

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

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.

6. Equivariant Localization

Motivated by our discussion of the Witten Lagrangian, its equivariant symmetries, and the localization of the corresponding partition function to the anti-self-dual moduli space to yield the Donaldson invariant, we return now to the finite-dimensional context and consider the general phenomenon of “equivariant localization.” In order to provide a relatively complete treatment and because it describes the finite-dimensional analogue of Witten’s partition function (the zero-dimensional Donaldson invariant) we will restrict our attention to the simplest (“discrete”) equivariant localization theorem.

We consider a compact, oriented, smooth manifold M of dimension $n = 2k$ and denote by ν a volume form on M . Suppose $H : M \rightarrow \mathbb{R}$ is a Morse function on M , i.e., a smooth function whose critical points p ($dH(p) = 0$) are all nondegenerate (this means that the Hessian $\mathcal{H}_p : T_p(M) \times T_p(M) \rightarrow \mathbb{R}$, defined by $\mathcal{H}_p(V_p, W_p) = V_p(W(H))$, where $V_p, W_p \in T_p(M)$ and W is a vector field on M with $W(p) = W_p$, is a nondegenerate bilinear form). Finally, let T denote some real parameter. We consider the integral

$$\int_M e^{iTH} \nu \tag{6.1}$$

and are especially interested in its asymptotic behavior as $T \rightarrow \infty$. The Stationary Phase Theorem (Chapter I of [22]) asserts roughly that, for large T , the dominant contributions to such an integral come from the critical points of H . More precisely, one has

$$\int_M e^{iTH} \nu = \sum_{\substack{p \in M \\ dH(p)=0}} \left(\frac{2\pi}{T}\right)^k e^{\pi i(\text{Sgn} \mathcal{H}_p)/4} |\det \mathcal{H}_p(e_i, e_j)|^{-\frac{1}{2}} e^{iTH(p)} + O(T^{-(k+1)}) \tag{6.2}$$

where $\text{Sgn} \mathcal{H}_p$ is the signature (number of positive eigenvalues minus the number of negative eigenvalues) of any matrix representing \mathcal{H}_p , $\{e_1, \dots, e_{2k}\}$ is a basis for $T_p(M)$ with $\nu_p(e_1, \dots, e_{2k}) = 1$, and $O(T^{-(k+1)})$ stands for terms which, in modulus, are bounded by C/T^{k+1} for some constant C and all T outside some compact set in \mathbb{R} . The terms preceding $O(T^{-(k+1)})$ on the right-hand side of (6.2) constitute the **stationary phase approximation** of the integral. These terms arise in the proof of (6.2) from writing H near p as a quadratic function in some coordinates (that this is possible is the content of the **Morse Lemma**) and computing directly the resulting Gaussian integral. It follows from the Morse Lemma that the critical points of a Morse function are isolated. Since M is compact, H can have only finitely many critical points so the sum in (6.2) is necessarily finite.

Let us write out a simple example used by Witten [53] to illustrate the phenomenon we wish to study. We take M to be the two-sphere S^2 in \mathbb{R}^3 and let ν be the usual metric volume form on S^2 (this is the restriction to S^2 of the two-form $x dy \wedge dz - y dx \wedge dz + z dx \wedge dy$ on \mathbb{R}^3). Let $H : S^2 \rightarrow \mathbb{R}$ be the ‘‘height function’’ ($H(x, y, z) = z$ for any $(x, y, z) \in S^2$). We claim that the critical points of H are the north and south poles, i.e., $N = (0, 0, 1)$ and $S = (0, 0, -1)$, and that both are nondegenerate so H is a Morse function on S^2 . For example, on $z > 0$ in S^2 , $(x, y, z) \rightarrow (x, y)$ is a chart with inverse $(x, y) \rightarrow (x, y, (1-x^2-y^2)^{\frac{1}{2}})$ and, in these coordinates, $H(x, y) = (1-x^2-y^2)^{\frac{1}{2}}$ so $dH(x, y) = -(1-x^2-y^2)^{-\frac{1}{2}}(x dx + y dy)$. Thus, the only critical point in $z > 0$ occurs when $(x, y) = (0, 0)$, i.e., at $N(0, 0, 1)$. Furthermore, the Hessian (which, in any coordinate system, is represented by the matrix of second order partial derivatives) is given by

$$-(1-x^2-y^2)^{-\frac{3}{2}} \begin{pmatrix} 1-y^2 & xy \\ xy & 1-x^2 \end{pmatrix}.$$

Thus, at $(x, y) = (0, 0)$ we obtain $\mathcal{H}_N = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ and this is, indeed, non-singular. The region $z < 0$ on S^2 is, of course, handled in the same way and projecting onto other coordinate planes shows that there are no critical points with $z = 0$.

Now we shall write out the stationary phase approximation for the integral

$$\int_{S^2} e^{iTz} \nu. \quad (6.3)$$

From (6.2) with $k = 1$ and $p = N, S$, this is

$$\begin{aligned} & \left(\frac{2\pi}{T} \right) e^{\pi i (\text{Sgn} \mathcal{H}_N) / 4} |\det \mathcal{H}_N(e_i, e_j)|^{-\frac{1}{2}} e^{iTz(N)} \\ & + \left(\frac{2\pi}{T} \right) e^{\pi i (\text{Sgn} \mathcal{H}_S) / 4} |\det \mathcal{H}_S(e_i, e_j)|^{-\frac{1}{2}} e^{iTz(S)}. \end{aligned} \quad (6.4)$$

Now, $\text{Sgn} \mathcal{H}_N = \text{Sgn} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -2$. Next note that evaluating $x dy \wedge dz - y dx \wedge dz + z dx \wedge dy$ at $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ gives z so, at N , $\nu_N(\frac{\partial}{\partial x}(N), \frac{\partial}{\partial y}(N)) = 1$. Thus, $\{e_1, e_2\} = \{\frac{\partial}{\partial x}(N), \frac{\partial}{\partial y}(N)\}$ is a basis of the required type for $T_N(S^2)$ so

$$|\det \mathcal{H}_N(e_i, e_j)|^{-\frac{1}{2}} = \left| \det \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right|^{-\frac{1}{2}} = 1.$$

Similarly, $\text{Sgn} \mathcal{H}_S = 2$ and $|\det \mathcal{H}_S(e_i, e_j)|^{-\frac{1}{2}} = 1$. Substituting all of this into (6.4) gives, for the stationary phase approximation to (6.3),

$$\frac{2\pi i}{T} (e^{-iT} - e^{iT}) = 4\pi \left(\frac{\sin T}{T} \right). \quad (6.5)$$

Next we observe that the integral (6.3) is actually easy to compute exactly. Let $\iota : S^2 \hookrightarrow \mathbb{R}^3$ be the inclusion map so that $\nu = \iota^*(x dy \wedge dz - y dx \wedge dz + z dx \wedge dy)$. Define an orientation preserving diffeomorphism φ of $(0, \pi) \times (-\pi, \pi)$ into S^2 by

$$(\iota \circ \varphi)(\phi, \theta) = (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi).$$

The image of this map covers all of S^2 except a set of measure zero. A simple computation shows that $\varphi^* \nu = \sin \phi d\phi \wedge d\theta$ and so

$$\varphi^*(e^{iTz} \nu) = e^{iT \cos \phi} \sin \phi d\phi \wedge d\theta.$$

Denoting by dm Lebesgue measure on the plane we therefore have

$$\begin{aligned}
 \int_{S^2} e^{iTz} \nu &= \int_{(0,\pi) \times (-\pi,\pi)} e^{iT \cos \phi} \sin \phi d\phi \wedge d\theta \\
 &= \int_{[0,\pi] \times [-\pi,\pi]} e^{iT \cos \phi} \sin \phi dm = \int_{-\pi}^{\pi} \int_0^{\pi} e^{iT \cos \phi} \sin \phi d\phi d\theta \\
 &= -\frac{2\pi}{iT} \left[e^{iT \cos \phi} \right]_0^{\pi} = \frac{2\pi i}{T} [e^{-iT} - e^{iT}] = 4\pi \left(\frac{\sin T}{T} \right).
 \end{aligned}$$

We find then that, in this particular case, the stationary phase approximation (6.5) to the integral (6.3) actually gives the exact value of the integral. Our goal now is to uncover the underlying features of this example which account for this exactness of the stationary phase approximation.

We begin with a few observations on the preceding example. Note that the volume form ν on S^2 is also a symplectic form, i.e., a closed, nondegenerate two-form. Indeed, any volume form ν on any orientable surface is a symplectic form (it is closed because it is a two-form on a two-dimensional manifold and nondegenerate because, at each point, an oriented basis $\{e_1, e_2\}$ for the tangent space satisfies $\nu(e_1, e_2) > 0$ so if $v = v^1 e_1 + v^2 e_2 \neq 0$ (say, $v^1 \neq 0$), then $\nu(v, e_2) = v^1 \nu(e_1, e_2) \neq 0$). When thinking of ν as a symplectic form on S^2 we will denote it ω . Now, the height function H , like any smooth, real-valued function on the symplectic manifold (S^2, ω) , determines a corresponding **Hamiltonian vector field** V_H on S^2 . This is defined to be the unique vector field on S^2 satisfying

$$dH = \iota_{V_H} \omega, \tag{6.6}$$

where ι_{V_H} is interior multiplication by V_H (so that, for any vector field W on S^2 , $dH(W) = \omega(V_H, W)$). We claim that if $\frac{\partial}{\partial \theta}$ is the θ -coordinate velocity field of the spherical coordinate chart on S^2 (taken to be zero at N and S), then

$$V_H = \frac{\partial}{\partial \theta}.$$

First note that (6.6) and the nondegeneracy of ω imply that V_H must vanish at the critical points N and S of H and so it agrees with $\frac{\partial}{\partial \theta}$ there. At any other point, $H(\phi, \theta) = \cos \phi$ so $dH(\phi, \theta) = -\sin \phi d\phi$. Since $\omega = \sin \phi d\phi \wedge d\theta$ (here and henceforth we adopt the time-honored custom of omitting references to the

diffeomorphism φ whenever it is convenient to do so), we have

$$\begin{aligned} \iota\left(\frac{\partial}{\partial\theta}\right)\omega &= \iota\left(\frac{\partial}{\partial\theta}\right)(\sin\phi d\phi \otimes d\theta - \sin\phi d\theta \otimes d\phi) \\ &= \sin\phi\left(d\phi\left(\frac{\partial}{\partial\theta}\right)\right)d\theta - \sin\phi\left(d\theta\left(\frac{\partial}{\partial\theta}\right)\right)d\phi \\ &= -\sin\phi d\phi = dH \end{aligned}$$

as required. The integral curves of $V_H = \frac{\partial}{\partial\theta}$ are then easily found. They are points at N and S and elsewhere they are “horizontal” circles traversed at speed one. The unique one through $p = \varphi(\phi, \theta)$ at time $t = 0$ is

$$\alpha_p(t) = (\sin\phi \cos(\theta + t), \sin\phi \sin(\theta + t), \cos\phi)$$

(we shall also omit references to the inclusion $\iota : S^2 \hookrightarrow \mathbb{R}^3$). These integral curves are therefore periodic with period 2π . The flow

$$\begin{aligned} \alpha : S^2 \times \mathbb{R} &\longrightarrow S^2 \\ \alpha(p, t) &= \alpha_p(t) \end{aligned}$$

is therefore also periodic in t . Finally, recall that any symplectic manifold (M^{2k}, ω) has a canonical orientation (volume form) ν_ω called the **Liouville form** and defined by

$$\nu_\omega = \frac{1}{k!} \omega \wedge \dots \wedge \omega = \frac{1}{k!} \omega^k.$$

For $k = 1$ this is just ω so, in our example on S^2 , ν , ω and ν_ω are all the same.

Duistermaat and Heckman [16] have shown that the exactness of the stationary phase approximation of $\int_{S^2} e^{iTz} \nu$ is a consequence of the fact that the Hamiltonian vector field of the height function on S^2 has a periodic flow. More generally, we have

Theorem 6.1 (Duistermaat-Heckman) *Let M be a compact manifold of dimension $n = 2k$ with symplectic form ω and oriented by the corresponding Liouville form $\nu_\omega = \frac{1}{k!} \omega^k$. Let $H \in C^\infty(M)$ be a Morse function on M and V_H its Hamiltonian vector field ($dH = \iota_{V_H} \omega$). If the flow of V_H is periodic, then, for any real number $T > 0$,*

$$\int_M e^{iTH} \nu_\omega = \sum_{\substack{p \in M \\ dH(p)=0}} \left(\frac{2\pi}{T}\right)^k e^{\pi i(\text{Sgn } \mathcal{H}_p)/4} |\det \mathcal{H}_p(e_i, e_j)|^{-\frac{1}{2}} e^{iTH(p)}$$

where $\mathcal{H}_p : T_p(M) \times T_p(M) \rightarrow \mathbb{R}$ is the Hessian of H at p and $\{e_1, \dots, e_{2k}\}$ is a basis for $T_p(M)$ with $\nu_\omega(e_1, \dots, e_{2k}) = 1$.

Note that the set of critical points must be nonempty since M is compact and so H must achieve maximum and minimum values. We intend to provide a complete proof of this result, but will proceed toward it in a rather roundabout fashion. First we return to our example on S^2 and isolate a group action which suggests a more general perspective on the Duistermaat-Heckman Theorem. We formulate this new perspective as a Generalized Duistermaat-Heckman Theorem that concerns itself with Hamiltonian actions on symplectic manifolds and show that this new result implies our Theorem 6.1. Still our perspective is not broad enough, however, and we focus our attention on general group actions on manifolds and their associated equivariant cohomologies. In this context we prove the simplest of the so-called Equivariant Localization Theorems and find that it has as a simple consequence the Generalized Duistermaat-Heckman Theorem and therefore also Theorem 6.1.

Let us then consider again the height function H on the symplectic manifold S^2 . Since the Hamiltonian vector field V_H has a periodic flow it gives rise to an obvious action of S^1 on S^2 (rotate points of S^2 around the integral curves containing them). Specifically, if $g = e^{iT} \in S^1$ and $p = \varphi(\phi, \theta) = (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi) \in S^2$, then we define

$$\begin{aligned} g \cdot p &= e^{iT} \cdot (\sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi) \\ &= (\sin \phi \cos(\theta + T), \sin \phi \sin(\theta + T), \cos \phi) \end{aligned}$$

(if $p = N$ or S we define $g \cdot p = p$ for all $g \in S^1$). This clearly defines a (left) action of S^1 on S^2 . As usual, we identify the Lie algebra of S^1 with $i\mathbb{R}$. Each $\xi = ia$ in the Lie algebra gives rise to an associated vector field $\xi^\#$ on S^2 defined by

$$\xi^\#(p) = \frac{d}{dt}(\exp(-t\xi) \cdot p)|_{t=0}. \quad (6.7)$$

It is a simple matter to compute $\xi^\#(p)$ explicitly. At $p = N, S$ it is zero and, otherwise,

$$\begin{aligned} \xi^\#(p) &= \frac{d}{dt}(\exp(-t\xi) \cdot p)|_{t=0} = \frac{d}{dt}(e^{-iat} \cdot p)|_{t=0} \\ &= \frac{d}{dt}(\sin \phi \cos(\theta - at), \sin \phi \sin(\theta - at), \cos \phi)|_{t=0} \\ &= -a(-\sin \phi \sin \theta, \sin \phi \cos \theta, 0) = -a \frac{\partial}{\partial \theta}(p) = -aV_H(p) = V_{-aH}(p) \end{aligned}$$

(the last equality is easy to check by verifying that $d(-aH) = \iota_{-aV_H}\omega$). Thus,

$$\xi = ia \implies \xi^\# = V_{-aH}.$$

In particular, every $\xi^\#$ is the Hamiltonian vector field of some smooth function on S^2 . We can therefore define a map

$$\mu : \text{Lie}(S^1) = i\mathbb{R} \longrightarrow C^\infty(S^2)$$

by

$$\mu(\xi) = \mu(ia) = -aH$$

which has the following properties:

1. μ is linear.
2. $\xi^\#$ is the Hamiltonian vector field on S^2 determined by $\mu(\xi)$.
3. μ is equivariant with the respect to the natural actions of S^1 on $\text{Lie}(S^1)$ and $C^\infty(S^2)$, i.e.,

$$\mu(g \cdot \xi) = g \cdot \mu(\xi).$$

Remark. Regarding (3), the natural action of S^1 on $\text{Lie}(S^1)$ is the adjoint action ($g \cdot \xi = g\xi g^{-1}$) which, in this case, is trivial since S^1 is abelian. Thus, $\mu(g \cdot \xi) = \mu(\xi)$. The action of S^1 on $C^\infty(S^2)$ is defined by $(g \cdot \psi)(p) = \psi(g^{-1} \cdot p)$ so $(g \cdot \mu(\xi))(p) = \mu(\xi)(g^{-1} \cdot p) = (-aH)(g^{-1} \cdot p) = (-aH)(p) = \mu(\xi)(p)$ because H is constant on the orbits. Thus, $g \cdot \mu(\xi) = \mu(\xi) = \mu(g \cdot \xi)$.

Now we abstract these properties of our example and formulate general definitions. Let (M, ω) be a compact symplectic manifold of dimension $n = 2k$ and G a compact Lie group (with Lie algebra \mathcal{G}) that acts smoothly on M on the left (we will write such an action as $\sigma : G \times M \rightarrow M$ with $\sigma(g, p) = g \cdot p = \sigma_g(p) = \sigma_p(g)$). The action is said to be **Hamiltonian** if there is a map

$$\mu : \mathcal{G} \longrightarrow C^\infty(M)$$

such that

1. μ is linear.
2. For each $\xi \in \mathcal{G}$ the vector field $\xi^\#$ on M (defined by (6.7)) is the Hamiltonian vector field associated with $\mu(\xi)$, i.e.,

$$d\mu(\xi) = \iota(\xi^\#)\omega. \tag{6.8}$$

3. μ is equivariant, i.e.,

$$\mu(g \cdot \xi) = g \cdot \mu(\xi).$$

The function $\mu(\xi)$ is called the **symplectic moment** of ξ and one defines the associated **moment map**

$$\Phi : M \longrightarrow \mathcal{G}^*$$

(where \mathcal{G}^* is the dual of the vector space \mathcal{G}) by

$$(\Phi(p))(\xi) = (\mu(\xi))(p).$$

Although they will play no role in our story here, these moment maps have many striking and beautiful properties (see, e.g., [2] and [23]).

Notice that it follows at once from (2) and the nondegeneracy of ω that *the critical points of $\mu(\xi)$ coincide with the zeros of $\xi^\#$* . Moreover, every fixed point of the G -action is, by (6.7), a zero of every $\xi^\#$ (and so, a critical point of every $\mu(\xi)$). If $\xi \in \mathcal{G}$ has the property that $\xi^\#$ vanishes *only* at the fixed points of the G -action, then ξ is said to be **nondegenerate** and, in this case, one can show that $\mu(\xi)$ is necessarily a Morse function (see [23]). We assume that some invariant Riemannian metric $\langle \cdot, \cdot \rangle_G$ on M has been selected and note that any vector field $\xi^\#$ defined by (6.7) for some $\xi \in \mathcal{G}$ is then necessarily a Killing vector field for $\langle \cdot, \cdot \rangle_G$, i.e.,

$$\mathcal{L}_{\xi^\#} \langle \cdot, \cdot \rangle_G = 0$$

where $\mathcal{L}_{\xi^\#}$ denotes the Lie derivative with respect to $\xi^\#$. This last condition can be written equivalently as

$$\xi^\#(\langle V, W \rangle_G) = \langle [\xi^\#, V], W \rangle_G + \langle V, [\xi^\#, W] \rangle_G \quad (6.9)$$

for all vector fields V and W on M .

Now, fix a $\xi \in \mathcal{G}$. We denote by $Z(\xi^\#)$ the set of zeros of the vector field $\xi^\#$. For each $p \in Z(\xi^\#)$ we define a linear transformation

$$L_p(\xi) : T_p(M) \longrightarrow T_p(M)$$

by

$$L_p(\xi)(V_p) = (\mathcal{L}_{\xi^\#} V)p = [\xi^\#, V]_p \quad (6.10)$$

where V is any vector field on M with $V(p) = V_p$. By writing out the definition of the Lie derivative explicitly one obtains the following alternative expression for $L_p(\xi)(V_p)$.

$$L_p(\xi)(V_p) = -\frac{d}{dt}(\sigma_{\exp(-t\xi)})_{*p}(V_p). \quad (6.11)$$

Note that, since $\xi^\#(p) = 0$, $\sigma_{\exp(-t\xi)}(p) = p$ for every t so $(\sigma_{\exp(-t\xi)})_{*p} : T_p(M) \rightarrow T_p(M)$ and the derivative in (6.11) is computed in the single tangent

space $T_p(M)$. We claim that $L_p(\xi)$ is *skew-symmetric* with respect to the inner product on $T_p(M)$ supplied by $\langle \cdot, \cdot \rangle_G$, i.e., that

$$\langle L_p(\xi)(V_p), W_p \rangle_G = -\langle V_p, L_p(\xi)(W_p) \rangle_G \quad (6.12)$$

for all $V_p, W_p \in T_p(M)$. To see this one simply evaluates (6.9) at p and uses the fact that $\xi^\#(p) = 0$ and the definition (6.10) of $L_p(\xi)$. Next we will require a lemma which follows from a simple manipulation of well-known identities from differential geometry, but, since we use the result several times, we supply a proof.

Lemma 6.2. *Let H be an arbitrary smooth function on the symplectic manifold (M, ω) and V_H its Hamiltonian vector field ($dH = \iota_{V_H}\omega$). Suppose $p \in M$ and $V_H(p) = 0$. Then, for any $V_p, W_p \in T_p(M)$,*

$$\mathcal{H}_p(V_p, W_p) = -\omega((\mathcal{L}_{V_H}V)_p, W_p) \quad (6.13)$$

where \mathcal{H}_p is the Hessian of H at p , \mathcal{L}_{V_H} is the Lie derivative with respect to V_H and V is any vector field on M with $V(p) = V_p$.

