathcal{F} = 0 \tag{1.18}$$

where  $d^{\mathcal{A}} * \mathcal{F} = d * \mathcal{F} + [\mathcal{A}, * \mathcal{F}]$  is the covariant exterior derivative of  $* \mathcal{F}$  associated with  $\mathcal{A}$ . Equations (1.17) are the **Yang-Mills equations** on  $\mathbb{R}^4$ . Quite independently of  $\mathcal{YM}$ , any field strength  $\mathcal{F}$  satisfies a purely geometrical condition called the **Bianchi identity**

$$d^{\mathcal{A}} \mathcal{F} = 0 \tag{1.19}$$

(page 268 of [34]). Now notice that if the field strength  $\mathcal{F}$  of  $\mathcal{A}$  happens to be ASD ( $* \mathcal{F} = -\mathcal{F}$ ), then (1.19) implies that (1.18) is automatically satisfied. Thus, a gauge potential  $\mathcal{A}$  on  $\mathbb{R}^4$  with ASD field strength  $\mathcal{F}$  is a solution to the Yang-Mills equations (it is shown on page 325 of [34] that these actually give absolute minima for the Yang-Mills action). This is the context in which the BPST instantons  $\mathcal{A}_{\lambda,n}$  were first discovered. Belavin, Polyakov, Schwarz and Tyupkin [4] sought finite action solutions to the Yang-Mills equations (1.18) on  $\mathbb{R}^4$  and found them via the simpler **anti-self-dual equations**

$$* \mathcal{F} = -\mathcal{F}. \tag{1.20}$$

A finite action gauge potential  $\mathcal{A}$  on  $\mathbb{R}^4$  with ASD field strength is called an **instanton** on  $\mathbb{R}^4$ .

We have described a family of instantons  $\mathcal{A}_{\lambda,n}$  parametrized by  $(\lambda, n) \in (0, \infty) \times \mathbb{R}^4$ . Now,  $\|\mathcal{F}\|^2$  is invariant under gauge transformation. This is essentially because it is defined in terms of the trace (1.14), which is invariant under conjugation, and a local gauge transformation  $g : U \rightarrow G$  (see the Remark preceding (1.9)) conjugates field strengths, i.e.,  $\mathcal{F}^g = g^{-1} \mathcal{F} g$ . Thus, we conclude from (1.15) that  $\mathcal{A}_{\lambda,n}$  and  $\mathcal{A}_{\lambda',n'}$  can be gauge equivalent if and only if  $(\lambda', n') = (\lambda, n)$ , i.e., distinct BPST instantons are gauge inequivalent. Nevertheless, they all have the same total field strength  $\mathcal{YM}(\mathcal{A}_{\lambda,n})$  and we must now investigate this “coincidence” more closely.

Lately we have been thinking of the  $\mathcal{A}_{\lambda,n}$  simply as Lie algebra-valued one-forms on  $\mathbb{R}^4$  and forgetting where they came from. They are, of course, much more. They are pullbacks to  $\mathbb{R}^4$  of the connections  $\omega_{\lambda,n}$  on  $Sp(1) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{H}P^1$ . Let us now identify  $\mathbb{H}P^1$  with  $\mathbb{S}^4$  in the manner described at the beginning of the Remark after (1.6). Each  $\omega_{\lambda,n}$  is then regarded as a connection one-form on

$$Sp(1) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{S}^4. \tag{1.21}$$

Their pullbacks to  $\mathbb{S}^4$  by the induced sections of (1.21), when written in coordinates on  $\mathbb{R}^4$  obtained by stereographic projection from the north pole of  $\mathbb{S}^4$ , are

the gauge potentials  $\mathcal{A}_{\lambda,n}$ . Summarizing, we find that the  $\mathcal{A}_{\lambda,n}$  (connection one-forms on the trivial  $Sp(1)$ -bundle over  $\mathbb{R}^4$ ) “come from” connection one-forms  $\omega_{\lambda,n}$  on the nontrivial Hopf bundle (1.21) over  $\mathbb{S}^4$ . Turning matters about, one might say that the connections on the trivial bundle over  $\mathbb{R}^4$  “extend to  $\mathbb{S}^4$ ” in the sense that  $\mathbb{S}^4 = \mathbb{R}^4 \cup \{\infty\}$  is the one-point compactification of  $\mathbb{R}^4$  and, due to their asymptotic behavior as  $\|x\| \rightarrow \infty$  in  $\mathbb{R}^4$ , the connections extend to the point at infinity. Note, however, that the extension process involves not only the connection, but the bundle on which it is defined as well. Now, a remarkable theorem of Karen Uhlenbeck [43] asserts that this interpretation is not as fanciful as it might sound. Indeed, a very special case of this result states that if  $\mathcal{A}$  is any finite action  $\text{Im}\mathbb{H}$ -valued gauge potential on  $\mathbb{R}^4$  with ASD field strength  $\mathcal{F}$ , then there exists a unique (up to equivalence)  $Sp(1)$ -bundle  $Sp(1) \hookrightarrow P \rightarrow \mathbb{S}^4$  over  $\mathbb{S}^4$  and a connection  $\omega$  on  $P$  whose pullback by some section  $s$  of  $P$  is  $\mathcal{A}$  when written in stereographic coordinates. Furthermore, the bundle to which  $\mathcal{A}$  “extends” is uniquely determined by the Yang-Mills action  $\mathcal{YM}(\mathcal{A})$  of  $\mathcal{A}$  as we shall now explain.

An  $Sp(1)$ -bundle  $Sp(1) \hookrightarrow P \xrightarrow{\pi} X$  over a compact, oriented, smooth four-manifold  $X$  is uniquely determined by a certain characteristic cohomology class, called the **second Chern class**  $c_2(P)$  and constructed as follows. Choose a connection  $\omega$  on  $P$ . The curvature  $\Omega$  of  $\omega$  is a Lie algebra-valued two-form on  $P$ . One can show (Section 6.3 of [35]) that  $\frac{1}{8\pi^2} \text{tr}(\Omega \wedge \Omega)$  is a real-valued four-form on  $P$  which descends to (i.e., is the pullback by  $\pi$  of) a closed four-form on  $X$  whose cohomology class  $c_2(P) \in H^4(X, \mathbb{R})$  does not depend on the choice of  $\omega$ . Two  $Sp(1)$ -bundles over  $X$  are known to be equivalent if and only if they have the same second Chern class and, indeed, if and only if they have the same **second Chern number**, defined by

$$c_2(P)[X] = \int_X c_2(P) \quad (1.22)$$

(which is always an integer). Now take  $X$  to be  $\mathbb{S}^4$ . Stereographic projection from the north pole of  $\mathbb{S}^4$  is an orientation preserving diffeomorphism from  $\mathbb{S}^4$  minus a point onto  $\mathbb{R}^4$  so  $c_2(P)[\mathbb{S}^4]$  can be computed by integrating pullbacks over  $\mathbb{R}^4$ , i.e.,

$$c_2(P)[\mathbb{S}^4] = \frac{1}{8\pi^2} \int_{\mathbb{R}^4} \text{tr}(\mathcal{F} \wedge \mathcal{F}) \quad (1.23)$$

where  $\mathcal{F}$  is the corresponding field strength.

Now let us consider an instanton  $\mathcal{A}$  on  $\mathbb{R}^4$  with field strength  $\mathcal{F}$ . Since  $\mathcal{F}$  is ASD,  $*\mathcal{F} = -\mathcal{F}$  and so  $-\operatorname{tr}(\mathcal{F} \wedge *\mathcal{F}) = \operatorname{tr}(\mathcal{F} \wedge \mathcal{F})$ . Uhlenbeck's Theorem guarantees that  $\mathcal{A}$  extends to a connection on some principal  $Sp(1)$ -bundle  $Sp(1) \hookrightarrow P \rightarrow \mathbb{S}^4$  over  $\mathbb{S}^4$  and a comparison of (1.17) and (1.23) shows that the second Chern number of this bundle is given by

$$c_2(P)[\mathbb{S}^4] = \frac{1}{8\pi^2} \mathcal{YM}(\mathcal{A}). \tag{1.24}$$

Thus, the Yang-Mills action of an instanton  $\mathcal{A}$  on  $\mathbb{R}^4$  is directly encoded in the topology of the bundle over  $\mathbb{S}^4$  to which  $\mathcal{A}$  extends. Notice, however, that the value of  $\mathcal{YM}(\mathcal{A})$  is entirely determined by the asymptotic behavior of the field strength  $\mathcal{F}$  on  $\mathbb{R}^4$  so it is this physical characteristic of the gauge field that is represented by the Chern number. Physicists call  $-c_2(P)[\mathbb{S}^4]$  the **instanton number**, or **topological charge**, of  $\mathcal{A}$ . The “reason” that all of the BPST instantons  $\mathcal{A}_{\lambda,n}$  have the same Yang-Mills action is now clear: they all extend to (i.e., come from) the same  $Sp(1)$ -bundle over  $\mathbb{S}^4$ , i.e., the Hopf bundle (1.21), which (1.16) now shows to have Chern number one (as promised in the Remark following (1.6)). Notice also that the topological charge of an instanton, being an integer, cannot be altered by a continuous variation of the field and so is “conserved”, but for purely topological reasons unlike the more common Noether conserved quantities. Such **topological conservation laws** play a crucial role in understanding modern gauge field theories.

**Remark.** For the complex Hopf bundle  $U(1) \hookrightarrow \mathbb{S}^3 \rightarrow \mathbb{C}P^1$  one defines the natural connection geometrically exactly as in the quaternionic case. The corresponding connection one-form  $\omega$  is the restriction to  $\mathbb{S}^3$  of the  $\operatorname{Im}\mathbb{C}$ -valued one-form  $i\operatorname{Im}(\bar{z}^1 dz^1 + \bar{z}^2 dz^2)$  on  $\mathbb{C}^2$ . Choosing a section and coordinates analogous to those in the quaternionic case gives the gauge potential  $i\operatorname{Im}\left(\frac{\bar{z}}{1+|z|^2} dz\right)$  on  $\mathbb{C} \cong \mathbb{R}^2$ . Identifying  $\mathbb{C}P^1$  with  $\mathbb{S}^2$  in the two ways indicated for  $\mathbb{H}P^1$  and  $\mathbb{S}^4$  gives two inequivalent  $U(1)$ -bundles over  $\mathbb{S}^2$  (those with “first Chern number”  $\pm 1$ ). On each of these one obtains an induced connection, uniquely determined by a gauge potential which, when written in spherical (rather than stereographic) coordinates, takes the form  $-\frac{in}{2}(1 - \cos \phi)d\theta$ , where  $n = \pm 1$ . More generally, one has, for each  $n \in \mathbb{Z}$ , a  $U(1)$ -bundle  $U(1) \hookrightarrow P_n \rightarrow \mathbb{S}^2$  over  $\mathbb{S}^2$  with first Chern number  $n$  and on it a connection uniquely determined by the gauge potential

$$\mathcal{A}_n = -\frac{in}{2}(1 - \cos \phi) d\theta.$$

*This potential (and the corresponding connection) represent the field of a **Dirac monopole of magnetic charge**  $-n$ . Just as for the instanton number, magnetic charge is “topological” in that it is encoded in the topology of the bundle on which the connection lives and is conserved for topological reasons. We will have more to say about Dirac monopoles in Section 7.*

There is another perspective on the topological nature of instantons which we now briefly describe. Observe that  $\mathcal{YM}(\mathcal{A}) < \infty$  implies that  $\|\mathcal{F}\|^2$  must approach zero sufficiently fast as  $\|x\| \rightarrow \infty$  in  $\mathbb{R}^4$ . This, together with  $\mathcal{F} = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}]$  would seem to require a similar decay for the components of  $\mathcal{A}$  and their first derivatives. However, due to the gauge freedom available in the choice of  $\mathcal{A}$  ( $\|\mathcal{F}\|^2$  is gauge invariant), this is not the case. All that is necessary is that there exist some local gauge transformation  $g : U \rightarrow Sp(1)$ , defined for sufficiently large  $\|x\|$ , such that the potential  $\mathcal{A}^g$  in this new gauge has components that decay sufficiently fast ( $g$  need only be defined for large  $\|x\|$  because the integral over any compact set in  $\mathbb{R}^4$  is necessarily finite). If such a  $g$  exists and  $\mathbb{S}_R^3$  is a three-sphere in  $\mathbb{R}^4$  of sufficiently large radius  $R$  that it is contained in the domain of  $g$ , then

$$g|_{\mathbb{S}_R^3} : \mathbb{S}_R^3 \longrightarrow Sp(1)$$

can be regarded as a map from the three-sphere to itself and so determines an element of the homotopy group  $\pi_3(\mathbb{S}^3)$ . But  $\pi_3(\mathbb{S}^3) \cong \mathbb{Z}$  and an isomorphism is provided by the Brouwer degree  $\deg$ . Thus, the asymptotic behavior of  $\mathcal{F}$  determines  $g$ , which determines  $[g|_{\mathbb{S}_R^3}] \in \pi_3(\mathbb{S}^3)$  and this gives an integer  $\deg(g|_{\mathbb{S}_R^3})$  (the restrictions of  $g$  to two such spheres are clearly homotopic and so have the same degree). Thus, the various possible asymptotic behaviors for finite action field strengths on  $\mathbb{R}^4$  fall into “homotopy classes”, each labeled by an integer. If  $\mathcal{F}$  is ASD so that  $\mathcal{A}$  extends to a principal  $Sp(1)$ -bundle over  $\mathbb{S}^4$ , then these integers also characterize the bundles.

**Remark.** *Briefly, the reason for this is as follows.  $\mathbb{S}^4 = \mathbb{R}^4 \cup \{\infty\}$  consists of two copies of the closed four-dimensional disc (upper and lower hemispheres) glued together along the equator which is a copy of  $\mathbb{S}^3$  and which we can take to be  $\mathbb{S}_R^3$ . The restriction of any bundle over  $\mathbb{S}^4$  to either of these discs is trivial because the disc is contractible. This provides a trivializing cover of  $\mathbb{S}^4$  consisting of just two trivializations and hence essentially one transition function  $g$ . This one transition function determines the bundle up to equivalence and its restriction to the equator is a map from  $\mathbb{S}^3$  to  $Sp(1) \cong \mathbb{S}^3$ . Furthermore, any map from  $\mathbb{S}^3$  to  $\mathbb{S}^3$  can be regarded as such a restriction and homotopic maps correspond to equivalent bundles. In particular, this is true of the restricted gauge transformation  $g|_{\mathbb{S}_R^3}$  so the integer  $\deg(g|_{\mathbb{S}_R^3})$  uniquely determines an  $Sp(1)$ -bundle over  $\mathbb{S}^4$ .*

Now let us consider somewhat more generally the Hopf bundle  $Sp(1) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{S}^4$  over  $\mathbb{S}^4$  with Chern number one. Any connection  $\omega$  on this bundle is uniquely determined by the gauge potential  $\mathcal{A}$  on  $\mathbb{R}^4$  obtained by pulling back by the natural section on  $\mathbb{S}^4$  minus the north pole and then again by the inverse of stereographic projection from the north pole. This stereographic projection is an orientation preserving conformal diffeomorphism and so preserves the Hodge dual. We will say that the connection  $\omega$  is **anti-self-dual (ASD)** if the gauge potential  $\mathcal{A}$  is ASD (we will see shortly how to extend this notion to bundles over more general four-manifolds). The set of ASD connections is invariant under global gauge transformations of the bundle so we may consider the set  $\mathcal{M}$  of gauge equivalence classes of ASD connections on  $Sp(1) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{S}^4$ . This then is the same as the set of gauge equivalence classes of ASD potentials  $\mathcal{A}$  on  $\mathbb{R}^4$  with  $\mathcal{YM}(\mathcal{A}) = 8\pi^2$  (regarded as connection forms on the trivial bundle).

Each BPST instanton  $\mathcal{A}_{\lambda,n}$  determines a point  $[\mathcal{A}_{\lambda,n}]$  in  $\mathcal{M}$  and we have already observed that distinct pairs  $(\lambda, n)$  give distinct points in  $\mathcal{M}$ . A remarkable, and very deep result of Atiyah, Hitchin and Singer [2] asserts that, in fact, *every* element of  $\mathcal{M}$  is represented by some  $\mathcal{A}_{\lambda,n}$  and so the map

$$(\lambda, n) \in (0, \infty) \times \mathbb{R}^4 \longrightarrow [\mathcal{A}_{\lambda,n}] \in \mathcal{M}$$

is a bijection. This picture of  $\mathcal{M}$  as the half-space  $(0, \infty) \times \mathbb{R}^4$  in  $\mathbb{R}^5$ , as simple and pleasing as it is, is not the most informative. An alternative arises from the fact that there is an orientation preserving conformal diffeomorphism of  $(0, \infty) \times \mathbb{R}^4$  onto the open five-dimensional ball  $B^5$  in  $\mathbb{R}^5$ . Indeed, one can (pages 337-341 of [34]) introduce “spherical coordinates” on  $\mathcal{M}$  that yield a picture of  $\mathcal{M}$  as  $B^5$  with  $[\mathcal{A}_{1,0}]$  at its center. Proceeding radially outward from  $[\mathcal{A}_{1,0}]$  toward a point on  $\partial B^5 = \mathbb{S}^4$  one encounters potentials all of which have the same center  $n$ , but which become more and more concentrated, i.e., for which the spread  $\lambda \rightarrow 0$ . A particularly pleasing aspect of this picture is that the base manifold  $\mathbb{S}^4$  of the bundle emerges as the boundary of the moduli space  $\mathcal{M}$  in a compactification of  $\mathcal{M}$  ( $\mathcal{M} \cong B^5 \hookrightarrow \bar{B}^5 = B^5 \cup \mathbb{S}^4$ ) and its points can be identified intuitively with “delta function” potentials.

One sees quite clearly in this example how the topologies of the underlying four-manifold  $\mathbb{S}^4$  and the moduli space  $\mathcal{M}$  of ASD connections on the bundle  $Sp(1) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{S}^4$  are inextricably linked. We will conclude by briefly describing an amazing generalization of this scenario due to Simon Donaldson [10].

We let  $X$  denote a compact, oriented, simply connected, smooth four-manifold.  $H_2(X, \mathbb{Z})$  will denote its second homology group (with integer coefficients) and

$$Q_X : H_2(X, \mathbb{Z}) \times H_2(X, \mathbb{Z}) \longrightarrow \mathbb{Z}$$

its intersection form.

**Remark.**  $H_2(X, \mathbb{Z})$  is a finitely generated, free Abelian group and each of its elements can be identified with a certain equivalence class represented by a smoothly embedded, oriented, closed surface (two-manifold with boundary)  $\Sigma$  in  $X$ . Very roughly, the definition of  $Q_X$  goes as follows: For  $\alpha_1, \alpha_2 \in H_2(X, \mathbb{Z})$  one can select surfaces  $\Sigma_1, \Sigma_2$  representing them that intersect transversely (i.e., at each intersection point the tangent spaces to  $\Sigma_1$  and  $\Sigma_2$  span the tangent space to  $X$ ). An intersection point  $p$  is assigned the value 1 if an oriented basis for  $T_p(\Sigma_1)$  together with an oriented basis for  $T_p(\Sigma_2)$  gives an oriented basis for  $T_p(X)$ ; otherwise it is assigned the value -1. Then  $Q_X(\alpha_1, \alpha_2)$  is the (necessarily finite) sum of these values over all the intersection points.  $Q_X$  is a symmetric, bilinear form and is, moreover, unimodular, i.e., if  $\alpha_1, \dots, \alpha_t$  is a basis for  $H_2(X, \mathbb{Z})$  over  $\mathbb{Z}$ , then the matrix  $(Q_X(\alpha_i, \alpha_j))$  has determinant  $\pm 1$ . Here are a few examples:

$X$	$H_2(X, \mathbb{Z})$	$Q_X$
$\mathbb{S}^4$	$0$	$\emptyset$
$\mathbb{C}\mathbb{P}^2$	$\mathbb{Z}$	$(1)$
$\overline{\mathbb{C}\mathbb{P}^2}$	$\mathbb{Z}$	$(-1)$
$\mathbb{S}^2 \times \mathbb{S}^2$	$\mathbb{Z} \oplus \mathbb{Z}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ .

Here  $\mathbb{C}\mathbb{P}^2$  is the orbit space of  $\mathbb{S}^5 = \{(z^1, z^2, z^3) \in \mathbb{C}^3 : |z^1|^2 + |z^2|^2 + |z^3|^2 = 1\}$  by the  $U(1)$ -action  $(z^1, z^2, z^3) \cdot g = (z^1 g, z^2 g, z^3 g)$ . It is naturally a complex two-manifold and so has a canonical orientation.  $\overline{\mathbb{C}\mathbb{P}^2}$  is the same manifold with the opposite orientation. A less pedestrian example is the **Kummer surface** which we will denote  $K3$  and which can be defined as the complex algebraic surface in  $\mathbb{C}\mathbb{P}^3$  (same definition as  $\mathbb{C}\mathbb{P}^2$ , but begin with  $\mathbb{S}^7 \subseteq \mathbb{C}^4$ ) whose homogeneous coordinates  $z^1, z^2, z^3, z^4$  satisfy  $(z^1)^4 + (z^2)^4 + (z^3)^4 + (z^4)^4 = 0$ . The rank of  $H_2(K3, \mathbb{Z})$  is 22 and the intersection form is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus (-E_8) \oplus (-E_8)$$

where  $E_8$  is given by

$$\begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 \end{pmatrix}$$

The intersection form can also be defined for topological four-manifolds, but we will not enter into this here.

It has been known for some time that the intersection form is a basic invariant for compact four-manifolds. In 1949, Whitehead proved that two compact, simply connected four-manifolds  $X_1$  and  $X_2$  have the same homotopy type if and only if their intersection forms are equivalent (i.e., there exist bases for  $H_2(X_1, \mathbb{Z})$  and  $H_2(X_2, \mathbb{Z})$  relative to which  $Q_{X_1}$  and  $Q_{X_2}$  have the same matrix). In 1982, Freedman [16] showed that *every* unimodular, symmetric, integer bilinear form is the intersection form of at least one (and at most two) compact, simply connected *topological* four-manifold(s). In particular, this is true of the vast, impenetrable maze of positive definite forms (when the rank is 40 there are at least  $10^{51}$  equivalence classes of definite forms). Donaldson has shown that the differential topologist need not venture into this maze because only *one* positive definite, unimodular, symmetric, integer bilinear form can arise as the intersection form of a compact, simply connected *smooth* four-manifold.

**Donaldson’s 1983 Theorem:** *be a compact, oriented, simply connected, smooth four-manifold with positive definite intersection form  $Q_X$ . Then  $Q_X$  is standard, i.e., there is a basis for  $H_2(X, \mathbb{Z})$  over  $\mathbb{Z}$  relative to which the matrix of  $Q_X$  is the identity matrix.*

Donaldson’s Theorem is remarkable, but still more remarkable is its proof, which is a byproduct of the analysis of an instanton moduli space for  $X$ . We (very, very) briefly sketch the idea. Consider the bundle over  $X$  analogous to the Hopf bundle over  $\mathbb{S}^4$ , i.e., the  $Sp(1)$ -bundle

$$Sp(1) \hookrightarrow P \longrightarrow X$$

over  $X$  with Chern number one. Next, choose a Riemannian metric  $g$  on  $X$ . Both the bundle and the metric are to be regarded as auxiliary structures to facilitate the study of  $X$ . From  $g$  and the given orientation for  $X$  one obtains a Hodge star operator and thereby a notion of ASD connection on  $P$ . In somewhat more detail, the definition is as follows. Let  $\omega$  be a connection on  $Sp(1) \hookrightarrow P \rightarrow X$  and  $\Omega$  and its curvature. Then  $\Omega$  is a globally defined Lie algebra-valued two-form on  $P$ . It is horizontal in the sense that it vanishes when either of its arguments is vertical (tangent to a fiber of the principal bundle). The corresponding local gauge field strengths  $\mathcal{F} = s^*\Omega$  on  $X$  are related by the adjoint representation of  $Sp(1)$  on its Lie algebra and so these patch together to give a globally defined two-form  $\mathcal{F}_\omega$  on  $X$  with values in the adjoint bundle  $\text{ad}(P)$  of  $P$  (the vector bundle associated to  $P$  by the adjoint representation). The two-form  $\mathcal{F}_\omega$  is very often also called the curvature of  $\omega$ . It's advantage is that it is defined on the four-manifold  $X$  so its Hodge dual two-form  $*\mathcal{F}_\omega$  is also a two-form and it makes sense to say that the connection  $\omega$  is **anti-self-dual (ASD)** if  $*\mathcal{F}_\omega = -\mathcal{F}_\omega$ .

**Remark.** *For the record we point out that anti-self-dual connections can exist only on bundles with positive Chern number, whereas self-dual connections ( $*\mathcal{F}_\omega = \mathcal{F}_\omega$ ) can exist only if the Chern number is negative. The discussion to follow can be carried at equally well with self-dual connections on the bundle with Chern number minus one.*

Now, in general, there is no reason to believe that such ASD connections exist, but a deep result of Taubes [41] asserts that, for manifolds satisfying the hypotheses we have assumed of  $X$ , the bundle  $Sp(1) \hookrightarrow P \rightarrow X$  actually admits ASD connections. Thus, we may introduce the moduli space  $\mathcal{M}(g)$  of ASD connections on  $P$ . This moduli space *does* depend on the choice of  $g$  and, for a randomly chosen Riemannian metric, little can be said about its structure. One can show, however, that, for some choice of  $g$  (indeed, for a “generic” choice of  $g$ ), the moduli space  $\mathcal{M}$  has all of the following properties.

1. If  $m$  denotes half the number of homology classes  $\alpha \in H_2(X, \mathbb{Z})$  for which  $Q_X(\alpha, \alpha) = 1$ , then there exist points  $p_1, \dots, p_m \in \mathcal{M}$  such that  $\mathcal{M} - \{p_1, \dots, p_m\}$  is a smooth, orientable five-manifold.
2. Each  $p_i$ ,  $i = 1, \dots, m$ , has a neighborhood in  $\mathcal{M}$  that is homeomorphic to a cone over  $\mathbb{C}\mathbb{P}^2$  with  $p_i$  at the vertex (the cone over  $\mathbb{C}\mathbb{P}^2$  is the quotient of  $\mathbb{C}\mathbb{P}^2 \times [0, 1]$  obtained by identifying all points of the form  $(p, 1)$ ).
3. There is a compact set  $K \subseteq \mathcal{M}$  such that  $\mathcal{M} - K$  is a submanifold of  $\mathcal{M} - \{p_1, \dots, p_m\}$  diffeomorphic to  $X \times (0, 1)$ .



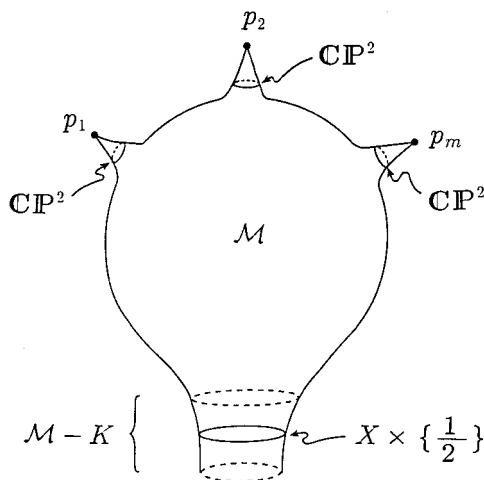


Figure 2. The Moduli Space

Now we build a new space  $\mathcal{M}_0$  from  $\mathcal{M}$  by cutting off the top half of each cone and the bottom half of the cylinder  $X \times (0, 1)$ .  $\mathcal{M}_0$  is compact (because  $K$  is compact). It is also a manifold with boundary whose boundary consists of the disjoint union of a copy of  $X$  and  $m$  copies of  $\mathbb{C}\mathbb{P}^2$ .

Now, in general, if  $X_1$  and  $X_2$  are two  $n$ -manifolds and if there exists an  $(n + 1)$ -manifold  $M$  with boundary whose boundary is a disjoint union of  $X_1$  and  $X_2$ , then  $M$  is called a **cobordism** between  $X_1$  and  $X_2$ .  $X_1$  and  $X_2$  are then said to be **cobordant**. Thus,  $\mathcal{M}_0$  is a cobordism between  $X$  and a disjoint union of  $\mathbb{C}\mathbb{P}^2$ 's. As it happens, the signature of the intersection form of a four-manifold is a cobordism invariant. This fact, together with the positive definiteness of  $Q_X$ , the known intersection form of  $\bigsqcup \mathbb{C}\mathbb{P}^2$  and a bit of integer linear algebra suffice to produce a basis for  $H_2(X, \mathbb{Z})$  relative to which the matrix for  $Q_X$  is the identity (page 347 of [34]).

Donaldson's 1983 Theorem was the first gauge-theoretic assault on a problem in the topology of smooth four-manifolds. Subsequent developments in what came to be known as **Donaldson Theory** yielded spectacular results, but at a cost in labor that seemed to grow exponentially with each new success (we will describe some of the most basic elements of Donaldson theory in Sections 4 and 8).

A breakthrough occurred in 1993 when Kronheimer and Mrowka isolated the (apparently large) class of four-manifolds of "simple type" and proved that, for these at least, Donaldson theory had some realistic chance of becoming effectively com-

putable. At precisely this moment, however, fate (or rather Ed Witten) intervened and the subject of smooth four-manifold topology took an entirely new turn. This is the story we would like to tell.

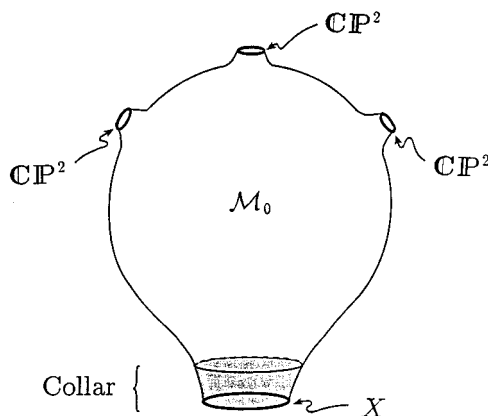


Figure 3. The Cobordism

## 2. $SU(2)$ Yang-Mills-Higgs Theory on $\mathbb{R}^3$

The notion of a “classical gauge theory”, which we have promised to define carefully in Section 3, is not adequately motivated by the examples of the preceding section (which would be classified by physicists as “pure” Yang-Mills theories). In general, gauge fields are coupled to (i.e., interact with) what we shall call “matter fields”. For example, electromagnetic fields ( $U(1)$ -gauge fields) are coupled to charged particles and, in the original proposal of Yang and Mills [49],  $SU(2)$ -gauge fields interact with nucleons and, at least in the absence of electromagnetic fields, govern the evolution of their isotopic spin (proton/neutron) state. In this section we will consider a concrete example which may seem a bit more abstruse, but which has proven to be very important and which has the added advantage of being derivable from either purely mathematical considerations (“dimensional reduction”) or in the manner more familiar to physicists (“field content plus action”). We will describe both.

Let us briefly return to the ASD equations on  $\mathbb{R}^4$  (it is traditional, and will be convenient in this section, to think in matrix terms so that we now identify  $Sp(1)$  with  $SU(2)$  and  $\text{Im}\mathbb{H}$  with  $sp(1)$  in the manner described in (1.2) and the Remark following (1.11)). Thus, we have a one-form  $\hat{\mathcal{A}} = \hat{\mathcal{A}}_1 dx^1 + \hat{\mathcal{A}}_2 dx^2 + \hat{\mathcal{A}}_3 dx^3 +$

$\hat{\mathcal{A}}_4 dx^4 = \hat{\mathcal{A}}_\alpha dx^\alpha$  on  $\mathbb{R}^4$  with each  $\hat{\mathcal{A}}_\alpha$  a smooth map on  $\mathbb{R}^4$  taking values in  $\mathfrak{su}(2)$ . The gauge potential  $\hat{\mathcal{A}}$  gives rise to a gauge field strength  $\hat{\mathcal{F}} = d\hat{\mathcal{A}} + \frac{1}{2}[\hat{\mathcal{A}}, \hat{\mathcal{A}}]$  which, in coordinates, is given by

$$\hat{\mathcal{F}} = \frac{1}{2} \hat{\mathcal{F}}_{\alpha\beta} dx^\alpha \wedge dx^\beta = \frac{1}{2} \left( \partial_\alpha \hat{\mathcal{A}}_\beta - \partial_\beta \hat{\mathcal{A}}_\alpha + [\hat{\mathcal{A}}_\alpha, \hat{\mathcal{A}}_\beta] \right) dx^\alpha \wedge dx^\beta \quad (2.1)$$

where  $\partial_\alpha$  means  $\partial/\partial x^\alpha$ . The Hodge dual of the field strength  $\hat{\mathcal{F}}$  is given by  $*\hat{\mathcal{F}} = \frac{1}{2} * \hat{\mathcal{F}}_{\alpha\beta} dx^\alpha \wedge dx^\beta$ , where  $*\hat{\mathcal{F}}_{\alpha\beta} = \frac{1}{2} \sum_{\gamma,\delta=1}^4 \varepsilon_{\alpha\beta\gamma\delta} \hat{\mathcal{F}}_{\gamma\delta}$  and  $\varepsilon_{\alpha\beta\gamma\delta}$  is the totally anti-symmetric Levi-Civita symbol with  $\varepsilon_{1234} = 1$ . Thus, the ASD equations (1.20) can be written

$$\hat{\mathcal{F}}_{\alpha\beta} = -\frac{1}{2} \sum_{\gamma,\delta=1}^4 \varepsilon_{\alpha\beta\gamma\delta} \hat{\mathcal{F}}_{\gamma\delta}, \quad \alpha, \beta = 1, 2, 3, 4. \quad (2.2)$$

There are many duplications in this list of equations (e.g.,  $\alpha = 3, \beta = 4$  and  $\alpha = 1, \beta = 2$  both reduce to  $\hat{\mathcal{F}}_{12} = -\hat{\mathcal{F}}_{34}$ ). Indeed, all of the equations in (2.2) are easily seen to be contained in

$$\hat{\mathcal{F}}_{ij} = -\sum_{k=1}^3 \varepsilon_{ijk} \hat{\mathcal{F}}_{k4}, \quad i, j = 1, 2, 3 \quad (2.3)$$

where  $\varepsilon_{ijk}$  is totally anti-symmetric and  $\varepsilon_{123} = 1$  (e.g.,  $\hat{\mathcal{F}}_{12} = -\hat{\mathcal{F}}_{34}$  is equivalent to  $\hat{\mathcal{F}}_{12} = -\varepsilon_{123} \hat{\mathcal{F}}_{34} = -\sum_{k=1}^3 \varepsilon_{12k} \hat{\mathcal{F}}_{k4}$ ).

Now, the finite action solutions to (2.3) are just what we have called instantons on  $\mathbb{R}^4$ . We wish now to abandon the finite action condition and seek solutions  $\hat{\mathcal{A}}$  to (2.3) that are **static**, i.e., independent of  $x^4$ . With this assumption, (2.1) gives  $\hat{\mathcal{F}}_{k4} = \partial_k \hat{\mathcal{A}}_4 + [\hat{\mathcal{A}}_k, \hat{\mathcal{A}}_4]$  so (2.3) becomes

$$\hat{\mathcal{F}}_{ij} = -\sum_{k=1}^3 \varepsilon_{ijk} \left( \partial_k \hat{\mathcal{A}}_4 + [\hat{\mathcal{A}}_k, \hat{\mathcal{A}}_4] \right), \quad i, j = 1, 2, 3. \quad (2.4)$$

Let us now “reduce to  $\mathbb{R}^3$ ” as follows: Fix (arbitrarily) some value  $x_0^4$  of  $x^4$  and consider the submanifold  $\mathbb{R}^3 \times \{x_0^4\}$  of  $\mathbb{R}^4$  (henceforth written simply  $\mathbb{R}^3$ ). For  $i = 1, 2, 3$  we let  $\mathcal{A}_i = \hat{\mathcal{A}}_i|_{\mathbb{R}^3}$  and then define  $\mathcal{A} = \mathcal{A}_1 dx^1 + \mathcal{A}_2 dx^2 + \mathcal{A}_3 dx^3 = \mathcal{A}_i dx^i$  on  $\mathbb{R}^3$ . The gauge potential  $\mathcal{A}$  on  $\mathbb{R}^3$  has a corresponding field strength  $\mathcal{F} = \frac{1}{2} \mathcal{F}_{ij} dx^i \wedge dx^j$ , with components  $\mathcal{F}_{ij} = \hat{\mathcal{F}}_{ij}|_{\mathbb{R}^3}$ . Note that  $\hat{\mathcal{A}}_4$  does not enter into either the potential or the field strength on  $\mathbb{R}^3$ . However, if we define  $\psi : \mathbb{R}^3 \rightarrow \mathfrak{su}(2)$  by

$$\psi = \hat{\mathcal{A}}_4|_{\mathbb{R}^3}$$

then the static ASD equations (2.4), when restricted to  $\mathbb{R}^3$ , become

$$\hat{\mathcal{F}}_{ij} = - \sum_{k=1}^3 \varepsilon_{ijk} (\partial_k \psi + [\mathcal{A}_k, \psi]), \quad i, j = 1, 2, 3 \quad (2.5)$$

or, in even more detail,

$$\partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i + [\mathcal{A}_i, \mathcal{A}_j] = - \sum_{k=1}^3 \varepsilon_{ijk} (\partial_k \psi + [\mathcal{A}_k, \psi]). \quad (2.6)$$

These we regard as field equations for an  $SU(2)$ -gauge potential  $\mathcal{A}$  coupled to a matter field  $\psi$  whose wavefunction takes values in  $\mathfrak{su}(2)$ . Somewhat more precisely,  $\mathcal{A}$  corresponds to a connection on the trivial  $SU(2)$ -bundle over  $\mathbb{R}^3$  and  $\psi$  is a section of the (likewise trivial) adjoint bundle. Thus,  $(\partial_k \psi + [\mathcal{A}_k, \psi]) dx^k$  is the corresponding covariant exterior derivative  $d^{\mathcal{A}}\psi$  of  $\psi$  and the sum on the right-hand side of (2.6) gives the components of the  $\mathbb{R}^3$ -Hodge dual of  $d^{\mathcal{A}}\psi$ . Consequently, (2.6) can be written

$$\mathcal{F} = - * d^{\mathcal{A}}\psi. \quad (2.7)$$

In whatever form they are written these are called the **Bogomolny monopole equations** (beginning instead with the self-dual equations  $*\hat{\mathcal{F}} = \hat{\mathcal{F}}$  on  $\mathbb{R}^4$  we would have arrived at  $\mathcal{F} = *d^{\mathcal{A}}\psi$  and these go by the same name). Before explaining the origin of the “monopole” terminology we will describe another path leading to the same set of equations.

We begin with the underlying base manifold  $\mathbb{R}^3$  with its usual Riemannian metric and orientation and with standard coordinates  $x^1, x^2$  and  $x^3$ . Consider the trivial  $SU(2)$ -bundle

$$SU(2) \hookrightarrow \mathbb{R}^3 \times SU(2) \longrightarrow \mathbb{R}^3$$

over  $\mathbb{R}^3$  (where the right action of  $SU(2)$  on  $\mathbb{R}^3 \times SU(2)$  being given by  $p \cdot g = (x, h) \cdot g = (x, hg)$ ). Thus, we have a natural global section

$$s : \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \times SU(2) \\ s(x) = (x, e)$$

where  $e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  is the identity element of  $SU(2)$ . Any other global section is then of the form

$$s^g : \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \times SU(2) \\ s^g(x) = s(x) \cdot g(x) = (x, g(x))$$

for some smooth map

$$g : \mathbb{R}^3 \longrightarrow SU(2)$$

which we identify with a global gauge transformation on the bundle. A connection on  $\mathbb{R}^3 \times SU(2)$  is an  $\mathfrak{su}(2)$ -valued one-form  $\omega$  on  $\mathbb{R}^3 \times SU(2)$ . Since the bundle is trivial,  $\omega$  is uniquely determined by the gauge potential

$$\mathcal{A} = s^* \omega.$$

Moreover, any  $\mathfrak{su}(2)$ -valued one-form on  $\mathbb{R}^3$  is the pullback by  $s$  of some connection on  $\mathbb{R}^3 \times SU(2)$  so we may restrict our attention entirely to globally defined gauge potentials  $\mathcal{A}$  on  $\mathbb{R}^3$ . A gauge transformation  $g : \mathbb{R}^3 \rightarrow SU(2)$  gives a new gauge representation

$$\mathcal{A}^g = (s^g)^* \omega = g^{-1} \mathcal{A} g + g^{-1} dg$$

where  $dg$  is the entrywise exterior derivative of  $g : \mathbb{R}^3 \rightarrow SU(2)$ . The curvature  $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$  is likewise determined by the field strength  $\mathcal{F} = s^* \Omega = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}]$  and a gauge transformation gives

$$\mathcal{F}^g = (s^g)^* \Omega = g^{-1} \mathcal{F} g.$$

We wish to construct a field theory in which one of the fields is an  $SU(2)$ -gauge potential  $\mathcal{A}$  on  $\mathbb{R}^3$  as described above and the other, to which  $\mathcal{A}$  will be coupled, is a so called ‘‘Higgs field’’. Now, in general, a ‘‘matter field’’, when quantized, represents a particle. The matter field itself is represented by a wavefunction which takes values in some vector space and which transforms under a gauge transformation by some representation of the structure group (in our case  $SU(2)$ ) on that vector space (we will expand upon these points in Section 3). More precisely, the matter field is a section of the vector bundle associated to the given principal bundle by some representation of the structure group on a vector space. When the vector space is the Lie algebra of the structure group and the representation is the adjoint representation (so that the vector bundle is the adjoint bundle) the matter field is called a Higgs field.

The adjoint bundle of the trivial bundle  $SU(2) \hookrightarrow \mathbb{R}^3 \times SU(2) \rightarrow \mathbb{R}^3$  is likewise trivial and any section of it can be identified with a smooth map  $\Psi : \mathbb{R}^3 \times SU(2) \rightarrow \mathfrak{su}(2)$  that is equivariant, i.e., satisfies

$$\begin{aligned} \Psi(p \cdot g) &= g^{-1} \cdot \Psi(p) \\ \Psi((x, h) \cdot g) &= g^{-1} \cdot \Psi(x, h) \\ \Psi(x, hg) &= g^{-1} \Psi(x, h) g. \end{aligned}$$

Since the bundle is trivial,  $\Psi$  is uniquely determined by

$$\begin{aligned}\psi &= s^* \Psi = \Psi \circ s \\ \psi(x) &= \Psi(x, e)\end{aligned}$$

(because  $\Psi(x, g) = \Psi((x, e) \cdot g) = g^{-1} \psi(x) g$ ) and we will focus our attention on  $\psi$ . A gauge transformation  $g : \mathbb{R}^3 \rightarrow SU(2)$  gives another gauge representation

$$\psi^g = (s^g)^* \Psi = g^{-1} \psi g$$

of the Higgs field.

We now write down an action (analogous to the Yang-Mills action (1.17)) the Euler-Lagrange equations of which will govern the interaction of  $\mathcal{A}$  and  $\psi$ . The respective integrand will contain three terms. The first is the **Yang-Mills term**  $-\text{tr}(\mathcal{F} \wedge * \mathcal{F})$  just as in (1.17). The second is called the **interaction term**  $-\text{tr}(d^{\mathcal{A}} \psi \wedge * d^{\mathcal{A}} \psi)$ , where  $d^{\mathcal{A}} \psi = d\psi + [\mathcal{A}, \psi]$  is the covariant exterior derivative of  $\psi$  (physicists would say that this term reflects the principle of minimal coupling). Finally, there is a term intended to describe the internal **self-interaction energy** of the matter field  $\psi$ . The precise form of this term must be postulated by choosing some non-negative, smooth, invariant, real-valued function  $V$  on  $\mathfrak{su}(2)$  and composing with  $\psi$ . We take  $V$  to be the familiar ‘‘Mexican hat’’ potential, i.e.,

$$V : \mathfrak{su}(2) \longrightarrow \mathbb{R} \quad V(A) = \frac{\lambda}{8} \left( \|A\|^2 - 1 \right)^2$$

where  $\lambda \geq 0$  is a constant and  $\|A\|^2 = \langle A, A \rangle = -2 \text{tr}(A^2)$ , as in the Remark following (1.11). We will write  $V \circ \psi$  as  $V(\psi) = \frac{\lambda}{8} (\|\psi\|^2 - 1)^2$  and think of it as a 0-form on  $\mathbb{R}^3$  so that its Hodge dual is  $\frac{\lambda}{8} * (\|\psi\|^2 - 1)^2 = \frac{\lambda}{8} (\|\psi\|^2 - 1)^2 \text{vol}$ . With this we can write down the so-called **Yang-Mills-Higgs action** on  $\mathbb{R}^3$ :

$$\begin{aligned}\mathcal{YMH}(\mathcal{A}, \psi) &= \int_{\mathbb{R}^3} \left( -\text{tr}(\mathcal{F} \wedge * \mathcal{F}) - \text{tr}(d^{\mathcal{A}} \psi \wedge * d^{\mathcal{A}} \psi) + \frac{\lambda}{8} * (\|\psi\|^2 - 1)^2 \right) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \left( \|\mathcal{F}\|^2 + \|d^{\mathcal{A}} \psi\|^2 + \frac{\lambda}{4} (\|\psi\|^2 - 1)^2 \right) \text{vol}. \quad (2.8)\end{aligned}$$

This action is gauge invariant, i.e.,  $\mathcal{A} \rightarrow \mathcal{A}^g$ ,  $\mathcal{F} \rightarrow \mathcal{F}^g$  and  $\psi \rightarrow \psi^g$  leaves the integral unchanged (we have already seen that  $\mathcal{F}^g = g^{-1} \mathcal{A} g$  and  $\psi^g = g^{-1} \psi g$  so  $\|\mathcal{F}^g\|^2 = \|\mathcal{F}\|^2$  and  $\|\psi^g\|^2 = \|\psi\|^2$  and a short calculation gives

$$d^{\mathcal{A}^g} \psi^g = g^{-1} (d^{\mathcal{A}} \psi) g$$

as well). We will generally be interested only in **finite action** field configurations  $(\mathcal{A}, \psi)$ , i.e., those for which (2.8) is finite. Now, the requirement that the action  $\mathcal{YM}\mathcal{H}(\mathcal{A}, \psi) < \infty$  implies that, as  $\|x\| \rightarrow \infty$  in  $\mathbb{R}^3$ ,  $\|\mathcal{A}\| \rightarrow 0$ ,  $\|d\mathcal{A}\psi\| \rightarrow 0$  and, at least if  $\lambda \neq 0$ ,

$$\|\psi\| \longrightarrow 1. \quad (2.9)$$

Indeed, it is shown in [22] that each of these limits is achieved uniformly on  $\mathbb{R}^3$ . Now, when  $\lambda = 0$  there is no reason to suppose that finite action implies  $\|\psi\| \rightarrow 1$  as  $\|x\| \rightarrow \infty$ . However, it is also shown in [22] that, for stationary configurations (i.e., those satisfying the Euler-Lagrange equations for  $\mathcal{YM}\mathcal{H}(\mathcal{A}, \psi)$ ) one loses nothing by restricting attention to those that satisfy (2.9), even when  $\lambda = 0$ . More precisely, we have the following: For any finite action critical point of  $\mathcal{YM}\mathcal{H}(\mathcal{A}, \psi)$  with  $\lambda = 0$  there exists a constant  $c \geq 0$  such that  $\|\psi\| \rightarrow c$  uniformly as  $\|x\| \rightarrow \infty$ . If  $c = 0$ , then  $(\mathcal{A}, \psi)$  is trivial. If  $c \neq 0$ , then one can rescale to obtain a new configuration  $(\mathcal{A}', \psi')$  given by  $(\mathcal{A}'(x), \psi'(x)) = (c^{-1}\mathcal{A}(c^{-1}x), c^{-1}\psi(c^{-1}x))$ . Then  $(\mathcal{A}', \psi')$  is also a finite action critical point for  $\mathcal{YM}\mathcal{H}(\mathcal{A}, \psi)$  with  $\lambda = 0$  and it satisfies  $\|\psi'\| \rightarrow 1$  uniformly as  $\|x\| \rightarrow \infty$ .

We intend to focus our attention on a certain limiting case of the Yang-Mills-Higgs action (that in which  $\lambda \rightarrow 0$  in (2.8)), but we retain a “virtual” self-interaction in the form of the boundary condition (2.9). First, however, we describe an important general feature of the full action (2.8). Notice that it has some obvious absolute minima. Indeed,  $\mathcal{YM}\mathcal{H}(\mathcal{A}, \psi)$ , which is non-negative, is zero when  $\mathcal{A} = 0$  and  $\psi = \psi_0$  is a constant in  $\mathfrak{su}(2)$  with  $\|\psi_0\| = 1$ . Such an absolute minimum is regarded as a **ground state** of the system. The corresponding quantum state of lowest energy is called a **vacuum state** and physicists perform perturbation calculations about such vacuum states. The point here is that these ground states are not unique. A specific choice of  $\psi_0$  is said to **break the symmetry** from  $SU(2)$  to  $U(1)$ . The rationale behind the terminology is as follows: A gauge transformation  $g : \mathbb{R}^3 \rightarrow SU(2)$  acts on  $\psi$  by  $\psi \rightarrow \psi^g = g^{-1}\psi g$ . If the ground state is to be gauge invariant, then we must have  $g^{-1}\psi_0 g = \psi_0$  and this occurs only if  $g$  is in the isotropy subgroup (stabilizer) of  $\psi_0$  in  $SU(2)$  under the adjoint action. We claim that this isotropy subgroup is a copy of  $U(1)$ . Briefly, the argument is as follows. Let  $H = \{g \in SU(2) : g^{-1}\psi_0 g = \psi_0\}$  be the isotropy subgroup. Obviously,  $-g \in H$  if and only if  $g \in H$ . Now, identifying  $SU(2)/\pm e$  with  $SO(3)$  (page 374 of [34]) and  $\mathfrak{su}(2)$  with  $\mathbb{R}^3$ , the adjoint action is just the natural action of  $SO(3)$  on  $\mathbb{R}^3$ , i.e., rotation (see the Appendix of [34]). This natural action of  $SO(3)$  is transitive on  $\mathbb{S}^2$  and the isotropy subgroup of  $\psi_0 \in \mathbb{S}^2$  is  $H' = H/\pm e$ . Since  $SO(3)$  is compact,  $SO(3)/H'$  is homeomorphic to  $\mathbb{S}^2$  (Theorem 1.6.6 of

[34]). Thus,  $(SU(2)/\pm e)/(H/\pm e) \cong \mathbb{S}^2$  so  $SU(2)/H \cong \mathbb{S}^2$ . Consequently,  $H$  is one-dimensional. Being closed in  $SU(2)$ ,  $H$  is also compact. Now, a compact, one-dimensional smooth manifold is a disjoint union of circles (Section 5.11 of [33]) so, being a subgroup of  $SU(2)$ ,  $H$  must be a single copy of the circle  $U(1)$ . What we have just witnessed is an instance of the phenomenon of **spontaneous symmetry breaking** in which a field theory with an exact symmetry group  $G$  gives rise to ground states that are invariant only under some proper subgroup  $H$  of  $G$ .

With this digression behind us we return to the limiting case of the Yang-Mills-Higgs action described above. Thus, we consider the action

$$\begin{aligned} A(\mathcal{A}, \psi) &= \int_{\mathbb{R}^3} \left( -\operatorname{tr}(\mathcal{F} \wedge *\mathcal{F}) - \operatorname{tr}(d^{\mathcal{A}}\psi \wedge *d^{\mathcal{A}}\psi) \right) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \left( \|\mathcal{F}\|^2 + \|d^{\mathcal{A}}\psi\|^2 \right) \mathbf{vol} \end{aligned} \quad (2.10)$$

and take as our configuration space

$$\mathcal{C} = \left\{ (\mathcal{A}, \psi) ; A(\mathcal{A}, \psi) < \infty, \lim_{R \rightarrow \infty} \sup_{\|x\| \geq R} |1 - \|\psi\|| = 0 \right\}. \quad (2.11)$$

The Euler-Lagrange equations for the action  $A(\mathcal{A}, \psi)$  are the **Yang-Mills-Higgs equations**

$$\begin{cases} *d^{\mathcal{A}} * \mathcal{F} = [d^{\mathcal{A}}\psi, \psi] \\ *d^{\mathcal{A}} * d^{\mathcal{A}}\psi = 0 \end{cases} \quad (2.12)$$

and we seek solutions to these in  $\mathcal{C}$ . Any configuration  $(\mathcal{A}, \psi)$  satisfying (2.12) also satisfies an analogue of the Bianchi identity (1.19) which we write as

$$\begin{cases} d^{\mathcal{A}}\mathcal{F} = 0 \\ d^{\mathcal{A}}d^{\mathcal{A}}\psi = [\mathcal{F}, \psi]. \end{cases} \quad (2.13)$$

Now, just as in the case of the Yang-Mills action, one can find a simpler set of first order equations whose solutions give absolute minima for the action  $A(\mathcal{A}, \psi)$  in (2.10) and so, in particular, satisfy the Yang-Mills-Higgs equations (2.12). To see how these arise we reason as follows. On  $\mathbb{R}^3$ , both  $\mathcal{F}$  and  $*d^{\mathcal{A}}\psi$  are two-forms and the Hodge dual is an isometry so  $\|*d^{\mathcal{A}}\psi\|^2 = \|d^{\mathcal{A}}\psi\|^2$ . Now observe that

$$\|\mathcal{F}\|^2 + \|d^{\mathcal{A}}\psi\|^2 = \|\mathcal{F}\|^2 + \|*d^{\mathcal{A}}\psi\|^2 = \|\mathcal{F} + *d^{\mathcal{A}}\psi\|^2 - 2\langle \mathcal{F}, *d^{\mathcal{A}}\psi \rangle \quad (2.14)$$



and similarly,

$$\|\mathcal{F}\|^2 + \|\mathrm{d}\mathcal{A}\psi\|^2 = \|\mathcal{F} - *\mathrm{d}\mathcal{A}\psi\|^2 + 2\langle\mathcal{F}, *\mathrm{d}\mathcal{A}\psi\rangle.$$

It follows that  $A(\mathcal{A}, \psi)$  will achieve its absolute minimum value (i.e., 0) when  $\mathcal{F} = \pm *\mathrm{d}\mathcal{A}\psi$  and these we recognize as the Bogomolny monopole equations introduced by quite different means earlier.

The appellation ‘‘monopole’’ derives from a certain exact solution to  $\mathcal{F} = -*\mathrm{d}\mathcal{A}\psi$  discovered by t’Hooft, Polyakov, Prasad and Sommerfeld. In spherical coordinates on  $\mathbb{R}^3$  this solution is given by  $\mathcal{A} = \mathcal{A}^1T^1 + \mathcal{A}^2T^2 + \mathcal{A}^3T^3$ ,  $\psi = \psi^1T^1 + \psi^2T^2 + \psi^3T^3$  (see the Remark following (1.11)), where

$$\mathcal{A}^1 = -\frac{\rho}{\sinh \rho} (\sin \theta \mathrm{d}\phi + \cos \theta \sin \phi \mathrm{d}\theta)$$

$$\mathcal{A}^2 = \frac{\rho}{\sinh \rho} (\cos \theta \mathrm{d}\phi - \sin \theta \sin \phi \mathrm{d}\theta)$$

$$\mathcal{A}^3 = -(1 - \cos \phi) \mathrm{d}\theta$$

$$\psi^1 = \psi^2 = 0$$

$$\psi^3 = \coth \rho - \frac{1}{\rho}$$

(the derivation of this solution is carried out in considerable detail on pages 141–150 of [35]). Notice that, despite appearances to the contrary, this configuration  $(\mathcal{A}, \psi)$  is a globally defined, smooth object on all of  $\mathbb{R}^3$  (the component functions are actually real analytic everywhere, even at  $\rho = 0$ ). Furthermore, when viewed from a distance (i.e., as  $\rho \rightarrow \infty$ ) the Higgs field approaches the constant value  $T_3$  (since  $\coth \rho - \frac{1}{\rho} \rightarrow 1$ ) and the first two components of  $\mathcal{A}$  approach zero (since  $\rho/\sinh \rho \rightarrow 0$ ). On the other hand,  $\mathcal{A}^3$  does not depend on  $\rho$  so it remains fixed at  $-(1 - \cos \phi)\mathrm{d}\theta$ . Thus, for large  $\rho$ ,  $\mathcal{A}$  is effectively  $-(1 - \cos \phi)\mathrm{d}\theta T^3$ . Under the isomorphism of  $\mathfrak{su}(2)$  onto  $\mathrm{Im}\mathbb{H}$  described in the Remark following (1.11), this becomes  $-\frac{\mathbf{k}}{2}(1 - \cos \phi)\mathrm{d}\theta$ . Since the span of  $\mathbf{k}$  in  $\mathrm{Im}\mathbb{H}$  is just a copy of  $\mathrm{Im}\mathbb{C}$  we recognize here just the potential for a Dirac monopole (see the Remark following (1.24)). Thus, the t’Hooft-Polyakov-Prasad-Sommerfeld monopole is a smooth field configuration of  $SU(2)$  Yang-Mills-Higgs theory which ‘‘looks like’’ a Dirac monopole from afar. The most interesting thing about the appearance of the Dirac monopole here is that it was entirely voluntary. In classical electromagnetic theory magnetic monopoles are, but certainly need not be, inserted by hand, whereas in  $SU(2)$  Yang-Mills-Higgs theory, they appear of their own accord (we return to this point at the end of the section). We remark that the same potential  $\mathcal{A}$  paired

with the Higgs field  $-\psi$  gives a solution to the “other” Bogomolny monopole equation  $\mathcal{F} = *d^{\mathcal{A}}\psi$ .

Thus motivated we will refer to any  $(\mathcal{A}, \psi) \in \mathcal{C}$  satisfying (2.7) as a **monopole** and will now associate with it a “monopole number”. Notice that if  $\mathcal{F} = -*d^{\mathcal{A}}\psi$ , then (2.14) becomes  $\|\mathcal{F}\|^2 + \|d^{\mathcal{A}}\psi\|^2 = -2\langle \mathcal{F}, *d^{\mathcal{A}}\psi \rangle$  so, for monopoles,

$$\begin{aligned} A(\mathcal{A}, \psi) &= - \int_{\mathbb{R}^3} \langle \mathcal{F}, *d^{\mathcal{A}}\psi \rangle \text{vol} = \int_{\mathbb{R}^3} 2\text{tr}(\mathcal{F} \wedge **d^{\mathcal{A}}\psi) \\ &= \int_{\mathbb{R}^3} 2\text{tr}(\mathcal{F} \wedge d^{\mathcal{A}}\psi) = \int_{\mathbb{R}^3} \text{Tr}(\mathcal{F} \wedge d^{\mathcal{A}}\psi) \end{aligned}$$

where we now use  $\text{Tr} = 2 \text{tr}$ . Computing this integral for the t’Hooft-Polyakov-Prasad-Sommerfeld monopole gives a value of  $4\pi$ . We normalize the action and define the **monopole number** of any  $(\mathcal{A}, \psi) \in \mathcal{C}$  satisfying (2.7) by

$$N(\mathcal{A}, \psi) = \frac{1}{4\pi} \int_{\mathbb{R}^3} \text{Tr}(\mathcal{F} \wedge d^{\mathcal{A}}\psi). \quad (2.15)$$

Like the instanton number introduced in Section 1, this monopole number is, in fact, an integer and one can see this in at least two different ways. Perhaps the easiest to describe is as follows (consult [35] for details on the rest that we have to say about  $N(\mathcal{A}, \psi)$ ): Since  $\|\psi\| \rightarrow 1$  as  $\|x\| \rightarrow \infty$  in  $\mathbb{R}^3$  there exists an  $R_0 > 0$  such that  $\|x\| > R_0$  implies  $\|\psi(x)\| > \frac{1}{2}$ . For  $\|x\| > R_0$  we can therefore define

$$\hat{\psi}(x) = \psi(x) / \|\psi(x)\|$$

and, for any  $R > R_0$ ,

$$\hat{\psi}_R = \hat{\psi}|_{\mathbb{S}_R^2}$$

where  $\mathbb{S}_R^2 = \{x \in \mathbb{R}^3; \|x\| = R\}$ . Now,  $\hat{\psi}_R$  can be regarded as a map from  $\mathbb{S}^2$  to  $\mathbb{S}^2$  and so determines an element  $[\hat{\psi}_R]$  of the homotopy group  $\pi_2(\mathbb{S}^2)$ . Moreover, since  $\hat{\psi}$  is smooth on  $\|x\| > R_0$ , its restrictions to any two such spheres are clearly homotopic so  $[\hat{\psi}_R]$  is independent of  $R > R_0$  and we will denote it simply  $[\hat{\psi}_\infty]$  (physicists would refer to  $\hat{\psi}_\infty$  as the restriction of  $\hat{\psi}$  to the “sphere at infinity”). Now,  $\pi_2(\mathbb{S}^2) \cong \mathbb{Z}$  and an isomorphism is provided by the Brouwer degree. One can show that the monopole number  $N(\mathcal{A}, \psi)$  is equal to the degree of any  $\hat{\psi}_R$ ,  $R > R_0$ , written

$$N(\mathcal{A}, \psi) = \text{deg}(\hat{\psi}_\infty) \quad (2.16)$$

and so, in particular, is an integer. Monopoles fall into distinct topological types according to the homotopy type of the (normalized) Higgs field on large spheres (these topological types are actually the connected components of the space of solutions to (2.7) in  $\mathcal{C}$  which are sometimes referred to as **topological sectors** in physics). This is, of course, entirely analogous to our earlier description of the instanton number in terms of the homotopy type of a gauge transformation on large (three-) spheres.

We will conclude by briefly sketching a description of the monopole number as the Chern number of a  $U(1)$ -bundle over  $\mathbb{S}^2$  obtained by breaking the  $SU(2)$  symmetry to  $U(1)$  through the selection of some ground state  $\psi_0$ . Fix some  $R > R_0$  as above.  $\psi$  is the pullback  $s^*\Psi$  by the standard section  $s$  of  $SU(2) \hookrightarrow \mathbb{R}^3 \times SU(2) \rightarrow \mathbb{R}^3$  of an equivariant map  $\Psi : \mathbb{R}^3 \times SU(2) \rightarrow \mathfrak{su}(2)$ . The restriction of this trivial  $SU(2)$ -bundle over  $\mathbb{R}^3$  to  $\mathbb{S}_R^2$  is the trivial  $SU(2)$ -bundle over  $\mathbb{S}_R^2$ :

$$SU(2) \hookrightarrow \mathbb{S}_R^2 \times SU(2) \xrightarrow{\pi} \mathbb{S}_R^2. \tag{2.17}$$

Now let  $\Psi_R = \Psi|_{\mathbb{S}_R^2 \times SU(2)}$  and  $\hat{\Psi}_R = \|\Psi_R\|^{-1}\Psi_R$ . Both are equivariant and  $\hat{\Psi}_R$  takes values in  $\mathbb{S}_{\mathfrak{su}(2)}^2 = \{A \in \mathfrak{su}(2); \|A\|^2 = 1\}$ . Furthermore,  $\hat{\psi}_R$  is the pullback of  $\hat{\Psi}_R$  by the standard section of the trivial bundle (2.17). Thus,  $\hat{\psi}_R$  is the standard gauge representation of a Higgs field on the bundle (2.17).

Now break the symmetry, i.e., select some ground state  $\psi_0 \in \mathbb{S}_{\mathfrak{su}(2)}^2$ . The isotropy subgroup of  $\psi_0$  (with respect to the adjoint action of  $SU(2)$  on  $\mathfrak{su}(2)$ ) is, as we have seen, a copy of  $U(1)$  and we will denote it simply  $U(1)$ . Now, one can show that  $\hat{\Psi}_R^{-1}(\psi_0)$  is a submanifold of  $\mathbb{S}_R^2 \times SU(2)$  which is invariant under the action of  $U(1)$  and that, moreover, the restriction of  $\pi$  to this submanifold gives a principal  $U(1)$ -bundle over  $\mathbb{S}_R^2$ :

$$U(1) \hookrightarrow \hat{\Psi}_R^{-1}(\psi_0) \xrightarrow{\pi|_{\hat{\Psi}_R^{-1}(\psi_0)}} \mathbb{S}_R^2. \tag{2.18}$$

The  $U(1)$ -bundle (2.18) is called a reduction of the structure group of (2.17) to  $U(1)$ . Now,  $U(1)$ -bundles over  $\mathbb{S}^2$  are classified by their first Chern number (the integral over  $\mathbb{S}^2$  of the first Chern class) which is always an integer. The result of interest to us here is that by choosing an appropriate connection on (2.18) and writing down the formula for the first Chern number using this connection one arrives at the expression (2.15) for the monopole number of  $(\mathcal{A}, \psi)$ .

We will conclude our discussion of  $SU(2)$ -monopoles by very briefly discussing an issue which must surely be troubling the reader. In classical electromagnetic theory magnetic monopoles must be inserted by hand. One of Maxwell’s equations explicitly forbids the existence of “magnetic charges” and, in order to understand the consequences of their possible existence, Dirac [9] was forced to

abandon (or, rather, modify) this equation and postulate the existence of a magnetic analogue of the electric charge. Certain  $SU(2)$ -monopoles “look like” Dirac monopoles from a distance. One might wonder as to the “source” of their magnetic charge.

A naive hint concerning the source of the magnetic charge of  $SU(2)$ -monopoles can be found in our earlier view of them as static, ASD potentials on  $\mathbb{R}^4$ . Recall that any solution to the ASD equations on  $\mathbb{R}^4$  also satisfies the full Yang-Mills equations  $d^{\mathcal{A}} * \mathcal{F} = 0$  and the Bianchi identity  $d^{\mathcal{A}} \mathcal{F} = 0$  and that these are regarded as a nonabelian generalization of Maxwell’s equations. The static version of Maxwell’s equations that contains both electric and magnetic charge densities ( $\rho_e$  and  $\rho_m$ , respectively) is, in appropriate units and on  $\mathbb{R}^3$ ,  $d * \mathbf{F} = 0$ ,  $d\mathbf{F} = * \rho_m$  and  $\nabla^2 \psi = \rho_e$ , where  $\mathbf{F}$  is the magnetic field two-form and  $\psi$  is the electric potential. Noting that  $d^{\mathcal{A}} \mathcal{F} = 0$  is equivalent to

$$d\mathcal{F} = -[\mathcal{A}, \mathcal{F}]$$

one can view the commutator term as playing the role of a magnetic charge density. The role of the Higgs field  $\psi$  (or, more to the point, the boundary condition  $\|\psi\| \rightarrow 1$  as  $\|x\| \rightarrow \infty$ ) is to break the symmetry at large distances from  $SU(2)$  down to  $U(1)$ , thus turning the  $SU(2)$  theory into a  $U(1)$ , i.e., electromagnetic, theory.

### 3. Classical Gauge Theories

Abstracting the salient features of the examples in the preceding sections, we now propose to enumerate a sequence of basic mathematical ingredients which together will serve as our working definition of a **classical gauge theory**.

- 1) A smooth, oriented, (semi-) Riemannian manifold  $X$ .

Generally, this will be space ( $\mathbb{R}^3$ ), a spacetime (e.g., Minkowski spacetime  $\mathbb{R}^{1,3}$ ), a Euclidean (“Wick rotated”) version of a spacetime (e.g.,  $\mathbb{R}^4$ ), a compactification of one of these (e.g.,  $S^4 = \mathbb{R}^4 \cup \{\infty\}$ ), an open submanifold of one of these (e.g.,  $\mathbb{R}^3 \setminus \{0\}$ ), or some homotopy equivalent (e.g.,  $S^2 \simeq \mathbb{R}^3 \setminus \{0\}$ ). The particles and fields which it is the ultimate goal of gauge theory to describe “live” in  $X$ .

- 2) A finite dimensional vector space  $\mathcal{V}$  equipped with an inner product  $\langle \cdot, \cdot \rangle$  (positive definite if  $\mathcal{V}$  is real and Hermitian if  $\mathcal{V}$  is complex).

The particles have wavefunctions that take values in  $\mathcal{V}$ . The choice of  $\mathcal{V}$  is dictated by the internal structure of the particle (charge, spin, isospin, etc.) and so  $\mathcal{V}$  is called the **internal space**. Typical examples are  $\mathbb{C}$  (spin zero charged particles),  $\mathbb{C}^4$  (Dirac electrons),  $\mathbb{C}^8$  (nucleons), or the Lie algebra  $\mathcal{G}$  of some Lie group  $G$  (Higgs fields). From the inner product  $\langle \cdot, \cdot \rangle$  one computes squared norms of  $\mathcal{V}$ -valued functions, forms, etc. and from these formulates action principles that govern the dynamics (see 8) below).

- 3) A matrix Lie group  $G$  and a representation  $\rho : G \rightarrow GL(\mathcal{V})$  of  $G$  on  $\mathcal{V}$  that is orthogonal with respect to the inner product  $\langle \cdot, \cdot \rangle$  i.e.,

$$\langle \rho(g)(v), \rho(g)(w) \rangle = \langle v, w \rangle$$

for all  $g \in G$  and  $v, w \in \mathcal{V}$ .

$G$  will generally be one of the classical groups (e.g.,  $U(1)$ ,  $SU(2)$ ,  $SO(4)$ ,  $SL(2, \mathbb{C})$ , etc.) or a product of these. In general,  $G$  describes a symmetry of the physical system under consideration, while  $\rho$  describes the particular type of invariance that a particle's wavefunction exhibits under this symmetry. More specifically, the Lie group  $G$  plays the following dual roles

- a) The inner product  $\langle \cdot, \cdot \rangle$  on  $\mathcal{V}$  determines a class of orthonormal bases, or frames, in  $\mathcal{V}$  and these are related by the elements of  $G$ , i.e., if  $P$  is the collection of all such frames, then there is a (right) action of  $G$  on  $P$  which sends any frame  $p \in P$  to a new frame  $p \cdot g \in P$ . By fixing (arbitrarily) some frame at the outset one can therefore identify the elements of  $G$  with the frames.
- b)  $G$  also acts on  $\mathcal{V}$  (on the left) via the representation  $\rho(v \rightarrow \rho(g)(v) = g \cdot v)$  and so acts on the wavefunction at each point. If  $\psi(p)$  is a value of the wavefunction described relative to the frame  $p \in P$ , then its description relative to the frame  $p \cdot g$  is

$$\psi(p \cdot g) = g^{-1} \cdot \psi(p). \tag{3.1}$$

The right action of  $G$  on  $P$  transforms frames in the internal space and the left action of  $G$  on  $\mathcal{V}$  describes the corresponding transformation law for the wavefunction.

- 4) A smooth principal  $G$ -bundle  $G \hookrightarrow P \xrightarrow{\pi} X$  over  $X$ .

Typical examples are trivial bundles (e.g.,  $SU(2) \hookrightarrow \mathbb{R}^4 \times SU(2) \rightarrow \mathbb{R}^4$ ) and Hopf bundles (e.g.,  $U(1) \hookrightarrow \mathbb{S}^3 \rightarrow \mathbb{S}^2$  and  $SU(2) \hookrightarrow \mathbb{S}^7 \rightarrow \mathbb{S}^4$ ). At each  $x \in X$  the fiber  $\pi^{-1}(x)$  is a copy of  $G$ , thought of as the set of all frames in the internal space at  $x \in X$ . A local section  $s : U \rightarrow \pi^{-1}(U) \subseteq P$  ( $U$  open in  $X$  and  $\pi \circ s = \text{id}_U$ ) is a smooth selection of an internal frame at each point of  $U$  relative to which wavefunctions can be described on  $U$ . Such a local section is also called a local **gauge**.

5) A connection  $\omega$  on  $G \hookrightarrow P \xrightarrow{\pi} X$  with curvature  $\Omega$ .

As motivation for 5) we recall that, in classical electrodynamics, an electromagnetic field is generally modeled by a two-form  $F$  defined on space or spacetime (i.e., on  $X$ ). The corresponding potential is a one-form  $A$  with  $F = dA$ .  $F$  is globally defined on  $X$ , but, in general, potentials are only locally defined so that a complete description of  $F$  will require a number of potentials with domains that cover  $X$ . In nonabelian gauge theories even the field strengths are, in general, only locally defined on  $X$ . However, by virtue of the manner in which these locally defined forms on  $X$  are related on the intersections of their domains (the local gauge transformation laws) one can piece them together into globally defined forms on the bundle space  $P$  of some principal bundle (characterized by transition functions that are simply read off from the transformation laws). These are the connection  $\omega$  and its curvature  $\Omega$ .