Proof: By definition, $\mathcal{H}_p(V_p, W_p) = V_p(W(H))$ and

$$W(H) = \mathcal{L}_W H = dH(W) = (\iota_{V_H}\omega)(W) = (\iota_W \circ \iota_{V_H})(\omega).$$

Now compute

$$\begin{aligned} V(W(H)) &= \mathcal{L}_V(W(H)) = \mathcal{L}_V \circ \iota_W \circ \iota_{V_H}(\omega) \\ &= (\iota_{[V,W]} + \iota_W \circ \mathcal{L}_V) \circ \iota_{V_H}(\omega) \\ &= \iota_{[V,W]} \circ \iota_{V_H}(\omega) + \iota_W \circ \mathcal{L}_V \circ \iota_{V_H}(\omega) \\ &= \iota_{[V,W]} \circ \iota_{V_H}(\omega) + \iota_W \circ (\iota_{[V,V_H]} + \iota_{V_H} \circ \mathcal{L}_V)(\omega) \\ &= \omega(V_H, [V, W]) - \omega([V_H, V], W) + (\mathcal{L}_V\omega)(V_H, W). \end{aligned}$$

Now, evaluate at p and use $V(p) = 0$ to obtain

$$V_p(W(H)) = 0 - \omega([V_H, V]_p, W_p) + 0$$

i.e.,

$$\mathcal{H}_p(V_p, W_p) = -\omega((\mathcal{L}_{V_H}V)_p, W_p)$$

as required. ■

Proposition 6.3. *Let (M, ω) be a symplectic manifold with a Hamiltonian action of the compact Lie group G . Let $\xi \in \mathcal{G}$ be nondegenerate and $p \in Z(\xi^\#)$. Then $L_p(\xi) : T_p(M) \rightarrow T_p(M)$ is nonsingular.*

Proof: We apply Lemma 6.2 to $H = \mu(\xi)$. Then $V_H = V_{\mu(\xi)} = \xi^\#$. For any $V_p \in T_p(M)$ we select a vector field V on M with $V(p) = V_p$. Then $(\mathcal{L}_{V_H} V)_p = (\mathcal{L}_{\xi^\#} V)_p = L_p(\xi)(V_p)$. Since ξ is nondegenerate, $\mu(\xi)$ is a Morse function so its Hessian \mathcal{H}_p is nondegenerate. Thus, the equality

$$\mathcal{H}_p(V_p, W_p) = -\omega(L_p(\xi)(V_p), W_p)$$

implies that $L_p(\xi)(V_p)$ cannot be zero unless $V_p = 0$. ■

Remark. *We will improve this result shortly by showing that if G is any compact Lie group acting on any (not necessarily symplectic) manifold M and if $\xi \in \mathcal{G}$ has the property that $\xi^\#$ (defined by (6.7)) has only isolated zeros, then each $L_p(\xi)$ (defined by (6.10)) is nonsingular.*

Now, let us assume that $\xi \in \mathcal{G}$ is nondegenerate and $p \in Z(\xi^\#)$. Then $L_p(\xi)$ is nonsingular and skew-symmetric with respect to $\langle \cdot, \cdot \rangle_G$. Thus, we can find a basis $\{e_1, \dots, e_{2k}\}$ for $T_p(M)$ that is orthonormal with respect to $\langle \cdot, \cdot \rangle_G$ and oriented (with respect to the Liouville form) and relative to which the matrix of $L_p(\xi)$ is of the form

$$\begin{pmatrix} 0 & \lambda_1 & 0 & 0 & \cdots & 0 & 0 \\ -\lambda_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda_k \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_k & 0 \end{pmatrix} \quad (6.14)$$

with $\lambda_j \in \mathbb{R} - \{0\}$ for $j = 1, \dots, k$. We define a square root of the determinant of $L_p(\xi)$ by taking

$$[\det(L_p(\xi))]^{\frac{1}{2}} = \lambda_1 \lambda_2 \dots \lambda_k. \quad (6.15)$$

Remark. *This is, in fact, the Pfaffian of the matrix (6.14). Although this observation will play no role in what we do here it is crucial in formulating more general localization theorems than the one we will prove since these involve the so-called equivariant Euler class of a certain (equivariant) vector bundle and this is constructed, a la Chern-Weil, from the Pfaffian.*

With this we are prepared to formulate what we will call the **Generalized Duistermaat-Heckman Theorem**.

Theorem 6.4. *Let (M, ω) be a compact, symplectic manifold of dimension $n = 2k$ with a Hamiltonian action of a compact Lie group G (and corresponding symplectic moments given by $\mu : \mathcal{G} \rightarrow C^\infty(M)$). Orient M with the Liouville form $\nu_\omega = \frac{1}{k!} \omega^k$ (and assume that a G -invariant Riemannian metric $\langle \cdot, \cdot \rangle_G$ on M has been chosen). If $\xi \in \mathcal{G}$ is nondegenerate, then*

$$\int_M e^{i\mu(\xi)} \nu_\omega = \sum_{\substack{p \in M \\ \xi^\#(p)=0}} (2\pi i)^k [\det(L_p(\xi))]^{-\frac{1}{2}} e^{i\mu(\xi)(p)}. \quad (6.16)$$

Remark. *Since ξ is nondegenerate, $\mu(\xi)$ is Morse and so has (at least two and at most) finitely many critical points. These critical points coincide with the zeros of $\xi^\#$ so the sum in (6.16) is (nonvacuous and) finite. Furthermore, $L_p(\xi)$ is nonsingular so $\det(L_p(\xi)) \neq 0$ and the right-hand side of (6.16) is meaningful.*

As we mentioned earlier we shall eventually derive Theorem 6.4 as a consequence of our equivariant localization theorem. Our task for the present is simply to show that Theorem 6.4 implies Theorem 6.1. Thus, we begin with a compact, symplectic manifold (M, ω) of dimension $n = 2k$ and oriented by the Liouville form ν_ω . We let $H \in C^\infty(M)$ be a Morse function and V_H the corresponding Hamiltonian vector field on M . The assumption of Theorem 6.1 is that the flow of V_H is periodic. By rescaling we may assume that the period is 2π . Now, just as for our example on S^2 , this gives rise to a circle action on M with the property that

$$\xi = ia \in \text{Lie}(S^1) \implies \xi^\# = V_{-aH}. \quad (6.17)$$

In somewhat more detail, the action moves $p \in M$ along the integral curve of V_H that begins at p so

$$(i(-1))^\#(p) = \left. \frac{d}{dt}(e^{-t(-i)} \cdot) \right|_{t=0} = \left. \frac{d}{dt}(e^{ti} \cdot) \right|_{t=0} = V_H(p).$$

Moreover, $V_{-aH} = -aV_H$ because $d(-aH) = -adH = -a\iota_{V_H}\omega = \iota_{-aV_H}\omega$ so

$$\xi^\#(p) = (ia)^\#(p) = -a(i(-1))^\#(p) = -aV_H(p) = V_{-aH}(p).$$

Consequently, the S^1 -action is Hamiltonian with symplectic moments given by

$$\mu(\xi) = \mu(ia) = -aH$$

(equivariance is proved in the same way as for the S^2 example). Next we record a simple, but crucial fact about S^1 -actions in general.

Lemma 6.5. *Let M be a smooth manifold and suppose S^1 acts smoothly on M on the left. Then, for any nonzero ξ in the Lie algebra of S^1 , the zero set $Z(\xi^\#)$ of the vector field $\xi^\#$ coincides with the fixed point set of the S^1 -action.*

Proof: Since $\xi \neq 0$ and S^1 is one-dimensional, ξ spans the Lie algebra of S^1 , i.e., $\text{Lie}(S^1) = \{-t\xi; t \in \mathbb{R}\}$. The exponential map of $\text{Lie}(S^1)$ to S^1 is onto so the orbit of any $p \in M$ coincides with $\{\exp(-t\xi) \cdot p; t \in \mathbb{R}\}$, i.e., with the integral curve of $\xi^\#$ through p . If $\xi^\#(p) = 0$, then this integral curve is a point and therefore the orbit of p is a point, i.e., p is a fixed point for the S^1 -action. Since a fixed point is obviously a zero of any $\xi^\#$, the result follows. \blacksquare

Returning to the derivation of Theorem 6.1 from Theorem 6.4 we now have that any nonzero ξ in the Lie algebra of S^1 is nondegenerate. In particular, for any $T > 0$, we may apply Theorem 6.4 to $\xi = i(-T)$ to obtain

$$\begin{aligned} \int_M e^{iTH} \nu_\omega &= \sum_{\substack{p \in M \\ \xi^\#(p)=0}} (2\pi i)^k [\det(L_p(-iT))]^{-\frac{1}{2}} e^{iTH(p)} \\ &= \sum_{\substack{p \in M \\ dH(p)=0}} \left(\frac{2\pi}{T}\right)^k (iT)^k [\det(L_p(-iT))]^{-\frac{1}{2}} e^{iTH(p)}. \end{aligned}$$

Comparing this with the conclusion of Theorem 6.1 we find that we need only show

$$(iT)^k [\det(L_p(-iT))]^{-\frac{1}{2}} = e^{\pi i(\text{Sgn} \mathcal{H}_p)/4} |\det \mathcal{H}_p(e_i, e_j)|^{-\frac{1}{2}} \quad (6.18)$$

to complete the proof (here $\{e_1, \dots, e_{2k}\}$ is such a basis for $T_p(M)$ that satisfies $\frac{1}{k!}(\omega \wedge \dots \wedge \omega)(e_1, \dots, e_{2k}) = 1$). While largely computational, the proof of (6.18) relies on one nontrivial result so we shall go through it in some detail. First note that

$$L_p(-iT)(V_p) = (\mathcal{L}_{(-iT)\#} V)_p = T(\mathcal{L}_{(i(-1))\#} V)_p = T(\mathcal{L}_{V_H} V)_p.$$

Thus, (6.13) can be written

$$T\mathcal{H}_p(V_p, W_p) = -\omega(L_p(-iT)(V_p), W_p). \quad (6.19)$$

In particular, if $\{e_1, \dots, e_{2k}\}$ is any basis for $T_p(M)$,

$$T\mathcal{H}_p(e_i, e_j) = -\omega(L_p(-iT)(e_i), e_j)$$

and if we write $L_p(-iT)(e_i) = L_i^l e_l$, then

$$T\mathcal{H}_p(e_i, e_j) = -L_i^l \omega(e_l, e_j)$$

for all $i, j = 1, \dots, 2k$. As a matrix product this is

$$\begin{pmatrix} T\mathcal{H}_p(e_1, e_1) & \cdots & T\mathcal{H}_p(e_1, e_{2k}) \\ \vdots & & \vdots \\ T\mathcal{H}_p(e_{2k}, e_1) & \cdots & T\mathcal{H}_p(e_{2k}, e_{2k}) \end{pmatrix} \quad (6.20)$$

$$= \begin{pmatrix} -L_1^1 & \cdots & -L_1^{2k} \\ \vdots & & \vdots \\ -L_{2k}^1 & \cdots & -L_{2k}^{2k} \end{pmatrix} \begin{pmatrix} \omega(e_1, e_1) & \cdots & \omega(e_1, e_{2k}) \\ \vdots & & \vdots \\ \omega(e_{2k}, e_1) & \cdots & \omega(e_{2k}, e_{2k}) \end{pmatrix}$$

Now we will make a particular choice of basis. The classical Darboux Theorem guarantees the existence of an oriented, orthonormal basis for $T_p(M)$ relative to which

$$(\omega(e_i, e_j)) = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 0 \end{pmatrix} \quad (6.21)$$

and $\frac{1}{k!}(\omega \wedge \dots \wedge \omega)(e_1, \dots, e_{2k}) = 1$. We know also that we can find an oriented, orthonormal basis in which the matrix of $L_p(-iT)$ has the form (6.14). It so happens that for circle actions (and, more generally, for torus actions) it is possible to do all of this simultaneously, i.e., to find one oriented, orthonormal basis $\{e_1, \dots, e_{2k}\}$ for $T_p(M)$ in which (6.14), (6.21) and $\nu_\omega(e_1, \dots, e_{2k}) = 1$ are all satisfied (see Section 32 of [23]). Making such a choice of basis, substituting (6.14) and (6.21) into (6.20) and taking determinants gives

$$T^{2k} \det(\mathcal{H}_p(e_i, e_j)) = \lambda_1^2 \dots \lambda_k^2.$$

Thus,

$$\begin{aligned} T^k |\det(\mathcal{H}_p(e_i, e_j))|^{\frac{1}{2}} &= \text{Sign}(\lambda_1 \dots \lambda_k) \lambda_1 \dots \lambda_k \\ &= \text{Sign}(\lambda_1 \dots \lambda_k) [\det(L_p(-iT))]^{\frac{1}{2}} \end{aligned}$$

(where $\text{Sign}(\lambda_1 \dots \lambda_k) = 1$ if $\lambda_1 \dots \lambda_k > 0$ and $\text{Sign}(\lambda_1 \dots \lambda_k) = -1$ if $\lambda_1 \dots \lambda_k < 0$) and so

$$T^k [\det(L_p(iT))]^{-\frac{1}{2}} = \text{Sign}(\lambda_1 \dots \lambda_k) |\det(\mathcal{H}_p(e_i, e_j))|^{-\frac{1}{2}}. \quad (6.22)$$

Comparing (6.22) and (6.18) we see that all that remains is to prove

$$\text{Sign}(\lambda_1 \dots \lambda_k) = (-i)^k e^{\pi i (\text{Sgn} \mathcal{H}_p(e_i, e_j)) / 4} \mathcal{S}. \quad (6.23)$$

This will follow easily by induction if we can show that it is true for $k = 1$. In this case, (6.20) gives

$$\begin{pmatrix} T\mathcal{H}_p(e_1, e_1) & T\mathcal{H}_p(e_1, e_2) \\ T\mathcal{H}_p(e_2, e_1) & T\mathcal{H}_p(e_2, e_2) \end{pmatrix} = \begin{pmatrix} 0 & -\lambda_1 \\ \lambda_1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and so

$$\begin{pmatrix} \mathcal{H}_p(e_1, e_1) & \mathcal{H}_p(e_1, e_2) \\ \mathcal{H}_p(e_2, e_1) & \mathcal{H}_p(e_2, e_2) \end{pmatrix} = \begin{pmatrix} \lambda_1/T & 0 \\ 0 & \lambda_1/T \end{pmatrix}.$$

Now, if $\lambda_1 > 0$ ($\text{Sign}(\lambda_1) = 1$), then $\text{Sgn}(\mathcal{H}_p(e_i, e_j)) = 2$ so

$$(-i)^1 e^{\pi i (\text{Sgn} \mathcal{H}_p(e_i, e_j))/4} = -ie^{\pi i/2} = 1 = \text{Sign}(\lambda_1)$$

and, similarly, if $\lambda_1 < 0$ ($\text{Sign}(\lambda_1) = -1$), then $\text{Sgn}(\mathcal{H}_p(e_i, e_j)) = -2$ so

$$(-i)^1 e^{\pi i (\text{Sgn} \mathcal{H}_p(e_i, e_j))/4} = -ie^{-\pi i/2} = -1 = \text{Sign}(\lambda_1).$$

Leaving the induction to the reader this completes the proof of (6.23) and therefore the derivation of the Duistermaat-Heckman Theorem 6.1 from the Generalized Duistermaat-Heckman Theorem 6.4. We shall find that Theorem 6.4 itself is a simple consequence of a beautiful localization theorem in equivariant cohomology.

The basic philosophical principle behind all of the equivariant localization theorems is that, in some sense, “ G -equivariant cohomology is determined by the fixed point set of the G -action”. Our first lemma is an initial indication of what this means and why it is true. Roughly, it says that if $\alpha \in \Omega_G^*(M)$ is G -equivariantly closed, then, for each $\xi \in \mathcal{G}$, $\alpha(\xi)_{[n]}$ ($n = \dim M$) is cohomologically trivial away from the zero set $Z(\xi^\#)$ (which contains the fixed point set of the G -action and, for S^1 -actions and $\xi \neq 0$, coincides with it by Lemma 6.5).

Lemma 6.6. *Let M be a smooth n -manifold and G a compact Lie group that acts smoothly on M on the left. Let $\alpha \in \Omega_G^*(M)$ be G -equivariantly closed. Then, for each $\xi \in \mathcal{G}$, $\alpha(\xi)_{[n]}$ is (de Rham) exact on $M - Z(\xi^\#)$.*

Proof: Fix a nonzero $\xi \in \mathcal{G}$ (the result is vacuous if $\xi = 0$).

Remark. *We will actually prove more than is asserted in the lemma. Since the additional strength will be required in the derivation of the Duistermaat-Heckman Theorem we will elaborate. Note that, with $\xi \in \mathcal{G}$ held fixed and for any $\alpha \in \Omega_G^*(M)$*

$$(d_G \alpha)(\xi) = (d - \iota(\xi^\#))\alpha(\xi).$$

Define

$$d_{\xi\#} = d - \iota(\xi\#). \quad (6.24)$$

Then $d_{\xi\#}$ acts on $\Omega^*(M)$, and, by (5.9)

$$d_{\xi\#} \circ d_{\xi\#} = -\mathcal{L}_{\xi\#}.$$

Consequently, on the subspace

$$\Omega_{\xi\#}^*(M) = \{\varphi \in \Omega^*(M); \mathcal{L}_{\xi\#}\varphi = 0\} \quad (6.25)$$

of $\xi\#$ -invariant forms

$$d_{\xi\#} \circ d_{\xi\#} = 0. \quad (6.26)$$

Applying analogous formulas for d and $\iota_{\xi\#}$ one obtains the Leibnitz Rule

$$d_{\xi\#}(\omega \wedge \eta) = (d_{\xi\#}\omega) \wedge \eta + [\omega_{[0]} - \omega_{[1]} + \cdots + (-1)^n \omega_{[n]}] \wedge d_{\xi\#}\eta \quad (6.27)$$

for any $\omega, \eta \in \Omega^*(M)$. The proof of Lemma 6.6 will rely only on the fact that $d_{\xi\#}(\alpha(\xi)) = 0$ and the properties of $d_{\xi\#}$ just described. In particular, the conclusion will also be true of any $\Omega^*(M)$ -valued map $\xi \rightarrow \alpha(\xi)$ on \mathcal{G} even if it is not polynomial in ξ , provided only that $d_{\xi\#}(\alpha(\xi)) = 0$.

Now we return to the proof of Lemma 6.6. Using the G -invariant Riemannian metric $\langle \cdot, \cdot \rangle_G$ on M we construct a one-form θ on M dual to $\xi\#$, i.e., we define

$$\theta(V) = \langle \xi\#, V \rangle_G \quad (6.28)$$

for each vector field V on M .

Claim #1: θ is $\xi\#$ -invariant, i.e., $\mathcal{L}_{\xi\#}\theta = 0$.

To see this we fix a $p \in M$ and $V_p \in T_p(M)$ and show that $(\mathcal{L}_{\xi\#}\theta)_p(V_p) = 0$. By definition,

$$\mathcal{L}_{\xi\#}\theta = \frac{d}{dt}(\sigma_{\exp(-t\xi)}^*\theta)|_{t=0}$$

so

$$\begin{aligned} (\mathcal{L}_{\xi\#}\theta)_p(V_p) &= \frac{d}{dt}((\sigma_{\exp(-t\xi)}^*\theta)_p(V_p))|_{t=0} \\ &= \frac{d}{dt}(\theta((\sigma_{\exp(-t\xi)})_*V_p))|_{t=0} \\ &= \frac{d}{dt}\langle \xi\#(\exp(-t\xi) \cdot p), (\sigma_{\exp(-t\xi)})_*V_p \rangle_G|_{t=0}. \end{aligned}$$

We leave it to the reader to verify that

$$\xi^\#(\exp(-t\xi) \cdot p) = (\sigma_{\exp(-t\xi)})_{*p}(\xi^\#(p))$$

which gives

$$\begin{aligned} (\mathcal{L}_{\xi^\#\theta})_p(V_p) &= \frac{d}{dt} \langle (\sigma_{\exp(-t\xi)})_{*p}(\xi^\#(p)), (\sigma_{\exp(-t\xi)})_{*p}(V_p) \rangle_G |_{t=0} \\ &= \frac{d}{dt} \langle \xi^\#(p), V_p \rangle_G |_{t=0} = 0 \end{aligned}$$

because $\langle \cdot, \cdot \rangle_G$ is invariant under the G -action. This proves Claim #1 and from it and (6.26) we conclude that

$$d_{\xi^\#\theta} = 0. \quad (6.29)$$

Now notice that

$$d_{\xi^\#\theta} = d\theta - \iota_{\xi^\#\theta} = d\theta - \langle \xi^\#, \xi^\# \rangle_G = -\|\xi^\#\|_G^2 + d\theta. \quad (6.30)$$

This is a (nonhomogeneous) element of $\Omega^*(M)$ whose scalar (i.e., $\Omega^0(M)$) part is $-\|\xi^\#\|_G^2$ and this scalar part is nonzero on $M - Z(\xi^\#)$.

Remark. A nonhomogeneous element of $\Omega^*(M)$ with nonzero scalar part always has a multiplicative inverse (relative to \wedge) obtained from the geometric series. Indeed, if we write such an element as $a + \alpha$ with $a \in \Omega^0(M)$, $a \neq 0$, and $\alpha \in \Omega^*(M)$ with $\alpha_{[0]} = 0$ and define

$$(a + \alpha)^{-1} = \frac{1}{a} \sum_{k=0}^{\infty} \left(-\frac{\alpha}{a}\right)^k$$

(a finite sum), then it is easy to verify that $(a + \alpha)^{-1} \wedge (a + \alpha) = (a + \alpha) \wedge (a + \alpha)^{-1} = 1 \in \Omega^0(M)$.

We conclude that, on $M - Z(\xi^\#)$, $d_{\xi^\#\theta} = -\|\xi^\#\|_G^2 + d\theta$ is invertible and

$$(d_{\xi^\#\theta})^{-1} = -\|\xi^\#\|_G^{-2}(1 + \|\xi^\#\|_G^{-2}d\theta + \dots). \quad (6.31)$$

Thus, on $M - Z(\xi^\#)$, we can define an element β of $\Omega^*(M)$ by

$$\beta = \theta \wedge (d_{\xi^\#\theta})^{-1}. \quad (6.32)$$

Claim #2: On $M - Z(\xi^\#)$, $d_{\xi^\#\beta} = 1$ and $\mathcal{L}_{\xi^\#\beta} = 0$.