On the other hand, given  $\omega$  and  $\Omega$  one retrieves the physical potentials and fields by choosing a local gauge/section  $s : U \rightarrow P$  and pulling back to  $X$ :  $\mathcal{A} = s^*\omega$  is the **local gauge potential** and  $\mathcal{F} = s^*\Omega$  is the **local gauge field strength** (both in gauge  $s$ ). Another local gauge  $s' : U' \rightarrow P$  with  $U \cap U' \neq \emptyset$  will be related to  $s$  by  $s'(x) = s(x) \cdot g(x)$ , where  $g : U \cap U' \rightarrow G$  and  $\cdot$  is the right action in the principal bundle. One generally writes  $s^g$  rather than  $s'$  to explicitly display the so-called **transition function**  $g$ . The corresponding potential and field strength are written  $\mathcal{A}^g = (s^g)^*\omega$  and  $\mathcal{F}^g = (s^g)^*\Omega$  and are given by

$$\mathcal{A}^g = g^{-1}\mathcal{A}g + g^{-1}dg \quad (3.2)$$

and

$$\mathcal{F}^g = g^{-1}\mathcal{F}g \quad (3.3)$$

on  $U \cap U'$ . The change of gauge  $s \rightarrow s^g = s \cdot g$  is a **local gauge transformation** and can be identified with the map  $g : U \cap U' \rightarrow G$ . The **gauge principle**, or **principle of local gauge invariance**, is a cornerstone of modern theoretical physics and asserts that such a gauge transformation alters only the appearance and not

the physics of a situation, e.g., that  $\mathcal{A}^g$  and  $\mathcal{F}^g$  represent the same potential and field strength as  $\mathcal{A}$  and  $\mathcal{F}$ , only written in different internal coordinates.

**Remark.** *Before recording the next item in our list of ingredients for a classical gauge theory we recall several facts from geometry (see Section 5.7 of [34] for more details). Given a principal  $G$ -bundle  $G \hookrightarrow P \xrightarrow{\pi} X$  and a left action of  $G$  on some manifold  $F$  one can construct a fiber bundle  $P \times_G F \xrightarrow{\pi_G} X$  associated to the principal bundle by the left action whose typical fiber is  $F$ . In particular, if  $F$  is a vector space  $\mathcal{V}$  and the left action of  $G$  on  $\mathcal{V}$  arises from a representation  $\rho : G \rightarrow GL(\mathcal{V})$  of  $G$  on  $\mathcal{V}$  one obtains an associated vector bundle, usually written  $P \times_\rho \mathcal{V}$ . A typical example is the adjoint bundle  $\text{ad}P = P \times_{\text{ad}} \mathcal{G}$ , where  $\mathcal{V}$  is the Lie algebra  $\mathcal{G}$  of the structure group  $G$  and  $\rho = \text{ad}$  is the adjoint representation of  $G$  on  $\mathcal{G}$  ( $\text{ad}(g)(A) = gAg^{-1}$ ). We will need to use the fact (page 356 of [34]) that there are two equivalent ways of viewing a section of an associated bundle  $P \times_G F$ , i.e., either as a map  $\psi$  from  $X$  to  $P \times_G F$  for which  $\pi_G \circ \psi$  is the identity, or as a map  $\psi$  from  $P$  to  $F$  that is equivariant ( $\psi(p \cdot g) = g^{-1} \cdot \psi(p)$ ). The latter view and (3.1) should motivate*

- 6) A global section  $\psi$  of the vector bundle  $P \times_\rho \mathcal{V}$  associated to  $G \hookrightarrow P \xrightarrow{\pi} X$  by the representation  $\rho : G \rightarrow GL(\mathcal{V})$  (or, equivalently, an equivariant map  $\psi : P \rightarrow \mathcal{V}$ ).

Particles coupled to (i.e., experiencing the effects of) the gauge field determined by  $\omega$  have locally defined wavefunctions taking values in  $\mathcal{V}$  that are obtained by solving field equations (see 8) below) that involve the local potentials  $\mathcal{A}$ . A change of gauge changes the wavefunction by the representation  $\rho$  (see (3.1)) so these local wavefunctions piece together into a globally defined object called a **matter field** that can be described in either of the two equivalent ways referred to in 6).

**Remark.** *It is entirely possible that more than one matter field is coupled to the gauge field, but we will phrase our basic scheme for classical gauge theories assuming that there is just one and leave it to the reader to add on more terms if necessary.*

- 7) A smooth, non-negative, real-valued function  $V : \mathcal{V} \rightarrow \mathbb{R}$  on  $\mathcal{V}$  that is invariant under the action of  $G$  on  $\mathcal{V}$  ( $V(g \cdot v) = V(v)$ ).

$V$  is regarded as a potential function with  $V \circ \psi = V(\psi)$  describing the **self-interaction energy** of the matter field  $\psi$ . Typically, this will depend only on

$\|v\|^2 = \langle v, v \rangle$ , e.g.,  $\frac{\lambda}{8}(\|v\|^2 - 1)^2$ , or  $\frac{1}{2}m\|v\|^2$ , where  $\lambda$  and  $m$  are non-negative constants.

- 8) An action (energy) functional  $A(\omega, \psi)$ , the stationary points of which are the physically significant field configurations  $(\omega, \psi)$ . The Euler-Langrange equations for  $A(\omega, \psi)$  are the **field equations** (or, **equations of motion**) for the classical gauge theory.

When  $X$  is Riemannian (as it is in cases of topological interest) one can generally expect an action of the form

$$A(\omega, \psi) = c \int_X \left( \|F_\omega\|^2 + c_1 \|d^\omega \psi\|^2 + c_2 V(\psi) \right) \text{vol} \quad (3.4)$$

where  $c$  is some normalizing constant,  $c_1$  and  $c_2$  are “coupling constants”,  $F_\omega$  is the globally defined two-form on  $X$  with values in the adjoint bundle  $\text{ad}(P)$  which locally reduces to the gauge field strengths  $\mathcal{F} = s^*\Omega$ ,  $d^\omega \psi$  is the covariant exterior derivative of the matter field  $\psi$  and the norms arise from the metric on  $X$ , the inner product on  $\mathcal{V}$  and some ad-invariant inner product on  $\mathcal{G}$ .

We have already seen several examples of classical gauge theories that are of particular interest to us because of the topological nature of certain solutions to their field equations. Later (Sections 5 and 7) we will see other, rather more complicated examples whose impact on topology and geometry has been much more profound. Of course, most examples of interest in physics are not topological in nature at all, but we will nevertheless pause briefly to describe one of the simplest of these (more details and still more examples are to be found in Chapter 2 of [35]). The situation we intend to model (at the classical level) is the interaction of an electromagnetic field with a charged, spin zero particle (e.g., a  $\pi^-$ -meson).

**Remark.** *Certain technical complications, which we do not wish to become involved in, arise for more familiar charged particles like the electron and proton. The reason is that these have spin  $\frac{1}{2}$  and so, according to Dirac, have wave-functions that transform under a certain representation of  $SL(2, \mathbb{C})$ , whereas the electromagnetic field to which it is coupled is a  $U(1)$ -gauge theory. To fit this interaction into the general framework we have described would require “splicing” a  $U(1)$ -bundle and an  $SL(2, \mathbb{C})$ -bundle together into a single  $U(1) \times SL(2, \mathbb{C})$ -bundle on which both objects may be thought to live. This can be done and the process is carried out in more detail in Section 2.4 of [35].*



The arena within which electrodynamics is done is **Minkowski spacetime**  $\mathbb{R}^{1,3}$ . As a differentiable manifold  $\mathbb{R}^{1,3}$  is just  $\mathbb{R}^4$ , but, rather than the usual Riemannian metric on  $\mathbb{R}^4$  we introduce the semi-Riemannian **Minkowski metric** given, relative to standard coordinates  $x^0, x^1, x^2, x^3$  by

$$\eta_{\alpha\beta} dx^\alpha \otimes dx^\beta$$

where

$$\eta_{\alpha\beta} = \begin{cases} 1, & \alpha = \beta = 0 \\ -1, & \alpha = \beta = 1, 2, 3 \\ 0, & \alpha \neq \beta. \end{cases}$$

One thinks of the elements of  $\mathbb{R}^{1,3}$  as **events** whose standard coordinates are the time ( $x^0$ ) and spatial ( $x^1, x^2, x^3$ ) coordinates by which the event is identified by some fixed, but arbitrary inertial observer. The entire history of a (point) object can then be identified with a continuous sequence of events (i.e., a curve) in  $\mathbb{R}^{1,3}$  called its **worldline**.

**Remark.** We will denote by  $\eta$  the  $4 \times 4$  matrix  $(\eta_{\alpha\beta})$  and, even though  $\eta^{-1}$  is actually equal to  $\eta$ , we will write  $\eta^{-1} = (\eta^{\alpha\beta})$  to facilitate use of the Einstein summation convention.

Now we let  $X$  denote some open submanifold of  $\mathbb{R}^{1,3}$  (the charges creating our electromagnetic field live in  $\mathbb{R}^{1,3}$  and we intend to carve out their worldlines and consider only the source free Maxwell equations on the resulting open submanifold of  $\mathbb{R}^{1,3}$ ). Traditionally, an electromagnetic field on  $X$  is modeled by a globally defined, real-valued two-form  $\mathbf{F}$  on  $X$  that satisfies the **source free Maxwell equations**

$$d\mathbf{F} = 0 \quad \text{and} \quad d*\mathbf{F} = 0$$

where  $*$  is the Hodge star on  $\mathbb{R}^{1,3}$  determined by the usual orientation of  $\mathbb{R}^{1,3}$  as  $\mathbb{R}^4$  and the Minkowski metric (specifically, if  $\mathbf{F} = \frac{1}{2}F_{\alpha\beta} dx^\alpha \wedge dx^\beta$ , then  $*\mathbf{F} = \frac{1}{2}*F_{\alpha\beta} dx^\alpha \wedge dx^\beta$ , where  $*F_{\alpha\beta} = \frac{1}{2}\varepsilon_{\alpha\beta\gamma\delta} F^{\gamma\delta}$  and  $F^{\gamma\delta} = \eta^{\mu\gamma}\eta^{\nu\delta} F_{\mu\nu}$ ). An electromagnetic potential for  $\mathbf{F}$  is a one-form  $\mathbf{A}$  (generally only locally defined) that satisfies  $d\mathbf{A} = \mathbf{F}$  on its domain. In the gauge-theoretic formulation we propose now these will both acquire a (trivial) Lie-algebra factor of  $-i$  (i.e., we will deal instead with  $\mathcal{F} = -i\mathbf{F}$  and  $\mathcal{A} = -i\mathbf{A}$ ).

Now we build the classical gauge theory model by introducing items 1) – 8).  $X$ , as we have said, will be an open submanifold of  $\mathbb{R}^{1,3}$ , with the induced orientation and semi-Riemannian metric. Since the particle we have in mind is charged and

has spin zero, physics dictates that its wavefunction should have one complex component so we take  $\mathcal{V}$  to be the (two-dimensional, real) vector space  $\mathbb{C}$  with the usual positive definite inner product  $\langle \cdot, \cdot \rangle$ , which can be written

$$\langle z_1, z_2 \rangle = \frac{1}{2} (z_1 \bar{z}_2 + \bar{z}_1 z_2). \quad (3.5)$$

The matrix Lie group  $G$  of 3) is taken to be  $U(1)$ . Now, every irreducible representation of  $U(1)$  on  $\mathbb{C}$  is of the form

$$\begin{aligned} \rho_n : U(1) &\longrightarrow GL(\mathbb{C}) \\ \rho_n(g)(z) &= g \cdot z = g^n z \end{aligned} \quad (3.6)$$

for some integer  $n$  and all of these are easily seen to be orthogonal with respect to  $\langle \cdot, \cdot \rangle$ . Since electric charge is quantized we can measure it in multiples of the electron's charge, i.e., by an integer. We identify the  $n$  in (3.6) with the charge of the spin zero particle we have under consideration.

Now let  $U(1) \hookrightarrow P \xrightarrow{\pi} X$  be a principal  $U(1)$ -bundle over  $X$  and  $\omega$  a connection on the bundle with curvature  $\Omega = d\omega$  (since  $U(1)$  is abelian, all brackets are zero). For any section  $s : U \rightarrow P$  we can write the corresponding gauge potential  $\mathcal{A}$  and field strength  $\mathcal{F}$  (which are  $u(1) = \text{Im}\mathbb{C}$ -valued) in terms of real-valued forms  $\mathbf{A}$  and  $\mathbf{F}$ , respectively, as follows:

$$\mathcal{A} = s^* \omega = \mathcal{A}_\alpha dx^\alpha = -iA_\alpha dx^\alpha = -i\mathbf{A}. \quad (3.7)$$

$$\mathcal{F} = s^* \Omega = \frac{1}{2} \mathcal{F}_{\alpha\beta} dx^\alpha \wedge dx^\beta = -\frac{1}{2} iF_{\alpha\beta} dx^\alpha \wedge dx^\beta = -i\mathbf{F}. \quad (3.8)$$

$$\mathcal{F}_{\alpha\beta} = \partial_\alpha \mathcal{A}_\beta - \partial_\beta \mathcal{A}_\alpha = -i(\partial_\alpha A_\beta - \partial_\beta A_\alpha).$$

If  $s' : U' \rightarrow P$  is another section with  $U \cap U' \neq \emptyset$  and if, on  $U \cap U'$ ,  $s' = s \cdot g$ , where  $g : U \cap U' \rightarrow U(1)$  is the local gauge transformation, then the corresponding potential and field strength are given by

$$\mathcal{A}^g = g^{-1} \mathcal{A} g + g^{-1} dg = \mathcal{A} + g^{-1} dg$$

and

$$\mathcal{F}^g = g^{-1} \mathcal{F} g = \mathcal{F}$$

on  $U \cap U'$  (again we use the fact that  $U(1)$  is abelian). Notice that  $\mathcal{F}^g = \mathcal{F}$  is the reason that field strengths in abelian gauge theories are globally defined on the base manifold  $X$ .

A matter field (item 6)) can be identified with a map  $\psi : P \rightarrow \mathbb{C}$  that is equivariant, i.e., satisfies

$$\psi(p \cdot g) = g^{-1} \cdot \psi(p) = g^{-n} \psi(p) \quad (3.9)$$

for all  $p \in P$  and all  $g \in U(1)$ , or, equivalently, with the corresponding section of the vector bundle  $P \times_{\rho_n} \mathbb{C}$  (we will use the same symbol  $\psi$  for both). As a potential function (item 7)) we take

$$V : \mathbb{C} \longrightarrow \mathbb{R} \quad (3.10)$$

$$V(z) = \frac{1}{2}m \langle z, z \rangle = \frac{1}{2}mz\bar{z} = \frac{1}{2}m|z|^2$$

where  $m > 0$  is a constant (ultimately identified with the mass of the particle). Since  $\rho_n$  is orthogonal with respect to  $\langle \cdot, \cdot \rangle$ ,  $V$  is invariant under the action of  $U(1)$  on  $\mathbb{C}$ , as required. Finally, we must specify an action (energy) functional (item 8)).

**Remark.** *Since the metric on  $X$  is now semi-Riemannian, inner products of forms need no longer be positive definite and we will refrain from writing norms as we did in (3.4). Since we have thus far dealt only with  $\mathfrak{su}(2)$  (i.e.,  $\text{Im}\mathbb{H}$ )-valued forms we briefly recall that if  $\alpha$  and  $\beta$  are two  $p$ -forms with values in some vector space with an inner product, then one defines the (pointwise) inner product of  $\alpha$  and  $\beta$  as follows: Select a basis  $\{T_a\}$  and write  $\alpha = \alpha^a T_a$  and  $\beta = \beta^b T_b$ , where  $\alpha^a$  and  $\beta^b$  are real-valued  $p$ -forms. These real-valued forms have (pointwise) inner products  $\langle \alpha^a, \beta^b \rangle$  defined by  $\alpha^a \wedge * \beta^b = \langle \alpha^a, \beta^b \rangle \mathbf{vol}$  and we define  $\langle \alpha, \beta \rangle$  by*

$$\langle \alpha, \beta \rangle = \langle \alpha^a T_a, \beta^b T_b \rangle = \langle \alpha^a, \beta^b \rangle \langle T_a, T_b \rangle$$

(we rely upon the reader to decide which inner product is intended by looking at what is inside). The result is independent of the choice of  $\{T_a\}$ . Applying this to the  $\text{Im}\mathbb{C}$ -valued two-form  $\mathcal{F}$  of (3.8) with the standard inner product (3.5) on  $\text{Im}\mathbb{C}$  reveals that

$$\mathcal{F} \wedge * \mathcal{F} = - \langle \mathcal{F}, \mathcal{F} \rangle \mathbf{vol} = \frac{1}{2} \mathcal{F}_{\alpha\beta} \mathcal{F}^{\alpha\beta} \mathbf{vol} = -\frac{1}{2} F_{\alpha\beta} F^{\alpha\beta} \mathbf{vol} . \quad (3.11)$$

Similarly, for any  $\mathbb{C}$ -valued  $p$ -form  $\mu$  one finds, again using the standard inner product (3.5) on  $\mathbb{C}$  and writing  $\mu = \mu^1 + \mu^2 i$ , that

$$\mu \wedge * \bar{\mu} = \mu^1 \wedge * \mu^1 + \mu^2 \wedge * \mu^2 = \langle \mu, \mu \rangle \mathbf{vol} . \quad (3.12)$$

Finally, we remark that the switch to Minkowski spacetime necessitates a sign change in the Yang-Mills term  $\mathcal{F} \wedge * \mathcal{F}$  of the action in order to ensure that the energy of the field (which is related to its spatial integrals) is positive.

Now, as was the case for  $SU(2)$  Yang-Mills-Higgs theory, our action will contain a Yang-Mills term, an interaction term and a potential term. Only the interaction

term remains to be discussed and it, once again, is determined by “minimal coupling”. In somewhat more detail, let us (temporarily) think of the matter field  $\psi$  as an equivariant  $\mathbb{C}$ -valued map on  $P$ . Then the covariant exterior derivative  $d^\omega \psi$  is just  $d\psi$  acting on  $\omega$ -horizontal parts of tangent vectors. As a section of the vector bundle  $P \times_{\rho_n} \mathbb{C}$ ,  $\psi$  is determined by the pullbacks of the equivariant map and the corresponding covariant exterior derivative is determined by the pullbacks of  $d^\omega \psi$ . These are given locally on  $X$  and in standard coordinates by

$$(\partial_\alpha + n\mathcal{A}_\alpha) \psi dx^\alpha = (\partial_\alpha - inA_\alpha) \psi dx^\alpha \quad (3.13)$$

where  $\mathcal{A} = \mathcal{A}_\alpha dx^\alpha = -iA_\alpha dx^\alpha$  is the corresponding gauge potential (we are thinking of the matter field as a section now and so write  $\psi$  rather than  $s^*\psi = \psi \circ s$ ). Now, (3.12) gives

$$d^\omega \psi \wedge \overline{*d^\omega \psi} = \langle d^\omega \psi, d^\omega \psi \rangle \mathbf{vol}$$

which, when written out locally in coordinates with (3.13) yields

$$\begin{aligned} \langle d^\omega \psi, d^\omega \psi \rangle &= (\partial_\alpha \psi + n\mathcal{A}_\alpha \psi) (\partial^\alpha \bar{\psi} - n\mathcal{A}^\alpha \bar{\psi}) \\ &= (\partial_\alpha \psi - inA_\alpha \psi) (\partial^\alpha \bar{\psi} + inA^\alpha \bar{\psi}) \end{aligned} \quad (3.14)$$

where  $\mathcal{A}^\alpha = \eta^{\alpha\beta} \mathcal{A}_\beta$ ,  $A^\alpha = \eta^{\alpha\beta} A_\beta$  and  $\partial^\alpha = \eta^{\alpha\beta} \partial_\beta$ .

With this we can write down a proposed action for our system consisting of a scalar field of mass  $m$  and charge  $n$  coupled to an electromagnetic field determined by the local gauge potential  $\mathcal{A} = \mathcal{A}_\alpha dx^\alpha = -iA_\alpha dx^\alpha$  as

$$\begin{aligned} A(\omega, \psi) &= \frac{1}{2} \int_X \left( \mathcal{F} \wedge *\mathcal{F} + d^\omega \psi \wedge \overline{*d^\omega \psi} + m *|\psi|^2 \right) \\ &= \int_X \left( -\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \right. \\ &\quad \left. + \frac{1}{2} (\partial_\alpha \psi - inA_\alpha \psi) (\partial^\alpha \bar{\psi} + inA^\alpha \bar{\psi}) + \frac{1}{2} m |\psi|^2 \right) \mathbf{vol}. \end{aligned} \quad (3.15)$$

The corresponding Euler-Lagrange equations are

$$(\partial_\alpha - inA_\alpha) (\partial^\alpha - inA^\alpha) \psi + m^2 \psi = 0 \quad (3.16)$$

$$d * \mathbf{F} = 0 \quad (3.17)$$

where  $*$  denotes the Minkowski spacetime Hodge star. Since  $-i\mathbf{F}$  is the pullback of a curvature form it satisfies the Bianchi identity and this gives

$$d\mathbf{F} = 0. \quad (3.18)$$

The last two equations are just the sourcefree Maxwell equations, while (3.16) is the **Klein-Gordon equation** coupling our scalar field to the electromagnetic field.

#### 4. The Zero-Dimensional Donaldson Invariant

We have seen in Section 1 that pure Yang-Mills theory, which arose from attempts by physicists to understand elementary particles, has deep consequences in differential topology (Donaldson’s 1983 Theorem). Coupling a gauge field to matter fields, as in  $SU(2)$  Yang-Mills-Higgs theory, also yields some rather tantalizing connections with topology, as we saw in Section 2. This is, however, just the beginning of our story. From 1983 to 1994 the study of smooth four-manifolds was dominated by the ideas of Simon Donaldson who showed how to extend the techniques behind his theorem on intersection forms to construct remarkably sensitive differential topological invariants for such manifolds (we describe the simplest of these in this section). In 1988, Witten [46], prompted by Atiyah, produced a classical gauge theory in the sense of Section 3 which, upon quantization, was found to contain certain observables whose expectation values were precisely these Donaldson invariants (the simplest of these invariants is the partition function of the quantum field theory and we will “derive” it in Section 5). This construction of Witten’s was a remarkable achievement and provided the most direct sort of link between topology and physics. However, the most extraordinary aspect of all of this did not emerge until the Fall of 1994 when his then recent work with Seiberg on supersymmetric gauge theories led Witten [48] to conjecture that all of the topological information contained in the Donaldson invariants could be extracted also from the vastly simpler set of invariants now known as Seiberg-Witten invariants (at least for a certain large class of four-manifolds). This part of the story will be related in Sections 7 and 8.

We begin our journey down this road by outlining the construction of the so-called zero-dimensional Donaldson invariant. Throughout this section  $B$  will denote a compact, simply connected, oriented, smooth four-manifold (when the need arises somewhat later we will recall the definition of  $b_2^+(B)$  and impose additional assumptions regarding it). Every  $SU(2)$ -bundle  $SU(2) \hookrightarrow P \xrightarrow{\pi} B$  over  $B$  has a second Chern number  $c_2(P)[B] \in \mathbb{Z}$  which can be written as

$$c_2(P)[B] = \frac{1}{8\pi^2} \int_B \text{tr}(\mathbf{F}\omega \wedge \mathbf{F}\omega) \tag{4.1}$$

where  $\omega$  is any connection on the bundle and  $\mathbf{F}\omega$  is its curvature (thought of as a two-form on  $B$  with values in the adjoint bundle  $\text{ad}(P)$ ). Such bundles are

characterized up to equivalence by this integer and we shall denote by

$$SU(2) \hookrightarrow P_k \xrightarrow{\pi_k} B$$

the bundle with  $c_2(P_k)[B] = k$ . Shortly we will explain why we are interested only in those bundles with  $k > 0$ .  $\mathcal{C}(P_k)$  will denote the set of all connection one-forms on  $P_k$  and  $\mathcal{G}(P_k)$  is the **gauge group** of all **(global) gauge transformations** of  $P_k$  (diffeomorphisms  $f$  of  $P_k$  onto itself satisfying  $\pi_k \circ f = \pi_k$  and  $f(p \cdot g) = f(p) \cdot g$  for all  $p \in P_k$  and  $g \in SU(2)$ ).  $\mathcal{G}(P_k)$  acts on  $\mathcal{C}(P_k)$  on the right by pullback ( $\omega \rightarrow \omega \cdot f = f^*\omega$ ). Two connections  $\omega, \omega' \in \mathcal{C}(P_k)$  are said to be **gauge equivalent** if there is an  $f \in \mathcal{G}(P_k)$  such that  $\omega' = f^*\omega$  and we will denote by  $[\omega]$  the gauge equivalence class of  $\omega$ . The set of all such gauge equivalence classes is called the **moduli space** of connections on  $P_k$  and written

$$\mathcal{B}(P_k) = \mathcal{C}(P_k) / \mathcal{G}(P_k) = \{[\omega]; \omega \in \mathcal{C}(P_k)\}.$$

It is this moduli space that we wish to study. Unfortunately, it has no reasonable mathematical structure in the smooth context in which we have just introduced it so one must replace the smooth objects just defined with appropriate Sobolev completions. This will require that some of the definitions be recast in other, but equivalent forms.

**Remark.** *Let us briefly recall a convenient means of defining Sobolev completions for a space of sections of a vector bundle. Begin with a compact Lie group  $G$  and a principal  $G$ -bundle  $G \hookrightarrow P \xrightarrow{\pi} X$  over some compact, oriented manifold  $X$ . Let  $\mathcal{V}$  be a finite-dimensional real vector space with a positive-definite inner product and  $\rho : G \rightarrow GL(\mathcal{V})$  an orthogonal representation of  $G$  on  $\mathcal{V}$ . Let  $E = P \times_{\rho} \mathcal{V}$  be the associated vector bundle (any vector bundle over  $X$  can be represented in this way). Let  $\Omega^i(X, E)$  be the space of  $i$ -forms on  $X$  with values in  $E$ . In particular,  $\Omega^0(X, E)$  is the space of sections of  $E$ . Choosing a Riemannian metric  $\mathbf{g}$  on  $X$  one obtains natural inner products on each  $\Omega^i(X, E)$ . Choosing a connection  $\omega$  on  $P$  induces covariant exterior differentiation operators*

$$\Omega^0(X, E) \xrightarrow{d^{\omega}} \Omega^1(X, E) \xrightarrow{d^{\omega}} \Omega^2(X, E) \xrightarrow{d^{\omega}} \dots \quad (4.2)$$

Now suppose  $\xi \in \Omega^0(X, E)$ . For each  $m = 0, 1, 2, \dots$  one defines the **Sobolev  $m$ -norm**  $\|\xi\|_m$  of  $\xi$  by

$$\|\xi\|_m^2 = \sum_{j=0}^m \int_X \|(d^{\omega} \circ \dots \circ d^{\omega})(\xi)\|^2 \mathbf{vol}$$

where  $\mathbf{vol}$  is the metric volume form of  $\mathbf{g}$ . This is, indeed, a norm on  $\Omega^0(X, E)$  and different choices of the Riemannian metric  $\mathbf{g}$ , the connection  $\omega$  and the inner product on  $\mathcal{V}$  give rise to equivalent norms. The completion of  $\Omega^0(X, E)$  relative to this norm is actually a Hilbert space  $L_m^2(E)$ . Sobolev embedding theorems guarantee that, by choosing  $m$  sufficiently large, one can achieve any desired degree of smoothness for the elements of  $L_m^2(E)$ . More precisely, if  $l$  is a non-negative integer and  $m > \frac{1}{2}(\dim X) + l$ , then  $L_m^2(E)$  embeds in the space  $C^l(X, E)$  of  $l$ -times continuously differentiable sections of  $E$ . Also note that each  $\Omega^i(X, E)$  is itself a space of sections of some vector bundle and so has Sobolev completions. Before returning to the main development we remark for future reference that, unlike the ordinary exterior derivative, the sequence (4.2) of covariant exterior derivatives is generally not a complex. Indeed, when  $E = \text{ad}P$  the composition of the first two

$$d^\omega \circ d^\omega : \Omega^0(X, \text{ad}P) \longrightarrow \Omega^2(X, \text{ad}P)$$

is given by

$$d^\omega \circ d^\omega (\cdot) = [F_\omega, \cdot]. \tag{4.3}$$

Now we refashion our earlier definitions in such a way that we can define their Sobolev completions in the manner described in the above Remark. Denote by  $\Omega^i(P_k, \mathfrak{su}(2))$  the vector space of  $i$ -forms on  $P_k$  with values in the Lie algebra  $\mathfrak{su}(2)$ . Then  $\Omega_{\text{ad}}^i(P_k, \mathfrak{su}(2))$  will denote the subspace consisting of all  $\varphi \in \Omega^i(P_k, \mathfrak{su}(2))$  that are **tensorial of type ad**, i.e., satisfy the following two conditions:

1.  $\varphi$  is **horizontal** in the sense that it vanishes whenever one of its arguments is vertical (tangent to a fiber in  $P_k$ ).
2. For each  $g \in SU(2)$   $\sigma_g^* \varphi = g^{-1} \cdot \varphi = g^{-1} \varphi g$ , where  $\sigma_g : P_k \rightarrow P_k$  is the diffeomorphism  $\sigma_g(p) = p \cdot g$ .

Finally, let  $\Omega^i(B, \text{ad}P_k)$  denote the space of  $i$ -forms on  $B$  with values in the adjoint bundle  $\text{ad}P_k$ . One easily shows that  $\Omega_{\text{ad}}^i(P_k, \mathfrak{su}(2))$  and  $\Omega^i(B, \text{ad}P_k)$  are isomorphic (pull back elements of  $\Omega_{\text{ad}}^i(P_k, \mathfrak{su}(2))$  by sections of  $P_k$  and show, using 1) and 2), that these piece together to give elements of  $\Omega^i(B, \text{ad}P_k)$ ). For example, the curvature  $\Omega$  of a connection  $\omega$  is an  $\mathfrak{su}(2)$ -valued two-form that is tensorial of type ad and the corresponding element of  $\Omega^2(B, \text{ad}P_k)$  is what we have been denoting  $F_\omega$ . Our interest in this vector space is accounted for by the following proposition.

**Proposition 4.1.**  $\mathcal{C}(P_k)$  is an affine space modeled on the vector space  $\Omega_{\text{ad}}^1(P_k, \mathfrak{su}(2)) \cong \Omega^1(B, \text{ad}P_k)$ , i.e., if  $\omega_0$  is any element of  $\mathcal{C}(P_k)$ , then

$$\mathcal{C}(P_k) = \{\omega_0 + \varphi ; \varphi \in \Omega_{\text{ad}}^1(P_k, \mathfrak{su}(2))\}. \quad (4.4)$$

The proof is simple since one need only show that if  $\omega$  and  $\omega_0$  are in  $\mathcal{C}(P_k)$ , then  $\omega - \omega_0$  is tensorial of type ad.

Now, each  $\Omega^i(B, \text{ad}P_k)$  is a space of sections of a vector bundle and so has Sobolev completions  $\Omega_m^i(B, \text{ad}P_k)$  for  $m = 0, 1, 2, 3, \dots$ . For sufficiently large  $m$  its elements are all continuous sections so the isomorphism  $\Omega^i(B, \text{ad}P_k) \cong \Omega_{\text{ad}}^i(P_k, \mathfrak{su}(2))$  serves to define the Sobolev completions  $\Omega_{\text{ad}}^i, m(P_k, \mathfrak{su}(2))$ . Thus, we can define a Sobolev space of connections on  $P_k$  for each such  $m$  by

$$\mathcal{C}_m(P_k) = \{\omega_0 + \varphi ; \varphi \in \Omega_{\text{ad}}^1, m(P_k, \mathfrak{su}(2))\},$$

where  $\omega_0$  is any fixed, smooth connection on  $P_k$ . For our purposes it will suffice to take  $m = 3$ .

$$\mathcal{C}_3(P_k) = \{\omega_0 + \varphi ; \varphi \in \Omega_{\text{ad}}^1, 3(P_k, \mathfrak{su}(2))\}. \quad (4.5)$$

To define Sobolev completions of the gauge group  $\mathcal{G}(P_k)$  we consider the nonlinear adjoint bundle  $\text{Ad}P_k$ . This is the fiber bundle associated to  $SU(2) \hookrightarrow P_k \rightarrow B$  by the adjoint (conjugation) action of  $SU(2)$  on itself. In particular, its typical fiber is the group  $SU(2)$ , although it is not a principal bundle. Let  $\Omega^0(B, \text{Ad}P_k)$  be the set of smooth sections of  $\text{Ad}P_k$ . It is a group under pointwise multiplication in the fibers and is easily seen to be isomorphic to the group  $\Omega_{\text{Ad}}^0(P_k, SU(2))$  of smooth maps  $\psi : P_k \rightarrow SU(2)$  that are equivariant, i.e., satisfy  $\sigma_g^* \psi = g^{-1} \cdot \psi$ , or, equivalently,  $\psi(p \cdot g) = g^{-1} \psi(p) g$  (here the group operation is pointwise multiplication in  $SU(2)$ ). We care about these groups for the following reason.

**Proposition 4.2.**  $\mathcal{G}(P_k) \cong \Omega_{\text{Ad}}^0(P_k, SU(2)) \cong \Omega^0(B, \text{Ad}P_k)$ .

Once again the proof is simple. A gauge transformation  $f : P_k \rightarrow P_k$  preserves the fibers of  $P_k$  and satisfies  $f(p \cdot g) = f(p) \cdot g$  so, for each  $p \in P_k$  there is a unique  $\psi(p) \in SU(2)$  for which  $f(p) = p \cdot \psi(p)$  and this defines the appropriate  $\psi \in \Omega_{\text{Ad}}^0(P_k, SU(2))$ . Now, unfortunately,  $\Omega^0(B, \text{Ad}P_k)$  consists of sections of a fiber bundle with fiber  $SU(2)$  and not a vector bundle so it is not immediately clear how to define its Sobolev completions. However, if we regard  $SU(2)$  as a subset of the vector space  $M_{2 \times 2}(\mathbb{C})$  of  $2 \times 2$ -complex matrices, then  $\text{Ad}P_k$  embeds in the vector bundle  $E = P_k \times_{\rho} M_{2 \times 2}(\mathbb{C})$ , where  $\rho$  is the representation



of  $SU(2)$  on  $M_{2 \times 2}(\mathbb{C})$  corresponding to conjugation. But the Sobolev spaces  $L_m^2(E)$  are defined (and  $C^1$  for sufficiently large  $m$ ) so we can take

$$\mathcal{G}_m(P_k) = \{s \in L_m^2(E); s(B) \subseteq \text{Ad}P_k\}$$

for such  $m$ . It will suffice for our purposes to take  $m = 4$  and we will abuse the notation somewhat and write

$$\mathcal{G}_4(P_k) = \Omega_4^0(B, \text{Ad}P_k) \tag{4.6}$$

(the Sobolev index for  $\mathcal{G}$  must be one greater than that for  $\mathcal{C}$  to ensure a smooth action of  $\mathcal{G}$  on  $\mathcal{C}$ ). One can show ([15] or [30]) that  $\mathcal{G}_4(P_k)$  is a Hilbert Lie group with Lie algebra (tangent space at the identity  $\mathbb{1}$ ) that can be identified with

$$T_{\mathbb{1}}(\mathcal{G}_4(P_k)) = \Omega_4^0(B, \text{ad}P_k) \tag{4.7}$$

(this is at least believable since the sections in  $\Omega_4^0(B, \text{ad}P_k)$  can be exponentiated pointwise to give elements of  $\Omega_4^0(B, \text{Ad}P_k)$ ).