To prove this we first compute

$$\begin{aligned} d_{\xi\#}\beta &= d_{\xi\#}(\theta \wedge (d_{\xi\#}\theta)^{-1}) = d_{\xi\#}\theta \wedge (d_{\xi\#}\theta)^{-1} - \theta \wedge d_{\xi\#}((d_{\xi\#}\theta)^{-1}) \\ &= 1 - \theta \wedge d_{\xi\#}((d_{\xi\#}\theta)^{-1}) \end{aligned}$$

and

$$\begin{aligned} \mathcal{L}_{\xi\#}\beta &= \mathcal{L}_{\xi\#}(\theta \wedge (d_{\xi\#}\theta)^{-1}) = \mathcal{L}_{\xi\#}\theta \wedge (d_{\xi\#}\theta)^{-1} + \theta \wedge \mathcal{L}_{\xi\#}((d_{\xi\#}\theta)^{-1}) \\ &= 0 + \theta \wedge \mathcal{L}_{\xi\#}((d_{\xi\#}\theta)^{-1}). \end{aligned}$$

Now we show that $d_{\xi\#}((d_{\xi\#}\theta)^{-1})$ and $\mathcal{L}_{\xi\#}((d_{\xi\#}\theta)^{-1})$ are both zero. Beginning with

$$(d_{\xi\#}\theta) \wedge (d_{\xi\#}\theta)^{-1} = 1$$

we compute $d_{\xi\#}$ of both sides to obtain

$$d_{\xi\#}(d_{\xi\#}\theta) \wedge (d_{\xi\#}\theta)^{-1} + [-\|\xi\#\|_G^2 + d\theta] \wedge d_{\xi\#}((d_{\xi\#}\theta)^{-1}) = 0$$

so, by (6.29) and (6.30)

$$(d_{\xi\#}\theta) \wedge d_{\xi\#}((d_{\xi\#}\theta)^{-1}) = 0.$$

Now multiply on both sides by $(d_{\xi\#}\theta)^{-1}$. The proof for $\mathcal{L}_{\xi\#}((d_{\xi\#}\theta)^{-1})$ is the same so this proves Claim #2.

Finally, we define $\lambda \in \Omega^*(M)$ by

$$\lambda = \beta \wedge \alpha(\xi) = (\theta \wedge (d_{\xi\#}\theta)^{-1}) \wedge \alpha(\xi)$$

and compute

$$\begin{aligned} d_{\xi\#}\lambda &= d_{\xi\#}(\beta \wedge \alpha(\xi)) = (d_{\xi\#}\beta) \wedge \alpha(\xi) + [\beta_{[0]} - \beta_{[1]} + \cdots] \wedge d_{\xi\#}(\alpha(\xi)) \\ &= 1 \wedge \alpha(\xi) + 0 = \alpha(\xi). \end{aligned}$$

Thus,

$$d\lambda - \iota_{\xi\#}\lambda = \alpha(\xi).$$

Now look at the top (n^{th}) degree parts. $\iota_{\xi\#}\lambda$ has none and $(d\lambda)_{[n]} = d(\lambda_{[n-1]})$ so

$$\alpha(\xi)_{[n]} = d(\lambda_{[n-1]})$$

and this completes the proof of Lemma 6.6. ■

For future reference we summarize what we have just proved.

$$\begin{aligned} d_{\xi^\#}(\alpha(\xi)) = 0 \text{ and } \boldsymbol{\theta} = \langle \xi^\#, \cdot \rangle_G &\implies \\ \alpha(\xi) = d_{\xi^\#}((\boldsymbol{\theta} \wedge (d_{\xi^\#}\boldsymbol{\theta})^{-1}) \wedge \alpha(\xi)) \end{aligned}$$

and

$$\alpha(\xi)_{[n]} = d(((\boldsymbol{\theta} \wedge (d_{\xi^\#}\boldsymbol{\theta})^{-1}) \wedge \alpha(\xi))_{[n-1]})$$

on $M - Z(\xi^\#)$.

In order to proceed further we must understand more about the structure of the set $Z(\xi^\#) = \{p \in M; \xi^\#(p) = 0\}$ of zeros of $\xi^\#$. Notice that it is clear from the definition (6.7) of $\xi^\#$ that any fixed point of the G -action on M is a zero of every $\xi^\#$ so every $Z(\xi^\#)$ contains the fixed point set

$$M^G = \{p \in M; g \cdot p = p, \quad \forall g \in G\}.$$

If $\xi \in \mathcal{G}$ has the property that $Z(\xi^\#) = M^G$ (i.e., $\xi^\#$ vanishes only at fixed points of the G -action), then ξ is said to be **nondegenerate** (for $G = S^1$ -actions, every ξ in the Lie algebra of S^1 is nondegenerate by Lemma 6.5). Notice that, in general, one can define, for any $\xi \in \mathcal{G}$, the subgroup

$$T_\xi = \text{closure}_G \{\exp(-t\xi); t \in \mathbb{R}\}$$

of G . Then $Z(\xi^\#)$ clearly coincides with the fixed point set of the action on M of T_ξ and, being compact, connected, and abelian, T_ξ is a torus. Thus, *the zero set of $\xi^\#$ is always the fixed point set of a torus action on M .*

As for Hamiltonian actions, we define, for each $p \in Z(\xi^\#)$ a linear transformation $L_p(\xi) : T_p(M) \rightarrow T_p(M)$ by (6.10) and note that it is skew-symmetric with respect to the G -invariant Riemannian metric $\langle \cdot, \cdot \rangle_G$ on M . Now we let \exp_p be the (metric) exponential map on $T_p(M)$ corresponding to $\langle \cdot, \cdot \rangle_G$. This carries a $V_p \in T_p(M)$ onto $\gamma_{V_p}(1)$, where γ_{V_p} is the geodesic of $\langle \cdot, \cdot \rangle_G$ with $\gamma'_{V_p}(0) = V_p$ and it is a local diffeomorphism of some neighborhood of 0 in $T_p(M)$ onto some neighborhood of p in M . The G -invariance of $\langle \cdot, \cdot \rangle_G$ implies that, on some neighborhood of 0 in $T_p(M)$

$$\exp_p(V_p + tL_p(\xi)(V_p)) = \exp(-t\xi) \cdot \exp_p(V_p). \quad (6.33)$$

Thus, if $\mathcal{L}_p(\xi)$ is the vector field on $T_p(M)$ corresponding to $L_p(\xi)$, i.e.,

$$\mathcal{L}_p(\xi) = \frac{d}{dt}[V_p + tL_p(\xi)(V_p)]|_{t=0}$$

then

$$\xi^\# = (\exp_p)_*(\mathcal{L}_p(\xi))$$

on some neighborhood of p in M . In particular, integral curves of $\mathcal{L}_p(\xi)$ are (locally) mapped by \exp_p to integral curves of $\xi^\#$.

Now, suppose $V_p \in \ker(L_p(\xi))$. Then $L_p(\xi)(V_p) = 0$ and this is the case if and only if the integral curve of $\mathcal{L}_p(\xi)$ through V_p is a point, i.e., the integral curve of $\xi^\#$ through $\exp_p(V_p)$ is a point. Since this is the case if and only if $\xi^\#(\exp_p(V_p)) = 0$ we conclude that, on some neighborhood of 0 in $T_p(M)$

$$V_p \in \ker(L_p(\xi)) \iff \exp_p(V_p) \in Z(\xi^\#).$$

Now $\ker(L_p(\xi))$ is a linear subspace (and therefore a submanifold) of $T_p(M)$ so the restriction of \exp_p to some open set in $\ker(L_p(\xi))$ maps diffeomorphically onto a neighborhood of p in $Z(\xi^\#)$. Thus, $Z(\xi^\#)$ has a local manifold structure near each of its points p whose dimension is $\dim(\ker(L_p(\xi)))$. This dimension need not be the same at each $p \in Z(\xi^\#)$, but is constant on the connected components of $Z(\xi^\#)$. Thus, we find that $Z(\xi^\#)$ is a disjoint union of submanifolds of M each of which has dimension $\dim(\ker(L_p(\xi)))$, where p is any point in the submanifold. In particular, we have the promised generalization of Proposition 6.3.

Proposition 6.7. *Let M be a smooth manifold, G a compact Lie group acting smoothly on M on the left, ξ an element of the Lie algebra \mathcal{G} of G and $p \in Z(\xi^\#)$ a zero of $\xi^\#$. Then p is an isolated point of $Z(\xi^\#)$ if and only if $L_p(\xi) : T_p(M) \rightarrow T_p(M)$ is invertible.*

Henceforth we assume that p is an isolated zero of $\xi^\#$. Then $L_p(\xi)$ is invertible and skew-symmetric with respect to $\langle \cdot, \cdot \rangle_G$. It follows that the dimension of M must be even, say

$$n = 2k$$

and that there exists an oriented, orthonormal basis $\{e_1, \dots, e_{2k}\}$ for $T_p(M)$ relative to which the matrix of $L_p(\xi)$ has the form (6.14) with $\lambda_j \in \mathbb{R} - \{0\}$ for $j = 1, \dots, k$. As before we define a square root of the determinant of $L_p(\xi)$ by (6.15). Now, if $V_p \in T_p(M)$ and we write $V_p = V_p^i e_i$ (summation convention), then

$$L_p(\xi)(V_p) = \lambda_1(V_p^2 e_1 - V_p^1 e_2) + \dots + \lambda_k(V_p^{2k} e_{2k-1} - V_p^{2k-1} e_{2k}).$$

If, as before, we identify $L_p(\xi)$ with a vector field $\mathcal{L}_p(\xi)$ on $T_p(M)$ and recall that, on some neighborhood of p in M , $\xi^\#$ agrees with $(\exp_p)_*(\mathcal{L}_p(\xi))$, then,

in normal coordinates x^1, \dots, x^{2k} on that neighborhood determined by \exp_p and $\{e_1, \dots, e_{2k}\}$, we have

$$\xi^\# = \lambda_1 \left(x^2 \frac{\partial}{\partial x^1} - x^1 \frac{\partial}{\partial x^2} \right) + \dots + \lambda_k \left(x^{2k} \frac{\partial}{\partial x^{2k-1}} - x^{2k-1} \frac{\partial}{\partial x^{2k}} \right). \quad (6.34)$$

Note that if p happens to be a fixed point of the G -action (e.g., if ξ is nondegenerate), then this neighborhood can be chosen G -invariant (restrict to some ϵ -ball relative to $\langle \cdot, \cdot \rangle_G$). With this we are finally prepared to prove our major result.

Theorem 6.8 (Equivariant Localization Theorem) *Let M be a compact, oriented manifold of dimension $n = 2k$ and G a compact Lie group acting smoothly on M on the left. Let α be a G -equivariantly closed differential form on M . Then, for any nondegenerate $\xi \in \mathcal{G}$ for which $\xi^\#$ has only isolated zeros,*

$$\int_M \alpha(\xi) = \sum_{\substack{p \in M \\ \xi^\#(p)=0}} (-2\pi)^k [\det L_p(\xi)]^{-\frac{1}{2}} \alpha(\xi)_{[0]}(p). \quad (6.35)$$

Remarks.

1. Since M is compact and $Z(\xi^\#)$ is discrete, the sum in (6.35) is finite. If $Z(\xi^\#)$ happens to be empty, then Lemma 6.6 implies that $\alpha(\xi)_{[n]}$ is exact on all of M so Stokes' Theorem gives $\int_M \alpha(\xi) = 0$ and (6.35) is vacuously satisfied.
2. For S^1 -actions Lemma 6.5 implies that the nondegeneracy assumption in Theorem 6.8 is unnecessary.
3. As was the case for Lemma 6.6 our proof of Theorem 6.8 will not use the full strength of the assumption that α is a G -equivariantly closed differential form on M , but only that $d_{\xi^\#}(\alpha(\xi)) = 0$ for the particular $\xi \in \mathcal{G}$ referred to in the Theorem.

Proof: By the first remark above we may assume $Z(\xi^\#) \neq \emptyset$. Let $p \in Z(\xi^\#)$. We have shown that we can find a G -invariant neighborhood U_p of p and (normal) coordinates x^1, \dots, x^{2k} on U_p such that $\xi^\#|_{U_p}$ is given by (6.34), where $[\det(L_p(\xi))]^{\frac{1}{2}} = \lambda_1 \cdots \lambda_{2k} \neq 0$. On U_p we define a one-form θ^p by

$$\theta^p = \lambda_1^{-1}(x^2 dx^1 - x^1 dx^2) + \dots + \lambda_k^{-1}(x^{2k} dx^{2k-1} - x^{2k-1} dx^{2k}). \quad (6.36)$$

Then a few simple computations show

$$\theta^p(\xi^\#) = (x^1)^2 + \dots + (x^{2k})^2 \quad (6.37)$$

$$d(\iota_{\xi^\#}\theta^p) = 2x^1 dx^1 + \dots + 2x^{2k} dx^{2k} \quad (6.38)$$

$$\iota_{\xi^\#}(d\theta^p) = -2x^1 dx^1 - \dots - 2x^{2k} dx^{2k} \quad (6.39)$$

and, from the last two of these

$$\mathcal{L}_{\xi^\#}\theta^p = (d \circ \iota_{\xi^\#} + \iota_{\xi^\#} \circ d)(\theta^p) = 0. \quad (6.40)$$

Now, each of the sets U_p , $p \in Z(\xi^\#)$, is G -invariant by construction and $M - Z(\xi^\#)$ is G -invariant because ξ is assumed nondegenerate (so $Z(\xi^\#) = M^G$ which is surely G -invariant). Thus

$$\{U_p\}_{p \in Z(\xi^\#)} \cup \{M - Z(\xi^\#)\}$$

is a G -invariant open cover of M . By choosing a partition of unity subordinate to this cover and averaging each of its elements over G (as we did to produce $\langle \cdot, \cdot \rangle_G$ and the map I in Section 5) one can produce a G -invariant partition of unity subordinate to the cover. With this and the one-forms θ^p on U_p and (as in the proof of Lemma 6.6) $\theta^0 = \langle \xi^\#, \cdot \rangle_G$ on $M - Z(\xi^\#)$, one can piece together a one-form θ on all of M with the following properties:

1. θ agrees with θ^p on some neighborhood of p .
2. $\mathcal{L}_{\xi^\#}\theta = 0$.
3. $d_{\xi^\#}\theta$ is invertible on $M - Z(\xi^\#)$.

Exactly as in the proof of Lemma 6.6, properties (2) and (3) together with $d_{\xi^\#}(\alpha(\xi)) = 0$ imply that

$$\alpha(\xi) = d_{\xi^\#} \left((\theta \wedge (d_{\xi^\#}\theta)^{-1}) \wedge \alpha(\xi) \right) \text{ on } M - Z(\xi^\#). \quad (6.41)$$

Now we compute the integral on the left-hand side of (6.35). For each $p \in Z(\xi^\#)$ and $\epsilon > 0$ sufficiently small we let

$$B_\epsilon(p) = \left\{ x = (x^1, \dots, x^{2k}); \|x\|_G^2 = (x^1)^2 + \dots + (x^{2k})^2 \leq \epsilon \right\} \subseteq U_p$$

and

$$S_\epsilon(p) = \{x; \|x\|_G^2 = \epsilon\}$$

and give both their usual orientations. Since $Z(\xi^\#)$ is a finite set

$$\begin{aligned}
\int_M \alpha(\xi) &= \int_{M-Z(\xi^\#)} \alpha(\xi) = \lim_{\epsilon \rightarrow 0} \int_{M-\cup_{p \in Z(\xi^\#)} B_\epsilon(p)} \alpha(\xi) \\
&= \lim_{\epsilon \rightarrow 0} \int_{M-\cup_{p \in Z(\xi^\#)} B_\epsilon(p)} d_{\xi^\#} \left((\boldsymbol{\theta} \wedge (d_{\xi^\#} \boldsymbol{\theta})^{-1}) \wedge \alpha(\xi) \right) \\
&= \lim_{\epsilon \rightarrow 0} \int_{M-\cup_{p \in Z(\xi^\#)} B_\epsilon(p)} d \left(\boldsymbol{\theta} \wedge (d_{\xi^\#} \boldsymbol{\theta})^{-1} \wedge \alpha(\xi) \right) \\
&\quad \text{(because the } \iota_{\xi^\#} \text{ term can have no top degree part)} \\
&= \lim_{\epsilon \rightarrow 0} \left(- \sum_{p \in Z(\xi^\#)} \int_{S_\epsilon(p)} (\boldsymbol{\theta} \wedge (d_{\xi^\#} \boldsymbol{\theta})^{-1}) \wedge \alpha(\xi) \right) \\
&\quad \text{(the minus sign being due to the switch from boundary to standard orientations)} \\
&= \sum_{\substack{p \in M \\ \xi^\#(p)=0}} \lim_{\epsilon \rightarrow 0} \left(- \int_{S_\epsilon(p)} (\boldsymbol{\theta} \wedge (d_{\xi^\#} \boldsymbol{\theta})^{-1}) \wedge \alpha(\xi) \right).
\end{aligned}$$

Comparing this with (6.35) we see that it remains only to prove that, for each $p \in Z(\xi^\#)$,

$$\lim_{\epsilon \rightarrow 0} \left(- \int_{S_\epsilon(p)} (\boldsymbol{\theta} \wedge (d_{\xi^\#} \boldsymbol{\theta})^{-1}) \wedge \alpha(\xi) \right) = (-2\pi)^k [\det L_p(\xi)]^{-\frac{1}{2}} \alpha(\xi)_{[0]}(p). \quad (6.42)$$

Thus, we fix a $p \in Z(\xi^\#)$. For each $\epsilon > 0$ sufficiently small, $\boldsymbol{\theta} = \boldsymbol{\theta}^p$ on $S_\epsilon(p)$. For such an $\epsilon > 0$ we introduce a change of coordinates on U_p by rescaling each x^i by a factor of $\sqrt{\epsilon}$, i.e., we replace x^i everywhere with $\sqrt{\epsilon}x^i$, $i = 1, \dots, 2k$:

$$x^i \longrightarrow \sqrt{\epsilon}x^i, \quad i = 1, \dots, 2k. \quad (6.43)$$

In the new coordinates, $S_\epsilon(p)$ becomes the unit sphere $S_1(p)$. Write $\alpha_\epsilon(\xi)$ for $\alpha(\xi)$ written in these new coordinates, i.e.,

$$\alpha_\epsilon(\xi)(x, dx) = \alpha(\xi)(\sqrt{\epsilon}x, \sqrt{\epsilon}dx).$$

Notice that, as $\epsilon \rightarrow 0$, all of the $\alpha_\epsilon(\xi)_{[i]}$ with $i > 0$ approach 0, whereas $\alpha_\epsilon(\xi)_{[0]} \rightarrow \alpha(\xi)_{[0]}(p)$ since $p = (0, \dots, 0)$.

Now we consider the effect of this substitution on $\theta \wedge (d_{\xi\#}\theta)^{-1}$. Near p , $\theta = \theta^p$ is given by (6.36) so (6.43) introduces an extra factor of ϵ . On the other hand,

$$d_{\xi\#}\theta = d\theta - \iota_{\xi\#}\theta = -2(\lambda_1^{-1}dx^1 \wedge dx^2 + \dots + \lambda_k^{-1}dx^{2k-1} \wedge dx^{2k}) - \|x\|_G^2$$

so this also picks up a factor of ϵ . Consequently, $(d_{\xi\#}\theta)^{-1}$ acquires a new factor of $\frac{1}{\epsilon}$ and, as a result, $\theta \wedge (d_{\xi\#}\theta)^{-1}$ is unaffected by the rescaling (6.43). Thus,

$$\int_{S_\epsilon(p)} (\theta \wedge (d_{\xi\#}\theta)^{-1}) \wedge \alpha(\xi) = \int_{S_1(p)} (\theta \wedge (d_{\xi\#}\theta)^{-1}) \wedge \alpha_\epsilon(\xi)$$

and so

$$\lim_{\epsilon \rightarrow 0} \left(- \int_{S_\epsilon(p)} (\theta \wedge (d_{\xi\#}\theta)^{-1}) \wedge \alpha(\xi) \right) = \left(- \int_{S_1(p)} (\theta \wedge (d_{\xi\#}\theta)^{-1}) \right) \alpha(\xi)_{[0]}(p). \quad (6.44)$$

We therefore compute

$$\begin{aligned} - \int_{S_1(p)} \theta \wedge (d_{\xi\#}\theta)^{-1} &= - \int_{S_1(p)} \theta \wedge (d\theta - 1)^{-1} (\|x\|_G^2 = 1 \text{ on } S_1(p)) \\ &= \int_{S_1(p)} \theta \wedge (1 - d\theta)^{-1} \\ &= \int_{S_1(p)} \theta \wedge (1 + d\theta + (d\theta)^2 + \dots + (d\theta)^{k-1} + (d\theta)^k) \\ &\quad \text{(where } (d\theta)^2 = d\theta \wedge d\theta, \text{ etc.)} \\ &= \int_{S_1(p)} \theta \wedge (d\theta)^{k-1} \text{ (since } \dim S_1(p) = 2k - 1) \\ &= \int_{B_1(p)} (d\theta)^k \text{ (by Stokes' Theorem since } d(\theta \wedge (d\theta)^{k-1}) \\ &= d\theta \wedge (d\theta)^{k-1} - \theta \wedge d((d\theta)^{k-1}) = (d\theta)^k - 0) \\ &= \int_{B_1(p)} ((-2)(\lambda_1^{-1}dx^1 \wedge dx^2 + \dots + \lambda_k^{-1}dx^{2k-1} \wedge dx^{2k}))^k \end{aligned}$$

$$\begin{aligned}
&= (-2)^k k! \lambda_1^{-1} \dots \lambda_k^{-1} \int_{B_1(p)} dx^1 \wedge \dots \wedge dx^{2k} \\
&= (-2)^k k! (\lambda_1 \dots \lambda_k)^{-1} \left(\frac{\pi^k}{k!} \right) = (-2\pi)^k [\det L_p(\xi)]^{-\frac{1}{2}}.
\end{aligned}$$

Substituting this into (6.44) yields (6.42) and so completes the proof of Theorem 6.8. \blacksquare

Finally we will derive the Generalized Duistermaat-Heckman Theorem 6.4 from our Localization Theorem 6.8. Recall that the scenario is as follows. We have a compact, symplectic manifold (M, ω) of dimension $2k$ and oriented by the Liouville form $\nu_\omega = \frac{1}{k!} \omega^k$. There is a Hamiltonian action of a compact Lie group G on M with corresponding equivariant moments given by $\mu : \mathcal{G} \rightarrow C^\infty(M)$. Finally, we have a $\xi \in \mathcal{G}$ which is nondegenerate. Notice that, because the action is Hamiltonian, nondegeneracy of ξ implies that $\xi^\#$ has isolated zeros ($\mu(\xi)$ is a Morse function and the zeros of $\xi^\#$ coincide with the critical points of $\mu(\xi)$). Our objective is to prove (6.16).

We consider a map $\mathcal{G} \rightarrow \Omega^*(M)$ called the **G -equivariant symplectic form** ω_G defined by

$$\omega_G = \mu + \omega,$$

i.e.,

$$\omega_G(\xi) = \mu(\xi) + \omega$$

for every $\xi \in \mathcal{G}$. We claim that

$$d_{\xi^\#}(\omega_G(\xi)) = 0 \tag{6.45}$$

for every $\xi \in \mathcal{G}$. Indeed,

$$\begin{aligned}
d_{\xi^\#}(\omega_G(\xi)) &= (d - \iota_{\xi^\#})(\mu(\xi) + \omega) \\
&= d(\mu(\xi) + \omega) - \iota_{\xi^\#}(\mu(\xi) + \omega) \\
&= d\mu(\xi) + 0 - 0 - \iota_{\xi^\#}\omega \\
&= 0 \quad (\text{by (6.8)}).
\end{aligned}$$

Now consider the element $e^{i\omega_G(\xi)} \in \Omega^*(M)$

$$e^{i\omega_G(\xi)} = 1 + i\omega_G(\xi) - \frac{1}{2}\omega_G(\xi) \wedge \omega_G(\xi) + \dots$$

(a finite sum). Since (6.45) and the Leibnitz Rule (6.27) imply that

$$d_{\xi^\#}(\omega_G(\xi) \wedge \dots \wedge \omega_G(\xi)) = 0$$

we conclude that

$$d_{\xi^\#} \left(e^{i\omega_G(\xi)} \right) = 0. \quad (6.46)$$

Remark (3) following Theorem 6.8 implies that (6.46) is sufficient to apply the Localization Theorem to $e^{i\omega_G(\xi)}$. Since

$$e^{i\omega_G(\xi)} = e^{i(\mu(\xi) + \omega)} = e^{i\mu(\xi)} e^{i\omega} = e^{i\mu(\xi)} \left(1 + i\omega - \frac{1}{2}\omega^2 + \dots \right)$$

we have

$$\left(e^{i\omega_G(\xi)} \right)_{[0]} = e^{i\mu(\xi)}.$$

Thus, (6.35) gives

$$\begin{aligned} \sum_{\substack{p \in M \\ \xi^\#(p)=0}} (-2\pi)^k [\det L_p(\xi)]^{-\frac{1}{2}} e^{i\mu(\xi)(p)} &= \int_M e^{i\omega_G(\xi)} = \int_M e^{i\mu(\xi)} e^{i\omega} \\ &= \int_M e^{i\mu(\xi)} \left(\frac{1}{k!} i^k \omega^k \right) = i^k \int_M e^{i\mu(\xi)} \nu_\omega \end{aligned}$$

which is (6.16).

7. Duality and Seiberg-Witten

In Section 5 we briefly described an argument which led Witten [52] to identify the partition function

$$Z_{DW} = \int e^{-S_{DW}[\Phi]/e^2} \mathcal{D}\Phi \quad (7.1)$$

of Donaldson-Witten theory with the zero-dimensional Donaldson invariant. Even more briefly, the idea is this: The symmetries built into the action $S_{DW}[\Phi]$ ensure that Z_{DW} is independent of the coupling constant e so that it can be computed in the limit $e \rightarrow 0$. For these small values of the coupling constant one has available the semiclassical (stationary phase) approximation which, again by virtue of the symmetries, one can show (formally) must be exact. In finite dimensions at least we saw in Section 6 that exactness of the stationary phase approximation is tantamount to the localization of the integral. Since $S_{DW}[\Phi]$ is gauge invariant, the integration in Z_{DW} can be understood over the space of field configurations modulo gauge transformations and, thought of in this way, the integral in (7.1) is found to localize to the moduli space of instantons, thus yielding Donaldson's definition of the zero-dimensional invariant.

Remarkable as it is that the subtle differential-topological invariants of Donaldson can be recast in these quantum field-theoretic terms, Witten's achievement would perhaps amount to no more than a fascinating novelty were it not for the fact that quantum field theory immediately suggests a new line of investigation. The partition function Z_{DW} is independent of e and all that we said above was obtained by analyzing it perturbatively in the so-called weak coupling limit $e \rightarrow 0$. Perhaps an entirely new perspective on Z_{DW} (i.e., on the Donaldson invariant) could be gleaned from an analysis in the strong coupling regime $e \rightarrow \infty$. Until relatively recently, however, such an analysis was out of the question because the strongly coupled, nonperturbative behavior of such quantum field theories was not well understood. In 1994, Seiberg and Witten [45] made an astonishing breakthrough in understanding this behavior for a certain class of quantum field theories ("N = 2 supersymmetric Yang-Mills theories"). Two key ingredients led to the unraveling of this behavior. The first was that an important part of the action depended *holomorphically* on the coupling constants. This puts severe restrictions on how the quantum theories can change as the coupling constants are varied. We will, however, have nothing further to say about this part of the story and will turn instead to that aspect of the Seiberg-Witten discovery that is more directly relevant to Witten's conjecture regarding the Donaldson invariants.

Since the earliest days of quantum mechanics it has been understood that a given classical system can admit more than one "quantization" (if for no other reason than the ordering ambiguity one encounters in replacing classical observables, which commute, with operators, which do not). More surprising, and a more recent discovery, is the fact that a given quantum system can result from the quantization of two quite distinct classical systems. Perhaps the best known example of this phenomenon is the quantization of the Sine-Gordon and Thirring Models in two spacetime dimensions (see, for example, [21]). When such a situation occurs the two classical models are said to be *dual* and one is presented with the possibility of adopting two, possibly quite different views of the quantum theory. For example, for a theory (such as SU(2) Yang-Mills-Higgs) in which magnetic monopoles appear as excitations of the fundamental fields (i.e., soliton-like solutions to the field equations), a "dual" description of the theory might have the monopoles themselves as the fundamental fields with the previously fundamental fields arising as excitations of these. The hallmark of such duality transformations is a symmetry that interchanges "electric" and "magnetic" aspects of the theory and therefore also (as we shall see) strong and weak coupling. When such a duality exists one can translate a question that is inherently nonperturbative (and therefore intractable) into a "dual" question in the weak coupling regime where

there is some hope of finding an answer. Seiberg and Witten [45] applied this strategy to the physical problem of *confinement* and Witten [54] used it to sort out the dual version of Donaldson theory.

Duality symmetries of this sort are exceedingly subtle and deep and we would not presume to offer an exegesis (see [1] for a rather detailed outline and Volume 2 of [11] for more details). However, we will pause to show that this notion of duality actually has its roots in classical electromagnetic theory and how a natural symmetry might interchange strong and weak coupling.

As in Section 3 we will let X denote some open submanifold of Minkowski space-time $\mathbb{R}^{1,3}$ with its usual orientation and semi-Riemannian metric and standard coordinates x^0, x^1, x^2 , and x^3 . We identify an electromagnetic field on X with a $U(1)$ -gauge field strength on X satisfying Maxwell's equations. In more detail, we consider a principal $U(1)$ -bundle $U(1) \hookrightarrow P \xrightarrow{\pi} X$ over X and a connection ω on it with curvature $\Omega = d\omega$ ($U(1)$ is abelian). Identifying the Lie algebra $\mathfrak{u}(1)$ of $U(1)$ with $i\mathbb{R}$ and letting $s : V \rightarrow P$ denote a section of the bundle we write the corresponding gauge potential \mathcal{A} and field strength \mathcal{F} as

$$\mathcal{A} = s^*\omega = -iA$$

and

$$\mathcal{F} = s^*\Omega = d\mathcal{A} = -idA = -iF$$

where A and F are the usual real-valued forms describing the potential and field in physics. Writing these out in standard coordinates gives

$$\mathcal{A} = A_\alpha dx^\alpha = -iA_\alpha dx^\alpha$$

and

$$\mathcal{F} = \frac{1}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta = -\frac{i}{2}F_{\alpha\beta}dx^\alpha \wedge dx^\beta.$$

Assuming (for the moment) that X contains none of the sources of the electromagnetic field, F satisfies the **(source free) Maxwell equations**

$$dF = 0 \tag{7.2}$$

$$d*F = 0 \tag{7.3}$$

where $*$ is the Hodge star on X determined by the Minkowski metric and the usual orientation. The first appearance of electromagnetic duality in classical physics is the obvious invariance of (7.2) and (7.3) under the symmetry $F \rightarrow *F$. To make contact with the notation commonly used in physics we define functions

E^1, E^2, E^3 and B^1, B^2, B^3 (thought of as components of the ordinary spatial electric \vec{E} and magnetic \vec{B} field vectors) by

$$F_{io} = E^i \text{ and } F_{ij} = \varepsilon_{ijk} B^k, \quad i, j, k = 1, 2, 3. \quad (7.4)$$

Thus,

$$\begin{aligned} \mathbf{F} &= (E^1 dx^1 + E^2 dx^2 + E^3 dx^3) \wedge dx^0 \\ &\quad + B^3 dx^1 \wedge dx^2 - B^2 dx^1 \wedge dx^3 + B^1 dx^2 \wedge dx^3 \end{aligned} \quad (7.5)$$

and a computation of the Hodge dual gives

$$\begin{aligned} *\mathbf{F} &= (-B^1 dx^1 - B^2 dx^2 - B^3 dx^3) \wedge dx^0 \\ &\quad + E^3 dx^1 \wedge dx^2 - E^2 dx^1 \wedge dx^3 + E^1 dx^2 \wedge dx^3. \end{aligned} \quad (7.6)$$

Computing $d\mathbf{F}$ and $d*\mathbf{F}$ one finds that Maxwell's equations (7.2) and (7.3) assume their more familiar forms

$$\vec{\nabla} \cdot \vec{B} = 0 \text{ and } \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial x^0} = \vec{0} \quad (d\mathbf{F} = 0) \quad (7.7)$$

$$\vec{\nabla} \cdot \vec{E} = 0 \text{ and } \vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial x^0} = \vec{0} \quad (d*\mathbf{F} = 0) \quad (7.8)$$

where $\vec{\nabla} = (\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3})$ is the usual spatial gradient operator and the \cdot and \times refer to the dot and cross products on \mathbb{R}^3 . Notice also that the symmetry $\mathbf{F} \rightarrow *\mathbf{F}$ of (7.2) and (7.3) becomes, from (7.5) and (7.6), $\vec{E} \rightarrow \vec{B}$ and $\vec{B} \rightarrow -\vec{E}$ so that, up to a sign, it interchanges electric and magnetic fields. Consequently, if we combine \vec{E} and \vec{B} into $\vec{E} + i\vec{B}$, $\mathbf{F} \rightarrow s*\mathbf{F}$ can be written

$$\vec{E} + i\vec{B} \longrightarrow \vec{B} - i\vec{E} = -i(\vec{E} + i\vec{B}) = e^{-\frac{\pi}{2}i}(\vec{E} + i\vec{B}). \quad (7.9)$$

Seen in this light one finds that the equations (7.7) and (7.8) actually admit a $U(1)$ symmetry

$$\vec{E} + i\vec{B} \longrightarrow e^{i\phi}(\vec{E} + i\vec{B}) = (\cos \phi \vec{E} - \sin \phi \vec{B}) + i(\sin \phi \vec{E} + \cos \phi \vec{B}) \quad (7.10)$$

where $e^{i\phi} \in U(1)$ is arbitrary (just substitute $\vec{E}' = \cos \phi \vec{E} - \sin \phi \vec{B}$ and $\vec{B}' = \sin \phi \vec{E} + \cos \phi \vec{B}$ into (7.7) and (7.8)). Finally note that, with this notation, the four equations in (7.7) and (7.8) can be written as

$$\vec{\nabla} \cdot (\vec{E} + i\vec{B}) = 0 \quad (7.11)$$

and

$$\frac{\partial}{\partial x^0}(\vec{E} + i\vec{B}) + i\vec{\nabla} \times (\vec{E} + i\vec{B}) = \vec{0}. \quad (7.12)$$

For regions X in which sources for the electromagnetic field are present equations (7.7) and (7.8) are modified by replacing $\vec{\nabla} \cdot \vec{E} = 0$ and $\frac{\partial \vec{E}}{\partial x^0} - \vec{\nabla} \times \vec{B} = \vec{0}$ by $\vec{\nabla} \cdot \vec{E} = q$ and $\frac{\partial \vec{E}}{\partial x^0} - \vec{\nabla} \times \vec{B} = \vec{j}_e$, respectively, where q is the electric charge density and \vec{j}_e is the electric current density. Thus, (7.11) and (7.12) become

$$\vec{\nabla} \cdot (\vec{E} + i\vec{B}) = q \quad (7.13)$$

$$\frac{\partial}{\partial x^0}(\vec{E} + i\vec{B}) + i\vec{\nabla} \times (\vec{E} + i\vec{B}) = \vec{j}_e. \quad (7.14)$$

These equations, of course, no longer admit even the special case (7.9) of the $U(1)$ symmetry (7.10). We will see that the full $U(1)$ symmetry can be restored if one is willing to hypothesize the presence of analogous magnetic charge g and current \vec{j}_m densities. Since such magnetic densities would presumably arise, as in the electric case, from a fundamental magnetically charged particle and since such a particle has never been observed in nature, we should pause to briefly discuss such “magnetic monopoles.”

Dirac [12] was the first to take seriously the implications, especially for quantum mechanics, of the possible existence of magnetic charges, which are explicitly forbidden ($\vec{\nabla} \cdot \vec{B} = 0$) by the traditional form (7.13) and (7.14) of Maxwell’s equations. Such a magnetic charge (or monopole) is a point object assumed to create the “magnetic analogue” of a Coulomb field and we wish to consider one of “magnetic charge” g at rest at the origin. Since our situation is entirely static we may restrict the analysis to a single spatial cross-section of Minkowski spacetime (say, $x^0 = 0$) which we will denote simply \mathbb{R}^3 and in which we will write standard rectangular and spherical coordinates (x, y, z) and (ρ, ϕ, θ) respectively (for the record, our ϕ is measured from the positive z -axis and takes values in $0 \leq \phi \leq \pi$). The field we wish to consider then is defined on $\mathbb{R}^3 - \{(0, 0, 0)\}$ and is given there by $\vec{E} = \vec{0}$ and

$$\vec{B} = \frac{g}{\rho^2} \vec{e}_\rho = \frac{g}{(x^2 + y^2 + z^2)^{3/2}} (x\vec{e}_x + y\vec{e}_y + z\vec{e}_z) \quad (7.15)$$

where \vec{e}_ρ is the outward unit radial field and \vec{e}_x , \vec{e}_y and \vec{e}_z are unit vectors in the positive x , y and z directions, respectively. One can write the field strength two-form F in either rectangular

$$F = g(x^2 + y^2 + z^2)^{-3/2} (x dy \wedge dz - y dx \wedge dz + z dx \wedge dy) \quad (7.16)$$

or spherical

$$\mathbf{F} = g \sin \phi d\phi \wedge d\theta \quad (7.17)$$

coordinates. Notice that (7.17) is independent of ρ and so may be regarded as a two-form on the unit sphere S^2 in \mathbb{R}^3 (we will return to this point shortly). It is easy to see that \mathbf{F} is not exact on $\{\mathbb{R}^3 - (0, 0, 0)\}$, i.e., that there does not exist a smooth one-form \mathbf{A} on $\{\mathbb{R}^3 - (0, 0, 0)\}$ for which $d\mathbf{A} = \mathbf{F}$ there ((7.17) is easily integrated over $S^2 \subseteq \mathbb{R}^3$ to give $\int_{S^2} \mathbf{F} = 4\pi g$, but the existence of such an \mathbf{A} would, by Stokes' Theorem, give a value of zero for the integral). Dirac was well aware of this, of course, and also knew (although perhaps not in these terms) that if one deletes from \mathbb{R}^3 not only the location of the monopole, but also some ray extending from it to infinity (a so-called **Dirac string**), then the result is a submanifold of \mathbb{R}^3 whose 2nd deRham cohomology is trivial so that every two-form on it (e.g., \mathbf{F}) is exact. Taking the Dirac string to be the non-negative z -axis one can explicitly write down a one-form \mathbf{A}_S on $\{\mathbb{R}^3 - (0, 0, z); z \geq 0\}$ with $d\mathbf{A}_S = \mathbf{F}$ there. One such is

$$\mathbf{A}_S = \frac{g}{\rho(\rho - z)}(ydx - xdy) = -g(1 + \cos \phi)d\theta. \quad (7.18)$$

Similarly, on $\{\mathbb{R}^3 - (0, 0, z); z \leq 0\}$,

$$\mathbf{A}_N = \frac{g}{\rho(\rho + z)}(ydx - xdy) = -g(1 - \cos \phi)d\theta \quad (7.19)$$

satisfies $d\mathbf{A}_N = \mathbf{F}$. Together the domains of \mathbf{A}_S and \mathbf{A}_N cover all of U . Thus, although \mathbf{F} has no globally defined potential on $\{\mathbb{R}^3 - (0, 0, 0)\}$, it does have two locally defined potentials whose domains exhaust the domain of \mathbf{F} .

Based on just what we have seen to this point, Dirac made a remarkable discovery. Roughly, his argument can be phrased in the following way. On the intersection of their domains the potentials \mathbf{A}_N and \mathbf{A}_S are related by

$$\mathbf{A}_N = \mathbf{A}_S + d(2g\theta). \quad (7.20)$$

Now consider an electric charge q in the field of the monopole. Each potential \mathbf{A}_N and \mathbf{A}_S gives rise (via the Schrödinger equation) to an expression for the wavefunction of q . Denote these ψ_N and ψ_S , respectively. It is a basic property of the Schrödinger equation that, because \mathbf{A}_N and \mathbf{A}_S differ by $d(2g\theta)$, ψ_N and ψ_S are related by

$$\psi_N = e^{2iqg\theta} \psi_S. \quad (7.21)$$

Now, the circle $\rho = 1$, $\phi = \frac{\pi}{2}$, is in the intersection of the domains of \mathbf{A}_N and \mathbf{A}_S so both ψ_N and ψ_S are defined (and single-valued) at points of this circle. Thus,

for such points, $\theta \rightarrow \theta + 2\pi$ must leave both ψ_N and ψ_S unchanged. But (7.21) then implies that

$$e^{2qg(\theta+2\pi)i} = e^{2qg\theta i}.$$

This, in turn, implies that $e^{4\pi qgi} = 1$ so $4\pi qg = 2n\pi$ for some $n \in \mathbb{Z}$, i.e.,

$$qg = \frac{1}{2}n \quad (7.22)$$

for some integer n . This is known as the **Dirac Quantization Condition** and it is interpreted to mean that if a magnetic monopole of charge g exists (or did at one time exist) somewhere in the universe, then all electric charges q are quantized in units of $\frac{1}{2g}$. Since no other plausible “explanation” for the quantization of charge has ever been put forward, this is often regarded as persuasive evidence that monopoles do (or did) exist. Of course, if we have a number of purely electric charges q_i and a number of purely magnetic charges g_j , then any pair of them will satisfy

$$q_i g_j = \frac{1}{2}n_{ij} \quad (7.23)$$

for some $n_{ij} \in \mathbb{Z}$.

If we are now prepared to accept the existence of magnetically charged particles, then one can define magnetic charge density g and magnetic current density \vec{j}_m just as in the electric case and introduce them into (7.13) and (7.14) to obtain

$$\vec{\nabla} \cdot (\vec{E} + i\vec{B}) = q + ig \quad (7.24)$$

$$\frac{\partial}{\partial x^0}(\vec{E} + i\vec{B}) + i\vec{\nabla} \times (\vec{E} + i\vec{B}) = \vec{j}_e + i\vec{j}_m. \quad (7.25)$$

For these equations the full $U(1)$ symmetry (7.10) is restored, provided we also rotate the electric and magnetic charges

$$q + ig \longrightarrow e^{i\phi}(q + ig) = (q \cos \phi - g \sin \phi) + i(q \sin \phi + g \cos \phi) \quad (7.26)$$

(just substitute (7.10) and $q' = q \cos \phi - g \sin \phi$, $g' = q \sin \phi + g \cos \phi$ into (7.24) and (7.25)).

Now notice that, in the special case $\phi = -\frac{\pi}{2}$, corresponding to $\vec{E} \rightarrow \vec{B}$ and $\vec{B} \rightarrow -\vec{E}$, the charges $q \rightarrow g$ and $g \rightarrow -q$ are interchanged. But these charges measure the strengths of the interactions, i.e., they play the role of coupling constants and, according to the Dirac Quantization Condition (7.22), they are inversely proportional. In particular, if one is “large”, then the other is “small” so that *our symmetry interchanges strong and weak coupling*.

We conclude our discussion of the Dirac monopole with one last observation. If we measure electric charge in multiples of the charge of the electron we may take $q = 1$ in (7.22) so that, for any magnetic monopole, $2g = n$ is an integer. Rewriting the potential (7.19) as

$$\mathbf{A}_N = \frac{1}{2}n(1 - \cos \phi) d\theta$$

and regarding this now as a form on the two-sphere $\rho = 1$ we obtain, except for the Lie algebra factor $-i$, the gauge potential \mathcal{A}_n for the connection on $U(1) \hookrightarrow P_n \rightarrow S^2$ described in the Remark following (1.24).

The duality symmetry of Donaldson-Witten theory is, of course, much more subtle and complex than the classical electromagnetic duality we have described and its consequences are much more profound. It has led, in particular, to a dual version of the Donaldson invariants, known as the Seiberg-Witten invariants, and these have precipitated a revolution in the differential topology of four-manifolds. Fundamentally, these are defined in a manner quite analogous to the Donaldson invariants from moduli spaces of classical fields, but, whereas in Donaldson theory these are pure $SU(2)$ Yang-Mills fields, the dual version has a spinor field coupled to a $U(1)$ connection. The relative simplicity of Seiberg-Witten over Donaldson theory resides in the shift from nonabelian ($SU(2)$) to Abelian ($U(1)$) and the fact that, quite unlike Donaldson theory, the moduli spaces of Seiberg-Witten theory are “usually” finite and always compact. The price one must pay for the eventual simplicity of the theory, however, is an initial expenditure of time and energy to assemble the rather substantial algebraic machinery required to write down the relevant equations. Our final objective is to lay this algebraic foundation and then describe, as we did for Donaldson theory in Section 4, the construction of the zero-dimensional Seiberg-Witten invariant.

Much of the algebraic background we require is most conveniently phrased in the language of Clifford algebras. We recall that any finite dimensional, real vector space V with an inner product $\langle \cdot, \cdot \rangle$ has a Clifford algebra $Cl(V)$ which can be described abstractly as the quotient of the tensor algebra $\mathcal{J}(V)$ by the two-sided ideal $\mathcal{I}(V)$ generated by elements of the form $v \otimes v + \langle v, v \rangle 1$ with $v \in V$. More concretely, if $\{e_1, \dots, e_n\}$ is an orthonormal basis for V , then $Cl(V)$ is the real associative algebra with unit 1 generated by $\{e_1, \dots, e_n\}$ and subject to the relations

$$e_i e_j + e_j e_i = -2\langle e_i, e_j \rangle 1, \quad i, j = 1, \dots, n. \quad (7.27)$$

We intend to be even more concrete and construct an explicit matrix model for the Clifford algebra $Cl(4) = Cl(\mathbb{R}^4)$ of \mathbb{R}^4 with its usual positive definite inner

product. The procedure will be to identify \mathbb{R}^4 with a real linear subspace of a matrix algebra, find an orthonormal basis for this copy of \mathbb{R}^4 satisfying the defining conditions (7.27), where the product is matrix multiplication and 1 is the identity matrix, and form the subalgebra it generates.