Now, the action of  $\mathcal{G}(P_k)$  on  $\mathcal{C}(P_k)$  extends to an action of  $\mathcal{G}_4(P_k)$  on  $\mathcal{C}_3(P_k)$  (same formulas since the elements of  $\mathcal{G}_4(P_k)$  are  $C^1$  and those of  $\mathcal{C}_3(P_k)$  are continuous). It is shown in [15] and [30] that this action is actually smooth and that, if  $\omega \in \mathcal{C}_3(P_k)$  is fixed, the map of  $\mathcal{G}_4(P_k)$  to  $\mathcal{C}_3(P_k)$  given by

$$f \longrightarrow \omega \cdot f$$

has a derivative at  $\mathbb{1}$  that can be identified with

$$d^\omega : \Omega_4^0(B, \text{ad}P_k) \longrightarrow \Omega_3^1(B, \text{ad}P_k). \tag{4.8}$$

**Remark.** *Differential operators extend to bounded operators on Sobolev completions and this is the meaning of  $d^\omega$  here and henceforth.*

In particular, the tangent space at  $\omega$  to the orbit  $\omega \cdot \mathcal{G}_4(P_k)$  of  $\omega$  under  $\mathcal{G}_4(P_k)$  is given by

$$T_\omega(\omega \cdot \mathcal{G}_4(P_k)) = \text{im}(d^\omega) = d^\omega(\Omega_4^0(B, \text{ad}P_k)). \tag{4.9}$$

Now, the **moduli space**

$$\mathcal{B}_3(P_k) = \mathcal{C}_3(P_k) / \mathcal{G}_4(P_k)$$

**of (Sobolev index 3) connections** on  $P_k$  is the set of gauge equivalence classes of the elements of  $\mathcal{C}_3(P_k)$  modulo the action of  $\mathcal{G}_4(P_k)$ . Since  $\mathcal{C}_3(P_k)$  is an affine space (by its very definition (4.5)) it has a natural topology and we provide  $\mathcal{B}_3(P_k)$  with the quotient topology, which one can show is Hausdorff ([15] or [30]).

Our next objective is to study the local structure of  $\mathcal{B}_3(P_k)$ . Ideally, we would like a local manifold structure at each  $[\omega]$ , but we will find that this is possible only for what are called “irreducible” connections  $\omega$ . The definition is as follows. For any  $\omega \in \mathcal{C}_3(P_k)$  the **stabilizer** (or **isotropy subgroup**) of  $\omega$  is the subgroup  $\text{stab}(\omega)$  of  $\mathcal{G}_4(P_k)$  that leaves  $\omega$  fixed, i.e.,

$$\text{stab}(\omega) = \{f \in \mathcal{G}_4(P_k); \omega \cdot f = \omega\}.$$

Any such stabilizer contains the subgroup  $\mathbb{Z}_2$  of  $\mathcal{G}_4(P_k)$  generated by  $\pm \mathbb{1}$  and if this is all it contains, i.e., if  $\text{stab}(\omega) = \mathbb{Z}_2$ , then  $\omega$  is said to be **irreducible**; otherwise  $\omega$  is **reducible**. The following characterization of reducibility is proved in [15] and [30] (indeed, these will be our references for everything further we have to say about the moduli spaces).

**Theorem 4.3.** *The following are equivalent for any  $\omega \in \mathcal{C}_3(P_k)$ .*

- a)  $\omega$  is reducible, i.e.,  $\text{stab}(\omega)/\mathbb{Z}_2$  is nontrivial.
- b)  $\text{stab}(\omega)/\mathbb{Z}_2 \cong U(1)$ .
- c)  $d\omega : \Omega_4^0(B, \text{ad}P_k) \longrightarrow \Omega_3^1(B, \text{ad}P_k)$  has nontrivial kernel.

We will denote by  $\hat{\mathcal{C}}_3(P_k)$  the subset of  $\mathcal{C}_3(P_k)$  of irreducible connections (it is, in fact, an open subset) and by

$$\hat{\mathcal{B}}_3(P_k) = \hat{\mathcal{C}}_3(P_k) / \mathcal{G}_4(P_k)$$

the **moduli space of irreducible (Sobolev index 3) connections** on  $P_k$ . The latter is an open subspace of  $\mathcal{B}_3(P_k)$ .

Now we turn to the local structure of these moduli spaces. First consider an  $\omega \in \hat{\mathcal{C}}_3(P_k)$  so that  $[\omega] \in \hat{\mathcal{B}}_3(P_k)$ . We will produce a “slice” of the  $\hat{\mathcal{G}}_4(P_k)$ -action on  $\mathcal{C}_3(P_k)$  near  $\omega$ , i.e., a submanifold  $\mathcal{O}$  of  $\mathcal{C}_3(P_k)$  such that

$$T_\omega(\mathcal{C}_3(P_k)) = T_\omega(\omega \cdot \mathcal{G}_4(P_k)) \oplus T_\omega(\mathcal{O}) \quad (4.10)$$

and such that the restriction to  $\mathcal{O}$  of the projection into the moduli space is injective near  $\omega$ . Then the local structure of the moduli space near  $[\omega]$  is the same as that of  $\mathcal{O}$  near  $\omega$ . To produce this  $\mathcal{O}$  we will first produce an “orthogonal decomposition” of  $T_\omega(\mathcal{C}_3(P_k))$  into  $T_\omega(\omega \cdot \mathcal{G}_4(P_k))$  plus “something” and then use the affine structure (4.5) of  $\mathcal{C}_3(P_k)$  to define a submanifold having this “something” as its tangent space at  $\omega$ .

Choosing a Riemannian metric  $\mathbf{g}$  on  $B$  and an ad-invariant inner product  $\langle \cdot, \cdot \rangle$  on the Lie algebra  $\mathfrak{su}(2)$  gives rise to natural inner products on all of the vector spaces  $\Omega^i(B, \text{ad}P_k)$  so that the operator  $d^\omega : \Omega^0(B, \text{ad}P_k) \longrightarrow \Omega^1(B, \text{ad}P_k)$  has a formal adjoint  $\delta^\omega$ :

$$\Omega^0(B, \text{ad}P_k) \begin{array}{c} \xrightarrow{d^\omega} \\ \xleftarrow{\delta^\omega} \end{array} \Omega^1(B, \text{ad}P_k)$$

(in fact,  $\delta^\omega = -*d^\omega*$ , where  $*$  is the Hodge dual corresponding to  $\mathbf{g}$  and the given orientation of  $B$ ). It turns out that

$$\delta^\omega \circ d^\omega : \Omega^0(B, \text{ad}P_k) \longrightarrow \Omega^0(B, \text{ad}P_k)$$

is a (formally self-adjoint) elliptic operator. We use the same symbols for the extensions of these operators to the Sobolev completions  $\Omega_4^0(B, \text{ad}P_k)$  and  $\Omega_3^1(B, \text{ad}P_k)$ . Elliptic theory (the generalized Hodge Decomposition Theorem) implies that

- a)  $\ker(\delta^\omega \circ d^\omega) = \ker(d^\omega)$  is finite-dimensional
- b)  $\text{im}(d^\omega) = \ker(\delta^\omega)^\perp$
- c)  $d^\omega$  has closed range and
- d) there is an orthogonal decomposition

$$\Omega_3^1(B, \text{ad}P_k) = \text{im}(d^\omega) \oplus \ker(\delta^\omega)$$

i.e.,

$$T_\omega(\mathcal{C}_3(P_k)) = T_\omega(\omega \cdot \mathcal{G}_4(P_k)) \oplus \ker(\delta^\omega) \quad (4.11)$$

(by (4.5) and (4.9)). Now, for any  $\varepsilon > 0$ , the submanifold

$$\mathcal{O}_{\omega, \varepsilon} = \{\omega + \mathbf{A}; \mathbf{A} \in \ker(\delta^\omega), \|\mathbf{A}\|_3 < \varepsilon\} \quad (4.12)$$

clearly satisfies

$$T_\omega(\mathcal{O}_{\omega, \varepsilon}) = \ker(\delta^\omega)$$

so

$$T_\omega(\mathcal{C}_3(P_k)) = T_\omega(\omega \cdot \mathcal{G}_4(P_k)) \oplus T_\omega(\mathcal{O}_{\omega, \varepsilon}). \quad (4.13)$$

We claim that, for sufficiently small  $\varepsilon > 0$ ,  $\mathcal{O}_{\omega, \varepsilon}$  projects injectively into the moduli space.

To prove this last claim one first observes that, since  $\omega \in \hat{\mathcal{C}}_3(P_k)$  and  $\hat{\mathcal{C}}_3(P_k)$  is open in  $\mathcal{C}_3(P_k)$  we can take  $\varepsilon > 0$  small enough to ensure that  $\mathcal{O}_{\omega, \varepsilon} \subseteq \hat{\mathcal{C}}_3(P_k)$ . Now consider the map

$$\Psi : \mathcal{O}_{\omega, \varepsilon} \times \mathcal{G}_4(P_k) \longrightarrow \hat{\mathcal{C}}_3(P_k), \quad \Psi(\omega', f) = \omega' \cdot f.$$

The derivative of  $\Psi$  at  $(\omega, \mathbb{1})$  is computed to be

$$\begin{aligned} (d\Psi)_{(\omega, \mathbb{1})} : \ker(\delta^\omega) \oplus \Omega_4^0(B, \text{ad}P_k) &\longrightarrow \ker(\delta^\omega) \oplus \text{im}(d^\omega) \\ (d\Psi)_{(\omega, \mathbb{1})} &= (\text{Id}_{\ker(\delta^\omega)}, d^\omega). \end{aligned}$$

This is certainly surjective and, because  $\omega$  is assumed irreducible, Theorem 4.3 c) implies that it is also injective. By the well known Open Mapping Theorem (a bounded, surjective, linear operator between Banach spaces is an open mapping),  $(d\Psi)_{(\omega, \mathbb{1})}$  is an isomorphism. Thus, the Inverse Function Theorem for Banach manifolds (see [29]) implies that, near  $(\omega, \mathbb{1})$ ,  $\Psi$  is a local diffeomorphism. More precisely, for some (perhaps smaller)  $\varepsilon > 0$  there is an open neighborhood  $U_\omega$  of  $\omega$  in  $\hat{\mathcal{C}}_3(P_k)$  and an open set  $U_{\mathbb{1}, \varepsilon} = \{f \in \mathcal{G}_4(P_k); \|\mathbb{1} - f\|_4 < \varepsilon\}$  in  $\mathcal{G}_4(P_k)$  such that the restriction

$$\Psi : \mathcal{O}_{\omega, \varepsilon} \times U_{\mathbb{1}, \varepsilon} \longrightarrow U_\omega$$

is a diffeomorphism. In particular, no two things in  $\mathcal{O}_{\omega, \varepsilon}$  are gauge equivalent by any gauge transformation that is within  $\varepsilon$  of  $\mathbb{1}$ . A “bootstrapping” argument then shows that, for a (possibly) still smaller  $\varepsilon > 0$ , no two things in  $\mathcal{O}_{\omega, \varepsilon}$  are gauge equivalent by any gauge transformation. For such an  $\varepsilon > 0$ ,  $\mathcal{O} = \mathcal{O}_{\omega, \varepsilon}$  projects injectively into  $\hat{\mathcal{B}}_3(P_k)$  and so is our slice and provides a local manifold structure for  $\hat{\mathcal{B}}_3(P_k)$  near  $[\omega]$ . In particular,  $\hat{\mathcal{B}}_3(P_k)$  has the structure of a smooth Hilbert manifold.

If  $\omega \in \mathcal{C}_3(P_k)$  is reducible the analysis is similar except that to get an injective projection into the moduli space one must first factor out the action of the stabilizer of  $\omega$ . More precisely, defining  $\mathcal{O}_{\omega, \varepsilon}$  as in (4.12) and  $\widetilde{\text{stab}}(\omega) = \text{stab}(\omega)/\mathbb{Z}_2 \cong U(1)$  one finds that, for sufficiently small  $\varepsilon > 0$ , the projection

$$\mathcal{O}_{\omega, \varepsilon}/\widetilde{\text{stab}}(\omega) \longrightarrow \mathcal{B}_3(P_k)$$

is a homeomorphism onto an open neighborhood of  $[\omega]$  in  $\mathcal{B}_3(P_k)$  which, in fact, is a diffeomorphism outside the fixed point set of  $\text{stab}(\omega)$ . There are generally singularities, where there is no local smooth structure, at the images of these fixed points (e.g., these account for the cones in our picture of the moduli space used in the proof of Donaldson’s 1983 Theorem).

Now, the objects of real interest in Donaldson theory are certain subspaces of  $\mathcal{B}_3(P_k)$  and  $\hat{\mathcal{B}}_3(P_k)$  which we now introduce. Begin by selecting some Riemannian metric  $g$  on  $B$ . Together with the orientation of  $B$  this gives a Hodge star operation  $*$  on smooth forms defined on  $B$ . Since  $B$  is four-dimensional,  $*$  :  $\Omega^2(B, \text{ad}P_k) \longrightarrow \Omega^2(B, \text{ad}P_k)$  and, since the elements of  $\Omega_2^2(B, \text{ad}P_k)$  are

continuous, this extends to

$$* : \Omega_2^2(B, \text{ad}P_k) \longrightarrow \Omega_2^2(B, \text{ad}P_k).$$

The curvature map  $\mathbf{F} : \mathcal{C}(P_k) \longrightarrow \Omega^2(B, \text{ad}P_k)$ ,  $\omega \rightarrow \mathbf{F}_\omega$ , also extends to a smooth map

$$\begin{aligned} \mathbf{F} : \mathcal{C}_3(P_k) &\longrightarrow \Omega_2^2(B, \text{ad}P_k) \\ \omega &\longrightarrow \mathbf{F}_\omega \end{aligned}$$

so we may say that an  $\omega \in \mathcal{C}_3(P_k)$  is ***g-anti-self-dual (g-ASD)*** if

$$*\mathbf{F}_\omega = -\mathbf{F}_\omega.$$

**Remark.** *The Chern number  $k$  of our bundle can be written as*

$$\begin{aligned} k = c_2(P_k)[B] &= \frac{1}{8\pi^2} \int_B \text{tr}(\mathbf{F}_\omega \wedge \mathbf{F}_\omega) \\ &= \frac{1}{8\pi^2} \int_B \left( \|\mathbf{F}_\omega^-\|^2 - \|\mathbf{F}_\omega^+\|^2 \right) \text{vol} \end{aligned}$$

where  $\mathbf{F}_\omega^\pm = \frac{1}{2}(\mathbf{F}_\omega \pm *\mathbf{F}_\omega)$  are the self-dual and anti-self-dual parts of  $\mathbf{F}_\omega$ . Consequently, when  $k < 0$  we must have  $\mathbf{F}_\omega^+ \neq 0$  and anti-self-dual connections cannot exist. When  $k = 0$  it is possible for anti-self-dual connections  $\omega$  to exist, but they must be flat ( $\mathbf{F}_\omega = 0$ ) because  $\mathbf{F}_\omega^+ = 0$  and  $k = 0$  gives  $\mathbf{F}_\omega^- = 0$  and  $\mathbf{F}_\omega = \mathbf{F}_\omega^+ + \mathbf{F}_\omega^-$ . We will see shortly that these are not particularly interesting and this will account for our restriction to bundles with  $k > 0$ .

Now we define

$$\text{Asd}_3(P_k, \mathfrak{g}) = \{\omega \in \mathcal{C}_3(P_k); \omega \text{ is } \mathfrak{g}\text{-ASD}\}$$

and

$$\widehat{\text{Asd}}_3(P_k, \mathfrak{g}) = \left\{ \omega \in \widehat{\mathcal{C}}_3(P_k); \omega \text{ is } \mathfrak{g}\text{-ASD} \right\}$$

and the corresponding **moduli spaces**

$$\mathcal{M}(P_k, \mathfrak{g}) = \text{Asd}_3(P_k, \mathfrak{g}) / \mathcal{G}_4(P_k)$$

$$\widehat{\mathcal{M}}(P_k, \mathfrak{g}) = \widehat{\text{Asd}}_3(P_k, \mathfrak{g}) / \mathcal{G}_4(P_k)$$

**of  $g$ -ASD and irreducible  $g$ -ASD connections.** Donaldson theory is built on the analysis of these moduli spaces.

**Remark.** *Note that the Sobolev indices have been dropped on  $\mathcal{M}(P_k, \mathfrak{g})$  and  $\hat{\mathcal{M}}(P_k, \mathfrak{g})$ . The reason is that, for any  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$ , elliptic regularity implies that there is an  $f \in \mathcal{G}_4(P_k)$  such that  $\omega \cdot f$  is a smooth connection (see Section 5 of [15]). Thus, these moduli spaces do not depend on the choice of (sufficiently large) Sobolev index.*

For a given  $B$ ,  $\mathfrak{g}$  and  $k$ ,  $\text{Asd}_3(P_k, \mathfrak{g})$  (and therefore  $\mathcal{M}(P_k, \mathfrak{g})$ ) might well be empty. This is the case, for example, when  $B$  is either  $\mathbb{S}^2 \times \mathbb{S}^2$  or  $\mathbb{C}\mathbb{P}^2$  with their standard orientations and metrics (Fubini-Study in the case of  $\mathbb{C}\mathbb{P}^2$ ) and  $k = 1$ . Changing the orientation of the manifold can have a dramatic effect, e.g., the  $k = 1$  bundle over  $\overline{\mathbb{C}\mathbb{P}^2}$  (also with the Fubini-Study metric) has a moduli space of ASD connections that one can describe as explicitly as we did for  $\mathbb{S}^4$  in Section 1 (for more details on this and many more examples, see [12]). A general result of considerable interest was proved by Taubes [41]. Through an ingenious “grafting” procedure using the  $k = 1$  instantons on  $\mathbb{S}^4$  described in Section 1 he was able to prove that the  $k = 1$  bundle over any  $B$  with  $b_2^+(B) = 0$  admits  $g$ -ASD connections for any Riemannian metric  $g$  (the definition of  $b_2^+(B)$  follows).

**Remark.** *Since it will play a recurrent role from this point on we recall the definition of  $b_2^+(B)$  for a compact, simply connected, oriented, smooth four-manifold  $B$ . In Section 1 we introduced the intersection form*

$$Q_B : H_2(B, \mathbb{Z}) \times H_2(B, \mathbb{Z}) \longrightarrow \mathbb{Z}.$$

*It is an integer-valued, symmetric, bilinear form on the finitely-generated, free abelian group  $H_2(B, \mathbb{Z})$ . If  $b_2(B)$  is the rank of  $H_2(B, \mathbb{Z})$ , then one can write*

$$b_2(B) = b_2^+(B) + b_2^-(B)$$

*where  $b_2^+(B)$  ( $b_2^-(B)$ ) is the maximal dimension of a subspace of  $H_2(B, \mathbb{Z})$  on which  $Q_B$  is positive (negative) definite. One can show that  $b_2^+(B)$  ( $b_2^-(B)$ ) is also the dimension of the space of self-dual (anti-self-dual) harmonic two-forms on  $B$  (for any choice of a Riemannian metric on  $B$ ) and this accounts for the role it plays in the study of ASD connections.*

Before proceeding with the study of  $\mathcal{M}(P_k, \mathfrak{g})$  and  $\hat{\mathcal{M}}(P_k, \mathfrak{g})$  for  $k > 0$  we explain our earlier comment that the  $k = 0$  case is “not particularly interesting” ( $k < 0$  is definitely not interesting since the moduli spaces are empty). The

$k = 0$  bundle  $SU(2) \hookrightarrow P_0 \rightarrow B$  is trivial and, as we observed earlier, any ASD connection on it is necessarily flat. Conversely, any flat connection is certainly ASD ( $F_\omega = 0$  implies  $F_\omega^+ = 0$ ). Since flat connections exist on any trivial bundle (page 92 of Vol I of [24]), the moduli space  $\mathcal{M}(P_0, \mathfrak{g})$  is nonempty (for any  $\mathfrak{g}$ ). Since  $B$  is simply connected, any two flat connections on  $B$  are gauge equivalent (Proposition 2.2.3 of Vol I of [24]) so  $\mathcal{M}(P_0, \mathfrak{g})$  is, in fact, just a single point.

Now we return to the general study of the moduli spaces  $\mathcal{M}(P_k, \mathfrak{g})$  and  $\hat{\mathcal{M}}(P_k, \mathfrak{g})$ . For this we consider the smooth map

$$\text{pr}_+ \circ \mathbf{F} : \mathcal{C}_3(P_k) \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k)$$

where  $\mathbf{F}$  is the curvature map ( $\mathbf{F}(\omega) = F_\omega$ ) and  $\text{pr}_+$  projects onto the self-dual part. Thus,

$$(\text{pr}_+ \circ \mathbf{F})(\omega) = F_\omega^+$$

and

$$\text{Asd}_3(P_k, \mathfrak{g}) = (\text{pr}_+ \circ \mathbf{F})^{-1}(0). \tag{4.14}$$

At any  $\omega \in \mathcal{C}_3(P_k, \mathfrak{g})$  the derivative of this map can be identified with

$$d_+^\omega = \text{pr}_+ \circ d^\omega : \Omega_3^1(B, \text{ad}P_k) \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k)$$

(see page 54 of [15]). Now, we have already observed that, in general,  $d^\omega \circ d^\omega$  is not zero, but it follows from (4.3) that, when  $\omega$  is ASD,

$$d_+^\omega \circ d^\omega = [F_\omega^+, \cdot] = [0, \cdot] = 0$$

so

$$\text{Im}(d^\omega) = d^\omega(\Omega_4^0(B, \text{ad}P_k)) \subseteq \ker(d_+^\omega). \tag{4.15}$$

Thus, we have associated with every  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$  a complex  $\mathcal{E}(\omega)$

$$0 \longrightarrow \Omega_4^0(B, \text{ad}P_k) \begin{array}{c} \xrightarrow{d^\omega} \\ \xleftarrow{\delta^\omega} \end{array} \Omega_3^1(B, \text{ad}P_k) \begin{array}{c} \xrightarrow{d_+^\omega} \\ \xleftarrow{\delta_+^\omega} \end{array} \Omega_{+,2}^2(B, \text{ad}P_k) \longrightarrow 0$$

where we have included also the adjoints  $\delta^\omega$  and  $\delta_+^\omega$  of  $d^\omega$  and  $d_+^\omega$ , respectively. This complex is, in fact, elliptic and the entire analysis of the local structure of the moduli space near  $[\omega]$  rests on an analysis of the structure of  $\mathcal{E}(\omega)$  (the so-called **fundamental elliptic complex** associated with  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$ ). We begin by simply enumerating some consequences of the generalized Hodge Decomposition Theorem for elliptic complexes.

## 1) The Laplacians

$$\begin{aligned}\Delta_0^\omega &= \delta^\omega \circ d^\omega \\ \Delta_1^\omega &= d^\omega \circ \delta^\omega + \delta_+^\omega \circ d_+^\omega \\ \Delta_2^\omega &= d_+^\omega \circ \delta_+^\omega\end{aligned}$$

are all self-adjoint, elliptic operators.

2) The spaces  $\ker(\Delta_k^\omega)$ ,  $k = 0, 1, 2$ , of harmonic forms are finite-dimensional and consist of smooth forms (smoothness follows from “elliptic regularity”).

3) Each of the cohomology groups

$$\begin{aligned}H^0(\omega) &= \ker(d^\omega) \\ H^1(\omega) &= \ker(d_+^\omega) / \text{im}(d^\omega) \\ H^2(\omega) &= \Omega_{+,2}^2(B, \text{ad}P_k) / \text{im}(d_+^\omega)\end{aligned}$$

associated with  $\mathcal{E}(\omega)$  contains a unique harmonic representative. In particular,

$$H^k(\omega) \cong \ker(\Delta_k^\omega), \quad k = 0, 1, 2$$

so all of these cohomology groups are finite-dimensional and we may define the **index** of the complex  $\mathcal{E}(\omega)$  by

$$\begin{aligned}\text{Ind}(\mathcal{E}(\omega)) &= \dim(H^0(\omega)) - \dim(H^1(\omega)) + \dim(H^2(\omega)) \\ &= \dim(\ker(\Delta_0^\omega)) - \dim(\ker(\Delta_1^\omega)) + \dim(\ker(\Delta_2^\omega)).\end{aligned}$$

4) There are orthogonal decompositions

$$\begin{aligned}\Omega_4^0(B, \text{ad}P_k) &\cong \text{im}(\delta^\omega) \oplus \ker(d^\omega) \\ \Omega_3^1(B, \text{ad}P_k) &\cong \text{im}(d^\omega) \oplus \ker(\delta^\omega) \\ \Omega_{+,2}^2(B, \text{ad}P_k) &\cong \text{im}(d_+^\omega) \oplus \ker(\delta_+^\omega).\end{aligned}$$

Now we put all of this information to use in the following way. Fix an element  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$  and restrict the map  $\text{pr}_+ \circ \mathbf{F}$  to one of the sets  $\mathcal{O}_{\omega, \varepsilon}$  (see (4.12))

$$\text{pr}_+ \circ \mathbf{F} \big|_{\mathcal{O}_{\omega, \varepsilon}} : \mathcal{O}_{\omega, \varepsilon} \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k).$$

Then

$$(\text{pr}_+ \circ \mathbf{F})^{-1}(0) = \text{Asd}_3(P_k, \mathfrak{g}) \cap \mathcal{O}_{\omega, \varepsilon} \quad (4.16)$$

and the derivative at  $\omega$  is

$$d_+^\omega \big|_{\ker(\delta^\omega)} : \ker(\delta^\omega) \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k). \quad (4.17)$$



Before proceeding we recall a few facts from analysis.

**Remark.** Recall that a bounded linear map  $T : H_1 \rightarrow H_2$  between two Hilbert spaces is said to be **Fredholm** if either of the following two equivalent conditions is satisfied:

- a)  $\dim(\ker T) < \infty$ ,  $\dim(\ker T^*) < \infty$  and  $\text{im}T$  is closed.
- b)  $H_1 \cong \ker T \oplus \text{im}T^*$  and  $H_2 \cong \ker T^* \oplus \text{im}T$ .

We will soon appeal to the following infinite-dimensional version of the **Implicit Function Theorem**: Let  $X$  and  $Y$  be Hilbert manifolds,  $F : X \rightarrow Y$  a smooth map and  $x_0 \in X$  a point at which the derivative  $Df_{x_0} : T_{x_0}(X) \rightarrow T_{F(x_0)}(Y)$  is a surjective Fredholm map. Then there exists an open neighborhood of  $x_0$  in  $F^{-1}(F(x_0))$  that is a smooth (finite-dimensional) manifold of dimension

$$\dim(\ker(Df_{x_0})).$$

We will show that the map  $d_+^\omega|_{\ker(\delta^\omega)}$  is always Fredholm and determine conditions under which it is surjective, thus setting up an application of the Implicit Function Theorem to obtain a smooth manifold structure for  $\text{Asd}_3(P_k, \mathbf{g}) \cap \mathcal{O}_{\omega, \varepsilon}$  near  $\omega$ . If, in addition, the projection of  $\text{Asd}_3(P_k, \mathbf{g}) \cap \mathcal{O}_{\omega, \varepsilon}$  into the moduli space is injective (i.e., if  $\omega$  is irreducible) this will give a finite-dimensional smooth manifold structure for the moduli space near  $[\omega]$ .

To see that  $d_+^\omega|_{\ker(\delta^\omega)}$  is Fredholm we reason as follows. First, we have that  $\ker(d_+^\omega|_{\ker(\delta^\omega)}) = \ker(d_+^\omega)/\text{im}(d^\omega) = H^1(\omega)$  by 4) and 3) so this is finite-dimensional by 3). Next observe that  $(d_+^\omega|_{\ker(\delta^\omega)})^* = \delta_+^\omega|_{\text{im}(d_+^\omega|_{\ker(\delta^\omega)})} = \delta_+^\omega|_{\text{im}(d_+^\omega)}$  so  $\ker((d_+^\omega|_{\ker(\delta^\omega)})^*) = \ker(\delta_+^\omega|_{\text{im}(d_+^\omega)})$  which is finite-dimensional because  $\ker(\delta_+^\omega) = H^2(\omega)$  is finite-dimensional. Finally,  $\text{im}(d_+^\omega|_{\ker(\delta^\omega)}) = \text{im}(d_+^\omega) = (\ker(\delta_+^\omega))^\perp$  by 4) so this is closed. We have verified the requirements of a) in the Remark above so  $d_+^\omega|_{\ker(\delta^\omega)}$  is Fredholm.

Now we determine when the map  $d_+^\omega|_{\ker(\delta^\omega)}$  is surjective. Since  $d_+^\omega$  acts on  $\Omega_3^1(B, \text{ad}P_k) \cong \text{im}(d^\omega) \oplus \ker(\delta^\omega)$  and since  $d_+^\omega$  vanishes identically on  $\text{im}(d^\omega)$  by (4.15),  $d_+^\omega|_{\ker(\delta^\omega)}$  is surjective if and only if  $d_+^\omega$  is surjective, i.e., if and only if  $\text{im}(d_+^\omega) = \Omega_{+,2}^2(B, \text{ad}P_k)$ . But, by 3), this is the case if and only if  $H^2(\omega)$  is trivial.

Noting that  $H^0(\omega)$  is trivial if and only if  $\ker(d^\omega)$  is trivial and, by Theorem 4.3 c), this is the case if and only if  $\omega$  is irreducible we arrive at the following interpretations of the cohomology groups of  $\mathcal{E}(\omega)$ .

$$H^0(\omega) = 0 \iff \omega \text{ is irreducible} \tag{4.18}$$

$$\begin{aligned} H^1(\omega) &= \ker(d_+^\omega|_{\ker(\delta^\omega)}) \\ &= \text{kernel of the derivative of } \text{pr}_+ \circ \mathbf{F}|_{\mathcal{O}_{\omega,\varepsilon}} \text{ at } \omega \end{aligned} \quad (4.19)$$

$$H^2(\omega) = 0 \iff d_+^\omega|_{\ker(\delta^\omega)} \text{ is surjective.} \quad (4.20)$$

Recalling that when  $\omega$  is irreducible the projection into the moduli space is injective near  $\omega$  we can summarize all of this in the following theorem.

**Theorem 4.4.** *At  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$  the map*

$$\text{pr}_+ \circ \mathbf{F}|_{\mathcal{O}_{\omega,\varepsilon}} : \mathcal{O}_{\omega,\varepsilon} \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k)$$

*has derivative*

$$d_+^\omega|_{\ker(\delta^\omega)} : \ker(\delta^\omega) \longrightarrow \Omega_{+,2}^2(B, \text{ad}P_k)$$

*that is Fredholm. The derivative is surjective if and only if  $H^2(\omega) = 0$  and, in this case,*

$$(\text{pr}_+ \circ \mathbf{F}|_{\mathcal{O}_{\omega,\varepsilon}})^{-1}(0) = \text{Asd}_3(P_k, \mathfrak{g}) \cap \mathcal{O}_{\omega,\varepsilon}$$

*is a smooth manifold of dimension*

$$\dim(\ker(d_+^\omega|_{\ker(\delta^\omega)})) = \dim(H^1(\omega))$$

*near  $\omega$ . If, in addition,  $H^0(\omega) = 0$ , then the projection into the moduli space  $\mathcal{M}(P_k, \mathfrak{g})$  gives a chart of dimension  $\dim(H^1(\omega))$  near  $[\omega]$ .*

Notice that if  $H^0(\omega) = 0$  and  $H^2(\omega) = 0$  are both trivial, then  $\text{Ind}(\mathcal{E}(\omega)) = -\dim(H^1(\omega))$  so, if we could calculate the index of the elliptic complex  $\mathcal{E}(\omega)$ , we would have (minus) the dimension of the moduli space near  $[\omega]$ . The Atiyah-Singer Index Theorem gives the index of  $\mathcal{E}(\omega)$  in terms of topological data on  $B$  and  $SU(2) \hookrightarrow P_k \rightarrow B$ . In our present context the result is

$$\text{Ind}(\mathcal{E}(\omega)) = -8k + 3(1 + b_2^+(B)) \quad (4.21)$$

(see pages 267-271 of [12]). Note, in particular, that the result is independent of  $\omega$  so we obtain the following consequence of Theorem 4.4.

**Corollary 4.5.** *If  $H^0(\omega) = 0$  and  $H^2(\omega) = 0$  for every  $\omega \in \text{Asd}_3(P_k, \mathfrak{g})$ , then the moduli space  $\mathcal{M}(P_k, \mathfrak{g}) = \hat{\mathcal{M}}(P_k, \mathfrak{g})$  is a smooth manifold of dimension  $8k - 3(1 + b_2^+(B))$ .*

For a given choice of the Riemannian metric  $g$  it may or may not be the case that  $H^0(\omega) = 0$  and  $H^2(\omega) = 0$  for every  $\omega \in \text{Asd}_3(P_k, g)$ . We will see shortly, however, that for “almost all” choices of  $g$ ,  $H^2(\omega)$  will be trivial for all  $\omega \in \text{Asd}_3(P_k, g)$  and, with one additional restriction on the topology of  $B$ , the same is true  $H^0(\omega)$ . First, however, we note that if  $\omega \in \text{Asd}_3(P_k, g)$  and  $H^2(\omega) = 0$ , then the slice  $\text{Asd}_3(P_k, g) \cap \mathcal{O}_{\omega, \varepsilon}$  still has a local manifold structure near  $\omega$ , but if  $H^0(\omega) \neq 0$  one can only obtain a one-to-one projection into the moduli space by first factoring out the action of the stabilizer of  $\omega$ . The consequence is that, near  $[\omega]$ ,  $\mathcal{M}(P_k, g)$  is not smooth but has a neighborhood homeomorphic to this quotient, which turns out to be a cone over  $\mathbb{C}\mathbb{P}^{4k-2}$  with  $[\omega]$  at the vertex. Reducible connections, when they exist, give rise to cone-like “singularities” in the moduli space.

Next we will require a brief discussion of various “generic metrics” theorems which assert that, under certain circumstances, “almost all” choices for the Riemannian metric  $g$  give rise to “nice” moduli spaces of ASD connections. Begin by considering the space  $\mathcal{R}$  of all Riemannian metrics on  $B$ . This is a space of sections of a fiber bundle over  $B$  and can be given the structure of a (pathwise connected) Hilbert manifold. With this structure one can show that

1. There is a dense  $G_\delta$ -set in  $\mathcal{R}$  such that, for every  $g$  in this set, any  $g$ -ASD connection  $\omega$  on  $P_k$ ,  $k > 0$ , satisfies  $H^2(\omega) = 0$ .
2. If  $b_2^+(B) > 0$ , then there is a dense  $G_\delta$ -set in  $\mathcal{R}$  such that, for every  $g$  in this set, any  $g$ -ASD connection  $\omega$  on  $P_k$ ,  $k > 0$ , satisfies  $H^2(\omega) = 0$  and  $H^0(\omega) = 0$ .

In short, for “generic”  $g$ ,  $\mathcal{M}(P_k, g) = \hat{\mathcal{M}}(P_k, g)$  is (either empty or) a smooth manifold of dimension  $8k - 3(1 + b_2^+(B))$ .

We will not attempt to sketch proofs of these last two results, but, very roughly, here is how one might go about showing that  $\hat{\mathcal{M}}(P_k, g)$  is smooth for a generic choice of  $g$ . Consider the so-called parametrized moduli space

$$\hat{\mathcal{M}}(P_k, \mathcal{R}) = \left\{ ([\omega], g) \in \hat{\mathcal{B}}_3(P_k) \times \mathcal{R} ; \omega \text{ is } g\text{-ASD} \right\}.$$

This is an infinite dimensional smooth submanifold of  $\hat{\mathcal{B}}_3(P_k) \times \mathcal{R}$ . One shows that the projection map

$$\hat{\mathcal{M}}(P_k, \mathcal{R}) \longrightarrow \mathcal{R}$$

is smooth with Fredholm derivative at each point. The Sard-Smale Theorem (infinite dimensional version of Sard’s Theorem) implies that the set of regular values

of the projection is a dense  $G_\delta$ -set in  $\mathcal{R}$ . But then, for any  $\mathbf{g}$  in this set, the inverse image

$$\left(\hat{\mathcal{B}}_3(P_k) \times \{\mathbf{g}\}\right) \cap \hat{\mathcal{M}}(P_k, \mathcal{R}) = \hat{\mathcal{M}}(P_k, \mathbf{g})$$

is a smooth submanifold of  $\hat{\mathcal{M}}(P_k, \mathcal{R})$ .