One can, of course, identify \mathbb{R}^4 with the algebra \mathbb{H} of quaternions $q = q^1 + q^2i + q^3j + q^4k$, but we wish to embed this into the real, associative algebra $\mathbb{H}^{2 \times 2}$ of 2×2 quaternionic matrices:

$$\mathbb{H}^{2 \times 2} = \left\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}; q_{ij} \in \mathbb{H}, \quad i, j = 1, 2 \right\}.$$

Specifically, we identify \mathbb{R}^4 with the real linear subspace of $\mathbb{H}^{2 \times 2}$ consisting of all elements of the form

$$x = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix}, \quad q \in \mathbb{H} \quad (7.28)$$

(this is, of course, *not* a subalgebra of $\mathbb{H}^{2 \times 2}$). Notice that $\det x = \|q\|^2$ so, defining a norm on the set of x given by (7.28) by

$$\|x\|^2 = \det x \quad (7.29)$$

and an inner product by polarization ($\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$) we find that the subspace of $\mathbb{H}^{2 \times 2}$ consisting of all x of the form (7.28) is isomorphic to \mathbb{R}^4 as an inner product space. One easily checks that $\{e_1, e_2, e_3, e_4\}$ given by

$$e_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix}, \quad e_4 = \begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix} \quad (7.30)$$

is an orthonormal basis and, moreover, satisfies

$$e_i e_j + e_j e_i = -2\langle e_i, e_j \rangle \mathbb{1}, \quad i, j = 1, 2, 3, 4 \quad (7.31)$$

where we use $\mathbb{1}$ generically for the identity matrix of any size (2×2 in this case). Note that it follows from (7.31) that

$$xy + yx = -2\langle x, y \rangle, \quad x, y \in \mathbb{R}^4. \quad (7.32)$$

The real subalgebra of $\mathbb{H}^{2 \times 2}$ generated by $\{e_1, e_2, e_3, e_4\}$ is the **real Clifford algebra** of \mathbb{R}^4 and is denoted $Cl(4)$. Writing out products of basis vectors and

using (7.31) to eliminate linear dependencies gives the following basis for $Cl(4)$:

$$\begin{aligned}
e_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1} \\
e_1 &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & e_2 &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} & e_3 &= \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix} & e_4 &= \begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix} \\
e_1e_2 &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} & e_1e_3 &= \begin{pmatrix} j & 0 \\ 0 & -j \end{pmatrix} & e_1e_4 &= \begin{pmatrix} k & 0 \\ 0 & -k \end{pmatrix} \\
e_2e_3 &= \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} & e_2e_4 &= \begin{pmatrix} -j & 0 \\ 0 & -j \end{pmatrix} & e_3e_4 &= \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix} & (7.33) \\
e_1e_2e_3 &= \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix} & e_1e_2e_4 &= \begin{pmatrix} 0 & -j \\ j & 0 \end{pmatrix} \\
e_1e_3e_4 &= \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} & e_2e_3e_4 &= \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \\
e_1e_2e_3e_4 &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

Thus, $\dim Cl(4) = 16$. Since $\mathbb{H}^{2 \times 2}$ itself has real dimension 16 we conclude that, in fact,

$$Cl(4) = \mathbb{H}^{2 \times 2}. \quad (7.34)$$

Notice that the basis (7.33) gives $Cl(4)$ a natural \mathbb{Z}_2 -grading

$$Cl(4) = Cl_0(4) \oplus Cl_1(4) \quad (7.35)$$

where $Cl_0(4)$ is spanned by $e_0, e_1e_2, e_1e_3, e_1e_4, e_2e_3, e_2e_4, e_3e_4$ and $e_1e_2e_3e_4$ and $Cl_1(4)$ is spanned by $e_1, e_2, e_3, e_4, e_1e_2e_3, e_1e_2e_4, e_1e_3e_4$ and $e_2e_3e_4$. The elements of $Cl_0(4)$ are said to be **even**, while those of $Cl_1(4)$ are **odd**. Regarding \mathbb{Z}_2 as $\{0, 1\}$ with addition modulo 2,

$$(Cl_i(4))(Cl_j(4)) \subseteq Cl_{i+j}(4) \quad (7.36)$$

for $i, j = 0, 1$, so $Cl(4)$ is a \mathbb{Z}_2 -graded algebra, i.e., a superalgebra. From (7.33) it is clear that the decomposition (7.35) corresponds simply to

$$\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} q_{11} & 0 \\ 0 & q_{22} \end{pmatrix} + \begin{pmatrix} 0 & q_{12} \\ q_{21} & 0 \end{pmatrix}.$$

Lemma 7.1. *The center $Z(Cl(4))$ of $Cl(4)$ is $\text{Span}\{e_0\} \cong \mathbb{R}$.*

Proof: Since $e_0 = \mathbb{1}$ it commutes with everything in $Cl(4)$, $\text{Span}\{e_0\} \subseteq Z(Cl(4))$ is clear. To complete the proof it will suffice to show that every

$$e_I = e_{i_1} \dots e_{i_k}, \quad 1 \leq k \leq 4, \quad 1 \leq i_1 < \dots < i_k \leq 4$$

fails to commute with something in $Cl(4)$. For $k = 1$ this is clear since $e_i e_j = -e_j e_i$ for $i \neq j$. For $k = 4$, $e_I = e_1 e_2 e_3 e_4$ so $e_1 e_I = (e_1 e_1) e_2 e_3 e_4 = -e_2 e_3 e_4$, whereas $e_I e_1 = (e_1 e_2 e_3 e_4) e_1 = (-1)^3 (e_1 e_1) e_2 e_3 e_4 = e_2 e_3 e_4$. Now suppose $1 < k < 4$. Then $e_I e_{i_1} = (-1)^{k-1} e_{i_1} e_I$ and, if e_l is not among e_{i_1}, \dots, e_{i_k} , $e_I e_l = (-1)^k e_l e_I$. Thus, e_I cannot commute with both e_{i_1} and e_l . ■

Lemma 7.2. *If $x \in \mathbb{R}^4 \subseteq Cl(4)$ and $\|x\| = 1$, then x is a unit in $Cl(4)$ (i.e., is invertible) and $x^{-1} = -x$.*

Proof: $\langle x, x \rangle = 1$ and $xx + xx = -2\langle x, x \rangle \mathbb{1}$ imply $xx = -\mathbb{1}$. ■

We denote by $Cl^\times(4)$ the multiplicative group of units in $Cl(4)$ and by $\text{Pin}(4)$ the subgroup of $Cl^\times(4)$ generated by all of the $x \in \mathbb{R}^4$ with $\|x\| = 1$ (see Lemma 7.2). Now, an x of the form (7.28) has $\|x\| = 1$ if and only if $q \in \text{Sp}(1)$ (the Lie group of unit quaternions) and the set of all such is closed under inversion ($x^{-1} = -x$). Thus, $\text{Pin}(4)$ is just the set of all products of such elements. The even elements of $\text{Pin}(4)$ are just its diagonal elements and they form a subgroup denoted

$$\begin{aligned} \mathfrak{Spin}(4) &= \text{Pin}(4) \cap Cl_0(4) \\ &= \left\{ \begin{pmatrix} u_1 & 0 \\ 0 & u_2 \end{pmatrix}; u_1, u_2 \in \text{Sp}(1) \right\} \cong \text{Sp}(1) \times \text{Sp}(1). \end{aligned} \quad (7.37)$$

The topology and differentiable structure $\mathfrak{Spin}(4)$ inherits from $\mathbb{H}^{2 \times 2} \cong \mathbb{H}^4 \cong \mathbb{R}^{16}$ are the product structures from $\text{Sp}(1)$ (which is diffeomorphic to S^3) so $\mathfrak{Spin}(4)$ is a compact, simply connected Lie group. Since the Lie algebra of $\text{Sp}(1)$ can be identified with the pure imaginary quaternions $\text{Im}\mathbb{H}$, the Lie algebra of $\mathfrak{Spin}(4)$ can be identified with

$$\mathfrak{spin}(4) = \left\{ \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}; Q_1, Q_2 \in \text{Im}\mathbb{H} \right\}. \quad (7.38)$$

The significance of $\mathfrak{Spin}(4)$ lies in the following theorem.

Theorem 7.3. *$\mathfrak{Spin}(4)$ is the universal double cover of $\text{SO}(4)$.*

Proof: Since $\mathfrak{Spin}(4)$ is a simply connected Lie group we need only show that it is a double cover of $SO(4)$. For this we consider the adjoint action of $Cl^\times(4)$ on $Cl(4)$, i.e., for each $u \in Cl^\times(4)$ we define a map

$$\text{ad}_u : Cl(4) \longrightarrow Cl(4)$$

by

$$\text{ad}_u(p) = upu^{-1} \tag{7.39}$$

for each $p \in Cl(4)$. This is clearly an algebra isomorphism that preserves the grading (7.35). Note that if $x \in \mathbb{R}^4 \subseteq Cl(4)$ has $\|x\| = 1$, then, for every $v \in \mathbb{R}^4 \subseteq Cl(4)$

$$\text{ad}_x(v) = xvx^{-1} = xv(-x) = -xvx$$

so the identity $vx + xv = -2\langle v, x \rangle \mathbb{1}$ implies

$$\begin{aligned} xvx + xxv &= -2\langle v, x \rangle x \\ xvx - v &= -2\langle v, x \rangle x \\ xvx &= v - 2\langle v, x \rangle x \\ xvx &= (v - \langle v, x \rangle x) - \langle v, x \rangle x. \end{aligned}$$

Now, $v - \langle v, x \rangle x$ is the projection of v into the hyperplane x^\perp orthogonal to x so xvx is the reflection of v through x^\perp , written $\text{Refl}_{x^\perp}(v)$. Thus,

$$\text{ad}_x(v) = -\text{Refl}_{x^\perp}(v). \tag{7.40}$$

In particular,

$$\text{ad}_x : \mathbb{R}^4 \longrightarrow \mathbb{R}^4, \quad x \in \mathbb{R}^4, \|x\| = 1.$$

But any element of $\text{Pin}(4)$ is a product of elements $x \in \mathbb{R}^4$ with $\|x\| = 1$ so

$$\text{ad}_u : \mathbb{R}^4 \longrightarrow \mathbb{R}^4, \quad u \in \text{Pin}(4). \tag{7.41}$$

Since any product of reflections is an orthogonal transformation, ad_u is an orthogonal transformation for each $u \in \text{Pin}(4)$. Since an element of $\mathfrak{Spin}(4)$ is a product of an even number of $x \in \mathbb{R}^4$ with $\|x\| = 1$ and any product of an even number of reflections is a rotation we have

$$\text{ad}_u \in SO(\mathbb{R}^4), \quad u \in \mathfrak{Spin}(4). \tag{7.42}$$

Thus, we have a map, called the **spinor map**,

$$\begin{aligned} \mathfrak{S} : \mathfrak{Spin}(4) &\longrightarrow SO(\mathbb{R}^4) \cong SO(4) \\ \mathfrak{S}(u) &= \text{ad}_u. \end{aligned} \tag{7.43}$$

Since any reflection can clearly be written as $-\text{ad}_x$ for some $x \in \mathbb{R}^4$ with $\|x\| = 1$ and since any rotation can be written as a product of an even number of reflections, the spinor map is a surjective group homomorphism. Finally, to see that $\ker(\mathfrak{S}) = \mathbb{Z}_2 = \{\pm \mathbb{1}\}$, so that it is precisely two-to-one, note that ad_u is the identity in $\text{SO}(4)$ if and only if $uxu^{-1} = x$ for each $x \in \mathbb{R}^4$. But then u must commute with everything in $Cl(4)$, i.e., $u \in Z(Cl(4))$. By Lemma 7.1, $u = ae_0 = a\mathbb{1}$ for some $a \in \mathbb{R}$. By (7.37), $u \in \text{Sp}(1)$ so $a^2 = 1$ and $u = \pm \mathbb{1}$. ■

Remark. *Globalizing these constructions leads to the notion of a “spin structure” on a manifold. We will briefly recall how this is done and then explain why we need the more general concept of a “spin^c structure”. We let B denote a compact, oriented, smooth four-manifold with a Riemannian metric \mathbf{g} . Let $\text{SO}(4) \hookrightarrow F_{\text{SO}}(B) \xrightarrow{\pi_{\text{SO}}} B$ denote the corresponding oriented, orthonormal frame bundle. A **spin structure** \mathcal{S} consists of a principal $\mathfrak{Spin}(4)$ -bundle*

$$\mathfrak{Spin}(4) \hookrightarrow S(B) \xrightarrow{\pi_S} B$$

over B and a smooth map

$$\lambda : S(B) \longrightarrow F_{\text{SO}}(B)$$

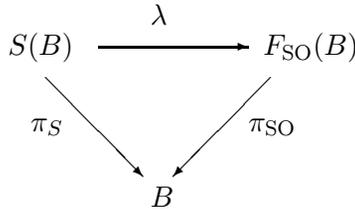
satisfying

$$\pi_{\text{SO}} \circ \lambda = \pi_S \tag{7.44}$$

and

$$\lambda(p \cdot u) = \lambda(p) \cdot \mathfrak{S}(u) \tag{7.45}$$

for each $p \in S(B)$ and each $u \in \mathfrak{Spin}(4)$.



The fibers of $F_{\text{SO}}(B)$ are copies of $\text{SO}(4)$ so (7.44) says that we have a copy of $\mathfrak{Spin}(4)$ “above” each of these and (7.45) says that the map λ of $S(B)$ onto $F_{\text{SO}}(B)$ is essentially the spinor map at each point of B . Now, unlike the frame bundle $F_{\text{SO}}(B)$, which exists for any manifold of the type we have described, there is an obstruction to the existence to a spin structure. This is most easily seen by rephrasing the definition in terms of transition functions. Let $\{U_\alpha\}$ be

any trivializing cover for $SO(4) \hookrightarrow F_{SO}(B) \xrightarrow{\pi_{SO}} B$ with transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow SO(4)$. Because $\mathfrak{Spin}(4)$ is the universal cover of $SO(4)$ each of the maps $g_{\alpha\beta}$ lifts to $\tilde{g}_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow \mathfrak{Spin}(4)$.

$$\begin{array}{ccc}
 & & \mathfrak{Spin}(4) \\
 & \nearrow \tilde{g}_{\alpha\beta} & \downarrow \mathfrak{Spin} \\
 U_\alpha \cap U_\beta & \xrightarrow{g_{\alpha\beta}} & SO(4)
 \end{array}$$

If these lifts happen to satisfy the cocycle condition ($\tilde{g}_{\alpha\beta}\tilde{g}_{\beta\gamma} = \tilde{g}_{\alpha\gamma}$ on $U_\alpha \cap U_\beta \cap U_\gamma$ whenever this is nonempty), then they determine a $\mathfrak{Spin}(4)$ -bundle over B which is easily seen to give rise to a spin structure for B . It may well happen, however, that no family of transition functions for a trivializing cover of the frame bundle has lifts that satisfy the cocycle condition and in this case no spin structure exists. In terms of the Čech cohomology of B with coefficients in \mathbb{Z}_2 it is easy to isolate a class $w_2(B) \in H^2(B, \mathbb{Z}_2)$, called the second Stiefel-Whitney class of B , whose vanishing is equivalent to the existence of the required lifts. Thus, B admits a spin structure if and only if its second Stiefel-Whitney class is trivial. Unfortunately, many interesting four-manifolds (e.g., $\mathbb{C}\mathbb{P}^2$) do not satisfy this condition and without a spin structure one cannot define “spinor fields” in the usual sense. Since spinor fields are crucial to Seiberg-Witten theory and since one would like this theory to apply to as many four-manifolds as possible we seek a generalized notion of both “spin structure” and “spinor field”. As it happens, there is a very natural generalization obtained by complexifying our previous algebraic considerations.

To define complex analogues of the algebraic objects we have introduced we will embed $Cl(4)$ into a complex algebra of matrices and form the complex subalgebra it generates. The basic tool we use is the usual matrix model of the quaternions. Specifically, we consider the map $\gamma : \mathbb{H} \rightarrow \mathbb{C}^{2 \times 2}$ from the quaternions to the 2×2 complex matrices given by

$$\gamma(q) = \gamma(\alpha + \beta j) = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \quad (7.46)$$

where we have written

$$q = q^1 + q^2 i + q^3 j + q^4 k = (q^1 + q^2 i) + (q^3 + q^4 i)j = \alpha + \beta j.$$

One easily verifies that γ is real linear, injective, preserves products, carries \bar{q} to $\overline{\gamma(q)}^\top$ and satisfies $\det(\gamma(q)) = \|q\|^2$ so that we can identify \mathbb{H} with the set all 2×2 complex matrices of the form $\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$. More specifically, if we let

$$\begin{aligned} \gamma(1) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{1} & \gamma(i) &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = I \\ \gamma(j) &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = J & \gamma(k) &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = K \end{aligned} \quad (7.47)$$

then we can identify $q = q^1 + q^2i + q^3j + q^4k$ with

$$q = q^1\mathbb{1} + q^2I + q^3J + q^4K. \quad (7.48)$$

Now we identify $Cl(4) = \mathbb{H}^{2 \times 2}$ with a subset of $\mathbb{C}^{4 \times 4}$. Define $\Gamma : \mathbb{H}^{2 \times 2} \longrightarrow \mathbb{C}^{4 \times 4}$ by

$$\Gamma \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} \gamma(q_{11}) & \gamma(q_{12}) \\ \gamma(q_{21}) & \gamma(q_{22}) \end{pmatrix} \quad (7.49)$$

where each $\gamma(q_{ij})$ is a 2×2 block in the matrix on the right-hand side. This map Γ is also real linear, injective and preserves products so we can identify the real algebra $Cl(4)$ with its image

$$Cl(4) = \Gamma(\mathbb{H}^{2 \times 2}).$$

The restriction of Γ to $\mathbb{R}^4 \subseteq Cl(4)$ is

$$x = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix} \longrightarrow \Gamma(x) = \begin{pmatrix} 0 & \gamma(q) \\ -\overline{\gamma(q)}^\top & 0 \end{pmatrix}. \quad (7.50)$$

Since $\det \Gamma(x) = \det x = \|x\|^2 = \|q\|^2$ we can define an inner product via polarization on this copy of \mathbb{R}^4 from $\|\Gamma(x)\|^2 = \det \Gamma(x)$ and then $\Gamma|_{\mathbb{R}^4}$ becomes an isometry. We now fully identify \mathbb{R}^4 with this copy and obtain the basis

$$\begin{aligned} E_1 &= \gamma(e_1) = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} & E_2 &= \gamma(e_2) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \\ E_3 &= \gamma(e_3) = \begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix} & E_4 &= \gamma(e_4) = \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \end{aligned} \quad (7.51)$$

satisfying

$$E_i E_j + E_j E_i = -2\langle E_i, E_j \rangle \mathbb{1}, \quad i, j = 1, 2, 3, 4. \quad (7.52)$$

In this context, $Cl(4)$ is the *real* subalgebra of $\mathbb{C}^{4 \times 4}$ generated by $\{E_1, E_2, E_3, E_4\}$ and a basis is as in (7.33), but with everything capitalized (and 1 changed to $\mathbb{1}$). Under γ , $Sp(1)$ is mapped to $SU(2)$ so, in our new model of $Cl(4)$ we have the identifications

$$\mathfrak{Spin}(4) = \left\{ \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}; U_1, U_2 \in SU(2) \right\} \cong SU(2) \times SU(2) \quad (7.53)$$

and

$$\mathfrak{spin}(4) = \left\{ \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}; A_1, A_2 \in \mathfrak{su}(2) \right\} \quad (7.54)$$

corresponding to (7.37) and (7.38).

Now we regard $\mathbb{C}^{4 \times 4}$ as a complex algebra and define the **complexified Clifford algebra** $Cl(4) \otimes \mathbb{C}$ to be the complex subalgebra generated by $\{E_1, E_2, E_3, E_4\}$, i.e., by $Cl(4)$. A basis over \mathbb{C} is given by (7.33), with all of the e_i capitalized. Since $\mathbb{C}^{4 \times 4}$ also has dimension 16 over \mathbb{C} we conclude that

$$Cl(4) \otimes \mathbb{C} = \mathbb{C}^{4 \times 4}. \quad (7.55)$$

Now let

$$S_{\mathbb{C}} = \mathbb{C}^4$$

be the complex vector space \mathbb{C}^4 with its usual Hermitian inner product ($\langle z, w \rangle = \bar{z}^1 w^1 + \bar{z}^2 w^2 + \bar{z}^3 w^3 + \bar{z}^4 w^4$) and identify

$$Cl(4) \otimes \mathbb{C} = \text{End}_{\mathbb{C}}(S_{\mathbb{C}}). \quad (7.56)$$

Thus, the elements of $Cl(4) \otimes \mathbb{C}$ (and therefore also $Cl(4)$, \mathbb{R}^4 and $\mathfrak{Spin}(4)$) act as endomorphisms of $S_{\mathbb{C}}$. This action is called **Clifford multiplication** and will be written with a dot \cdot . In particular, we have a representation of the real Clifford algebra by endomorphisms of $S_{\mathbb{C}}$

$$Cl(4) \longrightarrow \text{End}_{\mathbb{C}}(S_{\mathbb{C}})$$

(representations of algebras are by endomorphisms rather than isomorphisms since not all elements of an algebra are units). This representation of $Cl(4)$ is easily seen to be irreducible by writing out the real linear combinations of the basis $E_0, \dots, E_1 E_2 E_3 E_4$ for $Cl(4) \subseteq Cl(4) \otimes \mathbb{C}$. Restricting the Clifford action further to $\mathfrak{Spin}(4) \subseteq Cl(4) \otimes \mathbb{C}$ gives a *group* representation of $\mathfrak{Spin}(4)$ on $S_{\mathbb{C}}$

$$\Delta_{\mathbb{C}} : \mathfrak{Spin}(4) \longrightarrow \text{Aut}_{\mathbb{C}}(S_{\mathbb{C}})$$

(by automorphisms now since the elements of $\mathfrak{Spin}(4)$ are all units). This is called the **complex spin representation** and, as we shall now see, is *not* irreducible. Indeed, if we write

$$S_{\mathbb{C}} \cong S_{\mathbb{C}}^{+} \oplus S_{\mathbb{C}}^{-}$$

$$\begin{pmatrix} z^1 \\ z^2 \\ z^3 \\ z^4 \end{pmatrix} = \begin{pmatrix} z^1 \\ z^2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ z^3 \\ z^4 \end{pmatrix} \quad (7.57)$$

then Clifford multiplication by elements of $Cl_0(4)$, because they are block diagonal, preserves $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$, whereas Clifford multiplication by elements of $Cl_1(4)$, because they are block anti-diagonal, interchanges $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$. In particular, $\Delta_{\mathbb{C}}$ resolves into a direct sum

$$\Delta_{\mathbb{C}} = \Delta_{\mathbb{C}}^{+} \oplus \Delta_{\mathbb{C}}^{-}$$

where

$$\Delta_{\mathbb{C}}^{\pm} : \mathfrak{Spin}(4) \longrightarrow \mathrm{SU}(S_{\mathbb{C}}^{\pm})$$

(see (7.53) for the “SU”). $\Delta_{\mathbb{C}}^{+}$ and $\Delta_{\mathbb{C}}^{-}$ are inequivalent, irreducible representations of $\mathfrak{Spin}(4)$. Notice also that Clifford multiplication by the elements of \mathbb{R}^4 , which are odd, interchanges $S_{\mathbb{C}}^{+}$ and $S_{\mathbb{C}}^{-}$ (this will be crucial when we define the “Dirac operator” shortly).

Recall that $\mathfrak{Spin}(4)$ is the set of all even elements in the subgroup of multiplicative units in the Clifford algebra $Cl(4)$ generated by the unit sphere in $\mathbb{R}^4 \subseteq Cl(4)$. For the complex analogue we add to the generators the unit circle in \mathbb{C} . More precisely, we identify $U(1)$ with the subset

$$U(1) = \{e^{\theta i} \mathbb{1}; \theta \in \mathbb{R}\}$$

of $Cl(4) \otimes \mathbb{C}$ (often dropping the “ $\mathbb{1}$ ” sign and thinking of $e^{\theta i}$ as an element of $Cl(4) \otimes \mathbb{C}$). Then

$$\mathfrak{Spin}^c(4)$$

is defined to be the subgroup of the group of multiplicative units in $Cl(4) \otimes \mathbb{C}$ generated by $\mathfrak{Spin}(4)$ and $U(1)$. Notice that the elements of $\mathfrak{Spin}^c(4)$ are necessarily even, i.e., in $Cl_0(4) \otimes \mathbb{C}$. Since $U(1)$ is in the center of $Cl(4) \otimes \mathbb{C}$ we have

$$\begin{aligned} \mathfrak{Spin}^c(4) &= \{e^{\theta i} u; \theta \in \mathbb{R}, \quad u \in \mathfrak{Spin}(4)\} \\ &= \left\{ e^{\theta i} \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}; \theta \in \mathbb{R}, \quad U_1, U_2 \in \mathrm{SU}(2) \right\}. \end{aligned} \quad (7.58)$$

Note that $U_1, U_2 \in SU(2)$ implies $e^{\theta i}U_1, e^{\theta i}U_2 \in U(2)$ and

$$\det(e^{\theta i}U_1) = \det(e^{\theta i}U_2) = e^{2\theta i}. \quad (7.59)$$

Since every element of $U(2)$ can be written as $e^{\theta i}U$, $U \in SU(2)$ (uniquely up to a simultaneous change of sign for both $e^{\theta i}$ and U) we have

$$\mathfrak{Spin}^c(4) = \left\{ \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix}; U_{\pm} \in U(2), \det U_+ = \det U_- \right\}. \quad (7.60)$$

There is yet another useful way of looking at $\mathfrak{Spin}^c(4)$. The mapping

$$\begin{aligned} \mathfrak{Spin}(4) \times U(1) &\longrightarrow \mathfrak{Spin}^c(4) \\ (u, e^{\theta i}\mathbb{1}) &\longrightarrow e^{\theta i}u \end{aligned}$$

is a surjective homomorphism. Its kernel is the set of (α, α^{-1}) , where $\alpha \in \mathfrak{Spin}(4)$. But $\mathfrak{Spin}(4)$ intersects the scalars only in $\pm\mathbb{1}$ so this kernel is $\mathbb{Z}_2 = \pm(\mathbb{1}, \mathbb{1})$. Thus,

$$\mathfrak{Spin}^c(4) \cong \mathfrak{Spin}(4) \times U(1)/\mathbb{Z}_2. \quad (7.61)$$

Finally, notice that, from Lemma 7.1 and (7.58) it follows that the center of $\mathfrak{Spin}^c(4)$ is

$$Z(\mathfrak{Spin}^c(4)) = U(1). \quad (7.62)$$

Globalizing all of this to four-manifolds will require a few mappings which we now introduce. First define

$$\delta : \mathfrak{Spin}^c(4) \longrightarrow U(1)$$

as follows. For

$$\begin{aligned} \xi &= \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix} = \begin{pmatrix} e^{\theta i}U_1 & 0 \\ 0 & e^{\theta i}U_2 \end{pmatrix} \in \mathfrak{Spin}^c(4) \\ \delta(\xi) &= \det U_+ = \det U_- = e^{2\theta i}. \end{aligned} \quad (7.63)$$

Then δ is a surjective homomorphism with kernel $\mathfrak{Spin}(4)$. Next define

$$\pi : \mathfrak{Spin}^c(4) \longrightarrow SO(4)$$

as follows. The adjoint action of $\mathfrak{Spin}(4)$ on \mathbb{R}^4 extends to an adjoint action of $\mathfrak{Spin}^c(4)$ on \mathbb{R}^4 . Indeed, if $\xi = e^{\theta i}u \in \mathfrak{Spin}^c(4)$, then, for each $x \in \mathbb{R}^4$, $\text{ad}_{\xi}(x) = \xi x \xi^{-1} = u x u^{-1} = \text{ad}_u(x)$ so, on \mathbb{R}^4

$$\text{ad}_{\xi} = \text{ad}_u = \mathfrak{S}(u) \in SO(\mathbb{R}^4) \cong SO(4)$$

and we may take

$$\pi(\xi) = \text{ad}_\xi = \text{ad}_u = \mathfrak{S}(u). \quad (7.64)$$

Finally, define

$$\text{Spin}^c : \mathfrak{Spin}^c(4) \longrightarrow \text{SO}(4) \times \text{U}(1)$$

by

$$\mathfrak{S}^c(\xi) = \mathfrak{S}^c(e^{\theta i} u) = (\pi(\xi), \delta(\xi)) = (\mathfrak{S}(u), e^{2\theta i}). \quad (7.65)$$

Then \mathfrak{S}^c is a surjective homomorphism whose kernel is easily seen to be $\mathbb{Z}_2 = \pm \mathbb{1}$. It follows that $\mathfrak{Spin}^c(4)$ is a double cover of $\text{SO}(4) \times \text{U}(1)$ (it is not a universal cover, however, since one can show that $\pi_1(\mathfrak{Spin}^c(4)) \cong \mathbb{Z}$ so $\mathfrak{Spin}^c(4)$ is not simply connected). Thus, the Lie algebra $\mathfrak{spin}^c(4)$ is $\mathfrak{so}(4) \times \mathfrak{u}(1) \cong \mathfrak{spin}(4) \times \mathfrak{u}(1)$ and can be identified with the subset

$$\mathfrak{spin}^c(4) = \left\{ \left(\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} + ti \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} ; t \in \mathbb{R}, A_1, A_2 \in \mathfrak{su}(2) \right) \right\} \quad (7.66)$$

of $Cl(4) \otimes \mathbb{C}$ (in particular, $\mathfrak{spin}^c(4)$ also acts by Clifford multiplication on $S_{\mathbb{C}}$, preserving both $S_{\mathbb{C}}^{\pm}$).

Now, the identification $Cl(4) \otimes \mathbb{C} = \text{End}_{\mathbb{C}}(S_{\mathbb{C}})$ and the fact that the elements of $\mathfrak{Spin}^c(4)$ are all units implies that the complex spin representation $\Delta_{\mathbb{C}} : \mathfrak{Spin}(4) \rightarrow \text{Aut}_{\mathbb{C}}(S_{\mathbb{C}})$ extends to a representation

$$\hat{\Delta}_{\mathbb{C}} : \mathfrak{Spin}^c(4) \longrightarrow \text{Aut}_{\mathbb{C}}(S_{\mathbb{C}}).$$

Since the elements of $\mathfrak{Spin}^c(4)$ are block diagonal, $\hat{\Delta}_{\mathbb{C}}$ also splits into

$$\hat{\Delta}_{\mathbb{C}} = \hat{\Delta}_{\mathbb{C}}^+ \oplus \hat{\Delta}_{\mathbb{C}}^-$$

where

$$\hat{\Delta}_{\mathbb{C}}^{\pm} : \mathfrak{Spin}^c(4) \longrightarrow U(S_{\mathbb{C}}^{\pm}) \quad (7.67)$$

(see (7.60)).

One of the Seiberg-Witten equations relates the self-dual part of the curvature of a $\text{U}(1)$ -connection to a certain trace free endomorphism of a positive spinor bundle. The last of our algebraic preliminaries describes the relationship between two-forms and endomorphisms. We note that there is a natural linear isomorphism from the space $\Omega^2(\mathbb{R}^4)$ of (complex-valued) two-forms on \mathbb{R}^4 into $Cl(4) \otimes \mathbb{C}$. Indeed, if $\{e_1, e_2, e_3, e_4\}$ is the standard basis for \mathbb{R}^4 and $\{e^1, e^2, e^3, e^4\}$ is its dual, then we define

$$\rho : \Omega^2(\mathbb{R}^4) \longrightarrow Cl_0(4) \otimes \mathbb{C}$$

Proof: We give the argument for $\rho^+|\Omega_+^2(\mathbb{R}^4)$. The $\rho^-|\Omega_-^2(\mathbb{R}^4)$ case is analogous. A simple computation from (7.68) gives

$$\begin{aligned}\rho(e^1 \wedge e^2 + e^3 \wedge e^4) &= 2 \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \\ \rho(e^1 \wedge e^3 + e^4 \wedge e^2) &= 2 \begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix} \\ \rho(e^1 \wedge e^4 + e^2 \wedge e^3) &= 2 \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}\tag{7.72}$$

Since $\{e^1 \wedge e^2 + e^3 \wedge e^4, e^1 \wedge e^3 + e^4 \wedge e^2, e^1 \wedge e^4 + e^2 \wedge e^3\}$ spans the set of self-dual two-forms on \mathbb{R}^4 , it is clear that $\rho^+|\Omega_+^2(\mathbb{R}^4)$ is a linear, injective map to $\text{End}(S_{\mathbb{C}}^+)$. Because I, J and K are trace free, so is everything in the image of $\rho^+|\Omega_+^2(\mathbb{R}^4)$. Furthermore, one can show that every 2×2 complex, trace free matrix is a complex linear combination of I, J and K so $\rho^+|\Omega_+^2(\mathbb{R}^4)$ maps onto $\text{End}_0(S_{\mathbb{C}}^+)$. ■

It follows, in particular, from Lemma 7.4 that the map $\rho^+|\Omega_+^2(\mathbb{R}^4) : \Omega_+^2(\mathbb{R}^4) \rightarrow \text{End}_0(S_{\mathbb{C}}^+)$ has an inverse that we will simply denote

$$\sigma^+ : \text{End}_0(S_{\mathbb{C}}^+) \longrightarrow \Omega_+^2(\mathbb{R}^4).\tag{7.73}$$

One can compute this inverse explicitly, but we will content ourselves with describing its action on the particular type of trace free endomorphism that arises in the Seiberg-Witten equations. For this we consider an element

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}$$

of $S_{\mathbb{C}}^+$ (temporarily suppress the two zero components in $S_{\mathbb{C}}^+$). Define an endomorphism of $S_{\mathbb{C}}^+$ by the matrix

$$\psi \otimes \psi^* = \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} (\bar{\psi}^1 \bar{\psi}^2) = \begin{pmatrix} |\psi^1|^2 & \psi^1 \bar{\psi}^2 \\ \bar{\psi}^1 \psi^2 & |\psi^2|^2 \end{pmatrix}.\tag{7.74}$$

The trace free part of this endomorphism is

$$\begin{aligned}(\psi \otimes \psi^*)_0 &= \psi \otimes \psi^* - \frac{1}{2} \text{tr}(\psi \otimes \psi^*) \mathbf{1} \\ &= \begin{pmatrix} \frac{1}{2}(|\psi^1|^2 - |\psi^2|^2) & \psi^1 \bar{\psi}^2 \\ \bar{\psi}^1 \psi^2 & \frac{1}{2}(|\psi^2|^2 - |\psi^1|^2) \end{pmatrix}.\end{aligned}\tag{7.75}$$

One can verify directly that the image of $(\psi \otimes \psi^*)_0 \in \text{End}_0(S_{\mathbb{C}}^{\pm})$ under σ^+ is

$$\begin{aligned}
\sigma^+((\psi \otimes \psi^*)_0) &= -\frac{1}{4}i[(|\psi^1|^2 - |\psi^2|^2)(e^1 \wedge e^2 + e^3 \wedge e^4) \\
&\quad - 2\text{Im}(\psi^1 \bar{\psi}^2)(e^1 \wedge e^3 + e^4 \wedge e^2) \\
&\quad - 2\text{Re}(\psi^1 \bar{\psi}^2)(e^1 \wedge e^4 + e^2 \wedge e^3)] \\
&= -\frac{1}{4}[(\psi^* I \psi)(e^1 \wedge e^2 + e^3 \wedge e^4) \\
&\quad + (\psi^* J \psi)(e^1 \wedge e^3 + e^4 \wedge e^2) \\
&\quad + (\psi^* K \psi)(e^1 \wedge e^4 + e^2 \wedge e^3)]
\end{aligned} \tag{7.76}$$

(apply ρ^* to the right-hand side to get $(\psi \otimes \psi^*)_0$).

Finally, we have the algebraic spadework completed and we can proceed to the problem of globalizing all of these notions to manifolds and bundles. Henceforth, B will denote a compact, connected, simply connected, oriented smooth four-manifold (simple connectivity is not essential here, but will streamline some of what we have to say). For any choice of Riemannian metric g on B ,

$$\text{SO}(4) \hookrightarrow F_{\text{SO}}(B) \xrightarrow{\pi_{\text{SO}}} B$$

will denote the corresponding oriented, orthonormal frame bundle of B . Should B happen to admit a spin structure (see the Remark following the proof of Theorem 7.3), then the representations $\Delta_{\mathbb{C}}^{\pm} : \mathfrak{Spin}(4) \rightarrow \text{SU}(S_{\mathbb{C}}^{\pm})$ give associated spinor bundles $S(B) \times_{\Delta_{\mathbb{C}}^{\pm}} S_{\mathbb{C}}^{\pm}$ whose sections are spinor fields. Because such a spin structure need not exist (and because spinor fields are essential ingredients in Seiberg-Witten theory), we formulate a complex analogue of a spin structure, which always exists.

A **spin^c structure** \mathcal{L} on B consists of a principal $\mathfrak{Spin}^c(4)$ -bundle

$$\mathfrak{Spin}^c(4) \hookrightarrow S^c(B) \xrightarrow{\pi_{S^c}} B \tag{7.77}$$

over B and a smooth map

$$\Lambda : S^c(B) \longrightarrow F_{\text{SO}}(B) \tag{7.78}$$

satisfying

$$\pi_{\text{SO}} \circ \Lambda = \pi_{S^c} \tag{7.79}$$

and

$$\Lambda(p \cdot \xi) = \Lambda(p) \cdot \pi(\xi) \tag{7.80}$$

for each $p \in S^c(B)$ and each $\xi \in \mathfrak{Spin}^c(4)$. Here $\pi : \mathfrak{Spin}^c(4) \rightarrow \mathrm{SO}(4)$ is defined by (7.64).

$$\begin{array}{ccc}
 S^c(B) & \xrightarrow{\Lambda} & F_{\mathrm{SO}}(B) \\
 \searrow \pi_{S^c} & & \nearrow \pi_{\mathrm{SO}} \\
 & & B
 \end{array}$$

It is known that, for any B of the type we have described and any choice of the Riemannian metric g , spin^c structures exist. In terms of transition functions this means that for any trivializing cover $\{U_\alpha\}$ for the frame bundle with transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow \mathrm{SO}(4)$, there exist lifts $\tilde{g}_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow \mathfrak{Spin}^c(4)$

$$\begin{array}{ccc}
 & & \mathfrak{Spin}^c(4) \\
 & \nearrow \tilde{g}_{\alpha\beta} & \downarrow \pi \\
 U_\alpha \cap U_\beta & \xrightarrow{g_{\alpha\beta}} & \mathrm{SO}(4)
 \end{array}$$

satisfying the cocycle condition.

Given a spin^c structure \mathcal{L} on B each of the representations $\hat{\Delta}_{\mathbb{C}} : \mathfrak{Spin}^c(4) \rightarrow \mathrm{Aut}_{\mathbb{C}}(S_{\mathbb{C}})$, $\hat{\Delta}_{\mathbb{C}}^{\pm} : \mathfrak{Spin}^c(4) \rightarrow \mathrm{U}(S_{\mathbb{C}}^{\pm})$ and $\delta : \mathfrak{Spin}^c(4) \rightarrow \mathrm{U}(1)$ give rise to vector bundles associated to (7.77) which we will write as follows.

$$\begin{aligned}
 \mathcal{S}(\mathcal{L}) &= S^c(B) \times_{\hat{\Delta}_{\mathbb{C}}} S_{\mathbb{C}} \\
 \mathcal{S}^{\pm}(\mathcal{L}) &= S^c(B) \times_{\hat{\Delta}_{\mathbb{C}}^{\pm}} S_{\mathbb{C}}^{\pm} \\
 L(\mathcal{L}) &= S^c(B) \times_{\delta} \mathbb{C}
 \end{aligned}$$

$\mathcal{S}(\mathcal{L})$ is called the **spinor bundle** of \mathcal{L} , $\mathcal{S}^{\pm}(\mathcal{L})$ are the **positive** and **negative spinor bundles** of \mathcal{L} and $L(\mathcal{L})$ is the **determinant line bundle** of \mathcal{L} . The algebraic decomposition (7.57) persists in the bundle setting to give a Whitney sum decomposition

$$\mathcal{S}(\mathcal{L}) = \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L}).$$

We will also need the principal $\mathrm{U}(1)$ -bundle

$$\mathrm{U}(1) \hookrightarrow L^0(\mathcal{L}) \xrightarrow{\pi_{L^0}} B$$

associated to $L(\mathcal{L})$.

Remark. *This bundle can be described as follows. Choose a Hermitian fiber metric on the complex line bundle $L(\mathcal{L})$. Then $L^0(\mathcal{L})$ is the unit circle bundle in $L(\mathcal{L})$ i.e., it is the corresponding oriented, orthonormal frame bundle. One can retrieve $L(\mathcal{L})$ from $L^0(\mathcal{L})$ as the vector bundle associated to $L^0(\mathcal{L})$ by complex multiplication. One can show that $w_2(B) = c_1(L^0(\mathcal{L})) \pmod{2}$, where $w_2(B)$ is the second Stiefel-Whitney class of B , and that, conversely, given a $U(1)$ -bundle L^0 over B with $w_2(B) = c_1(L^0) \pmod{2}$ there is a spin^c structure \mathcal{L} on B with $L^0(\mathcal{L}) = L^0$.*

We will also require a bundle associated to the frame bundle that does not require a spin or spin^c structure. Notice that $\mathfrak{Spin}(4)$, being contained in $Cl^\times(4)$, acts on $Cl(4)$ by conjugation and, since $(-u)p(-u)^{-1} = upu^{-1}$, this gives an action of $SO(4) = \mathfrak{Spin}(4)/\mathbb{Z}_2$ on $Cl(4)$ which clearly preserves products $(u(pq)u^{-1} = (upu^{-1})(uqu^{-1}))$. The **Clifford bundle** $Cl(B)$ is the bundle with typical fiber $Cl(4)$ over B associated to the frame bundle by this action.

$$Cl(B) = F_{SO}(B) \times_{SO(4)} Cl(4)$$

Similarly, one has a **complexified Clifford bundle**

$$Cl(B) \otimes \mathbb{C} = F_{SO}(B) \times_{SO(4)} (Cl(4) \otimes \mathbb{C}).$$

These decompose into even and odd summands, e.g., $Cl(B) \cong Cl_0(B) \oplus Cl_1(B)$. Moreover, pointwise multiplication provides the spaces of sections of these bundles with algebra structures and such sections act on sections of the spinor bundle by pointwise Clifford multiplication.

Now, $\mathfrak{Spin}^c(4)$ double covers $SO(4) \times U(1)$ by the map \mathfrak{Spin}^c so $S^c(B)$ double covers the fiber product $F_{SO}(B) \dot{\times} L^0(\mathcal{L})$ (this is just that part of the product bundle $SO(4) \times U(1) \hookrightarrow F_{SO}(B) \times L^0(\mathcal{L}) \rightarrow B \times B$ above the diagonal in $B \times B$ with this diagonal identified with B in the obvious way). We will use the symbol \mathfrak{Spin}^c also for this double cover.

$$\begin{array}{ccc} \mathfrak{Spin}^c(4) \hookrightarrow S^c(B) & \longrightarrow & B \\ & \downarrow \mathfrak{S}^c & \\ SO(4) \times U(1) \hookrightarrow F_{SO}(B) \dot{\times} L^0(\mathcal{L}) & \longrightarrow & B \end{array} \quad (7.81)$$

Locally, this map is just

$$(b, \xi) \longrightarrow (b, \mathfrak{Spin}^c(\xi)) = (b, (\pi(\xi), \delta(\xi))) = (b, (\mathfrak{S}(u), e^{2\theta i})) \quad (7.82)$$

where $\xi = e^{\theta i} u \in \mathfrak{Spin}^c(4)$.

Next observe that the algebraic isomorphism σ^+ in (7.73) globalizes as follows. The map ρ of (7.68) is independent of the choice of oriented, orthonormal basis for \mathbb{R}^4 so, using local oriented, orthonormal frame fields on B (i.e., sections of $F_{\text{SO}}(B)$) it gives a map from two-forms on B to sections of $Cl(B) \otimes \mathbb{C}$. These sections in turn act on the spinor bundle $\mathcal{S}(\mathcal{L}) \cong \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$ of any spin^c structure. Since the action is fiberwise, the image of a self-dual two-form preserves $\mathcal{S}^+(\mathcal{L})$ and is, at each point, a trace free endomorphism of $S_{\mathbb{C}}^+$. Thus, a self-dual two-form on B gives rise to a section of the trace free endomorphism bundle $\text{End}_0(\mathcal{S}^+(\mathcal{L}))$ of $\mathcal{S}^+(\mathcal{L})$ and we have an isomorphism (also denoted σ^+)

$$\sigma^+ : \Gamma(\text{End}_0(\mathcal{S}^+(\mathcal{L}))) \longrightarrow \Omega_+^2(B). \quad (7.83)$$

In particular, if ψ is a positive spinor field on B (i.e., a section of $\mathcal{S}^+(\mathcal{L})$), then $(\psi \otimes \psi^*)_0 \in \Gamma(\text{End}_0(\mathcal{S}^+(\mathcal{L})))$ is given pointwise by (7.75) and $\sigma^+((\psi \otimes \psi^*)_0) \in \Omega_+^2(B)$ is given pointwise relative to a local oriented, orthonormal frame field on B by (7.76).