**Remark.** *The restriction on  $b_2^+(B)$  arises because the subset of  $\mathcal{R}$  consisting of those  $\mathbf{g}$  for which reducible  $\mathbf{g}$ -ASD connections on  $P_k$  exist is a countable union of smooth submanifolds of codimension  $b_2^+(B)$ . If  $b_2^+(B) = 0$ , then reducibles are generically unavoidable. As we saw in Section 1 this is a good, not a bad thing as it leads to Donaldson's theorem on definite intersection forms.*

Crudely put, the idea behind defining the Donaldson polynomial invariants is to regard the moduli spaces  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  as cycles over which to integrate certain carefully selected differential forms. In order to carry out such a program these moduli spaces must be orientable and, if the result is to be a differential topological invariant, the integrals must be independent of the choice of (generic)  $\mathbf{g}$ . We now record two results that guarantee this.

**Theorem 46.** *Suppose  $\mathbf{g}$  is a Riemannian metric on  $B$  for which  $H^2(\omega) = 0$  for every  $\mathbf{g}$ -ASD connection  $\omega$  on  $P_k$ ,  $k > 0$ . Then the moduli space  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  is orientable. An orientation for  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  can be uniquely specified by choosing an orientation for  $B$  and an orientation for the vector space  $H_+^2(B, \mathbb{R})$  of self-dual two-forms on  $B$ .*

The proof amounts to constructing an explicit model for the determinant line bundle (top exterior power) of  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  from a family of differential operators on  $B$  and exhibiting a nonzero section (see Sections 5.4.1 and 7.1.6 of [12]).

To state the final result of this section we consider two metric  $\mathbf{g}_0$  and  $\mathbf{g}_1$  in the dense  $G_\delta$ -set in  $\mathcal{R}$  on which  $H^2(\omega) = 0$  for all  $\mathbf{g}$ -ASD connections. Since  $\mathcal{R}$  is pathwise connected we can join them with a path  $\{\mathbf{g}_t; 0 \leq t \leq 1\}$  in  $\mathcal{R}$ . Define the parametrized moduli space

$$\hat{\mathcal{M}}(P_k, \{\mathbf{g}_t\}) = \left\{([\omega], t) \in \hat{\mathcal{B}}_3(P_k) \times [0, 1]; [\omega] \in \hat{\mathcal{M}}(P_k, \mathbf{g}_t)\right\}.$$

**Theorem 47.** *If  $\mathbf{g}_0$  and  $\mathbf{g}_1$  are in the dense  $G_\delta$ -set in  $\mathcal{R}$  on which  $H^2(\omega) = 0$  for all  $\mathbf{g}$ -ASD connections  $\omega$ , then, for a generic path  $\{\mathbf{g}_t; 0 \leq t \leq 1\}$  joining them,  $\hat{\mathcal{M}}(P_k, \{\mathbf{g}_t\})$  is a smooth, orientable submanifold of  $\hat{\mathcal{B}}_3(P_k) \times [0, 1]$  with boundary. A choice of orientation  $\mu$  for  $H_+^2(B, \mathbb{R})$  determines an orientation for  $\hat{\mathcal{M}}(P_k, \{\mathbf{g}_t\})$ . Moreover, the oriented boundary of  $\hat{\mathcal{M}}(P_k, \{\mathbf{g}_t\})$  is the disjoint*

union of  $\hat{\mathcal{M}}(P_k, \mathbf{g}_1)$  with the orientation induced by  $\mu$  and  $\hat{\mathcal{M}}(P_k, \mathbf{g}_0)$  with the orientation opposite to that induced by  $\mu$ .

In short, a generic variation of  $\mathbf{g}$  varies  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  within a single homology class. Even if  $\hat{\mathcal{M}}(P_k, \mathbf{g}_0)$  and  $\hat{\mathcal{M}}(P_k, \mathbf{g}_1)$  are smooth manifolds one cannot, in general, arrange that the intermediate moduli spaces  $\hat{\mathcal{M}}(P_k, \mathbf{g}_t)$  are all smooth. There may be finitely many values of  $t$  for which one encounters reducible connections. Even this can be avoided, however, if one is willing to assume of  $B$  that  $b_2^+(B) > 1$ . In this case a generic path  $\{\mathbf{g}_t\}$  from  $\mathbf{g}_0$  to  $\mathbf{g}_1$  has the property that each  $\hat{\mathcal{M}}(P_k, \mathbf{g}_t)$  is a smooth, orientable manifold.

With the apparatus we have now assembled one is almost (but not quite) in a position to begin building the Donaldson invariants. The remaining obstacle (and it is a serious one) is that the moduli spaces  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  are generally not compact so one cannot integrate over them (more precisely, they do not determine a fundamental homology class with which to pair cohomology classes). One overcomes this obstacle by replacing  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  with what is known as its ‘‘Uhlenbeck compactification’’. We intend to not overcome, but circumvent the obstacle by considering only a special case (an outline of the general situation is available in Section 8).

Notice that, by an appropriate arrangement of  $b_2^+(B)$  and the Chern number  $k$ , it is entirely possible for the dimension  $8k - 3(1 + b_2^+(B))$  of the generic moduli space  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  to come out just zero. In this case,  $\hat{\mathcal{M}}(P_k, \mathbf{g})$  is a 0-dimensional, oriented manifold (given an orientation of  $H_+^2(B, \mathbb{R})$ ), i.e., it is a set of isolated points  $[\omega]$  each equipped with a sign which we will write  $(-1)^{[\omega]}$ . As it happens, the moduli space is necessarily compact in this case so we can add these signs to obtain an integer

$$\gamma_0(B) = \sum_{[\omega] \in \hat{\mathcal{M}}(P_k, \mathbf{g})} (-1)^{[\omega]}. \tag{4.22}$$

One can show (from the homology result Theorem 4.7 and the remarks following it) that if  $b_2^+(B) > 1$ , then this integer does not depend on the choice of (generic) metric  $\mathbf{g}$  and is, in fact, an orientation preserving diffeomorphism invariant of  $B$  (assuming the orientation of  $H_+^2(B, \mathbb{R})$  is fixed). Under all of these circumstances the integer  $\gamma_0(B)$  is called the **zero-dimensional Donaldson invariant** of  $B$ .

Our concern here is not with calculating  $\gamma_0(B)$  nor with using it to obtain topological information (for a nontrivial calculation and application see Section 9.1 of [12]). Rather we would like to show that, by adopting a slightly different perspective,  $\gamma_0(B)$  is in many ways analogous to an Euler characteristic. This, in turn, will eventually lead us to formulas for  $\gamma_0(B)$  that evolve into the partition func-

tion for the topological quantum field theory of Witten referred to at the beginning of this section.

Before describing our new perspective on  $\gamma_0(B)$  we will briefly review some standard material on the Euler number of a vector bundle. We consider an oriented, real vector bundle  $\pi_E : E \rightarrow X$  of even rank (fiber dimension)  $2k$  over a compact, oriented manifold  $X$  of dimension  $2k$ , e.g., the tangent bundle of a compact, oriented  $2k$ -manifold. The typical fiber of  $E$  will generally be denoted  $V$  and will usually come equipped with a positive definite inner product (e.g., from a fiber metric on  $E$ ). The exterior algebra of  $V$  will be denoted  $\bigwedge V = \bigoplus_{p=0}^{\infty} \bigwedge^p V$  and should be thought of as the graded algebra of polynomials with real coefficients in the odd (anti-commuting) variables  $\psi^1, \dots, \psi^{2k}$ , where  $\{\psi^1, \dots, \psi^{2k}\}$  is some fixed oriented, orthonormal basis for  $V$ . The volume form for  $V$  corresponding to  $\{\psi^1, \dots, \psi^{2k}\}$  is **vol**  $= \psi^1 \cdots \psi^{2k}$ , where we omit the customary wedge  $\wedge$  and write the product in  $\bigwedge V$  by juxtaposition. There are a number of ways to approach the definition of the Euler number of the vector bundle  $E$ , several of which will be important to us. We proceed as follows. The **Euler number**  $\chi(E)$  of the vector bundle  $\pi_E : E \rightarrow X$  is defined by

$$\chi(E) = \int_X e(E) \tag{4.23}$$

where  $e(E) \in H^2(X, \mathbb{R})$  is the ‘‘Euler class’’ of  $E$ , for which we offer two (equivalent) definitions.

The Euler class of  $E$  can be defined by the Chern-Weil procedure in a manner entirely analogous to the familiar definitions of Chern and Pontryagin classes. We briefly review the ideas behind this procedure. Let  $G \hookrightarrow P \xrightarrow{\pi_P} X$  be a principal  $G$ -bundle with connection  $\omega$  and curvature  $\Omega$ . Let  $\{\xi_1, \dots, \xi_n\}$  be a basis for the Lie algebra  $\mathcal{G}$  of  $G$  and write  $\omega = \omega^a \xi_a$  and  $\Omega = \Omega^a \xi_a$ , where  $\omega^a \in \Omega^1(P)$  and  $\Omega^a \in \Omega^2(P)$ ,  $a = 1, \dots, n$ , are real-valued forms. Let  $\mathbb{C}[\mathcal{G}]^G$  be the algebra of complex-valued polynomials  $\mathcal{P}$  on  $\mathcal{G}$  that are  $\text{ad}G$ -invariant ( $\mathcal{P}(g^{-1}\xi g) = \mathcal{P}(\xi)$  for all  $g \in G$  and  $\xi \in \mathcal{G}$ ). One can realize  $\mathbb{C}[\mathcal{G}]^G$  concretely as follows: Let  $\{x^1, \dots, x^n\}$  be the basis for  $\mathcal{G}^*$  dual to  $\{\xi_1, \dots, \xi_n\}$  and think of the  $x^i$  as linear functions on  $\mathcal{G}$ . The symmetric algebra  $S(\mathcal{G}^*)$  can be identified with the polynomial algebra  $\mathbb{R}[x^1, \dots, x^n]$  and tensoring with  $\mathbb{C}$  gives  $\mathbb{C}[x^1, \dots, x^n]$  from which we select the  $\text{ad}$ -invariant elements to get  $\mathbb{C}[\mathcal{G}]^G$ . Next we denote by  $\Omega^*(P)_{BAS}$  the graded algebra of real-valued forms on  $P$  that are **basic**, i.e.,  $G$ -invariant ( $\sigma_g^* \varphi = \varphi$  for every  $g \in G$ , where  $\sigma_g(p) = p \cdot g$ ), and **horizontal** ( $\iota_W \varphi = 0$  for all vertical vector fields  $W$ , where  $\iota_W$  denotes interior multiplication by  $W$ ). These are precisely the forms on  $P$  which descend to  $X$ , i.e., for which there is a

$\bar{\varphi} \in \Omega^*(X)$  such that  $\pi_P^* \bar{\varphi} = \varphi$ . Now there is a map

$$CW_\omega : \mathbb{C}[\mathcal{G}]^G \longrightarrow \Omega^*(P)_{BAS}$$

called the **Chern-Weil homomorphism**, defined by “evaluating the polynomial on the curvature of  $\omega$ ”. More precisely, if  $\mathcal{P} \in \mathbb{C}[\mathcal{G}]^G$  has degree  $k$  and if we denote by  $\mathcal{P}$  also the corresponding  $k$ -multilinear map on  $\mathcal{G}$  (obtained by polarization), then

$$\begin{aligned} CW_\omega(\mathcal{P}) &= \mathcal{P}(\Omega) = \mathcal{P}(\Omega^{a_1} \xi_{a_1}, \dots, \Omega^{a_k} \xi_{a_k}) \\ &= \mathcal{P}(\xi_{a_1}, \dots, \xi_{a_k}) \Omega^{a_1} \wedge \dots \wedge \Omega^{a_k}. \end{aligned}$$

Being basic,  $\mathcal{P}(\Omega)$  is the pullback by  $\pi_P$  of a form  $\bar{\mathcal{P}}(\Omega)$  on  $X$  which can be shown to be closed and whose (deRham) cohomology class  $[\bar{\mathcal{P}}(\Omega)] \in H^{2k}(X, \mathbb{R})$  does not depend on the choice of  $\omega$ . Making specific choices for  $\mathcal{P}$  gives rise to various characteristic classes of the bundle. For example, if  $G = U(1)$ , then  $\frac{i}{2\pi}[\text{tr}(\Omega)]$  is the first Chern class and, if  $G = SU(2)$ , then  $-\frac{1}{8\pi^2}[\text{tr}(\Omega \wedge \Omega)]$  is the second Chern class (see Chapter XII, Vol. II, of [24] for more details).

Now, to define the Euler class of our vector bundle  $\pi_E : E \rightarrow X$  by this procedure we will require a principal bundle and an invariant polynomial. For the former we select a fiber metric on  $\pi_E : E \rightarrow X$  and consider the corresponding oriented, orthonormal frame bundle

$$SO(2k) \hookrightarrow F_{SO}(E) \xrightarrow{\pi_{SO}} X. \tag{4.24}$$

For the  $SO(2k)$ -invariant polynomial we select the **Pfaffian**

$$Pf : \mathfrak{so}(2k) \longrightarrow \mathbb{R}$$

defined as follows: To each skew-symmetric matrix  $Q = (q_{ij}) \in \mathfrak{so}(2k)$  we associate an element

$$\sum_{i < j} q_{ij} \psi^i \psi^j = \frac{1}{2} \psi^T Q \psi$$

in  $\wedge^2 V$  (here “ $T$ ” means transpose and  $\psi = (\psi^1 \dots \psi^{2k})^T$ ). Then  $(\frac{1}{2} \psi^T Q \psi)^k$  is in  $\wedge^{2k} V$  and so it is just a multiple of the volume form  $\mathbf{vol} = \psi^1 \dots \psi^{2k}$ .

We define  $Pf(Q)$  by

$$\frac{1}{k!} \left( \frac{1}{2} \psi^T Q \psi \right)^k = Pf(Q) \mathbf{vol}. \tag{4.25}$$

One can show that  $Pf$  is  $\text{ad}(SO(2k))$ -invariant (in fact, it is a square root of the determinant, i.e.,  $(Pf(Q))^2 = \det Q$  for each  $Q$  in  $\mathfrak{so}(2k)$ ). Now choose a connection  $\omega$  on the frame bundle (4.24), denote its curvature by  $\Omega$  and define the **Euler class**  $e(E)$  of  $\pi_E : E \rightarrow X$  by  $e(E) = (-2\pi)^{-k} [\overline{Pf}(\Omega)]$ , which we prefer to write as

$$e(E) = (2\pi)^{-k} [\overline{Pf}(-\Omega)]. \quad (4.26)$$

Locally,  $e(E)$  is given by

$$e(E) = (2\pi)^{-k} [Pf(-s^*\Omega)] \quad (4.27)$$

where  $s$  is any section of the frame bundle (4.24). With this our definition of the Euler number (4.23) is complete.

**Remark.** *The famous **Gauss-Bonnet-Chern Theorem** asserts that when  $E$  is the tangent bundle  $TX$  of  $X$ , then the Euler number  $\chi(TX)$  is, in fact, the Euler characteristic (alternating sum of Betti numbers) of  $X$  and so is a topological invariant of  $X$ .*

**Example.** *We consider the two-sphere  $\mathbb{S}^2$  with its usual orientation and Riemannian metric and its tangent bundle  $\pi : T\mathbb{S}^2 \rightarrow \mathbb{S}^2$ . The corresponding oriented, orthonormal frame bundle is  $SO(2) \hookrightarrow F_{SO}(T\mathbb{S}^2) \xrightarrow{\pi_{SO}} \mathbb{S}^2$ . If  $\theta$  and  $\phi$  are the usual spherical coordinates on  $\mathbb{S}^2$ , then  $\{e_1, e_2\} = \{\frac{\partial}{\partial\phi}, \frac{1}{\sin\phi} \frac{\partial}{\partial\theta}\}$  is an oriented, orthonormal frame field on  $\mathbb{S}^2$ , i.e., a section  $s$  of  $F_{SO}(\mathbb{S}^2)$ . The dual oriented, orthonormal field of one-forms is  $\{e^1, e^2\} = \{d\phi, \sin\phi d\theta\}$  so the metric volume form is  $e^1 \wedge e^2 = \sin\phi d\phi \wedge d\theta$ . One computes  $de^1 = 0 = 0(e^1 \wedge e^2)$  and  $de^2 = \cos\phi d\phi \wedge d\theta = \cot\phi(e^1 \wedge e^2)$  so  $0 \cdot e^1 + (\cot\phi)e^2 = \cot\phi d\theta$ . Thus, the Levi-Civita connection  $\omega$  on the frame bundle is determined by*

$$s^*\omega = \begin{pmatrix} 0 & \cos\phi d\theta \\ -\cos\phi d\theta & 0 \end{pmatrix}$$

so its curvature  $\Omega = d\omega$  (as  $SO(2)$  is abelian) is

$$s^*\Omega = \begin{pmatrix} 0 & -\sin\phi d\phi \wedge d\theta \\ \sin\phi d\phi \wedge d\theta & 0 \end{pmatrix}.$$

A representative of the Euler class is therefore

$$(2\pi)^{-1} Pf(-s^*\Omega) = \frac{1}{2\pi} \sin\phi d\phi \wedge d\theta.$$



Notice that the Euler number

$$\chi(T\mathbb{S}^2) = \frac{1}{2\pi} \int_{\mathbb{S}^2} \sin \phi d\phi \wedge d\theta = 2$$

which is, of course, the Euler characteristic of  $\mathbb{S}^2$ .

There is an alternative description of the Euler class that will be important to us soon. Denote by  $H_{CV}^*(E, \mathbb{R})$  the compact-vertical cohomology of  $E$  (generated by the differential forms on  $E$  whose restriction to each fiber of  $\pi_E : E \rightarrow X$  has compact support). One can show that there is a unique element  $U(E) \in H_{CV}^{2k}(E, \mathbb{R})$  whose integral over each fiber is 1. This is called the **Thom class** of  $\pi_E : E \rightarrow X$  and it has the property that, if  $s : X \rightarrow E$  is any section of the vector bundle, e.g., the zero-section, then

$$e(E) = s^*U(E). \tag{4.28}$$

It is not clear from either of these definitions, but the Euler number  $\chi(E)$  is actually an integer. An alternative description of  $\chi(E)$  in which its integrality is manifest is contained in the so-called **Poincaré-Hopf Theorem**. To state this we recall that, for any section  $s : X \rightarrow E$ , the image  $s(X)$  is a submanifold of  $E$  diffeomorphic to  $X$ . This is, in particular, true of the zero section  $s_0 : X \rightarrow E$  and one often identifies  $X$  with  $s_0(X)$ . Then  $s(X) \cap s_0(X)$  is the set of zeros of  $s$  (if  $E = TX$ ,  $s$  is a vector field on  $X$  and these are its singularities). We will say that  $s$  is **generic** if  $s(X)$  intersects  $s_0(X)$  transversely (meaning that, for any  $s(x) \in s(X) \cap s_0(X)$ ,  $T_{s(x)}(E) = T_{s(x)}(s(X)) \oplus T_{s(x)}(s_0(X))$ ). According to the Thom Transversality Theorem, generic sections are dense in the space of all sections. For such a section  $s$ ,  $s(X) \cap s_0(X)$  is necessarily a finite set of isolated points and we attach a sign to each such point  $p$  as follows:  $\text{sign}(p) = 1$  if an oriented basis for  $T_p(s(X))$  together with an oriented basis for  $T_p(s_0(X))$  is an oriented basis for  $T_p(E)$ ; otherwise,  $\text{sign}(p) = -1$ . The **intersection number** of  $s(X)$  and  $s_0(X)$  is the sum of these signs over all points in  $s(X) \cap s_0(X)$ . The Poincaré-Hopf Theorem asserts that *the Euler number  $\chi(E)$  is the intersection number of any generic section*.

With this digression behind us we may return to the new perspective on the Donaldson invariant  $\gamma_0(B)$  promised earlier. For this and all subsequent discussions we intend to employ a more economical notation, dropping all references to Sobolev indices, writing  $P$  for  $P_k$   $\mathcal{G}$  for  $\mathcal{G}_4(P_k)$ , etc.

The gauge group  $\mathcal{G}$  does not act freely on the space  $\hat{\mathcal{A}}$  of irreducible connections since even irreducible connections have a  $\mathbb{Z}_2$  stabilizer. However,  $\hat{\mathcal{G}} = \mathcal{G}/\mathbb{Z}_2$  does

act freely on  $\hat{\mathcal{A}}$  so we have an infinite-dimensional principal bundle

$$\hat{\mathcal{G}} \curvearrowright \hat{\mathcal{A}} \longrightarrow \hat{\mathcal{B}} \quad (4.29)$$

over the Banach manifold  $\hat{\mathcal{B}}$  (note that the orbits in  $\hat{\mathcal{A}}$  of the  $\hat{\mathcal{G}}$ -action are the same as those of the  $\mathcal{G}$ -action so the quotient is still  $\hat{\mathcal{B}}$ ). We build a vector bundle associated to this principal bundle as follows. Consider the (infinite-dimensional) vector space  $\Omega_+^2(B, \text{ad}P)$  of self-dual two-forms on  $B$  with values in the adjoint bundle. We claim that there is a smooth left action of  $\hat{\mathcal{G}}$  on  $\Omega_+^2(B, \text{ad}P)$ . To see this we think of  $\mathcal{G}$  as the group of sections of the nonlinear adjoint bundle  $\text{Ad}P$  under pointwise multiplication. Since the elements of  $\Omega_+^2(B, \text{ad}P)$  take values in the  $\mathfrak{su}(2)$  fibers of  $\text{ad}P$ ,  $\mathcal{G}$  acts on these values by conjugation. Moreover, conjugation has the same effect at  $\pm f \in \mathcal{G}$  so this  $\mathcal{G}$ -action on  $\Omega_+^2(B, \text{ad}P)$  descends to a  $\mathcal{G}/\mathbb{Z}_2 = \hat{\mathcal{G}}$ -action. Thus, we have an associated vector bundle

$$\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$$

the elements of which are equivalence classes  $[\omega, \gamma] = [\omega \cdot f, f^{-1} \cdot \gamma]$  with  $\omega \in \hat{\mathcal{A}}$ ,  $\gamma \in \Omega_+^2(B, \text{ad}P)$  and  $f \in \hat{\mathcal{G}}$ .

Now recall that sections of associated vector bundles can be identified with equivariant maps from the principal bundle space into the vector space. In our case we have an obvious map from  $\hat{\mathcal{A}}$  into  $\Omega_+^2(B, \text{ad}P)$ , i.e., the self-dual curvature map:

$$\begin{aligned} F^+ &: \hat{\mathcal{A}} \longrightarrow \Omega_+^2(B, \text{ad}P) \\ F^+(\omega) &= F_\omega^+ = \frac{1}{2}(F\omega + *F\omega). \end{aligned} \quad (4.30)$$

Since the action of  $\hat{\mathcal{G}}$  on  $\hat{\mathcal{A}}$  is by conjugation and curvatures transform by conjugation under a gauge transformation,  $F^+$  is equivariant:

$$F^+(\omega \cdot f) = F_{\omega \cdot f}^+ = f^{-1} F_\omega^+ f = f^{-1} \cdot F_\omega^+ = f^{-1} \cdot F^+(\omega)$$

$F^+$  can therefore be identified with a section

$$s_+ : \hat{\mathcal{B}} \longrightarrow \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$$

of our vector bundle, given explicitly by

$$s_+([\omega]) = [\omega, F_\omega^+]$$

for every  $[\omega] \in \hat{\mathcal{B}}$ . This section is **Fredholm** in the sense that its local representatives, thought of as sections of trivial bundles, i.e.,  $\Omega_+^2(B, \text{ad}P)$ -valued maps

on open subsets of the Banach manifold  $\hat{\mathcal{B}}$ , have derivatives at each point that are (linear) Fredholm maps. Notice now that the moduli space  $\hat{\mathcal{M}}$  of anti-self-dual connections ( $F_{\omega}^+ = 0$ ) is precisely the zero set of the section  $s_+$ . Identifying  $\hat{\mathcal{B}}$  with the image of the zero section

$$s_0 : \hat{\mathcal{B}} \longrightarrow \hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$$

we conclude that

$$\hat{\mathcal{M}} = s_+(\hat{\mathcal{B}}) \cap s_0(\hat{\mathcal{B}}). \tag{4.31}$$

In the case in which  $\hat{\mathcal{M}}$  is 0-dimensional so that each point of the intersection (4.31) acquires a sign and the Donaldson invariant  $\gamma_0(B)$  is the sum of these signs one sees quite clearly the sense in which  $\gamma_0(B)$  can be regarded (at least formally) as an infinite-dimensional analogue of the Poincaré-Hopf version of an Euler number.

Taking this analogy seriously would suggest the possibility of an integral representation of  $\gamma_0(B)$  modeled on our definition (4.23) of the Euler number. Notice, however, that such an “integral” would necessarily be over the infinite-dimensional moduli space  $\hat{\mathcal{B}}$  and such integrals are notoriously difficult to define rigorously. But, as Hitchin [21] has phrased it, “This is such stuff as quantum field theory is made of.” Indeed, it was Edward Witten who first produced such an integral representation of  $\gamma_0(B)$ , not directly, but as what is called the “partition function” of the quantum field theory introduced in [46]. We intend to produce Witten’s partition function, but not from the quantum field theory arguments of [46]. Rather we will follow Atiyah and Jeffrey [3] who showed that an integral formula for the Euler number of a (finite-dimensional) vector bundle proved by Mathai and Quillen [32], when formally applied to the vector bundle  $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$ , yields precisely this partition function.

## 5. Mathai-Quillen Formalism and Witten’s Partition Function

We begin by having a closer look at the expression (4.27) for the Euler class of the oriented, real vector bundle  $\pi_E : E \rightarrow X$ . Recall that we denote by  $V$  the typical fiber of the bundle, which we assume has dimension  $2k$  and a positive definite inner product. We fix, once and for all, an oriented, orthonormal basis  $\{\psi^1, \dots, \psi^{2k}\}$  for  $V$ . We regard the elements of the exterior algebra  $\bigwedge V$  as polynomials with real coefficients in the odd (anti-commuting) variables  $\psi^1, \dots, \psi^{2k}$  and provide  $\bigwedge V$  with its usual  $\mathbb{Z}_2$ -grading

$$\begin{aligned}\bigwedge V &= \bigoplus_{i=0}^{\infty} \bigwedge^i V \cong \left( \bigoplus_{i=0}^{\infty} \bigwedge^{2i} V \right) \oplus \left( \bigoplus_{i=0}^{\infty} \bigwedge^{2i+1} V \right) \\ &= (\bigwedge V)_0 \oplus (\bigwedge V)_1 .\end{aligned}$$

$\bigwedge V$  is therefore a supercommutative superalgebra. One can define the exponential map on  $\bigwedge V$  by the usual power series, noting that the series eventually terminates for any element of  $\bigwedge V$  due to the anti-commutativity of the multiplication.

The **Berezin (or fermionic) integral** of an element  $f$  of  $\bigwedge V$  is the (real) coefficient of  $\psi^1 \cdots \psi^{2k} = \mathbf{vol}$  in the polynomial  $f$  and we will write this as

$$\int f \mathcal{D}\psi = f_{\mathbf{vol}} .$$

For example, our definition (4.25) of the Pfaffian of  $Q \in \mathfrak{so}(2k)$  can be written

$$\int e^{\frac{1}{2} \psi^T Q \psi} \mathcal{D}\psi = Pf(Q) . \quad (5.1)$$

In particular, (4.27) now gives representatives of the Euler class as Berezin integrals of the form

$$(2\pi)^{-k} \int e^{\frac{1}{2} \psi^T (-s^* \mathbf{\Omega}) \psi} \mathcal{D}\psi . \quad (5.2)$$

We will also need to extend this notion of Berezin integration in the following way. Let  $A$  be any other supercommutative superalgebra and consider the (super) tensor product  $A \otimes \bigwedge V$ . Regard the elements of  $A \otimes \bigwedge V$  as polynomials in the odd variables  $\psi^1, \dots, \psi^{2k}$  with coefficients in  $A$  and define the Berezin integral of such an  $F \in A \otimes \bigwedge V$  to be the coefficient (in  $A$ ) of  $\psi^1 \cdots \psi^{2k} = \mathbf{vol}$

$$\int F \mathcal{D}\psi = F_{\mathbf{vol}} .$$

As an example we introduce coordinates  $u_1, \dots, u_{2k}$  on  $V$  corresponding to the basis  $\psi^1, \dots, \psi^{2k}$ . Thus,  $\{u_1, \dots, u_{2k}\}$  is the basis for  $V^*$  dual to  $\{\psi^1, \dots, \psi^{2k}\}$ . Let  $A = \Omega^*(V)$  be the algebra of differential forms on  $V$  (which, throughout this section, we take to be complex-valued). Thus, each  $du_j$  is in  $\Omega^1(V)$  and  $-idu_j \psi^j = i\psi^j du_j$  (sum over  $j = 1, \dots, 2k$ ) is in  $\Omega^*(V) \otimes \bigwedge V$ . Writing  $du$  for  $(du_1 \cdots du_{2k})^\top$  we will show that

$$\int e^{i\psi^T du} \mathcal{D}\psi = \int e^{i\psi^j du_j} \mathcal{D}\psi = du_1 \cdots du_{2k} \quad (5.3)$$

(by which we mean  $du_1 \wedge \cdots \wedge du_{2k} \in \Omega^{2k}(V)$ ). Indeed,

$$\begin{aligned}
 \int e^{i\psi^T du} \mathcal{D}\psi &= \int e^{i\psi^j du_j} \mathcal{D}\psi = \int e^{i(\psi^1 du_1 + \cdots + \psi^{2k} du_{2k})} \mathcal{D}\psi \\
 &= \int e^{i\psi^1 du_1} \cdots e^{i\psi^{2k} du_{2k}} \mathcal{D}\psi \\
 &\quad \text{(the elements } \psi^j du_j \text{ are even in } \Omega^*(V) \otimes \wedge V \text{ and so commute)} \\
 &= \int (1 + i\psi^1 du_1) \cdots (1 + i\psi^{2k} du_{2k}) \mathcal{D}\psi \\
 &= \int (i\psi^1 du_1) \cdots (i\psi^{2k} du_{2k}) \mathcal{D}\psi \\
 &\quad \text{(only this product contributes to the coefficient of } \psi^1 \cdots \psi^{2k}\text{)} \\
 &= (i)^{2k} \int (-1)^{\frac{1}{2}(2k)(2k+1)} du_1 \cdots du_{2k} \psi^1 \cdots \psi^{2k} \mathcal{D}\psi \\
 &= du_1 \cdots du_{2k}.
 \end{aligned}$$

Notice that if we write  $\|u\|^2 = u_1^2 + \cdots + u_{2k}^2 \in \Omega^0(V)$  and identify this with  $\|u\|^2 \otimes 1 \in \Omega^*(V) \otimes \wedge V$ , then

$$(2\pi)^{-k} \int e^{-\frac{1}{2}\|u\|^2 + i\psi^T du} \mathcal{D}\psi = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} du_1 \cdots du_{2k} \quad (5.4)$$

which is a form on  $V$  that integrates to 1 over  $V$ . It does not have compact support on  $V$ , but one can think of it as a ‘‘Gaussian representative’’ of the Thom class of the vector bundle over a point whose fiber is  $V$  (the compact vertical cohomology of a vector bundle is isomorphic to the cohomology of forms that are ‘‘rapidly decreasing’’ in the fibers and the usual discussion of the Thom class extends easily to this context). Shortly we will introduce the so-called ‘‘universal Thom form’’ of Mathai and Quillen [32] which adds one more term to the exponent in (5.4) to produce what is called an ‘‘equivariant differential form.’’ For this though we require a brief digression.

Equivariant cohomology arose from attempts to understand the topology of the orbit space  $M/G$  of a topological space on which some topological group  $G$  acts. We will be concerned only with the case in which  $M$  is a smooth manifold and  $G$  is a compact, connected matrix Lie group acting smoothly on  $M$  on the left. For this action we will write

$$\begin{aligned}
 \sigma &: G \times M \longrightarrow M \\
 \sigma(g, m) &= g \cdot m = \sigma_g(m) = \sigma_m(g).
 \end{aligned}$$

We denote by  $\mathcal{G}$  the Lie algebra of  $G$ ,  $\mathbb{C}[\mathcal{G}]$  the algebra of complex-valued polynomials on  $\mathcal{G}$  and  $\Omega^*(M)$  the algebra of complex-valued differential forms on  $M$ . We consider the tensor product  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$ , every element of which is a sum of terms of the form  $\alpha = \mathcal{P} \otimes \varphi$ , where  $\mathcal{P} \in \mathbb{C}[\mathcal{G}]$  and  $\varphi \in \Omega^*(M)$ . These are best thought of as  $\Omega^*(M)$ -valued polynomials on  $\mathcal{G}$  (e.g.,  $\alpha(\xi) = (\mathcal{P} \otimes \varphi)(\xi) = \mathcal{P}(\xi)\varphi$  for each  $\xi \in \mathcal{G}$ ). Rather than the usual tensor product grading on  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  we will, for reasons that will become clear shortly, “double the degrees” in  $\mathbb{C}[\mathcal{G}]$ . More precisely, if  $\alpha = \mathcal{P} \otimes \varphi$  we define

$$\deg \alpha = \deg (\mathcal{P} \otimes \varphi) = 2 \deg \mathcal{P} + \deg \varphi \quad (5.5)$$

where  $\deg \mathcal{P}$  is the algebraic degree of the polynomial  $\mathcal{P}$  and  $\deg \varphi$  is the cohomological degree of the form  $\varphi$ . Thus,

$$\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M) = \bigoplus_{2i+j=k} \mathbb{C}^i[\mathcal{G}] \otimes \Omega^j(M).$$

The action of  $G$  on  $M$  together with the adjoint action of  $G$  on  $\mathcal{G}$  give a natural action of  $G$  on  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$ , i.e., if  $\alpha = \mathcal{P} \otimes \varphi$  and  $g \in G$ , then  $g \cdot \alpha$  is the element of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  whose value at any  $\xi \in \mathcal{G}$  is

$$(g \cdot \alpha)(\xi) = (g \cdot (\mathcal{P} \otimes \varphi))(\xi) = \mathcal{P}(g^{-1}\xi g) \sigma_{g^{-1}}^* \varphi.$$

An element  $\alpha$  of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  is said to be  **$G$ -invariant** if  $g \cdot \alpha = \alpha$  for every  $g \in G$ . This is easily seen to be equivalent to  $\alpha(g\xi g^{-1}) = \sigma_{g^{-1}}^* \alpha(\xi)$  for every  $g \in G$  and every  $\xi \in \mathcal{G}$ . The algebra of all  $G$ -invariant elements of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  is denoted

$$\Omega_G^*(M) = [\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)]^G$$

and its elements are called  **$G$ -equivariant differential forms** on  $M$ . Our grading of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  gives

$$\Omega_G^*(M) = \bigoplus_{k=0}^{\infty} \Omega_G^k(M) = \bigoplus_{2i+j=k} [\mathbb{C}^i[\mathcal{G}] \otimes \Omega^j(M)]^G$$

and we will take  $\Omega_G^k(M)$  to be trivial for  $k < 0$ . If  $\alpha \in \Omega_G^*(M)$ , then, for each  $\xi \in \mathcal{G}$ ,  $\alpha(\xi)$  is an element of  $\Omega^*(M)$  and so can be written

$$\alpha(\xi) = \alpha(\xi)_{[0]} + \alpha(\xi)_{[1]} + \cdots + \alpha(\xi)_{[n]}$$

where  $\alpha(\xi)_{[k]} \in \Omega^k(M)$  and  $n = \dim M$ . Similarly, we will write  $\alpha_{[k]}$  for the part of  $\alpha$  in  $\Omega_G^k(M)$ . Notice that both  $\mathbb{C}[\mathcal{G}]^G$  (those  $\mathcal{P} \in \mathbb{C}[\mathcal{G}]$  satisfying  $\mathcal{P}(g^{-1}\xi g) =$

$\mathcal{P}(\xi)$  for all  $g \in G$  and  $\xi \in \mathcal{G}$ ) and  $\Omega^*(M)^G$  (those  $\varphi \in \Omega^*(M)$  satisfying  $\sigma_{g^{-1}}^* \varphi = \varphi$  for all  $g \in G$ ) can be identified with subalgebras of  $\Omega_G^*(M)$  via  $\mathcal{P} \rightarrow \mathcal{P} \otimes 1$  and  $\varphi \rightarrow 1 \otimes \varphi$ , respectively.