Remark. *At this point we can begin to give some idea of where all of this is going. To write the Seiberg-Witten equations for B one chooses a Riemannian metric \mathbf{g} and a spin^c structure \mathcal{L} . The field content of the theory consists of a $U(1)$ -connection \mathbf{A} on $L^0(\mathcal{L})$ (the gauge field) and a positive spinor field ψ . They are related by two equations, one of which requires that the self-dual two-form $\sigma^+((\psi \otimes \psi^*)_0)$ should coincide with the self-dual part of the curvature of \mathbf{A} i.e., $\mathbf{F}_{\mathbf{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0)$. The other equation still requires a bit of preparation, however.*

Recall that the frame bundle $\text{SO}(4) \hookrightarrow F_{\text{SO}}(B) \rightarrow B$ has a distinguished (Levi-Civita) connection which we will denote ω_{LC} . This can be characterized locally as follows. If $\{e_1, e_2, e_3, e_4\}$ is a local oriented, orthonormal frame field on B (i.e., a section of $F_{\text{SO}}(B)$) with dual one-form field $\{e^1, e^2, e^3, e^4\}$, then ω_{LC} is represented by a skew-symmetric matrix (ω_j^i) of \mathbb{R} -valued one-forms satisfying $de^i = -\omega_j^i \wedge e^j$, $i = 1, 2, 3, 4$. Now notice that if B had a spin structure $\mathfrak{Spin}(4) \hookrightarrow S(B) \rightarrow B$, then the map $\lambda : S(B) \rightarrow F_{\text{SO}}(B)$ is a double cover that respects the group actions so that any connection on $F_{\text{SO}}(B)$, e.g., ω_{LC} , automatically lifts to a connection on $S(B)$ (think of the connection as a distribution of horizontal spaces). However, if B has only a spin^c structure $\mathfrak{Spin}^c(4) \hookrightarrow S^c(B) \rightarrow B$, then the map Λ of (7.78) is not a finite covering so ω_{LC} alone will not determine a connection on $S^c(B)$. However, $\mathfrak{Spin}^c : S^c(B) \rightarrow F_{\text{SO}}(B) \times L^0(\mathcal{L})$ is a double cover ((7.81)) and if \mathbf{A} is any connection

on $L^0(\mathcal{L})$, then \mathbf{A} and ω_{LC} together determine a connection on $F_{\text{SO}}(B) \dot{\times} L^0(\mathcal{L})$ which will then lift to a connection on $S^c(B)$. Specifically, if pr_F and pr_{L^0} denote the restrictions to $F_{\text{SO}}(B) \dot{\times} L^0(\mathcal{L})$ of the projections of $F_{\text{SO}}(B) \times L^0(\mathcal{L})$ onto $F_{\text{SO}}(B)$ and $L^0(\mathcal{L})$, respectively, then

$$\text{pr}_F^* \omega_{LC} \oplus \text{pr}_{L^0}^* \mathbf{A}$$

is a connection on the fiber product and, identifying $\mathfrak{spin}^c(4)$ with the subset of $Cl(4) \otimes \mathbb{C}$ given in (7.66),

$$\omega_{\mathbf{A}} = (\mathfrak{S}^c)^*(\text{pr}_F^* \omega_{LC} \oplus \text{pr}_{L^0}^* \mathbf{A}) \quad (7.84)$$

is a connection on $S^c(B)$. Any such connection $\omega_{\mathbf{A}}$ is called a **spin^c connection** for \mathcal{L} and with one of these we can introduce the basic differential operator of Seiberg-Witten theory.

Remark. *The Levi-Civita connection ω_{LC} on $F_{\text{SO}}(B)$ is to be regarded as fixed. The $U(1)$ -connection \mathbf{A} on $L^0(\mathcal{L})$, on the other hand, is the gauge field of Seiberg-Witten theory and will be constrained only by the field equations we eventually write down. The task of \mathbf{A} is to produce, with ω_{LC} , a spin^c connection $\omega_{\mathbf{A}}$ on $S^c(B)$ which in turn induces covariant derivatives on all of the associated spinor bundles $\mathcal{S}(\mathcal{L})$, $\mathcal{S}^+(\mathcal{L})$ and $\mathcal{S}^-(\mathcal{L})$. It is in terms of these covariant derivatives that we will build a “Dirac operator” on spinor fields and then the second Seiberg-Witten equation.*

Now recall that if $G \hookrightarrow P \rightarrow X$ is a principal G -bundle and F is a manifold on which G acts on the left, then sections of the corresponding associated fiber bundle can be identified with maps from P to F that are equivariant ($\psi(p \cdot g) = g^{-1} \cdot \psi(p)$). We will generally use the same symbol for both and will adopt whichever view is convenient in each context. A spin^c connection $\omega_{\mathbf{A}}$ on $\mathfrak{Spin}^c(4) \hookrightarrow S^c(B) \xrightarrow{\pi_{S^c}} B$ induces covariant derivatives on the associated spinor bundles $\mathcal{S}(\mathcal{L})$, $\mathcal{S}^+(\mathcal{L})$ and $\mathcal{S}^-(\mathcal{L})$, all of which will be denoted $\nabla_{\mathbf{A}}$ when thought of as operating on sections of vector bundles and $d_{\mathbf{A}}$ when operating on $S_{\mathbb{C}}$ -valued equivariant maps on $S^c(B)$. We will use

$$\nabla_{\mathbf{A}} : \Gamma(\mathcal{S}(\mathcal{L})) \longrightarrow \Omega^1(B) \otimes \Gamma(\mathcal{S}(\mathcal{L}))$$

to define the **(coupled) Dirac operator**

$$\tilde{\mathcal{D}}_{\mathbf{A}} : \Gamma(\mathcal{S}(\mathcal{L})) \longrightarrow \Gamma(\mathcal{S}(\mathcal{L}))$$

as follows. Let $\{e_1, e_2, e_3, e_4\}$ be a local oriented, orthonormal frame field on $U \subseteq B$ (i.e., a local section of $F_{\text{SO}}(B)$). Each e_i can be regarded as a vector

field on U and also as a section of the Clifford bundle $Cl(B)$ which therefore acts by Clifford multiplication on sections of $\mathcal{S}(\mathcal{L})$ defined on U . Thus, for each $\Psi \in \Gamma(\mathcal{S}(\mathcal{L}))$ we can define $\tilde{\mathcal{D}}_{\mathbf{A}}\Psi$ on U by

$$\tilde{\mathcal{D}}_{\mathbf{A}}\Psi = \sum_{i=1}^4 e_i \cdot \nabla_{\mathbf{A}}\Psi(e_i). \quad (7.85)$$

One shows that this is independent of the choice of $\{e_1, e_2, e_3, e_4\}$ and so defines $\nabla_{\mathbf{A}}\Psi$ globally. We will write out a concrete example shortly.

Since $\mathcal{S}(\mathcal{L}) = \mathcal{S}^+(\mathcal{L}) \oplus \mathcal{S}^-(\mathcal{L})$ we may restrict $\tilde{\mathcal{D}}_{\mathbf{A}}$ to sections of either $\mathcal{S}^+(\mathcal{L})$ or $\mathcal{S}^-(\mathcal{L})$. Since Clifford multiplication by e_i switches $\mathcal{S}^{\pm}(\mathcal{L})$, so will these restrictions. We will write these as

$$\mathcal{D}_{\mathbf{A}} : \Gamma(\mathcal{S}^+(\mathcal{L})) \longrightarrow \Gamma(\mathcal{S}^-(\mathcal{L})) \quad (7.86)$$

and

$$\mathcal{D}_{\mathbf{A}}^* : \Gamma(\mathcal{S}^-(\mathcal{L})) \longrightarrow \Gamma(\mathcal{S}^+(\mathcal{L})) \quad (7.87)$$

(these are, in fact, adjoints relative to the L^2 inner product on sections induced by the pointwise Hermitian inner product on fibers). We will also follow the custom in mathematics of referring to $\mathcal{D}_{\mathbf{A}}$ also as a **(coupled) Dirac operator**.

With this we can (at last) formulate the Seiberg-Witten equations. Thus, we let B denote a compact, connected, simply connected, oriented, smooth four-manifold. Select a Riemannian metric g for B and then a spin^c structure \mathcal{L} for the corresponding oriented, orthonormal frame bundle $F_{\text{SO}}(B)$. A pair (\mathbf{A}, ψ) consisting of a $U(1)$ -connection \mathbf{A} on $U(1) \hookrightarrow L^0(B) \rightarrow B$ and a positive spinor field $\psi \in \Gamma(\mathcal{S}^+(\mathcal{L}))$ satisfies the **Seiberg-Witten (SW) equations** if

$$\mathcal{D}_{\mathbf{A}}\psi = 0 \quad (\text{DiracEquation}) \quad (7.88)$$

and

$$\mathbf{F}_{\mathbf{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0) \quad (\text{Curvature Equation}) \quad (7.89)$$

where $\mathbf{F}_{\mathbf{A}}^+$ is the g -self-dual part of the curvature of \mathbf{A} .

Remark. *The curvature of \mathbf{A} is actually a $\mathfrak{u}(1)$ -valued two-form $\Omega_{\mathbf{A}} = d\mathbf{A}$ on $L^0(\mathcal{L})$, but, since $U(1)$ is abelian, this projects to a $\mathfrak{u}(1)$ -valued two-form on B and this is what we mean by $\mathbf{F}_{\mathbf{A}}$.*

To gain some sense of what these equations actually look like we will write them out explicitly in local coordinates. More precisely, we consider \mathbb{R}^4 with its usual Riemannian metric and orientation. Since \mathbb{R}^4 is contractible, all of the relevant

bundles over it are trivial and we will work with explicit trivializations. Thus, the oriented, orthonormal frame bundle is

$$\mathrm{SO}(4) \hookrightarrow \mathbb{R}^4 \times \mathrm{SO}(4) \longrightarrow \mathbb{R}^4$$

and there is an essentially unique spin^c structure

$$\begin{array}{ccc} \mathbb{R}^4 \times \mathfrak{Spin}^c(4) & \xrightarrow{\Lambda} & \mathbb{R}^4 \times \mathrm{SO}(4) \\ & \searrow & \swarrow \\ & \mathbb{R}^4 & \end{array}$$

where $\Lambda(b, \xi) = (b, \pi(\xi))$ with π given by (7.64). The spinor bundles are therefore also trivial so their sections can be identified with globally defined functions on \mathbb{R}^4 which we will write

$$\Psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \mathbb{R}^4 \longrightarrow S_{\mathbb{C}} \cong \mathbb{C}^4$$

$$\psi = \begin{pmatrix} \psi^1 \\ \psi^2 \\ 0 \\ 0 \end{pmatrix} : \mathbb{R}^4 \longrightarrow S_{\mathbb{C}}^+ \cong \mathbb{C}^2 \quad \phi = \begin{pmatrix} 0 \\ 0 \\ \psi^3 \\ \psi^4 \end{pmatrix} : \mathbb{R}^4 \longrightarrow S_{\mathbb{C}}^- \cong \mathbb{C}^2.$$

For convenience we will often abuse the notation and write $\Psi = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$ by suppressing the zero components. We use x^1, x^2, x^3, x^4 for the standard coordinates on \mathbb{R}^4 and write ∂_i for $\frac{\partial}{\partial x^i}$, $i = 1, 2, 3, 4$ (these being applied componentwise to spinor fields).

The determinant line bundle is likewise trivial, as is the corresponding principal $U(1)$ -bundle

$$U(1) \hookrightarrow \mathbb{R}^4 \times U(1) \longrightarrow \mathbb{R}^4.$$

A connection on this $U(1)$ -bundle is then uniquely determined by a globally defined $u(1) = \mathrm{Im}\mathbb{C}$ -valued one-form on \mathbb{R}^4

$$\mathbf{A} = A_i dx^i \quad A_i : \mathbb{R}^4 \longrightarrow \mathrm{Im}\mathbb{C}, \quad i = 1, 2, 3, 4.$$

In any orthonormal coordinates the covariant exterior derivative induced by the Levi-Civita connection is just ordinary (componentwise) exterior differentiation

(Christoffel symbols are zero) so the covariant derivative $\nabla_{\mathbf{A}}$ induced by it and the $U(1)$ -connection \mathbf{A} takes the form $\nabla_{\mathbf{A}} = d + \mathbf{A}$, i.e.,

$$\nabla_{\mathbf{A}} = \nabla_i dx^i = (\partial_i + A_i) dx^i$$

so that

$$\nabla_{\mathbf{A}} \Psi = (\partial_i \Psi + A_i \Psi) dx^i = \begin{pmatrix} (\partial_i \psi^1 + A_i \psi^1) dx^i \\ (\partial_i \psi^2 + A_i \psi^2) dx^i \\ (\partial_i \psi^3 + A_i \psi^3) dx^i \\ (\partial_i \psi^4 + A_i \psi^4) dx^i \end{pmatrix}.$$

Thus, with $\{e_i\} = \{\partial_i\}$ the standard oriented, orthonormal frame field on \mathbb{R}^4 , we have $\nabla_{\mathbf{A}} \Psi(e_i) = \partial_i \Psi + A_i \Psi$ and, for convenience, we will write this as

$$\nabla_i \Psi = (\partial_i + A_i) \Psi = \begin{pmatrix} \partial_i \psi^1 + A_i \psi^1 \\ \partial_i \psi^2 + A_i \psi^2 \\ \partial_i \psi^3 + A_i \psi^3 \\ \partial_i \psi^4 + A_i \psi^4 \end{pmatrix}.$$

The Dirac operator $\tilde{\mathcal{D}}_{\mathbf{A}} \Psi = \sum_{i=1}^4 e_i \cdot \nabla_i \Psi$ requires that we Clifford multiply by the basis elements e_i , i.e., matrix multiply by $E_i = \Gamma(e_i) \in Cl(4) \otimes \mathbb{C}$ as in (7.51).

For this we write $\Psi = \begin{pmatrix} \psi \\ \phi \end{pmatrix}$ so that

$$\tilde{\mathcal{D}}_{\mathbf{A}} \Psi = \sum_{i=1}^4 e_i \cdot \nabla_i \Psi = \sum_{i=1}^4 E_i \nabla_i \Psi = \begin{pmatrix} \nabla_1 \phi + I \nabla_2 \phi + J \nabla_3 \phi + K \nabla_4 \phi \\ -\nabla_1 \psi + I \nabla_2 \psi + J \nabla_3 \psi + K \nabla_4 \psi \end{pmatrix}.$$

Note that, as expected, $\tilde{\mathcal{D}}_{\mathbf{A}}$ sends positive spinor fields $\Psi = \begin{pmatrix} \psi \\ 0 \end{pmatrix}$ to negative spinor fields and vice versa. The restriction of $\tilde{\mathcal{D}}_{\mathbf{A}}$ to positive spinor fields will be written

$$\mathcal{D}_{\mathbf{A}} \psi = -\nabla_1 \psi + I \nabla_2 \psi + J \nabla_3 \psi + K \nabla_4 \psi \quad (7.90)$$

by suppressing the zero components (but now one must remember that $\mathcal{D}_{\mathbf{A}}$ is a negative spinor field). The first Seiberg-Witten equation (7.88) then becomes

$$\nabla_1 \psi = I \nabla_2 \psi + J \nabla_3 \psi + K \nabla_4 \psi \quad (7.91)$$

or, in complete detail,

$$\begin{pmatrix} -(\partial_1 + A_1) + i(\partial_2 + A_2) & (\partial_3 + A_3) + i(\partial_4 + A_4) \\ -(\partial_3 + A_3) + i(\partial_4 + A_4) & -(\partial_1 + A_1) - i(\partial_2 + A_2) \end{pmatrix} \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (7.92)$$

This part is linear, of course.

For the second Seiberg-Witten equation (7.89) we use the (pointwise) expressions (7.76) for $\sigma^+(\psi \otimes \psi^*)_0$ and the following local description of $F_{\mathbf{A}}^+$. With $\mathbf{A} = A_i dx^i$ we have $F_{\mathbf{A}} = d\mathbf{A} = \sum_{i < j} F_{ij} dx^i \wedge dx^j$, where $F_{ij} = \partial_i A_j - \partial_j A_i$.

A basis for the self-dual two-forms is given by $\{dx^1 \wedge dx^2 + dx^3 \wedge dx^4, dx^1 \wedge dx^3 + dx^4 \wedge dx^2, dx^1 \wedge dx^4 + dx^2 \wedge dx^3\}$ so

$$\begin{aligned} F_{\mathbf{A}}^+ &= \frac{1}{2}(F_{\mathbf{A}} + *F_{\mathbf{A}}) = \frac{1}{2}(F_{12} + F_{34})(dx^1 \wedge dx^2 + dx^3 \wedge dx^4) \\ &\quad + \frac{1}{2}(F_{13} + F_{42})(dx^1 \wedge dx^3 + dx^4 \wedge dx^2) \\ &\quad + \frac{1}{2}(F_{14} + F_{23})(dx^1 \wedge dx^4 + dx^2 \wedge dx^3). \end{aligned}$$

Thus, (7.89) becomes

$$\begin{aligned} F_{12} + F_{34} &= -\frac{1}{2}\psi^* I \psi \\ F_{13} + F_{42} &= -\frac{1}{2}\psi^* J \psi \\ F_{14} + F_{23} &= -\frac{1}{2}\psi^* K \psi \end{aligned} \tag{7.93}$$

or, in more detail,

$$\begin{aligned} (\partial_1 A_2 - \partial_2 A_1) + (\partial_3 A_4 - \partial_4 A_3) &= -\frac{1}{2}i(|\psi^1|^2 - |\psi^2|^2) \\ (\partial_1 A_3 - \partial_3 A_1) + (\partial_4 A_2 - \partial_2 A_4) &= -i\text{Re}(\bar{\psi}^1 \psi^2) \\ (\partial_1 A_4 - \partial_4 A_1) + (\partial_2 A_3 - \partial_3 A_2) &= -i\text{Im}(\bar{\psi}^1 \psi^2). \end{aligned} \tag{7.94}$$

Note that these are only rather mildly nonlinear.

Remarks. *It is perhaps worth pointing out that these equations do have non-trivial solutions. Indeed, if $\psi \equiv 0$, the Dirac equation is satisfied identically and \mathbf{A} can be taken to be any “U(1)-instanton”, i.e., anti-self-dual connection on $U(1) \hookrightarrow \mathbb{R}^4 \times U(1) \rightarrow \mathbb{R}^4$ and these are abundant. Witten has shown, moreover, that any solution (\mathbf{A}, ψ) to (7.91) and (7.93) for which $\psi \in L^2(\mathbb{R}^4)$ is necessarily of this type (for a proof see pages 416-420 of [41]). Finally, we mention that solutions to the Seiberg-Witten equations can, like solutions to the anti-self-dual equations, be viewed as the absolute minima of a certain action functional and so fit nicely into our general framework for classical gauge theories.*

We return now to the general development so that B is a compact, connected, simply connected, oriented, smooth four-manifold. Choosing a Riemannian metric g on B gives an oriented, orthonormal frame bundle and one can then select a spin^c structure \mathcal{L} . The corresponding **Seiberg-Witten configuration space** $\mathcal{A}(\mathcal{L})$ consists of all pairs (\mathbf{A}, ψ) , where \mathbf{A} is a connection on the principal $U(1)$ -bundle $L^0(\mathcal{L})$ and $\psi \in \Gamma(S^+(\mathcal{L}))$ is a positive spinor field. An $(\mathbf{A}, \psi) \in \mathcal{A}(\mathcal{L})$ is a **Seiberg-Witten monopole (SW monopole)** if it satisfies (7.88) and (7.89) (these two equations together will henceforth be denoted simply (SW)). As in the case of Donaldson theory, our real interest is in a moduli space of SW monopoles so we begin by isolating the appropriate gauge group. This will be a subgroup of the group of automorphisms of the spin^c bundle (diffeomorphisms σ of $S^c(B)$ onto itself satisfying $\sigma(p \cdot \xi) = \sigma(p) \cdot \xi$ for each $p \in S^c(B)$ and each $\xi \in \mathfrak{Spin}^c(4)$ and $\pi_{S^c} \circ \sigma = \pi_{S^c}$).

Recall that $S^c(B)$ double covers the fiber product $F_{\text{SO}}(B) \dot{\times} L^0(\mathcal{L})$ via the map \mathfrak{S}^c . Letting pr_F and pr_{L^0} be the projections of $F_{\text{SO}}(B) \dot{\times} L^0(\mathcal{L})$ onto $F_{\text{SO}}(B)$ and $L^0(\mathcal{L})$ we obtain maps

$$\begin{array}{ccc} \mathfrak{Spin}^c(4) \hookrightarrow S^c(B) & \longrightarrow & B \\ & \downarrow \text{pr}_F \circ \mathfrak{S}^c & \\ \text{SO}(4) \hookrightarrow F_{\text{SO}}(B) & \longrightarrow & B \end{array} \quad (7.95)$$

and

$$\begin{array}{ccc} \mathfrak{Spin}^c(4) \hookrightarrow S^c(B) & \longrightarrow & B \\ & \downarrow \text{pr}_{L^0} \circ \mathfrak{S}^c & \\ U(1) \hookrightarrow L^0(\mathcal{L}) & \longrightarrow & B \end{array} \quad (7.96)$$

We will say that an automorphism $\sigma : S^c(B) \rightarrow S^c(B)$ **covers the identity on** $F_{\text{SO}}(B)$ if

$$\text{pr}_F \circ \mathfrak{S}^c \circ \sigma = \text{pr}_F \circ \mathfrak{S}^c. \quad (7.97)$$

The collection of all such is a group $\mathcal{G}(\mathcal{L})$ under composition which we call the **(Seiberg-Witten) gauge group** and which we will show acts naturally on the solutions to (SW).

Lemma 7.5. *If $\gamma \in C^\infty(B, U(1))$ is any smooth map of B into $U(1) \subseteq \mathfrak{Spin}^c(4)$, then the map*

$$\begin{aligned} \sigma_\gamma : S^c(B) &\longrightarrow S^c(B) \\ \sigma_\gamma(p) &= p \cdot \gamma(\pi_{S^c}(p)) \end{aligned}$$

is an automorphism of $S^c(B)$ that covers the identity on $F_{\text{SO}}(B)$. Conversely, every element of $\mathcal{G}(\mathcal{L})$ is σ_γ for some $\gamma \in C^\infty(B, \text{U}(1))$ and

$$\mathcal{G}(\mathcal{L}) \cong C^\infty(B, \text{U}(1))$$

where the group operation in $C^\infty(B, \text{U}(1))$ is pointwise multiplication in $\text{U}(1)$.