Next we define the  **$G$ -equivariant exterior derivative**  $d_G$  on  $\Omega_G^*(M)$  as follows: For any  $\alpha \in \mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  and any  $\xi \in \mathcal{G}$  we define

$$(d_G \alpha)(\xi) = d(\alpha(\xi)) - \iota_{\xi^\#}(\alpha(\xi)) \quad (5.6)$$

where  $\xi^\#$  is the vector field on  $M$  defined, at each  $m \in M$ , by

$$\xi^\#(m) = \frac{d}{dt} (\exp(-t\xi) \cdot m)|_{t=0} \quad (5.7)$$

and  $\iota_{\xi^\#}$  denotes interior multiplication by  $\xi^\#$  (the minus sign in (5.7) is due to the fact that  $G$  acts on  $M$  on the left and ensures that the map  $\xi \rightarrow \xi^\#$  is a homomorphism of Lie algebras). Alternatively, if  $\{\xi_1, \dots, \xi_n\}$  is a basis for  $\mathcal{G}$  and if we write  $\iota_a$  for  $\iota_{\xi_a^\#}$ , then

$$d_G = 1 \otimes d - x^a \otimes \iota_a \quad (5.8)$$

where  $\{x^1, \dots, x^n\}$  is a basis for  $\mathcal{G}^*$  dual to  $\{\xi_1, \dots, \xi_n\}$  and we regard each  $x^a$  as an element of  $\mathbb{C}[\mathcal{G}]$ . It is enough to verify (5.8) for elements of the form  $\alpha = \mathcal{P} \otimes \varphi$  and this is straightforward. Now, for any element of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  of the form  $\alpha = \mathcal{P} \otimes \varphi$  we have  $\deg \alpha = 2 \deg \mathcal{P} + \deg \varphi$  and so

$$\deg((1 \otimes d)(\alpha)) = 2 \deg \mathcal{P} + (\deg \varphi + 1) = \deg \alpha + 1$$

and

$$\deg((x^a \otimes \iota_a)(\alpha)) = 2(\deg \mathcal{P} + 1) + (\deg \varphi - 1) = \deg \alpha + 1$$

imply that  $\deg(d_G \alpha) = \deg \alpha + 1$  (this is the reason for the peculiar grading on  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$ ). One also verifies that  $d_G$  preserves the subalgebra  $\Omega_G^*(M)$  of invariant elements and satisfies, for any  $\alpha \in \mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  and any  $\xi \in \mathcal{G}$

$$((d_G \circ d_G)(\alpha))(\xi) = -\mathcal{L}_{\xi^\#}(\alpha(\xi)) \quad (5.9)$$

where  $\mathcal{L}_{\xi^\#}$  is the Lie derivative with respect to the vector field  $\xi^\#$ . Since an invariant element  $\alpha$  of  $\mathbb{C}[\mathcal{G}] \otimes \Omega^*(M)$  satisfies  $\mathcal{L}_{\xi^\#}(\alpha(\xi)) = 0$  for every  $\xi \in \mathcal{G}$  we obtain from (5.9) that

$$d_G \circ d_G = 0 \quad \text{on} \quad \Omega_G^*(M). \quad (5.10)$$

Thus

$$(\Omega_G^*(M), d_G)$$

is a cochain complex. The cohomology of this complex is called the **Cartan model** of the  **$G$ -equivariant cohomology** of  $M$  and is denoted  $H_G^*(M)$ . In somewhat more detail, an element  $\alpha$  of  $\Omega_G^*(M)$  is said to be  **$G$ -equivariantly closed** if  $d_G \alpha = 0$  and  **$G$ -equivariantly exact** if  $\alpha = d_G \beta$  for some  $\beta \in \Omega_G^*(M)$ . Writing  $d_G^k$  for the restriction of  $d_G$  to  $\Omega_G^k(M)$  we have

$$\Omega_G^{k-1}(M) \xrightarrow{d_G^{k-1}} \Omega_G^k(M) \xrightarrow{d_G^k} \Omega_G^{k+1}(M)$$

with  $d_G^k \circ d_G^{k-1} = 0$  so that

$$H_G^*(M) = \bigoplus_{k=0}^{\infty} H_G^k(M) = \bigoplus_{k=0}^{\infty} \ker(d_G^k) / \text{im}(d_G^{k-1}).$$

Notice that if  $M$  is a single point (connected zero-dimensional manifold), then every element of  $\Omega_G^*(M)$  is of the form  $\mathcal{P} \otimes 1$  for some  $\mathcal{P} \in \mathbb{C}[\mathcal{G}]^G$ . Each of these is  $G$ -equivariantly closed, but none is  $G$ -equivariantly exact so

$$H_G^*(\text{pt}) \cong \mathbb{C}[\mathcal{G}]^G. \quad (5.11)$$

Notice also that if  $G$  is trivial, then so is the Lie algebra  $\mathcal{G}$  so there are only constant polynomials on  $\mathcal{G}$ . Everything is  $G$ -invariant so one can identify  $\Omega_G^*(M)$  with  $\Omega^*(M)$ . Furthermore,  $\iota_{\xi\#} = \iota_0 = 0$  so  $d_G$  agrees with  $d$  and we conclude that

$$G = \{1\} \implies H_G^*(M) \cong H_{\text{de Rham}}^*(M). \quad (5.12)$$

**Example.** *To gain some familiarity with these definitions we will compute just one equivariant cohomology group from scratch. We consider the standard action of  $G = \mathbb{S}^1$  on  $M = \mathbb{S}^3$  that gives rise to the complex Hopf bundle. Specifically, we consider*

$$\mathbb{S}^3 = \left\{ (z^1, z^2) \in \mathbb{C}^2; |z^1|^2 + |z^2|^2 = 1 \right\}$$

and define a left action of  $\mathbb{S}^1 = \{e^{i\theta}; \theta \in \mathbb{R}\}$  on  $\mathbb{S}^3$  by

$$e^{i\theta} \cdot (z^1, z^2) = (e^{i\theta} z^1, e^{i\theta} z^2).$$

The action is clearly free and the orbit space  $\mathbb{S}^3/\mathbb{S}^1$  is, by definition, the complex projective line  $\mathbb{C}\mathbb{P}^1$ , which is diffeomorphic to  $\mathbb{S}^2$ . Since  $\mathbb{S}^1$  is one-dimensional, its



Lie algebra has a single generator. Choose one such and denote it  $\xi_1$ . We denote by  $x^1$  the corresponding dual basis vector so that  $\mathbb{C}[\mathcal{G}]$  can be identified with the algebra  $\mathbb{C}[x^1]$  of polynomials with complex coefficients in the single “variable”  $x^1$ . Since  $\mathbb{S}^1$  is abelian, all of these polynomials are  $\mathbb{S}^1$ -invariant so

$$[\mathbb{C}[x^1] \otimes \Omega^*(\mathbb{S}^3)]^{\mathbb{S}^1} = \mathbb{C}[x^1] \otimes \Omega^*(\mathbb{S}^3)^{\mathbb{S}^1}. \quad (5.13)$$

We will leave it to the reader to show that  $H_{\mathbb{S}^1}^0(\mathbb{S}^3) \cong \mathbb{C}$  and  $H_{\mathbb{S}^1}^1(\mathbb{S}^3)$  is trivial so that we may turn our attention to  $H_{\mathbb{S}^1}^2(\mathbb{S}^3)$ . Thus, we consider

$$\Omega_{\mathbb{S}^1}^1(\mathbb{S}^3) \xrightarrow{d_{\mathbb{S}^1}^1} \Omega_{\mathbb{S}^1}^2(\mathbb{S}^3) \xrightarrow{d_{\mathbb{S}^1}^2} \Omega_{\mathbb{S}^1}^3(\mathbb{S}^3)$$

and compute  $\ker(d_{\mathbb{S}^1}^2)/\text{im}(d_{\mathbb{S}^1}^1)$ . Now notice that (5.13) together with the grading we have defined on  $\Omega_{\mathbb{S}^1}^*(\mathbb{S}^3)$  imply that every element of  $\Omega_{\mathbb{S}^1}^1(\mathbb{S}^3)$  can be written in the form

$$1 \otimes \boldsymbol{\eta}, \quad (5.14)$$

where  $\boldsymbol{\eta} \in \Omega^1(\mathbb{S}^3)^{\mathbb{S}^1}$ , and every element of  $\Omega_{\mathbb{S}^1}^2(\mathbb{S}^3)$  can be written in the form

$$1 \otimes \boldsymbol{\omega} + x^1 \otimes f \quad (5.15)$$

where  $\boldsymbol{\omega} \in \Omega^2(\mathbb{S}^3)^{\mathbb{S}^1}$  and  $f$  is a complex-valued function on  $\mathbb{S}^3$  that is constant on each  $\mathbb{S}^1$ -orbit. Now, let  $\tilde{\boldsymbol{\omega}} \in \Omega_{\mathbb{S}^1}^2(\mathbb{S}^3)$  be  $\mathbb{S}^1$ -equivariantly closed. Write  $\tilde{\boldsymbol{\omega}} = 1 \otimes \boldsymbol{\omega} + x^1 \otimes f$  as in (5.15). Then

$$\begin{aligned} 0 = d_{\mathbb{S}^1} \tilde{\boldsymbol{\omega}} &= (1 \otimes d - x^1 \otimes \iota_1) (1 \otimes \boldsymbol{\omega} + x^1 \otimes f) \\ &= 1 \otimes d\boldsymbol{\omega} + x^1 \otimes (df - \iota_1 \boldsymbol{\omega}) \end{aligned}$$

implies that

$$d\boldsymbol{\omega} = 0 \quad \text{and} \quad df = \iota_1 \boldsymbol{\omega}. \quad (5.16)$$

We show first that there exists an  $a \in \mathbb{C}$  and an  $\boldsymbol{\eta} \in \Omega^1(\mathbb{S}^3)^{\mathbb{S}^1}$  such that

$$(1 \otimes \boldsymbol{\omega} + x^1 \otimes f) - a(x^1 \otimes 1) = d_{\mathbb{S}^1}^1(1 \otimes \boldsymbol{\eta}) \quad (5.17)$$

i.e.,

$$1 \otimes \boldsymbol{\omega} + x^1 \otimes (f - a) = 1 \otimes d\boldsymbol{\eta} - \iota_1 \boldsymbol{\eta}. \quad (5.18)$$

Now, in order for (5.18) to be satisfied we must have  $d\boldsymbol{\eta} = \boldsymbol{\omega}$  and  $\iota_1 \boldsymbol{\eta} = a - f$  so we will simply solve these equations. Since  $d\boldsymbol{\omega} = 0$  by (5.16) and since  $H_{\text{de Rham}}^*(\mathbb{S}^3) = 0$ ,  $\boldsymbol{\omega}$  must be exact in the de Rham sense, i.e.,  $\boldsymbol{\omega}$  is a differential of something in  $\Omega^1(\mathbb{S}^3)$ . To see that we can choose this element of  $\Omega^1(\mathbb{S}^3)$  to be  $\mathbb{S}^1$ -invariant we require a general lemma.

**Lemma 5.1.** *If  $\omega \in \Omega^{k+1}(M)^G$  is (de Rham) exact, then there exists an  $\eta \in \Omega^k(M)^G$  with  $d\eta = \omega$ .*

The proof of the lemma proceeds in the following way. One shows that any  $\alpha \in \Omega^*(M)$  can be “ $G$ -invariantized” in the sense that there is a cochain map  $I : \Omega^*(M) \rightarrow \Omega^*(M)^G$  which reduces to the identity on  $\Omega^*(M)^G \subseteq \Omega^*(M)$  (“cochain map” means  $d \circ I = I \circ d$ ). This map is constructed by “averaging over the group  $G$ ”. In somewhat more detail, one chooses an invariant measure  $dG$  on  $G$  and, for  $\alpha \in \Omega^k(M)$ ,  $p \in M$  and  $v_1, \dots, v_k \in T_p(M)$ , defines

$$\begin{aligned} (I(\alpha))_p(v_1, \dots, v_k) &= \int_G (\sigma_g^* \alpha)_p(v_1, \dots, v_k) dG \\ &= \int_G \alpha_{g \cdot p}((\sigma_g)_{*p}(v_1), \dots, (\sigma_g)_{*p}(v_k)) dG. \end{aligned}$$

Now, if  $\omega \in \Omega^{k+1}(M)^G$  is exact there is an  $\alpha \in \Omega^k(M)$  with  $d\alpha = \omega$ . But  $I(\alpha) \in \Omega^k(M)^G$  and  $\omega = I(\omega) = I(d\alpha) = d(I(\alpha))$  as required.

Returning to the proof of (5.18), we can now select an  $\eta \in \Omega^1(\mathbb{S}^3)^{\mathbb{S}^1}$  with  $\omega = d\eta$  and so the first equation is satisfied. Furthermore, since  $\eta$  is  $\mathbb{S}^1$ -invariant,

$$0 = \mathcal{L}_1 \eta = (d \circ \iota_1 + \iota_1 \circ d) \eta = d(\iota_1 \eta) + \iota_1 \omega = d(\iota_1 \eta + f).$$

Since  $\mathbb{S}^3$  is connected this implies that  $\iota_1 \eta + f$  is some constant function  $a$ , i.e.,  $\iota_1 \eta = a - f$  so, for this  $a$ , the second condition is satisfied as well. This completes the proof of (5.18) and therefore of (5.17).

To understand the conclusion to be drawn from (5.17) we observe that  $x^1 \otimes 1$  is  $\mathbb{S}^1$ -equivariantly closed and so determines an  $\mathbb{S}^1$ -equivariant cohomology class in  $H_{\mathbb{S}^1}^2(\mathbb{S}^3)$ . Thus, (5.17) implies that the cohomology class of  $\tilde{\omega} = 1 \otimes \omega + x^1 \otimes f$  is a multiple (by  $a$ ) of the cohomology class of  $x^1 \otimes 1$ . Since  $\tilde{\omega}$  was an arbitrary  $\mathbb{S}^1$ -equivariantly closed element of  $\Omega_{\mathbb{S}^1}^2(\mathbb{S}^3)$  we conclude that  $H_{\mathbb{S}^1}^2(\mathbb{S}^3)$  is generated by the class of  $x^1 \otimes 1$ . We conclude by showing that  $x^1 \otimes 1$  is not  $d_{\mathbb{S}^1}$ -exact so that this class is nontrivial and therefore

$$H_{\mathbb{S}^1}^2(\mathbb{S}^3) \cong \mathbb{C}. \quad (5.19)$$

To prove this we assume to the contrary that there is an element  $\tilde{\eta}$  in  $\Omega_{\mathbb{S}^1}^1(\mathbb{S}^3)$  with  $d_{\mathbb{S}^1}^1 \tilde{\eta} = x^1 \otimes 1$ .  $\tilde{\eta}$  can be written as  $\tilde{\eta} = 1 \otimes \eta$ , where  $\eta \in \Omega^1(\mathbb{S}^3)^{\mathbb{S}^1}$ . Thus,  $d_{\mathbb{S}^1}^1(1 \otimes \eta) = x^1 \otimes 1$ , i.e.,

$$1 \otimes d\eta - x^1 \otimes \iota_1 \eta = x^1 \otimes 1$$

so we must have

$$d\eta = 0 \quad \text{and} \quad \iota_1 \eta = -1. \tag{5.20}$$

But  $d\eta = 0$  and  $H_{\text{de Rham}}^1(\mathbb{S}^3) = 0$  imply that  $\eta$  is de Rham exact and Lemma 5.1 implies that there is an  $f \in \Omega^0(\mathbb{S}^3)^{\mathbb{S}^1}$  with  $\eta = df$ . Thus,

$$\iota_1 \eta = \iota_1(df) = \mathcal{L}_1 f - d(\iota_1 f) = 0 - d(0) = 0$$

and so the second condition in (5.20) could not be satisfied. Thus,  $x^1 \otimes 1$  cannot be  $d_{\mathbb{S}^1}$ -exact and the proof of (5.19) is complete.

We should point out that, for each of the examples we have described for  $H_{\mathbb{S}^1}^*(\mathbb{S}^3)$ , the  $\mathbb{S}^1$ -equivariant cohomology group of  $\mathbb{S}^3$  agrees with the corresponding ordinary de Rham cohomology group (with complex coefficients) of the orbit space  $\mathbb{S}^3/\mathbb{S}^1 \cong \mathbb{S}^2$ . That this is no accident is the content of a beautiful theorem of Henri Cartan (see [20] for a proof of a much more general result).

**Theorem 5.2 (Cartan)** *Let  $M$  be a smooth manifold and  $G$  a compact, connected Lie group. Suppose there is a smooth, free action of  $G$  on  $M$  on the left. Then the  $G$ -equivariant cohomology algebra  $H_G^*(M)$  is isomorphic to the de Rham cohomology  $H_{\text{de Rham}}^*(M/G)$  with complex coefficients of the orbit manifold  $M/G$ .*

Finally, we must introduce a notion of integration for equivariant forms and cohomology classes. For this we now assume that  $M$  is compact and oriented and that the  $G$ -action on  $M$  preserves the orientation (each diffeomorphism  $\sigma_g : M \rightarrow M$  is orientation preserving). For each  $\alpha \in \Omega_G^*(M)$  we define an element  $\int_M \alpha \in$

$\mathbb{C}[\mathcal{G}]^G$  by setting, for each  $\xi \in \mathcal{G}$

$$\left( \int_M \alpha \right) (\xi) = \int_M \alpha(\xi) \stackrel{\text{def}}{=} \int_M \alpha(\xi)_{[n]}$$

where  $n = \dim M$ . Note that  $\int_M \alpha$  really is  $G$ -invariant since

$$\begin{aligned} \left( \int_M \alpha \right) (g\xi g^{-1}) &= \int_M \alpha(g\xi g^{-1})_{[n]} = \int_M \sigma_{g^{-1}}^* (\alpha(\xi)_{[n]}) \\ &= \int_M \alpha(\xi)_{[n]} = \left( \int_M \alpha \right) (\xi). \end{aligned}$$

Notice also that if  $\alpha$  is  $d_G$ -exact (say,  $\alpha = d_G\beta$ ), then, for each  $\xi \in \mathcal{G}$ ,  $\alpha(\xi)_{[n]} = d(\beta(\xi)_{[n-1]})$  so, by Stokes' Theorem,  $\int_M \alpha = 0 \in \mathbb{C}[\mathcal{G}]^G$ . The conclusion is that

the integration map

$$\int_M : \Omega_G^*(M) \longrightarrow \mathbb{C}[\mathcal{G}]^G$$

descends to cohomology:

$$\int_M : H_G^*(M) \longrightarrow \mathbb{C}[\mathcal{G}]^G.$$

Now we return to the general development. We wish to write out a specific representative of an equivariant cohomology class called the “universal Thom form” (for vector bundles with typical fiber  $V$ ). For this we take  $M = V$  (our real vector space of dimension  $2k$  with a positive definite inner product) and  $G = \mathbb{S}\mathbb{O}(V)$  with its defining action on  $V$ . Thus, we seek an element of

$$\Omega_{\mathbb{S}\mathbb{O}(V)}^*(V) = [\mathbb{C}[\mathfrak{so}(V)] \otimes \Omega^*(V)]^{\mathbb{S}\mathbb{O}(V)}$$

and we will obtain it as the Berezin integral of an element of  $A \otimes \bigwedge V$ , where  $A = \mathbb{C}[\mathfrak{so}(V)] \otimes \Omega^*(V)$ . Recalling the notation introduced earlier,  $\{\psi^1, \dots, \psi^{2k}\}$  is a fixed orthonormal basis for  $V$ ,  $\{u_1, \dots, u_{2k}\}$  is its dual basis of coordinate functions on  $V$ ,  $\{\xi_1, \dots, \xi_n\}$  is a basis for the Lie algebra  $\mathcal{G} = \mathfrak{so}(V)$  and  $\{x^1, \dots, x^n\}$  is its dual basis, regarded as linear functions on  $\mathfrak{so}(V)$ , i.e., as elements of  $\mathbb{C}[\mathfrak{so}(V)]$ . Now define, for each  $\xi \in \mathfrak{so}(V)$ , a linear transformation

$$M_\xi : V \longrightarrow V$$

by

$$M_\xi(\psi) = \left. \frac{d}{dt} (\exp(t\xi) \cdot \psi) \right|_{t=0}$$

for each  $\psi \in V$ . Write  $M_a$  for  $M_{\xi_a}$  and notice that, for each  $\xi = x^a(\xi)\xi_a \in \mathfrak{so}(V)$ ,

$$M_\xi = x^a(\xi) M_a.$$

Furthermore, if  $(M_\xi)$  denotes the matrix of  $M_\xi$  relative to  $\{\psi^1, \dots, \psi^{2k}\}$ , then  $(M_\xi) \in \mathfrak{so}(2k)$  and

$$-\frac{1}{2} \sum_{l=1}^{2k} \psi^l x^a(\xi) M_a \psi^l = \frac{1}{2} \psi^\top (M_\xi) \psi$$

so (5.1) gives

$$\int e^{-\frac{1}{2} \sum_l \psi^l x^a (\xi) M_a \psi^l} \mathcal{D}\psi = Pf(M_\xi). \quad (5.21)$$

Now notice that  $-\frac{1}{2} \sum_l \psi^l x^a M_a \psi^l = \sum_l (-\frac{1}{2} x^a) \otimes 1 \otimes (\psi^l (M_a \psi^l))$  can be regarded as an element of  $\mathbb{C}[\mathfrak{so}(V)] \otimes \Omega^*(V) \otimes \wedge V$  and therefore so can

$$e^{-\frac{1}{2} \sum_l \psi^l x^a M_a \psi^l}.$$

We now intend to include this factor in the integrand on the left-hand side of (5.4) to obtain what is called the **Mathai-Quillen universal Thom form**  $\nu$  for  $V$ , defined by

$$\begin{aligned} \nu &= (2\pi)^{-k} \int e^{-\frac{1}{2} \|u\|^2 + i\psi^j du_j - \frac{1}{2} \sum_l \psi^l x^a M_a \psi^l} \mathcal{D}\psi \\ &= (2\pi)^{-k} e^{-\frac{1}{2} \|u\|^2} \int \exp\left(i\psi^j du_j - \frac{1}{2} \sum_l \psi^l x^a M_a \psi^l\right) \mathcal{D}\psi. \end{aligned} \quad (5.22)$$

For example, if one carries out the Berezin integration in (5.22) for  $V = \mathbb{R}^2$  (usual orientation and inner product) and  $\mathbb{S}\mathbb{O}(V) = SO(2)$  the result is

$$\nu = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} du_1 du_2 + (2\pi)^{-1} x^1 e^{-\frac{1}{2}(u_1^2 + u_2^2)}. \quad (5.23)$$

Note that each term in (5.23) is  $SO(2)$ -invariant, has degree  $2 = \dim V$  and the first (in  $\mathbb{C}^0[\mathfrak{so}(2)] \otimes \Omega^2(\mathbb{R}^2)$ ) integrates to 1 over  $\mathbb{R}^2$ . In general, one can verify the following properties of the form  $\nu$  given by (5.22).

1.  $\nu$  is an  $\mathbb{S}\mathbb{O}(V)$ -invariant element of  $\mathbb{C}[\mathfrak{so}(V)] \otimes \Omega^*(V)$  of degree  $2k = \dim V$ .
2.  $\nu$  is  $\mathbb{S}\mathbb{O}(V)$ -equivariantly closed ( $d_{\mathbb{S}\mathbb{O}(V)} \nu = 0$ ) and so determines an equivariant cohomology class

$$[\nu] \in H_{\mathbb{S}\mathbb{O}(V)}^{2k}(V).$$

3. The integral of (the  $\mathbb{C}^0[\mathfrak{so}(V)] \otimes \Omega^{2k}(V)$ -part of)  $\nu$  over  $V$  is 1.

$\nu$  is called a universal Thom form because, as we now show, one can produce from it a (Gaussian) representative of the Thom class for any vector bundle whose typical fiber is  $V$ . We recall that any such vector bundle can be regarded as the vector bundle associated to some principal  $G$ -bundle by a representation of  $G$  on  $V$ . Thus, let us suppose  $G$  is a Lie group and

$$\rho : G \longrightarrow \mathbb{S}\mathbb{O}(V)$$

is a representation of  $G$  on  $V$ . Then  $\rho$  induces a Lie algebra homomorphism

$$\rho_* : \mathcal{G} \longrightarrow \mathfrak{so}(V)$$

(just the derivative of  $\rho$  at the identity in  $G$ ). It then follows easily from the fact that  $\nu$  is an  $\mathbb{S}\mathbb{O}(V)$ -equivariantly closed form on  $V$  that

$$\nu_G = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp \left( i\psi^j du_j - \frac{1}{2} \sum_l \psi^l (x^a \circ \rho_*) M_a \psi^l \right) \mathcal{D}\psi \quad (5.24)$$

is a  $G$ -equivariantly closed form on  $V$ .

Now suppose that  $G \hookrightarrow P \xrightarrow{\pi_P} X$  is a principal  $G$ -bundle over a compact, smooth manifold of dimension  $2k = \dim V$ . Then the representation  $\rho$  determines an associated vector bundle  $P \times_\rho V \xrightarrow{\pi_\rho} X$  over  $X$  with typical fiber  $V$ . We show now that there is a generalization of the Chern-Weil map  $CW_\omega : \mathbb{C}[\mathcal{G}]^G \rightarrow \Omega^*(P)_{\text{BAS}} \cong \Omega^*(X)$  (Section 4) which associates with every  $G$ -equivariantly closed form on  $V$  an ordinary form on the vector bundle space  $P \times_\rho V$  and that, when applied to  $\nu_G$ , one obtains a (Gaussian) representative of the Thom class of  $P \times_\rho V$ .

Begin by considering the commutative diagram

$$\begin{array}{ccc} P \times V & \xrightarrow{\text{pr}_P} & P \\ q \downarrow & & \downarrow \pi_P \\ P \times_\rho V & \xrightarrow{\pi_\rho} & X \end{array}$$

where  $\text{pr}_P$  is the projection onto the first factor and  $q$  is the map

$$q(p, \psi) = q(p \cdot g, g^{-1} \cdot \psi) = [p, \psi]$$

which projects  $P \times V$  onto the orbit space  $P \times_\rho V$  of the  $G$ -action  $(p, \psi) \cdot g = (p \cdot g, g^{-1} \cdot \psi) = (p \cdot g, \rho(g^{-1})(\psi))$ . Since the action of  $G$  on  $P$  is free, so is this action of  $G$  on  $P \times V$  and we may regard

$$G \hookrightarrow P \times V \xrightarrow{q} P \times_\rho V$$

as a principal  $G$ -bundle. In particular, we have an isomorphism

$$\Omega^*(P \times V)_{\text{BAS}} \cong \Omega^*(P \times_\rho V)$$

between the algebras of ordinary forms on  $P \times_\rho V$  and the forms on  $P \times V$  that are basic with respect to the action of  $G$  on  $P \times V$ . Thus, to specify a form on  $P \times_\rho V$  (e.g., a Thom form) it is enough to specify a basic form on  $P \times V$ .

Now choose a connection  $\omega$  on  $G \hookrightarrow P \rightarrow X$ . Then  $\omega' = \text{pr}_P^* \omega$  is a connection on  $G \hookrightarrow P \times V \rightarrow P \times_\rho V$ . Identifying  $T_{(p,\psi)}(P \times V)$  with  $T_p(P) \oplus T_\psi(V)$ , the  $\omega'$ -horizontal spaces are clearly given by  $\text{Hor}_{(p,\psi)}(\omega') \cong \text{Hor}_p(\omega) \oplus T_\psi(V)$ . The decomposition

$$T_{(p,\psi)}(P \times V) \cong (\text{Hor}_p(\omega) \oplus T_\psi(V)) \oplus \text{Vert}_{(p,\psi)}(P \times V)$$

determines a projection  $h_{\omega'}$  of forms on  $P \times V$  to  $\omega'$ -horizontal forms on  $P \times V$  (evaluate on  $\omega'$ -horizontal parts of tangent vectors). Now, for any  $\alpha = \mathcal{P} \otimes \varphi \in \Omega_G^*(V)$  one can evaluate the polynomial part  $\mathcal{P}$  on the curvature  $\Omega$  of the connection  $\omega$  as in the ordinary Chern-Weil map to obtain  $\mathcal{P}(\Omega) \otimes \varphi \in \Omega^*(P) \otimes \Omega^*(V)$ . This one can identify with a form  $\mathcal{P}(\Omega) \wedge \varphi$  on  $P \times V$  which, because  $\mathcal{P} \otimes \varphi$  is  $G$ -invariant, is in  $\Omega^*(P \times V)^G$ . It is generally not horizontal, however, so we compose with the horizontal projection  $h_{\omega'}$  to define the **generalized Chern-Weil homomorphism**, also denoted

$$CW_\omega : \Omega_G^*(V) \longrightarrow \Omega^*(P \times V)_{BAS} \cong \Omega^*(P \times_\rho V)$$

by

$$CW_\omega(\alpha) = CW_\omega(\mathcal{P} \otimes \varphi) = h_{\omega'}(\mathcal{P}(\Omega) \wedge \varphi) \quad (5.25)$$

for elements of the form  $\mathcal{P} \otimes \varphi$  and then by linearity on all of  $\Omega_G^*(V)$ . One can show that  $CW_\omega$  is actually a cochain map

$$d \circ CW_\omega = CW_\omega \circ d_G \quad (5.26)$$

and so carries  $G$ -equivariantly closed forms on  $V$  to ordinary closed forms on  $P \times V$  which then descend to closed forms on  $P \times_\rho V$ . Applying this procedure to the  $G$ -equivariantly closed form  $\nu_G$  of (5.24) gives a closed, basic form  $CW_\omega(\nu_G)$  on  $P \times V$  which one can write formally as the horizontal projection of

$$(2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp \left( i\psi^j du_j - \frac{1}{2} \sum_l \psi^l (x^a (\rho_* \Omega)) M_a \psi^l \right) \mathcal{D}\psi \quad (5.27)$$

which it is customary to write more compactly as

$$(2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp \left( i\psi^T du + \frac{1}{2} \psi^T (\rho_* \Omega) \psi \right) \mathcal{D}\psi. \quad (5.28)$$

In this last expression  $(\rho_* \Omega)$  is to be interpreted as the skew-symmetric matrix image of the  $G$ -curvature under (the derivative of) the representation when  $\mathfrak{so}(V)$  is identified with  $\mathfrak{so}(2k)$ . We generally work directly with (5.28), but one obtains

a (Gaussian) representative of the Thom class (which pulls back to an Euler form) by replacing  $\Omega$  by  $-\Omega$ , or, what amounts to the same thing, the transpose of  $\Omega$ , cf. (4.26).

As an example of what the result of such a calculation might look like we return to the universal Thom form (5.23) for  $V = \mathbb{R}^2$  and  $\mathbb{S}\mathbb{O}(V) = SO(2)$ . For the vector bundle with fiber  $\mathbb{R}^2$  we take the tangent bundle  $T\mathbb{S}^2$  of the two-sphere. This we describe as an associated bundle in the following way. The usual orientation and Riemannian metric on  $\mathbb{S}^2$  give an oriented, orthonormal frame bundle

$$SO(2) \hookrightarrow F_{SO}(\mathbb{S}^2) \xrightarrow{\pi_{SO}} \mathbb{S}^2.$$

If  $\rho : SO(2) \rightarrow SO(2)$  is the identity representation, i.e.,  $\rho = \text{id}_{SO(2)}$ , then  $F_{SO}(\mathbb{S}^2) \times_{\rho} \mathbb{R}^2$  is the tangent bundle of  $\mathbb{S}^2$ . Moreover,  $\rho_* : \mathfrak{so}(2) \rightarrow \mathfrak{so}(2)$  is also the identity so  $x^1 \circ \rho_* = x^1$  and

$$\nu_{SO(2)} = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} (x^1 + du_1 du_2).$$

Choosing a connection  $\omega = \omega^1 \xi_1$  with curvature  $\Omega = \Omega^1 \xi_1$  on the frame bundle  $F_{SO}(\mathbb{S}^2)$  we have  $x^1(\rho_* \Omega) = \Omega^1$  so  $CW_{\omega}(\nu_{SO(2)})$  is the horizontal projection (determined by  $\omega' = \text{pr}_{F_{SO}(\mathbb{S}^2)}^* \omega$ ) of

$$\mu = (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} (\Omega^1 + du_1 du_2).$$

The horizontal projection of this form on  $F_{SO}(\mathbb{S}^2) \times \mathbb{R}^2$  can either be described by evaluating  $\mu$  on  $\omega'$ -horizontal parts or explicitly calculated from the easily verified formula  $h_{\omega'}(\mu) = \mu - (\text{pr}_{F_{SO}(\mathbb{S}^2)}^* \omega^1) \wedge \iota_{\xi_1} \mu$ . Performing this latter calculation gives

$$\begin{aligned} CW_{\omega}(\nu_{SO(2)}) &= (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} (\Omega^1 + du_1 du_2) \\ &+ (\text{pr}_{F_{SO}(\mathbb{S}^2)}^* \omega^1) \wedge (u_1 du_1 + u_2 du_2). \end{aligned} \quad (5.29)$$

In general, we will denote by  $U$  the horizontal projection of the form (5.28) and will, at least temporarily, write this as

$$U = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_* \Omega) \psi\right) \mathcal{D}\psi \quad (5.30)$$

(evaluated on horizontal parts).

Since our primary concern, however, is with Euler forms we will want to pull the form to which  $U$  descends on  $P \times_{\rho} V$  back by a section of the vector bundle.



Now, any section of the associated bundle  $P \times_{\rho} V$  can be written as

$$x \xrightarrow{(s, S \circ s)} (s(x), S(s(x))) \xrightarrow{q} [s(x), S(s(x))]$$

where  $s$  is a section of  $G \hookrightarrow P \rightarrow X$  and  $S : P \rightarrow V$  is an equivariant map ( $S(p \cdot g) = \rho(g^{-1})(s(p))$ ). Thus, if we temporarily write  $U = q^* \tilde{U}$ , then

$$(q \circ (s, S \circ s))^* \tilde{U} = ((1, S) \circ s)^* (q^* \tilde{U}) = s^* ((1, S)^* U).$$

Thus, to pull  $\tilde{U}$  back by a section of  $P \times_{\rho} V$  we compute  $(1, S)^* U$ , which simply pulls the  $V$ -factors of  $U$  back by the equivariant map  $S$ , and then pull this form on  $P$  back by a section of the principal bundle. We will illustrate with an example (taken from [28]). Begin with the form  $CW_{\omega}(\nu_{SO(2)})$  in (5.29). As a section of the principal bundle  $F_{SO}(\mathbb{S}^2)$  we choose the oriented, orthonormal frame field corresponding to the spherical coordinate chart:

$$s(\phi, \theta) = \left( \phi, \theta, \partial_{\phi}, \frac{1}{\sin \phi} \partial_{\theta} \right).$$

We choose an equivariant map  $F_{SO}(\mathbb{S}^2) \xrightarrow{S} \mathbb{R}^2$  by beginning with a vector field on  $\mathbb{S}^2$  (section of  $T\mathbb{S}^2$ ). This can be chosen arbitrarily and we will take  $V = \gamma \sin \theta \partial_{\phi} + \gamma \cos \theta \cot \phi \partial_{\theta}$ , where  $\gamma$  is an arbitrary real parameter (with  $\gamma = 1$  this is the infinitesimal generator for rotations about the  $x$ -axis). Relative to the frame field introduced above the components of  $V$  are  $\gamma \sin \theta$  and  $\gamma \cos \theta \cot \phi$  so we define  $S$  on the image of  $s$  by

$$S \circ s(\phi, \theta) = (\gamma \sin \theta, \gamma \cos \theta \cot \phi)$$

and elsewhere by equivariance. Pull back the  $\mathbb{R}^2$ -parts of  $CW_{\omega}(\nu_{SO(2)})$  by  $S \circ s$  by substituting  $u_1 = \gamma \sin \theta$  and  $u_2 = \gamma \cos \theta \cot \phi$ . One finds that

$$\begin{aligned} (2\pi)^{-1} e^{-\frac{1}{2}(u_1^2 + u_2^2)} &= (2\pi)^{-1} e^{-\frac{1}{2}\gamma^2(\sin^2 \theta + \cos^2 \theta \cot^2 \phi)} \\ du_1 du_2 &= du_1 \wedge du_2 = \gamma^2 \cos^2 \theta \sin \phi d\phi \wedge d\theta \end{aligned}$$

and

$$u_1 du_1 + u_2 du_2 = \gamma^2 (\sin \theta \cos \phi \sin^2 \phi d\theta - \cos^2 \theta \sin \phi \cot \phi d\phi).$$

As in our earlier computation of the Euler number of  $T\mathbb{S}^2$  we substitute the (transposed) Levi-Civita connection  $s^* \omega$  and curvature  $s^* \Omega$  so that  $s^* \Omega^1 = \sin \phi d\phi \wedge$

$d\theta$  and  $s^*(\text{pr}_{F_{SO}^*}^*(\mathbb{S}^2)\omega^1) = -\cos\phi d\theta$ . The result of all of these substitutions is the following representative of the Euler class of  $\mathbb{S}^2$ :

$$(2\pi)^{-1} e^{-\frac{1}{2}\gamma^2(\sin^2\theta+\cos^2\theta\cos^2\phi)} \sin\phi (1 + \gamma^2 \cos^2\theta \sin^2\phi) d\phi \wedge d\theta. \quad (5.31)$$

In particular, one obtains the not altogether obvious integral formula

$$\frac{1}{2\pi} \int_0^{2\pi} \int_0^\pi e^{-\frac{1}{2}\gamma^2(\sin^2\theta+\cos^2\theta\cos^2\phi)} \sin\phi (1 + \gamma^2 \cos^2\theta \sin^2\phi) d\phi d\theta = 2$$

(note that, for  $\gamma = 0$ , this reduces to our earlier computation of the Euler characteristic of  $\mathbb{S}^2$ ).

We recall now that our interest in the Mathai-Quillen formalism stems from the fact that Atiyah and Jeffrey [3] have shown how it can be adapted and formally applied to the infinite-dimensional vector bundle  $\hat{\mathcal{A}} \times_{\hat{G}} \Omega_+^2(B, \text{ad}P)$  to yield an integral representation of the 0-dimensional Donaldson invariant which coincides with the partition function of Witten's topological quantum field theory [46]. We begin now with the appropriate adaptation of the formula (5.30), still working in the finite-dimensional context. In addition to the assumptions we have made thus far we will henceforth assume that  $P$  is oriented and that the action of  $G$  on  $P$  preserves orientation (i.e., each of the diffeomorphisms  $\sigma_g$  is orientation preserving). We now make a specific choice of connection on  $G \hookrightarrow P \rightarrow X$ .

**Remark.** *Before proceeding we must recall that for any action of a compact Lie group  $G$  on a manifold  $M$  it is always possible to construct a Riemannian metric  $\langle \cdot, \cdot \rangle_G$  on  $M$  that is  $G$ -invariant, i.e., for which the diffeomorphisms  $\sigma_g: M \rightarrow M$  are all isometries. Roughly, this is done by selecting some Riemannian metric  $\langle \cdot, \cdot \rangle$  on  $M$  and, at each point  $p \in M$ , averaging over  $G$  relative to some invariant measure  $dG$  on  $G$ , i.e., defining, for all  $V_p, W_p \in T_p(M)$ ,*

$$\langle V_p, W_p \rangle_G = \int_G \langle (\sigma_g)_* V_p, (\sigma_g)_* W_p \rangle dG.$$

We fix, once and for all, a  $G$ -invariant Riemannian metric, denoted simply  $\langle \cdot, \cdot \rangle$ , on  $P$ . At each  $p \in P$  this Riemannian metric defines an orthogonal complement to the vertical subspace of  $T_p(P)$  (tangent space to the  $G$ -orbit at  $p$ ) and, since  $G$  acts by isometries, these orthogonals are invariant under the action of  $G$  and so they determine a connection  $\omega$  on  $P$ . Henceforth, we will use this connection on

$P$  exclusively. In particular, the pullback connection  $\omega' = \text{pr}_P^* \omega$  on  $P \times V$  has  $\text{Hor}_{(p,\psi)}(\omega') \cong T_p(p \cdot G)^\perp \oplus T_\psi(V)$  at each point.

Now we proceed with some cosmetic surgery on (5.30). First recall the Cartan formula  $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$  and notice that the second term vanishes on  $\omega$ -horizontal vectors by definition. Since the formula for  $U$  in (5.30) is to be evaluated on  $\omega'$ -horizontal parts and since  $\text{Hor}_{(p,\psi)}(\omega') \cong \text{Hor}_p(\omega) \oplus T_\psi(V)$ , the result will be the same whether or not  $\frac{1}{2}[\omega, \omega]$  is present. Thus, we may write

$$U = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(d\omega)) \psi\right) \mathcal{D}\psi \quad (5.32)$$

(evaluated on horizontal parts).

For the next manipulation of  $U$  we will require a few preliminaries. Begin by defining, at each,  $p \in P$ , a linear map

$$C_p : \mathcal{G} \longrightarrow \text{Vert}_p(P) \subseteq T_p(P)$$

by

$$C_p(\xi) = \xi^\#(p) = \frac{d}{dt} (p \cdot \exp(t\xi))|_{t=0}.$$

This is an isomorphism onto  $\text{Vert}_p(P)$ , but we wish to regard it as a map into  $T_p(P)$ . Now choose some ad-invariant inner product  $(\ , \ )$  on  $\mathcal{G}$ .  $T_p(P)$  has an inner product  $\langle \ , \ \rangle_p$  arising from the Riemannian metric on  $P$ . Thus,  $C_p$  has an adjoint

$$C_p^* : T_p(P) \longrightarrow \mathcal{G}$$

defined by

$$\langle w, C_p(\eta) \rangle_p = (C_p^*(w), \eta)$$

for all  $w \in T_p(P)$  and  $\eta \in \mathcal{G}$ . In particular,

$$\langle C_p(\xi), C_p(\eta) \rangle_p = (C_p^*(C_p(\xi)), \eta) = (R_p(\xi), \eta)$$

where

$$R_p = C_p^* \circ C_p : \mathcal{G} \longrightarrow \mathcal{G}.$$

It is easy to see that  $R_p$  is self-adjoint and has trivial kernel so we have an inverse

$$R_p^{-1} : \mathcal{G} \longrightarrow \mathcal{G}.$$

Now, since  $C_p$  carries  $\mathcal{G}$  isomorphically onto  $\text{Vert}_p(P)$  there is also an inverse

$$C_p^{-1} : \text{Vert}_p(P) \longrightarrow \mathcal{G}$$

and we claim that this agrees with  $\omega_p$  on  $\text{Vert}_p(P)$ , i.e.,

$$C_p^{-1}(w) = \omega_p(w), w \in \text{Vert}_p(P). \quad (5.33)$$

Indeed, if  $w \in \text{Vert}_p(P)$ , then  $w = \eta^\#(p) = C_p(\eta)$  for a unique  $\eta \in \mathcal{G}$ .  $\omega$  is a connection form so  $\omega_p(\eta^\#(p)) = \eta$  for every  $\eta \in \mathcal{G}$ . Thus,  $\omega_p(w) = \eta = C_p^{-1}(w)$ .

Now we define a one-form  $\theta \in \Omega^1(P, \mathcal{G}^*)$  on  $P$  with values in the dual  $\mathcal{G}^*$  of  $\mathcal{G}$  as follows: For  $p \in P$  and  $w \in T_p(P)$ ,  $\theta_p(w) \in \mathcal{G}^*$  is the map  $\theta_p(w) : \mathcal{G} \rightarrow \mathbb{R}$  given by

$$(\theta_p(w))(\xi) = \langle C_p(\xi), w \rangle_p = (\xi, C_p^*(w)). \quad (5.34)$$

Note that  $\theta_p$  vanishes on horizontal vectors at  $p$  because  $\omega$ -horizontal means  $\langle \cdot, \cdot \rangle$ -orthogonal to the  $G$ -orbit through  $p$ , i.e., to  $\text{Vert}_p(P)$ , and  $C_p(\xi)$  is in  $\text{Vert}_p(P)$ . Now, if we use the inner product  $(\cdot, \cdot)$  to identify  $\mathcal{G}^*$  and  $\mathcal{G}$ , the last equality in (5.34) shows that  $\theta_p(w)$  is identified with  $C_p^*(w)$ . Thus, regarded as a  $\mathcal{G}$ -valued one-form on  $P$ ,  $\theta$  is just  $C^*$  so, in particular,  $C^* \in \Omega^1(P, \mathcal{G})$  vanishes on horizontal vectors. We claim that

$$C^* = R \circ \omega \quad (5.35)$$

i.e.,  $C_p^*(w) = R_p(\omega_p(w))$  for every  $w \in T_p(P)$ . Since both sides vanish on horizontal vectors one need only verify (5.35) when  $w \in \text{Vert}_p(P)$ . But then (5.33) gives

$$R_p(\omega_p(w)) = R_p(C_p^{-1}(w)) = (C_p^* \circ C_p)(C_p^{-1}(w)) = C_p^*(w).$$

Fixing a basis for  $\mathcal{G}$  we can identify each  $R_p$  with an invertible matrix and  $\omega$  with a matrix of real-valued one-forms on  $P$  so (5.35) becomes a matrix equation  $C^* = R\omega$  which we write as

$$\omega = R^{-1}C^*.$$

From this we compute

$$d\omega = dR^{-1} \wedge C^* + R^{-1}dC^*.$$

Noting that the first term vanishes on horizontal vectors we find that in the expression (5.32) for  $U$  we may replace  $d\omega$  by  $R^{-1}dC^*$  to obtain

$$U = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_* (R^{-1}dC^*)) \psi\right) \mathcal{D}\psi \quad (5.36)$$

(evaluated on horizontal parts).

The next objective is to remove the explicit appearance of the inverse in (5.36) by using the Fourier inversion formula. We begin with a brief review of the Fourier transform. Let  $W$  be an oriented real vector space of dimension  $n$  with volume element  $dw \in \bigwedge^n W^*$  and let  $w_1, \dots, w_n$  be coordinates on  $W$  with  $dw = dw_1 \cdots dw_n$ . Let  $y_1, \dots, y_n$  be coordinates on  $W^*$  dual to  $w_1, \dots, w_n$  and  $dy = dy_1 \cdots dy_n \in \bigwedge^n W$  the volume element for  $W^*$ . Let  $S(W)$  and  $S(W^*)$  be the Schwartz spaces of rapidly decreasing functions in  $w_1, \dots, w_n$  and  $y_1, \dots, y_n$ , respectively. Finally, let  $\langle \cdot, \cdot \rangle$  denote the natural pairing between  $W$  and  $W^*$ . The Fourier transform of  $f \in S(W)$  is  $\hat{f} \in S(W^*)$  defined by

$$\hat{f}(y) = (2\pi)^{-n/2} \int_W e^{-i\langle w, y \rangle} f(w) dw.$$

The Fourier inversion formula then asserts that

$$f(w) = (2\pi)^{-n/2} \int_{W^*} e^{i\langle w, y \rangle} \hat{f}(y) dy.$$

Combining these two formulas gives

$$f(w) = (2\pi)^{-n} \int_{W^*} \int_W e^{i\langle w, y \rangle} e^{-i\langle z, y \rangle} f(z) dz dy.$$

Assuming now that  $W$  and  $W^*$  are identified via some inner product we will write this simply as

$$f(w) = (2\pi)^{-n} \iint e^{i\langle w, y \rangle} e^{-i\langle z, y \rangle} f(z) dz dy \tag{5.37}$$

with the understanding that both integrations are over  $W$  and the exponents are inner products.

The situation to which we would like to apply (5.37) is as follows. If  $R$  is a positive self-adjoint matrix, then one can use the formula to compute  $f(R^{-1}w)$ . To get an integral that does not explicitly involve the inverse, however, we also make the change of variable  $y \rightarrow Ry$ . Then  $\langle R^{-1}w, Ry \rangle = \langle w, y \rangle$  and  $d(Ry) = \det R dy$  so

$$f(R^{-1}w) = (2\pi)^{-n} \iint e^{i\langle w, y \rangle} e^{-i\langle z, Ry \rangle} f(z) \det R dz dy. \tag{5.38}$$

Now we return to our last expression (5.36) for  $U$ . Letting  $\phi = (\phi_1, \dots, \phi_n)$  denote a Lie algebra variable in  $\mathcal{G}$  we consider the function on  $\mathcal{G}$  defined by

$$f(\phi) = (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \int \exp\left(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(\phi)) \psi\right) \mathcal{D}\psi.$$

Each value of  $f$  is an element of  $\Omega^*(V)$  whose components (relative to  $du_1, \dots, du_k$ ) are polynomials in  $\phi$ .

**Remark.** *These polynomials are not in the Schwartz space  $S(\mathcal{G})$ , but we nevertheless propose to apply the Fourier formula (5.38) componentwise to  $f(\phi)$ . This is rather sloppy, of course, but could be made more precise by inserting a rapidly decaying test function  $e^{-\epsilon(\phi, \phi)}$  and taking the limit as  $\epsilon \rightarrow 0$ . Since our objective is a formula to be applied formally in an infinite-dimensional situation where complete rigor is (for the time being, at least) out of the question anyway, we will not be scrupulous about such details.*

Letting  $\lambda = (\lambda_1, \dots, \lambda_n)$  be another Lie algebra variable in  $\mathcal{G}$  we apply (5.38) with  $w = dC^*$  (i.e., with  $w = dC^*(\chi_1, \chi_2)$  for each pair  $(\chi_1, \chi_2)$  of tangent vectors) to get

$$\begin{aligned}
 U &= f(R^{-1}dC^*) = (2\pi)^{-n} \iint e^{i(dC^*, \lambda)} e^{-i(\phi, R\lambda)} f(\phi) \det R d\phi d\lambda \\
 &= (2\pi)^{-n} (2\pi)^{-k} \iint e^{i(dC^*, \lambda)} e^{-i(\phi, R\lambda)} e^{-\frac{1}{2}\|u\|^2} \int \exp(i\psi^T du \\
 &\quad + \frac{1}{2}\psi^T (\rho_*(\phi)) \psi) \mathcal{D}\psi \det R d\phi d\lambda \\
 U &= (2\pi)^{-n} (2\pi)^{-k} e^{-\frac{1}{2}\|u\|^2} \iiint \exp(i\psi^T du + \frac{1}{2}\psi^T (\rho_*(\phi)) \psi \\
 &\quad + i(dC^*, \lambda) - i(\phi, R\lambda)) \det R \mathcal{D}\psi d\lambda d\phi \text{ (evaluated on horizontal parts)}.
 \end{aligned} \tag{5.39}$$

Notice that this expression contains one fermionic and two ordinary (“bosonic”) integrations.

Next we would like to include the parenthetical remark “evaluated on horizontal parts” in (5.39) directly into the integral expression for  $U$ . For this we require the notion of a “normalized vertical volume form” on a principal bundle, which is essentially an analogue of a Thom form on a vector bundle. Thus, we consider a principal  $G$ -bundle  $G \hookrightarrow Q \xrightarrow{\pi_Q} M$  with  $M$  compact and orientable and  $\dim G = n$ . We assume the bundle itself to be orientable in the sense that there exists an  $n$ -form  $\Psi$  on  $Q$  such that, if  $m \in M$  and  $\iota_m : \pi_Q^{-1}(m) \hookrightarrow Q$  is the inclusion, then  $\iota_m^* \Psi$  is an orientation form for the submanifold  $\pi_Q^{-1}(m) \cong G$ . It follows that  $Q$  is orientable and we assume it to be oriented by the so-called local product orientation  $\pi_Q^* \omega_M \wedge \Psi$ , where  $\omega_M$  is an orientation form for  $M$ . One

can assume also that the action of  $G$  on  $Q$  is orientation preserving (these matters are discussed in detail in Chapter VII, Vol I, of [24]). We will henceforth make these assumptions of our underlying principal bundle  $G \hookrightarrow P \xrightarrow{\pi_P} X$  as well. A **normalized vertical volume form** for the bundle is an  $n$ -form  $\mathbf{W}$  on  $Q$  such that, if  $\iota_m : \pi_Q^{-1}(m) \hookrightarrow Q$  is the inclusion of a fiber, then

$$\int_{\pi_Q^{-1}(m)} \iota_m^* \mathbf{W} = 1. \tag{5.40}$$

It is not difficult to show that one can construct such a form  $\mathbf{W}$  from a connection  $\omega$  on  $Q$  as follows. Choose a positive definite ad-invariant inner product  $(\ , \ )$  on  $\mathcal{G}$ , normalized so that the volume of  $G$  (arising from the corresponding bi-invariant Riemannian metric on  $G$ ) is 1. Let  $\{\xi_1, \dots, \xi_n\}$  be an orthonormal basis for  $\mathcal{G}$  relative to  $(\ , \ )$  and consistent with the orientation  $G$  inherits as a fiber of  $Q$ . Write  $\omega = \omega^a \xi_a$ , where  $\omega^a \in \Omega^1(Q)$ ,  $a = 1, \dots, n$ . One then shows that

$$\mathbf{W} = \omega^1 \wedge \dots \wedge \omega^n \tag{5.41}$$

is a normalized vertical volume form for  $G \hookrightarrow Q \xrightarrow{\pi_Q} M$ . Such a form has a number of properties of interest to us. For any top rank form  $\beta$  on  $M$  one has

$$\int_M \beta = \int_Q \pi_Q^* \beta \wedge \mathbf{W} \tag{5.42}$$

(essentially Fubini's Theorem together with (5.40)). Furthermore, the process of evaluating an element  $\varphi$  of  $\Omega^*(Q)$  on  $\omega$ -horizontal parts (i.e., of computing the  $\omega$ -horizontal projection  $h_\omega(\varphi)$  of  $\varphi$ ) can be accomplished as follows. An explicit formula for  $h_\omega(\varphi)$  reads

$$\begin{aligned} h_\omega(\varphi) &= \varphi - \omega^1 \wedge \iota_1 \varphi - \omega^2 \wedge \iota_2 \varphi - \dots - \omega^n \wedge \iota_n \varphi \\ &+ \sum_{\substack{1 \leq a_1 < \dots < a_r \leq n \\ r > 1}} (-1)^{r(r+1)/2} \omega^{a_1} \wedge \dots \wedge \omega^{a_r} \wedge (\iota_{a_1} \circ \dots \circ \iota_{a_r})(\varphi) \end{aligned} \tag{5.43}$$

where  $\iota_a = \iota_{\xi_a^\#}$ . Thus  $h_\omega(\varphi) \wedge \mathbf{W} = \varphi \wedge \mathbf{W}$  (by (5.41)) so one can arrive at  $h_\omega(\varphi)$  by wedge multiplying  $\varphi$  by  $\mathbf{W}$  and integrating over the fibers to obtain

$$h_\omega(\varphi) \int_{\pi_Q^{-1}(m)} \iota_m^* \mathbf{W} = h_\omega(\varphi).$$

Notice that what is really going on here is that  $\varphi \wedge \mathbf{W}$  kills all of the terms in  $\varphi$  with vertical parts (because  $\mathbf{W}$  has a full contingent of  $n$  vertical coordinate differentials) and the surviving terms are just those of  $h_\omega(\varphi)$  with an extra factor of  $\mathbf{W}$  which we integrate out (i.e., ignore). The conclusion is that “evaluating on  $\omega$ -horizontal parts” can be accomplished by “multiplying by the vertical volume form (5.41) and integrating over the fibers”.

We wish to apply this observation to the form  $U$  of (5.39), where the principal bundle is  $G \hookrightarrow P \times V \xrightarrow{q} P \times_\rho V$ . However, we would like to include the procedure as part of the integration formula so we begin by showing that  $\mathbf{W}$  can be written as a Berezin integral.

Denote by  $\{\eta_1, \dots, \eta_n\}$  an orthonormal basis for  $\mathcal{G}$  relative to the normalized  $G$ -invariant inner product  $(\ , \ )$  on  $\mathcal{G}$  introduced above (we will be more specific about the choice of this basis shortly). Regard these as odd generators of  $\wedge(\mathcal{G})$  and consider the following element of  $\Omega^*(Q) \otimes \wedge(\mathcal{G})$ :

$$e^{\sum_{a=1}^n \omega_a \eta_a} = e^{\omega_1 \eta_1} \dots e^{\omega_n \eta_n} = (1 + \omega_1 \eta_1) \dots (1 + \omega_n \eta_n). \quad (5.44)$$

Performing a Berezin integration with respect to  $\eta$  gives

$$\begin{aligned} \int e^{\sum_{a=1}^n \omega_a \eta_a} \mathcal{D}\eta &= \int (1 + \omega_1 \eta_1) \dots (1 + \omega_n \eta_n) \mathcal{D}\eta = \int \omega_1 \eta_1 \dots \omega_n \eta_n \mathcal{D}\eta \\ &= (-1)^{n(n-1)/2} \int \omega_1 \dots \omega_n \eta_1 \dots \eta_n \mathcal{D}\eta. \end{aligned}$$

If we now choose  $\{\eta_1, \dots, \eta_n\}$  to be the same as  $\{\xi_1, \dots, \xi_n\}$  if  $n(n-1)/2$  is even and an odd permutation of  $\{\xi_1, \dots, \xi_n\}$  if  $n(n-1)/2$  is odd, this gives

$$\int e^{\sum_{a=1}^n \omega_a \eta_a} \mathcal{D}\eta = \omega_1 \dots \omega_n = \omega_1 \wedge \dots \wedge \omega_n = \mathbf{W} \quad (5.45)$$

for the normalized vertical volume form of  $G \hookrightarrow Q \xrightarrow{\pi_Q} M$ .

We would now like to express the Berezin integral representation (5.45) for  $\mathbf{W}$  without explicit reference to the connection forms  $\omega_a$ . For this we assume, as for  $G \hookrightarrow P \xrightarrow{\pi_P} X$  earlier, that  $Q$  is a Riemannian manifold and that  $\omega$  is the connection on  $Q$  whose horizontal spaces are the orthogonal complements to the  $G$ -orbits. Thus, we have available the maps  $C$ ,  $C^*$  and  $R$  and all of the results we have proved about them (note that this is true of the bundle  $G \hookrightarrow P \times V \xrightarrow{q} P \times_\rho V$  when the metric, connection and orientation are taken to be the pullbacks by  $\text{pr}_P : P \times V \rightarrow P$  of those we have chosen for  $P$ ). Now consider the element



of  $\mathbb{C}^1[\mathcal{G}] \otimes \Omega^1(Q)$  whose value at any  $\eta \in \mathcal{G}$  is the one-form  $(C^*, \eta)$  on  $Q$  defined by

$$(C^*, \eta)(\chi) = (C^* \chi, \eta) = \langle \chi, C\eta \rangle = \langle \chi, \eta^\# \rangle \quad (5.46)$$

for any vector field  $\chi$  on  $Q$ . Now let  $\{A^1, \dots, A^n, \dots, A^d\}$  be local coordinates on  $Q$  (where  $n = \dim G$ ). Then

$$(C^*, \eta) = \sum_{j=1}^d \left( C^* \left( \frac{\partial}{\partial A^j} \right), \eta \right) dA^j.$$

Write  $C^* \left( \frac{\partial}{\partial A^j} \right) = \sum_{i=1}^n a_{ij} \eta_i$  so that

$$(C^*, \eta) = \sum_{j=1}^d \sum_{i=1}^n a_{ij} (\eta_i, \eta) dA^j = \sum_{i=1}^n \left( \sum_{j=1}^d a_{ij} dA^j \right) (\eta_i, \eta).$$

Define one-forms  $\beta_i$  on  $Q$  by  $\beta_i = \sum_{j=1}^d a_{ij} dA^j$ . Then

$$(C^*, \eta) = \sum_{i=1}^n \beta_i (\eta_i, \eta) = \sum_{i=1}^n \beta_i (\eta_i, \cdot) (\eta) = \sum_{i=1}^n \beta_i \eta^i (\eta)$$

where  $\{\eta^1, \dots, \eta^n\} = \{(\eta_1, \cdot), \dots, (\eta_n, \cdot)\}$  is the basis for  $\mathcal{G}^*$  dual to  $\{\eta_1, \dots, \eta_n\}$ . Now identify  $\sum_{i=1}^n \beta_i \eta^i \in \Omega^*(Q) \otimes \Lambda(\mathcal{G}^*)$  with  $\sum_{i=1}^n \beta_i \eta_i \in \Omega^*(Q) \otimes \Lambda(\mathcal{G})$  and compute the Berezin integral

$$\begin{aligned} \int e^{\sum_{i=1}^n \beta_i \eta_i} \mathcal{D}\eta &= (-1)^{n(n-1)/2} \beta_1 \wedge \dots \wedge \beta_n \\ &= (-1)^{n(n-1)/2} \left( a_{11} dA^1 + \dots + a_{1d} dA^d \right) \wedge \dots \wedge \left( a_{n1} dA^1 + \dots + a_{nd} dA^d \right) \\ &= (-1)^{n(n-1)/2} \sum_H \det(a_H) dA^H \end{aligned}$$

where  $H = \{h_1, \dots, h_n\}$ ,  $1 \leq h_1 < \dots < h_n \leq d$ ,  $dA^H = dA^{h_1} \wedge \dots \wedge dA^{h_n}$  and

$$a_H = \begin{pmatrix} a_{1h_1} & \cdots & a_{1h_n} \\ \vdots & & \vdots \\ a_{nh_1} & \cdots & a_{nh_n} \end{pmatrix}.$$

We compute the determinants  $\det(a_H)$  as follows. For each  $j = 1, \dots, n$ ,  $C^* \left( \frac{\partial}{\partial A^j} \right) = \sum_{i=1}^n a_{ij} \eta_i$  and so

$$a_{ij} = \left( C^* \left( \frac{\partial}{\partial A^j} \right), \eta_i \right) = \left( R \left( \omega \left( \frac{\partial}{\partial A^j} \right) \right), \eta_i \right)$$

by (5.35). Now let us write  $\omega = \tilde{\omega}^1 \eta_1 + \cdots + \tilde{\omega}^n \eta_n$ . By the way we have chosen  $\{\eta_1, \dots, \eta_n\}$ ,  $\{\tilde{\omega}^1, \dots, \tilde{\omega}^n\}$  is at worst a permutation of  $\{\omega^1, \dots, \omega^n\}$  and, in any case,

$$\tilde{\omega}^1 \wedge \cdots \wedge \tilde{\omega}^n = (-1)^{n(n-1)/2} \omega^1 \wedge \cdots \wedge \omega^n = (-1)^{n(n-1)/2} \mathbf{W}.$$

Now we have

$$\omega \left( \frac{\partial}{\partial A^j} \right) = \tilde{\omega}^1 \left( \frac{\partial}{\partial A^j} \right) \eta_1 + \cdots + \tilde{\omega}^n \left( \frac{\partial}{\partial A^j} \right) \eta_n$$

and

$$R \left( \omega \left( \frac{\partial}{\partial A^j} \right) \right) = \tilde{\omega}^1 \left( \frac{\partial}{\partial A^j} \right) R(\eta_1) + \cdots + \tilde{\omega}^n \left( \frac{\partial}{\partial A^j} \right) R(\eta_n)$$

so that

$$a_{ij} = \tilde{\omega}^1 \left( \frac{\partial}{\partial A^j} \right) (R(\eta_1), \eta_i) + \cdots + \tilde{\omega}^n \left( \frac{\partial}{\partial A^j} \right) (R(\eta_n), \eta_i).$$

Consequently,

$$a_H = \begin{pmatrix} (R(\eta_1), \eta_1) \cdots (R(\eta_n), \eta_1) \\ \vdots \\ (R(\eta_1), \eta_n) \cdots (R(\eta_n), \eta_n) \end{pmatrix} \begin{pmatrix} \tilde{\omega}^1 \left( \frac{\partial}{\partial A^{h_1}} \right) \cdots \tilde{\omega}^1 \left( \frac{\partial}{\partial A^{h_n}} \right) \\ \vdots \\ \tilde{\omega}^n \left( \frac{\partial}{\partial A^{h_1}} \right) \cdots \tilde{\omega}^n \left( \frac{\partial}{\partial A^{h_n}} \right) \end{pmatrix}. \quad (5.47)$$

Thus,

$$\det(a_H) = \det R \det(\tilde{\omega}_H) \quad (5.48)$$

where  $\tilde{\omega}_H$  is the matrix shown in (5.47). Since

$$\tilde{\omega}^j = \sum_{i=1}^d \tilde{\omega}^j \left( \frac{\partial}{\partial A^i} \right) dA^i, \quad j = 1, \dots, n$$

we find that

$$\begin{aligned}
 \int e^{\sum_{i=1}^n \beta_i \eta_i} \mathcal{D}\eta &= (-1)^{n(n-1)/2} \sum_H \det(a_H) dA^H \\
 &= (-1)^{n(n-1)/2} \sum_H \det R \det(\tilde{\omega}_H) dA^H \\
 &= (-1)^{n(n-1)/2} \det R \sum_H \det(\tilde{\omega}_H) dA^H \\
 &= (-1)^{n(n-1)/2} \det R \tilde{\omega}^1 \wedge \dots \wedge \tilde{\omega}^n \\
 &= (\det R) \mathbf{W}.
 \end{aligned}$$

Thus,

$$\mathbf{W} = (\det R)^{-1} \int e^{\sum_{i=1}^n \beta_i \eta_i} \mathcal{D}\eta. \quad (5.49)$$

We have already seen that, with our identification of  $\mathcal{G}^*$  with  $\mathcal{G}$  via  $(\cdot, \cdot)$ ,  $(C^*, \cdot)$  is identified with  $\sum_{i=1}^n \beta_i \eta_i$ , i.e.,  $(C^*, \eta) = \sum_{i=1}^n \beta_i \eta^i(\eta)$  for each  $\eta \in \mathcal{G}$ . Substituting this into the Berezin integral (5.49) it is best (notationally) to retain reference to the fermionic variable  $\eta$  and write

$$\mathbf{W} = (\det R)^{-1} \int e^{(C^*, \eta)} \mathcal{D}\eta. \quad (5.50)$$

Now we return to the expression (5.39) for  $U$  and enforce the horizontal projection by multiplying by  $\mathbf{W}$  in the form (5.50). In this way the  $\det R$  cancels and we simply add the term  $(C^*, \eta)$  to the exponent to obtain

$$\begin{aligned}
 U &= (2\pi)^{-n} (2\pi)^{-k} \iiint \exp \left( -\frac{1}{2} \|u\|^2 + i\psi^T du + \frac{1}{2} \psi^T (\rho_*(\phi)) \psi \right. \\
 &\quad \left. + i(dC^*, \lambda) - i(\phi, R\lambda) + (C^*, \eta) \right) \mathcal{D}\eta \mathcal{D}\psi d\lambda d\phi.
 \end{aligned} \quad (5.51)$$

**Remarks.** *One should keep in mind that we are here applying the result (5.50) to the principal bundle  $G \hookrightarrow P \times V \xrightarrow{q} P \times_\rho V$ . Moreover, in (5.51) there is also an implicit integration over the fibers to remove the vertical volume form after it has served its purpose of killing the vertical parts.*

The form  $U$  in (5.51), after integrating out the vertical parts, is a basic form on  $P \times V$  which, regarded as a form on  $P \times_\rho V$ , represents the Thom class. Pulling  $U$  back by a section of  $P \times_\rho V$  gives a representative of the Euler class which,

when integrated over  $X$ , gives the Euler number. We have already observed that every section of  $P \times_{\rho} V$  is of the form  $x \rightarrow [s(x), S(s(x))]$ , where  $s$  is a section of  $P$  and  $S : P \rightarrow V$  is an equivariant map and that pulling back by such a section amounts to pulling back the  $V$ -factors of  $\mathbf{U}$  by  $S$  and then pulling back the resulting form on  $P$  by  $s$ . Thus, our Euler form is the pullback by  $s$  of

$$(2\pi)^{-n} (2\pi)^{-k} \iiint \exp \left( -\frac{1}{2} \|S\|^2 + i\psi^T dS + \frac{1}{2} \psi^T (\rho_*(\phi)) \psi \right. \\ \left. + i(dC^*, \lambda) - i(\phi, R\lambda) + (C^*, \eta) \right) \mathcal{D}\eta \mathcal{D}\psi d\lambda d\phi \quad (5.52)$$

where we have written simply  $S$  for  $S^*u = u \circ S$ . Integrating this over  $X$  gives the Euler number. On the other hand, if we refrain from pulling back by  $s$  (and from integrating out the vertical volume form), (5.52) gives a form on  $P$  whose integral over  $P$  is also the Euler number of  $P \times_{\rho} V$ .

We have one last bit of cosmetic surgery to perform on  $\mathbf{U}$ . There is a common notational device in (supersymmetric) physics whereby the integral of a top rank form on a manifold is written as two successive integrations, one fermionic and one bosonic. Recall that the integral of a (properly decaying) *function*  $\varphi$  on an oriented, Riemannian manifold  $P$  is defined by multiplying the volume form  $d\omega$  of  $P$  by  $\varphi$  and integrating this over  $P$

$$\int_P \varphi d\omega.$$

Now, if  $\alpha$  is any (properly decaying) form on  $P$  written in terms of local coordinates  $x^i$  on  $P$  ( $\alpha = \alpha(x^i, dx^i)$ ) and if one introduces odd variables  $\chi^i$  (generators for some exterior algebra), then one can define an element  $\alpha(x^i, \chi^i)$  of this exterior algebra by formally making the substitutions  $dx^i \rightarrow \chi^i$ . Then the fermionic integral

$$\int \alpha(x^i, \chi^i) \mathcal{D}\chi$$

is precisely the function one integrates (next to  $d\omega$  as above) to get the integral of  $\alpha$  over  $P$

$$\int_P \alpha = \int_P \int \alpha(x^i, \chi^i) \mathcal{D}\chi d\omega.$$

Applying this convention to the expression for the Euler number obtained by integrating over  $P$  gives the final formula toward which all of this has been leading us. For this we will explicitly indicate all of the dependences on the three fermionic

$(\chi, \eta, \psi)$  and three bosonic  $(\lambda, \phi, \omega)$  variables (in particular, all of the terms giving rise to forms on  $P$  ( $i\psi^T dS, i(dC^*, \lambda)$ , and  $(C^*, \eta)$ ) are regarded as functions of the new fermionic variable  $\chi$ ). We will also (finally!) suppress all but one of the integral signs.

$$\begin{aligned}
 & (2\pi)^{-n} (2\pi)^{-k} \int \exp \left\{ -\frac{1}{2} \|S(\omega)\|^2 + \frac{1}{2} \psi^T (\rho_*(\phi)) \psi \right. \\
 & \quad + i\psi^T dS_\omega(\chi) + i(dC_\omega^*(\chi, \chi), \lambda) \\
 & \quad \left. - i(\phi, R_\omega \lambda) + (C_\omega^* \chi, \eta) \right\} \mathcal{D}\chi \mathcal{D}\eta \mathcal{D}\psi d\lambda d\phi d\omega
 \end{aligned} \tag{5.53}$$

This we will call the **Atiyah-Jeffrey formula** for the Euler number of  $P \times_\rho V$ . Our objective now is to formally apply it to the infinite-dimensional vector bundle  $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$  of Donaldson theory with  $S = F_+$  as the equivariant map. The result will be, formally at least, an expression for an ‘‘Euler number’’ for the bundle (which, however, depends on the choice of  $S$ ) and also, as it happens, the partition function for Witten’s topological quantum field theory (i.e., the zero-dimensional Donaldson invariant of  $B$ ). We must emphasize at the outset that what we intend to do here is not mathematics (and certainly not physics). Our objective is to find, within the context of the infinite-dimensional vector bundle  $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$  associated with the Donaldson invariant, formal field-theoretic analogues of the various bosonic and fermionic variables appearing in (5.53) and natural identifications of the terms in the exponent of (5.53) with functions of these variables. In the process the (perfectly well-defined) bosonic and fermionic integrals in (5.53) will metamorphose into Feynman integrals over spaces of fields with all of their attendant mathematical difficulties. The purist will argue that this is meaningless manipulation of symbols and we can offer no credible defense against the charge. The only mitigating circumstance is that such formal manipulations have proved extraordinarily productive for both physics and mathematics and promise to be even more so in the future as the two subjects continue to re-establish lines of communication.

We begin with a brief summary of the notation accumulated in Section 4. Throughout the remainder of this section  $B$  will denote a compact, simply connected, oriented, smooth four-manifold with  $b_2^+(B) > 1$  and we will consider only the structure group  $G = SU(2)$  with Lie algebra  $\mathfrak{su}(2)$ .  $g$  will denote a generic Riemannian metric on  $B$  and  $SU(2) \hookrightarrow P \xrightarrow{\pi} B$  a smooth principal  $SU(2)$ -bundle over  $B$ .  $\hat{\mathcal{A}}$  is the space of irreducible connections on  $P$ ,  $\mathcal{G}$  is the group of gauge transformations and  $\hat{\mathcal{G}} = \mathcal{G}/\mathbb{Z}_2$  is  $\mathcal{G}$  modulo its center. Then  $\hat{\mathcal{B}} = \hat{\mathcal{A}}/\hat{\mathcal{G}} \cong \widehat{\mathcal{A}}/\hat{\mathcal{G}}$  is the moduli space of irreducible connections on  $P$ . Next,  $\text{Asd}(P, g) = \widehat{\text{Asd}}(P, g)$

is the space of  $\mathfrak{g}$ -ASD connections on  $P$  and  $\mathcal{M} = \text{Asd}(P, \mathfrak{g})/\mathcal{G} \cong \text{Asd}(P, \mathfrak{g})/\hat{\mathcal{G}}$  is the moduli space of gauge equivalence classes of (irreducible)  $\mathfrak{g}$ -ASD connections on  $P$ . Then  $\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \rightarrow \hat{\mathcal{B}}$  is a principal  $\hat{\mathcal{G}}$ -bundle and  $\hat{\mathcal{G}}$  acts on  $\Omega_+^2(B, \text{ad}P)$  on the left so we have an associated vector bundle  $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P) \rightarrow \hat{\mathcal{B}}$ . A section is determined by the equivariant map  $S = \mathbf{F}^+ : \hat{\mathcal{A}} \rightarrow \Omega_+^2(B, \text{ad}P)$  defined by  $S(\omega) = \mathbf{F}^+(\omega) = \mathbf{F}\omega = \frac{1}{2}(\mathbf{F}\omega + *\mathbf{F}\omega)$ .  $\mathcal{M}$  is identified with the zero set of this section, i.e., with the intersection of the images of the corresponding section and the zero-section. We assume for the remainder of this section that the Chern number of the bundle  $SU(2) \hookrightarrow P \xrightarrow{\pi} B$  has been fixed so that the dimension of  $\mathcal{M}$  is zero. Then the Donaldson invariant  $\gamma_0(B)$  given by (4.22) can be viewed as the intersection number of the section corresponding to  $S$ , i.e., as an ‘‘Euler number’’ for the bundle  $\hat{\mathcal{A}} \times_{\hat{\mathcal{G}}} \Omega_+^2(B, \text{ad}P)$ . The formal extension of (5.53) to this new infinite-dimensional context ‘‘should’’ provide an integral (Gauss-Bonnet-Chern) representation of the Donaldson invariant.

We begin by recalling that our derivation of the Atiyah-Jeffrey formula (5.53) assumes the existence of a Riemannian metric on the principal bundle space (in our case,  $\hat{\mathcal{A}}$ ) for which the group ( $\hat{\mathcal{G}}$ ) acts by isometries. Such a metric is easy to produce. Since  $\hat{\mathcal{A}}$  is open in  $\mathcal{A}$  (the space of all connections on  $P$ ) and  $\mathcal{A}$  is an affine space modeled on  $\Omega^1(B, \text{ad}P)$

$$T_\omega(\hat{\mathcal{A}}) \cong \Omega^1(B, \text{ad}P)$$

for each  $\omega \in \hat{\mathcal{A}}$ . Now, all of the spaces  $\Omega^k(B, \text{ad}P)$  have natural  $L^2$ -inner products arising from the metric  $\mathfrak{g}$  on  $B$  (and the corresponding Hodge star  $*$ ) and an invariant inner product on the Lie algebra. Taking the inner product on  $\mathfrak{su}(2)$  to be  $(A, B) = -\text{tr}(AB)$  this is given by

$$\langle \alpha, \beta \rangle_k = - \int_B \text{tr}(\alpha \wedge *\beta). \quad (5.54)$$

In particular, this is true for  $T_\omega(\hat{\mathcal{A}})$  and this defines a metric on  $\hat{\mathcal{A}}$ . Since the inner product is invariant under the action of  $\hat{\mathcal{G}}$  (pointwise conjugation by an element of  $P \times_{\text{Ad}} SU(2)$ ),  $\hat{\mathcal{G}}$  acts by isometries on  $\hat{\mathcal{A}}$ . This metric defines a connection on  $\hat{\mathcal{G}} \hookrightarrow \hat{\mathcal{A}} \rightarrow \hat{\mathcal{B}}$  whose horizontal spaces are the orthogonal complements to the gauge orbits. Indeed, we already know these horizontal spaces since (4.11) gives the orthogonal decomposition

$$T_\omega(\hat{\mathcal{A}}) \cong T_\omega(\omega \cdot \hat{\mathcal{G}}) \oplus \ker(\delta^\omega) = \text{im}(d^\omega) \oplus \ker(\delta^\omega) \quad (5.55)$$

where  $d^\omega : \Omega^0(B, \text{ad}P) \rightarrow \Omega^1(B, \text{ad}P)$  is the covariant exterior derivative and  $\delta^\omega : \Omega^1(B, \text{ad}P) \rightarrow \Omega^0(B, \text{ad}P)$  is its formal adjoint relative to the natural inner products (5.54) on the spaces of forms.

The first term we must contend with in the exponent of the Atiyah-Jeffrey formula (5.53) is  $-\frac{1}{2}\|S(\omega)\|^2$ , where  $S = \mathbf{F}^+ : \hat{\mathcal{A}} \rightarrow \Omega_+^2(B, \text{ad}P)$  and the norm is computed in the natural inner product  $\langle \cdot, \cdot \rangle_2$  on  $\Omega^2(B, \text{ad}P)$  at each  $\omega \in \hat{\mathcal{A}}$ . Thus,

$$-\frac{1}{2}\|S(\omega)\|^2 = -\frac{1}{2}\|\mathbf{F}^+\|^2 = \frac{1}{2}\int_B \text{tr}(\mathbf{F}^+ \wedge * \mathbf{F}^+) = \frac{1}{2}\int_B \text{tr}(\mathbf{F}^+ \wedge \mathbf{F}^+).$$

Using the orthogonality of the Hodge decomposition one finds that this can be written as

$$-\frac{1}{2}\|S(\omega)\|^2 = \frac{1}{4}\int_B \text{tr}(\mathbf{F}\omega \wedge * \mathbf{F}\omega) + \frac{1}{4}\int_B \text{tr}(\mathbf{F}\omega \wedge \mathbf{F}\omega). \quad (5.56)$$

The first term is of the typical Yang-Mills variety for a classical gauge theory, whereas the second Witten [46] calls a topological term because it is, up to a constant, the Chern number of the underlying  $SU(2)$ -bundle.

**Remark.** Witten [46] employs the notation more common in physics whereby everything is written in such a way as to appear local. We will not attempt to translate all that we do into this language, but will illustrate with (5.56). Let  $\{T_a\}$  be an orthonormal basis for  $\mathfrak{su}(2)$  relative to  $(A, B) = -\text{tr}(AB)$ , e.g.,  $T_a = -\frac{1}{\sqrt{2}}i\sigma_a$ ,  $a = 1, 2, 3$ , and  $\sigma_1, \sigma_2, \sigma_3$  are the Pauli spin matrices. Write  $\mathbf{F}\omega = \frac{1}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta$ , where  $F_{\alpha\beta} = F_{\alpha\beta}^a T_a$  and  $*\mathbf{F}\omega = \frac{1}{2}\tilde{F}_{\alpha\beta}dx^\alpha \wedge dx^\beta$ , where  $\tilde{F}_{\alpha\beta} = \tilde{F}_{\alpha\beta}^a T_a$ . Raise indices with  $g$  to get  $F^{\alpha\beta} = g^{\alpha\alpha'}g^{\beta\beta'}F_{\alpha'\beta'}$  and  $\tilde{F}^{\alpha\beta} = g^{\alpha\alpha'}g^{\beta\beta'}\tilde{F}_{\alpha'\beta'}$ . A quick computation shows that  $\frac{1}{4}\text{tr}(\mathbf{F}\omega \wedge * \mathbf{F}\omega) = \frac{1}{4}\text{tr}(F_{\alpha\beta}F^{\alpha\beta})\text{vol}g$  and  $\frac{1}{4}\text{tr}(\mathbf{F}\omega \wedge \mathbf{F}\omega) = \frac{1}{4}\text{tr}(F_{\alpha\beta}\tilde{F}^{\alpha\beta})\text{vol}g$ . Writing  $\text{vol}g$  as  $\sqrt{g}d^4x$  one obtains the two terms corresponding to (5.56) in (2.41) of [46]:

$$-\frac{1}{2}\|S(\omega)\|^2 = \int_B \sqrt{g}d^4x \text{tr}\left(\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \frac{1}{4}F_{\alpha\beta}\tilde{F}^{\alpha\beta}\right). \quad (5.57)$$

To proceed we must sort out the appropriate analogues, in the Donaldson theory context, of the maps  $C$ ,  $C^*$  and  $R$ . At each point  $\omega$  in the principal bundle space  $\hat{\mathcal{A}}$ ,  $C_\omega$  is the map from the Lie algebra of  $\hat{\mathcal{G}}$ , which we have seen can be identified with  $\Omega^0(B, \text{ad}P)$ , to the tangent space  $T_\omega(\hat{\mathcal{A}}) \cong \Omega^1(B, \text{ad}P)$  defined, for each  $\xi \in \Omega^0(B, \text{ad}P)$  by

$$C_\omega(\xi) = \frac{d}{dt}(\omega \cdot \exp(t\xi))|_{t=0}.$$

Computing this derivative locally gives

$$C_\omega(\xi) = d^\omega \xi. \quad (5.58)$$

Consequently,  $C_\omega^*$  is the formal adjoint

$$C_\omega^* = \delta^\omega : \Omega^1(B, \text{ad}P) \longrightarrow \Omega^0(B, \text{ad}P) \quad (5.59)$$

of  $d^\omega$  relative to the natural inner products on the spaces of forms and so

$$R_\omega = C_\omega^* \circ C_\omega = \Delta_0^\omega : \Omega^0(B, \text{ad}P) \longrightarrow \Omega^0(B, \text{ad}P) \quad (5.60)$$

is the scalar Laplacian corresponding to  $\omega$ .

With this information in hand we consider the term  $-i(\phi, R_\omega \lambda)$  in (5.53). Both  $\phi$  and  $\lambda$  are in the Lie algebra so we introduce two “bosonic” fields

$$\phi, \lambda \in \Omega^0(B, \text{ad}P)$$

and interpret  $(\ , \ )$  as the natural inner product  $\langle \ , \ \rangle_0$  on  $\Omega^0(B, \text{ad}P)$ .

**Remark.** We apply the adjectives “bosonic” and “fermionic” to the fields we introduce only because of the type of integral these variables correspond to in (5.53). We do not claim to have justified any physical connotations associated with the terms.

Thus, the term  $-i(\phi, R_\omega \lambda)$  is to be interpreted as

$$\begin{aligned} -i(\phi, R_\omega \lambda) &= -i \langle \phi, \Delta_0^\omega \lambda \rangle_0 = i \int_B \text{tr}(\phi \wedge *(\Delta_0^\omega \lambda)) \\ &= i \int_B \text{tr}(*(\phi \Delta_0^\omega \lambda)). \end{aligned} \quad (5.61)$$

**Remark.** As Atiyah and Jeffrey [3] point out, the real field  $\phi$  must be replaced by  $i\phi$  and  $\lambda$  must be replaced by  $\frac{1}{2}\lambda$  to conform to Witten’s notation. In physics notation, the corresponding term in [46] is

$$\int_B \sqrt{g} d^4x \text{tr}\left(\frac{1}{2}\phi D_\alpha D^\alpha \lambda\right).$$



Next we consider the term  $(C_\omega^* \chi, \eta)$  in (5.53). Since  $C_\omega^*$  maps  $\Omega^1(B, \text{ad}P)$  to  $\Omega^0(B, \text{ad}P)$  we will need two fermionic fields

$$\eta \in \Omega^0(B, \text{ad}P)$$

and

$$\chi \in \Omega^1(B, \text{ad}P)$$

and, as above,  $(, ) = \langle , \rangle_0$ . Thus, we find that

$$(C_\omega^* \chi, \eta) = \langle \delta^\omega \chi, \eta \rangle_0 = \langle \chi, d^\omega \eta \rangle_1 = - \int_B \text{tr}(\chi \wedge *d^\omega \eta). \quad (5.62)$$

The fermionic variable  $\psi$  in the Mathai-Quillen formalism arises from the odd generators of the exterior algebra of the fiber vector space  $V$ . In our case this vector space is  $\Omega_+^2(B, \text{ad}P)$  so we introduce a fermionic field

$$\psi \in \Omega_+^2(B, \text{ad}P).$$

Now consider the term  $i\psi^T dS_\omega(\chi)$  in (5.53).  $S$  is the self-dual curvature map  $S = \mathbf{F}^+ : \hat{\mathcal{A}} \rightarrow \Omega_+^2(B, \text{ad}P)$  and we noted in Section 4 that the derivative of this map at  $\omega \in \hat{\mathcal{A}}$  is identified with

$$d_\omega^+ : \Omega^1(B, \text{ad}P) \longrightarrow \Omega_+^2(B, \text{ad}P)$$

so

$$dS_\omega(\chi) = d_\omega^+ \chi \quad (5.63)$$

for each  $\chi \in \Omega^1(B, \text{ad}P)$ . We will interpret finite-dimensional expressions such as

$$A^T B = (A^1 \dots A^r) \begin{pmatrix} B^1 \\ \vdots \\ B^r \end{pmatrix} = A^1 B^1 + \dots + A^r B^r$$

in terms of the appropriate field-theoretic inner product so that  $i\psi^T dS_\omega(\chi)$  becomes

$$i \langle \psi, d_\omega^+ \chi \rangle_2 = i \langle \psi, d^\omega \chi \rangle_2$$

(because  $\psi$  is self-dual and the Hodge decomposition is orthogonal)

$$= i \langle d^\omega \chi, \psi \rangle_2 = -i \int_B \text{tr}(d^\omega \chi \wedge \psi)$$

$$i\psi^T dS_\omega(\chi) = -i \int_B \text{tr}(d^\omega \chi \wedge \psi). \quad (5.64)$$

Next we consider the term  $\frac{1}{2}\psi^T(\rho_*(\phi))\psi$  in (5.53). We know already that  $\phi \in \Omega^0(B, \text{ad}P)$  and  $\psi \in \Omega_+^2(B, \text{ad}P)$ . In the Mathai-Quillen form,  $\rho$  corresponds to the action of  $G$  on  $V$  that gives rise to the associated vector bundle. In our case,  $\hat{G}$  (regarded as sections of the nonlinear adjoint bundle) acts on  $\Omega_+^2(B, \text{ad}P)$  pointwise by conjugation. At each point this is just the ordinary adjoint action of  $SU(2)$  on its Lie algebra for which the infinitesimal action is just bracket. Thus, for each  $\phi \in \Omega^0(B, \text{ad}P)$ ,  $\rho_*(\phi)$  acts on  $\psi \in \Omega_+^2(B, \text{ad}P)$  by

$$\rho_*(\phi)\psi = [\phi, \psi]$$

so  $\frac{1}{2}\psi^T(\rho_*(\phi))\psi$  is interpreted as  $\frac{1}{2}\psi^T(\rho_*(\phi))\psi = \frac{1}{2}\langle \psi, [\phi, \psi] \rangle_2$  which we rearrange as follows

$$\begin{aligned} \frac{1}{2}\psi^T(\rho_*(\phi))\psi &= \frac{1}{2}\langle \psi, [\phi, \psi] \rangle_2 = \frac{1}{2}\langle [\phi, \psi], \psi \rangle_2 \\ &= -\frac{1}{2}\int_B \text{tr}([\phi, \psi] \wedge *\psi) = -\frac{1}{2}\int_B \text{tr}([\phi, \psi] \wedge \psi) \end{aligned}$$

because  $\psi$  is self-dual. Now, any ad-invariant inner product  $(\ , \ )$  on any Lie algebra satisfies  $(x, [y, z]) = ([x, y], z)$  so in this last integral we may replace  $[\phi, \psi] \wedge \psi$  by  $\phi[\psi, \psi]$ . Moreover,  $\text{tr}(AB) = \text{tr}(BA)$  so we conclude that

$$\frac{1}{2}\psi^T(\rho_*(\phi))\psi = -\frac{1}{2}\int_B \text{tr}([\psi, \psi]\phi). \quad (5.65)$$

The only remaining term in (5.53) is  $i(dC_\omega^*(\chi, \chi), \lambda)$  and this requires a bit more work.  $C^*$  is a one-form on  $\hat{A}$  with values in  $\Omega^0(B, \text{ad}P)$ . We compute  $dC^*$  at  $\omega \in \hat{A}$  as follows. Fix  $\chi_1, \chi_2 \in T_\omega \hat{A}$ . Since  $\mathcal{A}$  is an affine space and  $\hat{A}$  is open in  $\mathcal{A}$  we may regard  $\chi_1$  and  $\chi_2$  as constant vector fields on  $\hat{A}$ . Thus,

$$\begin{aligned} dC_\omega^*(\chi_1, \chi_2) &= \chi_1(C^*\chi_2) - \chi_2(C^*\chi_1) - C^*([\chi_1, \chi_2]) \\ &= \chi_1(C^*\chi_2) - \chi_2(C^*\chi_1) \end{aligned}$$

where  $C^*\chi_i$  is the function on  $\theta \rightarrow C_\theta^*\chi_i = \delta^\theta \chi_i$  for  $i = 1, 2$ . Now

$$\begin{aligned} (\chi_1(C^*\chi_2))(\omega) &= \chi_1(\omega)(C^*\chi_2) = \chi_1(C^*\chi_2) \\ &= \frac{d}{dt}C_{\omega+t\chi_1}^*(\chi_2)|_{t=0} = \frac{d}{dt}\delta^{\omega+t\chi_1}(\chi_2)|_{t=0}. \end{aligned} \quad (5.66)$$

We compute  $\delta^{\omega+t\chi_1}$  as follows: For any  $\lambda \in \Omega^0(B, \text{ad}P)$ ,

$$d^{\omega+t\chi_1}(\lambda) = d^\omega \lambda + t[\chi_1, \lambda] = d^\omega \lambda + tB_{\chi_1}(\lambda)$$

where  $B_{\chi_1} : \Omega^0(B, \text{ad}P) \rightarrow \Omega^1(B, \text{ad}P)$  is given by  $B_{\chi_1}(\lambda) = [\chi_1, \lambda]$ . Thus,

$$\delta^{\omega+t\chi_1}(\chi_2) = \delta^\omega \chi_2 + B_{\chi_1}^*(\chi_2)$$

where  $B_{\chi_1}^* : \Omega^1(B, \text{ad}P) \rightarrow \Omega^0(B, \text{ad}P)$  is the adjoint of  $B_{\chi_1}$ . We claim that

$$B_{\chi_1}^*(\chi_2) = -*[\chi_1, *\chi_2]. \quad (5.67)$$

Indeed, for any  $\lambda \in \Omega^0(B, \text{ad}P)$ ,

$$\begin{aligned} \langle B_{\chi_1}(\lambda), \chi_2 \rangle_1 &= \langle [\chi_1, \lambda], \chi_2 \rangle_1 = - \int_B \text{tr}([\chi_1, \lambda] \wedge *\chi_2) \\ &= \int_B \text{tr}([\lambda, \chi_1] \wedge *\chi_2) = \int_B \text{tr}(\lambda \wedge [\chi_1, *\chi_2]) \\ &= \int_B \text{tr}(\lambda \wedge **[\chi_1, *\chi_2]) = - \langle \lambda, *[\chi_1, *\chi_2] \rangle_0 \\ &= \langle \lambda, -*[\chi_1, *\chi_2] \rangle_0 \end{aligned}$$

which establishes (5.67). Thus,  $\delta^{\omega+t\chi_1}(\chi_2) = \delta^\omega \chi_2 - t*[\chi_1, *\chi_2]$  and computing the derivative at  $t = 0$  gives, from (5.66),

$$(\chi_1(C^*\chi_2))(\omega) = -*[\chi_1, *\chi_2].$$

Interchanging  $\chi_1$  and  $\chi_2$  gives

$$(\chi_2(C^*\chi_1))(\omega) = -*[\chi_2, *\chi_1].$$

Since these are independent of  $\omega$  we have  $\chi_1(C^*\chi_2) = -*[\chi_1, *\chi_2]$  and  $\chi_2(C^*\chi_1) = -*[\chi_2, *\chi_1]$ . One can verify that, for any  $\alpha, \beta \in \Omega^1(B, \text{ad}P)$ ,  $*[\beta, *\alpha] = -*[\alpha, *\beta]$  so we may write  $*[\chi_2, *\chi_1] = -*[\chi_1, *\chi_2]$  and thereby obtain

$$dC_\omega^*(\chi_1, \chi_2) = -2*[\chi_1, *\chi_2] \quad (5.68)$$

for any  $\omega \in \hat{A}$ . Thus

$$\begin{aligned} i \langle dC_\omega^*(\chi_1, \chi_2), \lambda \rangle_0 &= i \langle \lambda, dC_\omega^*(\chi_1, \chi_2) \rangle_0 \\ &= -i \int_B \text{tr}(\lambda \wedge *(-2*[\chi_1, *\chi_2])) \\ &= 2i \int_B \text{tr}(\lambda \wedge [\chi_1, *\chi_2]) = 2i \int_B \text{tr}([\chi_1, *\chi_2] \lambda) \end{aligned}$$

and finally

$$i(dC_{\omega}^*(\chi, \chi), \lambda) = 2i \int_B \text{tr}([\chi, * \chi] \lambda). \quad (5.69)$$

With this we have identified all of the terms in the exponent in (5.53). Each is the integral over  $B$  of a trace and so we may collect them all together into

$$\begin{aligned} \int_B \text{tr} \left( \frac{1}{4} F \omega \wedge * F \omega + \frac{1}{4} F \omega \wedge F \omega - \frac{1}{2} [\psi, \psi] \phi - \text{id}^{\omega} \chi \wedge \psi \right. \\ \left. + 2i [\chi, * \chi] \lambda + i * (\phi \Delta_0^{\omega} \lambda) - \chi \wedge * d^{\omega} \eta \right). \end{aligned} \quad (5.70)$$

Now we introduce some of the terminology used in physics. Each fixed choice of the three bosonic ( $\omega \in \hat{A}, \phi, \lambda \in \Omega^0(B, \text{ad}P)$ ) and three fermionic ( $\eta \in \Omega^0(B, \text{ad}P), \chi \in \Omega^1(B, \text{ad}P), \psi \in \Omega_+^2(B, \text{ad}P)$ ) fields will be called a **field configuration** and will be denoted

$$\Phi = (\omega, \phi, \lambda, \eta, \chi, \psi).$$

For each such choice the expression (5.70) is a number so this integral can be regarded as a function of  $\Phi$ . Minus this function is the **Donaldson-Witten action functional**

$$\begin{aligned} S_{DW}[\Phi] = \int_B \text{tr} \left( -\frac{1}{4} F \omega \wedge * F \omega - \frac{1}{4} F \omega \wedge F \omega + \frac{1}{2} [\psi, \psi] \phi + \text{id}^{\omega} \chi \wedge \psi \right. \\ \left. - 2i [\chi, * \chi] \lambda - i * (\phi \Delta_0^{\omega} \lambda) + \chi \wedge * d^{\omega} \eta \right) \end{aligned} \quad (5.71)$$

Thus, in our present infinite-dimensional context the integral in (5.53) can be written

$$\int e^{-S_{DW}[\Phi]} \mathcal{D}\Phi \quad (5.72)$$

where we have abbreviated  $\mathcal{D}\chi \mathcal{D}\eta \mathcal{D}\psi d\lambda d\phi d\omega$  as simply  $\mathcal{D}\Phi$ . It is only in this last expression that we leave the world of mathematically well-defined objects and proceed “formally”.

In (5.72) we have omitted the constant  $(2\pi)^{-n}(2\pi)^{-k}$  in (5.53) since, in our present circumstances, both  $n$  and  $k$  would be infinite. In the physics literature one often sees the integral (5.72) normalized with a factor of  $1/\text{vol}(\mathcal{G})$ , where  $\text{vol}(\mathcal{G})$  is intended to represent the “volume” of the gauge group  $\mathcal{G}$ . About this we will have nothing further to say, but, for certain remarks we wish to make here and in Section 7, we point out that the physicists often include in the exponent in (5.72),

or directly in the action (5.71), a factor of  $1/e^2$ , where  $e$  is a so-called **coupling constant**. Mathematically, one can view the inclusion of such a factor in  $S_{DW}[\Phi]$  as simply a different choice of invariant inner product on the Lie algebra  $\mathfrak{su}(2)$ , which is determined only up to a positive constant multiple. Classically, one can rescale and give this factor any convenient value. However, upon quantization the different values of the coupling constant give rise to an entire one-parameter family of quantum field theories and the computability (“renormalizability”) of the theory generally depends on this value. Since this dependence on the coupling constant (or, rather, a lack thereof in the cases of topological interest) is relevant to a few comments we will make here and somewhat later, we record the following alternative to (5.72)

$$\int e^{-S_{DW}[\Phi]/e^2} \mathcal{D}\Phi. \quad (5.73)$$

This integral represents the **partition function** of the quantum field theory constructed by Witten in [46]. Of course, Witten arrived at the action (5.71) and therefore the partition function by quite a different route than the one we have followed. We began with the zero-dimensional Donaldson invariant, regarded it as an “Euler number” and massaged the Mathai-Quillen integral representation of this Euler number until it could be formally applied in our infinite-dimensional context to yield (5.71) and (5.72). Witten’s arguments leading to  $S_{DW}[\Phi]$  were physical, but the objective was to describe a quantum field theory in which the Donaldson invariants appeared as expectation values of certain observables and, in particular, the zero-dimensional invariant was the partition function. How then did Witten uncover the Donaldson invariants in the field theory with action  $S_{DW}[\Phi]$ ?

Witten chose the field content  $\Phi = (\omega, \phi, \lambda, \eta, \chi, \psi)$  and action  $S_{DW}[\Phi]$  in order to ensure the presence of certain symmetries (gauge invariance and “BRST-like” symmetries). The BRST symmetries are expressed in terms of a certain operator  $Q$  on the fields which squares to zero ( $Q \circ Q = 0$ ) and so determines cohomology classes that are taken to represent the physical states of the theory. The energy-momentum tensor of the theory (defined in terms of the variation of the action under an infinitesimal change in the Riemannian metric  $g$  of the underlying four-manifold  $B$ ) turns out to be  $Q$ -exact (and so cohomologically trivial). With this, certain formal manipulations with functional integrals imply that the partition function of the theory is independent of both the metric  $g$  and the coupling constant  $e$  in the sense that its infinitesimal variation with respect to either  $g$  or  $e$  is zero.

**Note:** These are hallmarks of what are today called **cohomological field theories**. In such field theories the expectation values of observables are also “in-

dependent of  $\mathfrak{g}$ ” in the same sense. This has led the physicists to refer to such field theories as “topological quantum field theories” and the expectation values as “topological invariants”. However, these are very different uses of the terms “independent of  $\mathfrak{g}$ ” and “topological invariant” than one would encounter in mathematics. For example, we have seen in Section 4 that the zero-dimensional Donaldson invariant  $\gamma_0(B)$  is only independent of a *generic* choice of  $\mathfrak{g}$  and even this is true only when  $b_2^+(B) > 1$ . Even granting this,  $\gamma_0(B)$  is an invariant of the *differentiable structure* of  $B$  and certainly not of its topology.

The fact that the partition function is independent of the coupling constant  $e$  is particularly significant since one is then free to compute it in the limit of either small ( $e \rightarrow 0$ ) or large ( $e \rightarrow \infty$ ) values, whichever is most convenient or most informative. For small values of  $e$ , physicists employ a technique known as semi-classical approximation which, again because of the symmetries of  $S_{DW}[\Phi]$ , one can show is actually exact in our case. This phenomenon is an infinite-dimensional analogue of a well-known finite-dimensional theorem on the exactness of the stationary phase approximation due to Duistermaat and Heckman [13]. As we shall see in the next section this theorem is most properly understood within the context of equivariant cohomology and the localization of certain integrals of equivariant differential forms to the fixed point set of the group action. Similarly, the BRST operator  $Q$  can be viewed as the equivariant exterior derivative in a model of the  $\mathcal{G}$ -equivariant cohomology of  $\mathcal{A}$  and Witten shows that, for certain well-chosen observables in his field theory, the path integral representations of their expectation values localize to the (finite-dimensional) moduli spaces of anti-self-dual connections thus yielding integral formulas for the Donaldson invariants. For the particular case we have under consideration, the partition function (which, being invariant, descends to the moduli space of fields) localizes to a sum over the zero-dimensional moduli space of ASD connections yielding  $\gamma_0(B)$ .

In the following Part II of this survey we will take up our story at this point with a rather detailed discussion of the simplest of the “Equivariant Localization” theorems and its relation to the theorem of Duistermaat and Heckman on exact stationary phase approximation. This done, we will turn to the question of what can be learned by examining the partition function in the limit  $e \rightarrow \infty$  of large coupling constants in a section on “Duality and Seiberg-Witten”. Finally then we will consider “The Witten Conjecture” itself.

## References

- [1] Atiyah M. and Bott R., *The Moment Map and Equivariant Cohomology*, *Topology* **23** (1984) 1-28.
- [2] Atiyah M., Hitchin N. and Singer I., *Self-Duality in Four-Dimensional Riemannian Geometry*, *Proc. Roy. Soc. Lond. A* **362** (1978) 425-461.
- [3] Atiyah M. and Jeffrey L., *Topological Lagrangians and Cohomology*, *J. Geom. Phys.* **7** (1990) 119-136.
- [4] Belavin A., Polyakov A., Schwarz A. and Tyupkin Y., *Pseudoparticle Solutions of the Yang-Mills Equations* *Phys. Lett.* **59 B** (1975) 85-87.
- [5] Berline N., Getzler E. and Vergne M., *Heat Kernels and Dirac Operators*, Springer, New York, 1996.
- [6] Blau M., *The Mathai-Quillen Formalism and Topological Field Theory*, *J. Geom. Phys.* **11** (1993) 95-127.
- [7] Blau M. and Thompson G., *Localization and Diagonalization*, *J. Math. Phys.* **36** (1995) 2192-2236.
- [8] Cordes S., Moore G. and Ramgoolam S., *Lectures on 2D Yang-Mills Theory, Equivariant Cohomology and Topological Field Theory*, arXiv: hep-th/9402107.
- [9] Dirac P.A.M., *Quantised Singularities in the Electromagnetic Field*, *Proc. Roy. Soc. A* **133** (1931) 60-72.
- [10] Donaldson S., *An Application of Gauge Theory to Four-Dimensional Topology*, *J. Diff. Geom.* **18** (1983) 279-315.
- [11] Donaldson S., *Polynomial Invariants for Smooth four-Manifolds*, *Topology* **29** (1990) 257-315.
- [12] Donaldson S. and Kronheimer P., *The Geometry of Four-Manifolds*, Oxford University, Oxford, 1990.
- [13] Duistermaat J. and Heckman G., *On the Variation in the Cohomology of the Symplectic Form of the Reduced Phase Space*, *Invent. Math.* **69** (1982) 250-268. Addendum **72** (1983) 153-158.
- [14] Feehan P. and Leness T., *On Donaldson and Seiberg-Witten Invariants*, arXiv: math.DG/0106221.
- [15] Freed D. and Uhlenbeck K., *Instantons and Four-Manifolds*, Springer, New York, 1984.
- [16] Freedman M., *The Topology of Four-Dimensional Manifolds*, *J. Diff. Geom.* **17** (1982) 357-454.

- 
- [17] Friedman R. and Morgan J., *Smooth Four-Manifolds and Complex Surfaces*, Springer, New York, 1994.
- [18] Guillemin V. and Sternberg S., *Geometric Asymptotics*, Amer. Math. Soc., Providence, RI, 1977.
- [19] Guillemin V. and Sternberg S., *Symplectic Techniques in Physics*, Cambridge University, Cambridge, England, 1984.
- [20] Guillemin V. and Sternberg S., *Supersymmetry and Equivariant De Rham Theory*, Springer, New York, 1999.
- [21] Hitchin N., *The Geometry and Topology of Moduli Spaces*. In: *Global Geometry and Mathematical Physics*, Springer Lecture Notes in Mathematics 1451, Springer, New York, 1990, 1-48.
- [22] Jaffe A. and Taubes C., *Vortices and Monopoles*, Birkhauser, Boston, MA, 1980.
- [23] Kalkman J., *BRST Model for Equivariant Cohomology and Representatives for the Equivariant Thom Class*, Comm. Math. Phys. **153** (1993) 447-463.
- [24] Kobayashi S. and Nomizu K., *Foundations of Differential Geometry*, Vols. I, II, Wiley Interscience, 1963 and 1969.
- [25] Kronheimer P. and Mrowka T., *Recurrence Relations and Asymptotics for Four-Manifold Invariants*, Bull. Amer. Math. Soc. **30** (1994) 215-221.
- [26] Kronheimer P. and Mrowka T., *Embedded Surfaces and the Structure of Donaldson's Polynomial Invariants*, J. Diff. Geom. **3** (1995) 573-734.
- [27] Labastida J. and Marino M., *A Topological Lagrangian for Monopoles on Four-Manifolds*, Phys. Lett. **B 351** (1995) 146-152.
- [28] Labastida J. and Lozano C., *Lectures on Topological Quantum Field Theory*, arXiv: hep-th/9709192.
- [29] Lang S., *Introduction to Differentiable Manifolds*, John Wiley & Sons, New York, 1962.
- [30] Lawson H., *The Theory of Gauge Fields in Four Dimensions*, Amer. Math. Soc., Providence, RI, 1985.
- [31] Lawson H. and Michelsohn M-L., *Spin Geometry*, Princeton University, Princeton, NJ, 1989.
- [32] Mathai V. and Quillen D., *Superconnections, Thom Classes and Equivariant Differential Forms*, Topology **25** (1986) 85-110.
- [33] Naber G., *Topological Methods in Euclidean Spaces*, Dover Publications, Mineola, NY, 2000.
- [34] Naber G., *Topology, Geometry, and Gauge Fields: Foundations*, Springer, New York, 1997.



- [35] Naber G., *Topology, Geometry, and Gauge Fields: Interactions*, Springer, New York, 2000.
- [36] O'Grady K., *Donaldson's Polynomials for K3 Surfaces*, J. Diff. Geom. **35** (1992) 415-427.
- [37] Parker T., *Gauge Theories on Four-Dimensional Riemannian Manifolds* Comm. Math. Phys. **85** (1982) 563-602.
- [38] Pidstrigach V. and Tyurin A., *Localization of Donaldson's Invariants along Seiberg-Witten Classes*, arXiv: dg-ga/9507004.
- [39] Seiberg N. and Witten E., *Monopoles, Duality and Chiral Symmetry Breaking in  $N = 2$  Supersymmetric QCD*, Nucl. Phys. **B 431** (1994) 581-640.
- [40] Szabo R., *Equivariant Localization of Path Integrals*, arXiv: hep-th/9608068.
- [41] Taubes, C., *Self-Dual Connections on Non-Self-Dual four-Manifolds*, J. Diff. Geom. **17** (1982) 139-170.
- [42] Taubes, C., *The Seiberg-Witten Invariants* (Videotape), AMS Selected Lectures in Mathematics, Amer. Math. Soc., Providence, RI, 1995.
- [43] Uhlenbeck K., *Removable Singularities in Yang-Mills Fields*, Comm. Math. Phys. **83** (1982) 11-30.
- [44] Vajiac A., *A Generalization of Witten's Conjecture Relating Donaldson and Seiberg-Witten Invariants*, arXiv: hep-th/0003214.
- [45] Warner F., *Foundations of Differentiable Manifolds and Lie Groups*, Springer, New York, 1983.
- [46] Witten E., *Topological Quantum Field Theory*, Comm. Math. Phys. **117** (1988) 353-386.
- [47] Witten E., *Two Dimensional Gauge Theories Revisited*, J. Geom. Phys. **9** (1992) 303-368.
- [48] Witten E., *Monopoles and Four-Manifolds*, Math. Res. Lett. **1** (1994) 769-796.
- [49] Yang C. and Mills R., *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191-195.

Gregory L. Naber  
 Department of Mathematics  
 Drexel University  
 PA 19104, Philadelphia, U.S.A.  
 gregory.l.naber@drexel.edu