Proof: First note that σ_γ is an automorphism since it is clearly a diffeomorphism and, because $\text{U}(1) = Z(\mathfrak{Spin}^c(4))$, $\sigma_\gamma(p \cdot \xi) = (p \cdot \xi) \cdot \gamma(\pi_{S^c}(p)) = (p \cdot \gamma(\pi_{S^c}(p))) \cdot \xi = \sigma_\gamma(p) \cdot \xi$. It covers the identity because, locally, $\text{pr}_F \circ \mathfrak{S}^c$ is given by $(b, \xi) = (b, e^{\theta i} u) \rightarrow (b, \mathfrak{S}(u))$ so p and $p \cdot e^{\phi i}$ always have the same image.

For the converse, it is easy to verify that $p_1, p_2 \in S^c(B)$ have the same image under $\text{pr}_F \circ \mathfrak{S}^c$ if and only if they differ by the action of something in $\text{U}(1)$, i.e., $p_2 = p_1 \cdot e^{\phi i}$ for some $\phi \in \mathbb{R}$. Thus, an automorphism $\sigma : S^c(B) \rightarrow S^c(B)$ that covers the identity on $F_{\text{SO}}(B)$ must satisfy $\sigma(p) = p \cdot (\text{something in } \text{U}(1))$ for each $p \in S^c(B)$. We claim that this “something” must be the same for all points in the same fiber of π_{S^c} . Indeed, $\pi_{S^c}(p_1) = \pi_{S^c}(p_2)$ implies $p_2 = p_1 \cdot \xi$ for some $\xi \in \mathfrak{Spin}^c(4)$ and if $\sigma(p_1) = p_1 \cdot e^{\phi i}$, then $\sigma(p_2) = \sigma(p_1 \cdot \xi) = \sigma(p_1) \cdot \xi = (p_1 \cdot e^{\phi i}) \cdot \xi = (p_1 \cdot \xi) \cdot e^{\phi i} = p_2 \cdot e^{\phi i}$ also. Thus, $\sigma(p) = p \cdot \gamma(\pi_{S^c}(p))$ for some $\gamma \in C^\infty(B, \text{U}(1))$ as required and the rest is clear. ■

We will use whichever view of the gauge group $\mathcal{G}(\mathcal{L})$ is most convenient in any particular situation.

Our goal now is to show that $\mathcal{G}(\mathcal{L})$ acts naturally on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ and, indeed, preserves the set of solutions to (SW). We begin by defining the action of $\mathcal{G}(\mathcal{L})$ on positive spinor fields $\psi \in \Gamma(S^+(\mathcal{L}))$. For this we identify ψ with an equivariant $S_{\mathbb{C}}^+$ -valued map on $S^c(B)$ and define the action of $\sigma_\gamma \in \mathcal{G}(\mathcal{L})$ by pullback, i.e.,

$$\psi \cdot \sigma_\gamma = \psi \circ \gamma = \sigma_\gamma^* \psi = \psi \circ \sigma_\gamma. \quad (7.98)$$

Thus, at each $p \in S^c(B)$,

$$(\psi \cdot \sigma_\gamma)(p) = \psi(\sigma_\gamma(p)) = \psi(p \cdot \gamma(\pi_{S^c}(p))) = (\gamma(\pi_{S^c}(p)))^{-1} \psi(p). \quad (7.99)$$

Thus, if we think instead of ψ as a section of $S^+(\mathcal{L})$ we have

$$\psi \cdot \gamma = (\gamma \circ \pi_{S^c})^{-1} \psi. \quad (7.100)$$

The same formulas define the action of $\mathcal{G}(\mathcal{L})$ on negative spinor fields.

Turning next to the connection \mathbf{A} on $U(1) \hookrightarrow L^0(\mathcal{L}) \rightarrow B$ we note that the automorphism σ_γ of $S^c(B)$ induces an automorphism σ'_γ of $L^0(\mathcal{L})$ as follows:

$$\begin{array}{ccc} S^c(B) & \xrightarrow{\sigma_\gamma} & S^c(B) \\ \text{pr}_{L^0} \circ \mathfrak{S}^c \downarrow & & \downarrow \text{pr}_{L^0} \circ \mathfrak{S}^c \\ L^0(\mathcal{L}) & \xrightarrow{\sigma'_\gamma} & L^0(\mathcal{L}) \end{array} \quad (7.101)$$

$$\sigma'_\gamma \circ \text{pr}_{L^0} \circ \mathfrak{S}^c = \text{pr}_{L^0} \circ \mathfrak{S}^c \circ \sigma_\gamma$$

(we will write out an explicit local expression for σ'_γ shortly). Now we define the action of $\sigma_\gamma \in \mathcal{G}(\mathcal{L})$ on \mathbf{A} by

$$\mathbf{A} \cdot \sigma_\gamma = \mathbf{A} \cdot \gamma = (\sigma'_\gamma)^* \mathbf{A}. \quad (7.102)$$

It will be convenient to have (7.102) expressed locally in terms of gauge potentials. Thus, we let s be a local section of $L^0(\mathcal{L})$ and write $\mathcal{A} = s^* \mathbf{A}$. Also define

$$\mathcal{A} \cdot \gamma = s^*(\mathbf{A} \cdot \gamma) = s^*((\sigma'_\gamma)^* \mathbf{A}) = (\sigma'_\gamma \circ s)^* \mathbf{A}. \quad (7.103)$$

Now, since $\text{pr}_{L^0} \circ \mathfrak{Spin}^c$ is locally given by

$$(b, \xi) = (b, e^{\theta i} u) \longrightarrow (b, \delta(u)) = (b, e^{2\theta i})$$

it satisfies

$$(\text{pr}_{L^0} \circ \mathfrak{S}^c)(p \cdot \xi_0) = ((\text{pr}_{L^0} \circ \mathfrak{S}^c)(p)) \cdot \delta(\xi_0) \quad (7.104)$$

so

$$\begin{aligned} \sigma'_\gamma((\text{pr}_{L^0} \circ \mathfrak{S}^c)(p)) &= (\text{pr}_{L^0} \circ \mathfrak{S}^c)(\sigma_\gamma(p)) \\ &= (\text{pr}_{L^0} \circ \mathfrak{S}^c)(p \cdot \gamma(\pi_{S^c}(p))) \\ &= ((\text{pr}_{L^0} \circ \mathfrak{S}^c)(p)) \cdot \delta(\gamma(\pi_{S^c}(p))). \end{aligned}$$

Thus, we may write

$$\sigma'_\gamma(x) = x \cdot \delta(\gamma(\pi_{L^0}(x))) = x \cdot (\gamma(\pi_{L^0}(x)))^2 \quad (7.105)$$

for each $x \in L^0(\mathcal{L})$. In particular,

$$(\sigma'_\gamma \circ s)(b) = s(b) \cdot (\gamma(b))^2. \quad (7.106)$$

It then follows from (7.102) that

$$\mathcal{A} \cdot \gamma = (\gamma^2)^{-1} \mathcal{A} (\gamma^2) + (\gamma^2)^{-1} d(\gamma^2) = \mathcal{A} + (\gamma^2)^{-1} (2\gamma d\gamma)$$

and therefore

$$\mathbf{A} \cdot \gamma = \mathbf{A} + 2\gamma^{-1}d\gamma. \quad (7.107)$$

Applying $\pi_{L^0}^*$ to both sides of (7.106) gives

$$\mathbf{A} \cdot \gamma = \mathbf{A} + \pi_{L^0}^*(2\gamma^{-1}d\gamma) = \mathbf{A} + 2(\gamma \circ \pi_{L^0})^{-1}d(\gamma \circ \pi_{L^0}). \quad (7.108)$$

We now have an action of the group $\mathcal{G}(\mathcal{L})$ on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ given by

$$\begin{aligned} (\mathbf{A}, \psi) \cdot \sigma_\gamma &= (\mathbf{A}, \psi) \cdot \gamma = ((\sigma'_\gamma)^* \mathbf{A}, \sigma_\gamma^* \psi) \\ &= (\mathbf{A} + 2(\gamma \circ \pi_{L^0})^{-1}d(\gamma \circ \pi_{L^0}), (\gamma \circ \pi_{S^c})^{-1}\psi) \end{aligned}$$

or, locally on B

$$(\mathbf{A}, \psi) \cdot \gamma = (\mathbf{A} + 2\gamma^{-1}d\gamma, \gamma^{-1}\psi). \quad (7.109)$$

In order to show that this action preserves solutions to (SW) we first observe that the spin^c connection corresponding to $\mathbf{A} \cdot \gamma$ is the pullback by σ_γ of that corresponding to \mathbf{A} , i.e.,

$$\omega_{\mathbf{A} \cdot \gamma} = \sigma_\gamma^* \omega_{\mathbf{A}}. \quad (7.110)$$

To see this we note, from (7.84), that

$$\sigma_\gamma^* \omega_{\mathbf{A}} = (\mathfrak{S}^c \circ \sigma_\gamma)^*(\text{pr}_F^* \omega_{LC} + \text{pr}_{L^0}^* \mathbf{A})$$

and

$$\begin{aligned} \omega_{\mathbf{A} \cdot \gamma} &= (\mathfrak{S}^c)^*(\text{pr}_F^* \omega_{LC} + \text{pr}_{L^0}^*((\sigma'_\gamma)^* \mathbf{A})) \\ &= (\text{pr}_F \circ \mathfrak{S}^c)^* \omega_{LC} + (\sigma'_\gamma \circ \text{pr}_{L^0} \circ \mathfrak{S}^c)^* \mathbf{A} \\ &= (\text{pr}_F \circ \mathfrak{S}^c \circ \sigma_\gamma)^* \omega_{LC} + (\text{pr}_{L^0} \circ \mathfrak{S}^c \circ \sigma_\gamma)^* \mathbf{A} \\ &\quad \text{(by (7.97) and (7.101))} \\ &= (\mathfrak{S}^c \circ \sigma_\gamma)^*(\text{pr}_F^* \omega_{LC} + \text{pr}_{L^0}^* \mathbf{A}) \\ &= \sigma_\gamma^* \omega_{\mathbf{A}}. \end{aligned}$$

Another computation of $\omega_{\mathbf{A} \cdot \gamma}$ using (7.107) gives

$$\begin{aligned} \omega_{\mathbf{A} \cdot \gamma} &= (\mathfrak{S}^c)^*(\text{pr}_F^* \omega_{LC} + \text{pr}_{L^0}^*(\mathbf{A} + \pi_{L^0}^*(2\gamma^{-1}d\gamma))) \\ &= \omega_{\mathbf{A}} + (\mathfrak{S}^c)^*(\text{pr}_{L^0}^*(\pi_{L^0}^*(2\gamma^{-1}d\gamma))) \\ &= \omega_{\mathbf{A}} + (\pi_{L^0} \circ \text{pr}_{L^0} \circ \mathfrak{S}^c)^*(2\gamma^{-1}d\gamma) \\ &= \omega_{\mathbf{A}} + \pi_{S^c}^*(2\gamma^{-1}d\gamma) \\ \omega_{\mathbf{A} \cdot \gamma} &= \omega_{\mathbf{A}} + 2(\gamma \circ \pi_{S^c})^{-1}d(\gamma \circ \pi_{S^c}). \end{aligned} \quad (7.111)$$

Theorem 7.6. *The action of $\mathcal{G}(\mathcal{L})$ on the Seiberg-Witten configuration space $\mathcal{A}(\mathcal{L})$ carries solutions to (SW) onto other solutions to (SW). More precisely, if $(\mathbf{A}, \psi) \in \mathcal{A}(\mathcal{L})$ satisfies*

$$\begin{cases} \mathcal{D}_{\mathbf{A}}\psi = 0 \\ \mathbf{F}_{\mathbf{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0) \end{cases}$$

then, for any $\sigma_\gamma \in \mathcal{G}(\mathcal{L})$, $(\mathbf{A}, \psi) \cdot \gamma = (\mathbf{A} \cdot \gamma, \psi \cdot \gamma)$ satisfies

$$\begin{cases} \mathcal{D}_{\mathbf{A} \cdot \gamma}(\psi \cdot \gamma) = 0 \\ \mathbf{F}_{\mathbf{A} \cdot \gamma}^+ = \sigma^+(((\psi \cdot \gamma) \otimes (\psi \cdot \gamma)^*)_0) \end{cases}$$

Proof: For the curvature equation we observe that (7.105), the usual transformation equation for the curvature and the fact that $U(1)$ is abelian imply that

$$\mathbf{F}_{\mathbf{A} \cdot \gamma}^+ = \gamma^2 \mathbf{F}_{\mathbf{A}}^+ (\gamma^2)^{-1} = \mathbf{F}_{\mathbf{A}}^+.$$

Similarly, the commutativity of $U(1)$ gives

$$\begin{aligned} (\psi \cdot \gamma) \otimes (\psi \cdot \gamma)^* &= (\gamma^{-1}\psi) \otimes (\gamma^{-1}\psi)^* = (\gamma^{-1}\psi) \otimes (\gamma\psi^*) \\ &= (\gamma^{-1}\gamma)(\psi \otimes \psi^*) = \psi \otimes \psi^*. \end{aligned}$$

Thus, $\mathbf{F}_{\mathbf{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0)$ implies $\mathbf{F}_{\mathbf{A} \cdot \gamma}^+ = \sigma^+(((\psi \cdot \gamma) \otimes (\psi \cdot \gamma)^*)_0)$. To verify the analogous statement for the Dirac equation it will surely be enough to show that

$$\mathcal{D}_{\mathbf{A} \cdot \gamma}(\psi \cdot \gamma) = (\mathcal{D}_{\mathbf{A}}\psi) \cdot \gamma. \quad (7.112)$$

For this it will be convenient to identify ψ with an equivariant $S_{\mathbb{C}}^+$ -valued map on $S^c(B)$ and compare the covariant exterior derivatives $d_{\mathbf{A}}\psi$ and $d_{\mathbf{A} \cdot \gamma}(\psi \cdot \gamma)$. There are standard formulas for such derivatives (see, for example, Theorem 3.1.5 of [9]) which, in our case, give

$$d_{\mathbf{A}}\psi = d\psi + \frac{1}{2}\omega_{\mathbf{A}}\psi$$

and

$$d_{\mathbf{A} \cdot \gamma}(\psi \cdot \gamma) = d(\psi \cdot \gamma) + \frac{1}{2}\omega_{\mathbf{A} \cdot \gamma}(\psi \cdot \gamma)$$

where, e.g., $\omega_{\mathbf{A}}$ takes values in $\mathfrak{spin}^c(4)$, identified with a subset of $Cl(4) \otimes \mathbb{C}$ as in (7.66), and so $\omega_{\mathbf{A}}\psi$ is a matrix product.

Remark. *The factor $\frac{1}{2}$ arises from the actual identification of $\mathfrak{spin}^c(4)$ with $\mathfrak{so}(4) \oplus \mathfrak{u}(1)$ via the derivative at the identity of the double cover map $\mathfrak{S}^c : \mathfrak{Spin}^c(4) \longrightarrow \mathrm{SO}(4) \times \mathrm{U}(1)$.*

Now we compute

$$\begin{aligned}
d_{\mathbf{A}\cdot\gamma}(\psi \cdot \gamma) &= d_{\mathbf{A}\cdot\gamma}((\gamma \circ \pi_{S^c})^{-1}\psi) \\
&= d((\gamma \circ \pi_{S^c})^{-1}\psi) + \frac{1}{2}\boldsymbol{\omega}_{\mathbf{A}\cdot\gamma}((\gamma \circ \pi_{S^c})^{-1}\psi) \\
&= (\gamma \circ \pi_{S^c})^{-1}d\psi - (\gamma \circ \pi_{S^c})^{-2}d(\gamma \circ \pi_{S^c})\psi \\
&\quad + \frac{1}{2}(\boldsymbol{\omega}_{\mathbf{A}} + 2(\gamma \circ \pi_{S^c})^{-1}d(\gamma \circ \pi_{S^c}))((\gamma \circ \pi_{S^c})^{-1}\psi) \\
&= (\gamma \circ \pi_{S^c})^{-1}(d\psi + \frac{1}{2}\boldsymbol{\omega}_{\mathbf{A}}\psi) - (\gamma \circ \pi_{S^c})^{-2}d(\gamma \circ \pi_{S^c})\psi \\
&\quad + (\gamma \circ \pi_{S^c})^{-2}d(\gamma \circ \pi_{S^c})\psi \\
&= (\gamma \circ \pi_{S^c})^{-1}(d_{\mathbf{A}}\psi)
\end{aligned}$$

and from this conclude that

$$\nabla_{\mathbf{A}\cdot\gamma}(\psi \cdot \gamma) = (\gamma \circ \pi_{S^c})^{-1}\nabla_{\mathbf{A}}\psi. \quad (7.113)$$

Finally,

$$\begin{aligned}
\mathcal{D}_{\mathbf{A}\cdot\gamma}(\psi \cdot \gamma) &= \sum_{i=1}^4 e_i \cdot \nabla_{\mathbf{A}\cdot\gamma}(\psi \cdot \gamma)(e_i) = \sum_{i=1}^4 e_i \cdot (\psi \circ \pi_{S^c})^{-1}\nabla_{\mathbf{A}}\psi(e_i) \\
&= (\psi \circ \pi_{S^c})^{-1} \sum_{i=1}^4 e_i \cdot \nabla_{\mathbf{A}}\psi(e_i) = (\psi \circ \pi_{S^c})^{-1} \mathcal{D}_{\mathbf{A}}\psi = (\mathcal{D}_{\mathbf{A}}\psi) \cdot \gamma
\end{aligned}$$

which proves (7.112) and therefore Theorem 7.6. \blacksquare

Thus, the space of solutions to the Seiberg-Witten equations is invariant under the action of the gauge group $\mathcal{G}(\mathcal{L})$ and we may, as for the anti-self-dual equations, consider the **moduli space** $\mathcal{M}(\mathcal{L})$ of gauge equivalence classes of Seiberg-Witten monopoles:

$$\mathcal{M}(\mathcal{L}) = \{(\mathbf{A}, \psi) \in \mathcal{A}(\mathcal{L}); \mathcal{D}_{\mathbf{A}}\psi = 0, \mathbf{F}_{\mathbf{A}}^+ = \sigma^+((\psi \otimes \psi^*)_0)\}/\mathcal{G}(\mathcal{L}). \quad (7.114)$$

The ensuing analysis required to manufacture a differential topological invariant of B from $\mathcal{M}(\mathcal{L})$ is in many ways analogous to that which we outlined for Donaldson theory in Section 4. For this reason we will simply sketch with much broader strokes those aspects of the construction that are much the same and linger a bit longer over those that present something new.

Seiberg-Witten theory is in many ways technically much simpler than Donaldson theory and very often the simplifications can be attributed to the fact that $U(1)$, unlike $SU(2)$, is abelian. Our first manifestation of this is the ease with which we identify the reducible elements of the configuration space.

Lemma 7.7. *An element (\mathcal{A}, ψ) of $\mathcal{A}(\mathcal{L})$ is left fixed by some non-identity element σ_γ of $\mathcal{G}(\mathcal{L})$ if and only if $\psi \equiv 0$ and, in this case, $\gamma : B \rightarrow \mathrm{U}(1)$ must be a constant map.*

Proof: It will suffice to argue locally so we let \mathcal{A} be a gauge potential for \mathcal{A} . Then $(\mathcal{A}, \psi) \cdot \gamma = (\mathcal{A}, \psi)$ if and only if

$$(\mathcal{A} + 2\gamma^{-1}d\gamma, \gamma^{-1}\psi) = (\mathcal{A}, \psi)$$

and this, in turn, is the case if and only if

$$\gamma^{-1}\psi = \psi \text{ and } 2\gamma^{-1}d\gamma = 0.$$

Since $\gamma \neq 1$, the first of these can be true if and only if $\psi \equiv 0$. The second implies $d\gamma = 0$ and, since B is connected, γ must be constant. ■

$(\mathcal{A}, \psi) \in \mathcal{A}(\mathcal{L})$ is said to be **reducible** if $\psi \equiv 0$ and **irreducible** otherwise.

As was the case in Donaldson theory, the configuration space $\mathcal{A}(\mathcal{L})$ is an affine space and therefore an infinite-dimensional manifold. The tangent space at any $(\mathcal{A}, \psi) \in \mathcal{A}(\mathcal{L})$ can be identified with

$$T_{(\mathcal{A}, \psi)}(\mathcal{A}(\mathcal{L})) = \Omega^1(B, \mathrm{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^+(\mathcal{L})) \quad (7.115)$$

(no adjoint bundle required in the first summand because $\mathrm{U}(1)$ is abelian). The gauge group $\mathcal{G}(\mathcal{L})$ has the structure of a Hilbert Lie group whose Lie algebra can be identified with $\Omega^0(B, \mathrm{Im}\mathbb{C})$. Fixing $(\mathcal{A}, \psi) \in \mathcal{A}(\mathcal{L})$, the action of $\mathcal{G}(\mathcal{L})$ on (\mathcal{A}, ψ) gives a map $\mathcal{G}(\mathcal{L}) \rightarrow \mathcal{A}(\mathcal{L})$ whose derivative at the identity is

$$\begin{aligned} \Omega^0(B, \mathrm{Im}\mathbb{C}) &\longrightarrow \Omega^1(B, \mathrm{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^+(\mathcal{L})) \\ a &\longrightarrow (2da, -a \cdot \psi) \end{aligned} \quad (7.116)$$

where $a \cdot \psi$ is the rotation of ψ obtained by regarding each value of a as an element of the complexified Clifford algebra and Clifford multiplying by a .

Remark. *Here is a formal, i.e., C^∞ , local argument to persuade you of (7.116). The map is $\gamma \in \mathcal{G}(\mathcal{L}) \rightarrow (\mathcal{A}, \psi) \cdot \gamma = (\mathcal{A} + 2\gamma^{-1}d\gamma, \gamma^{-1}\psi)$. Let $a \in \Omega^0(B, \mathrm{Im}\mathbb{C})$ and write $a(x) = i\theta(x)$, where θ is a real-valued function on B . Then $\alpha(t) = e^{it\theta(x)}$ is a curve in $\mathcal{G}(\mathcal{L}) = C^\infty(B, \mathrm{U}(1))$ with $\alpha(0) = 1$ and $\alpha'(0) = a$. Thus, the derivative at $t = 0$ is*

$$\begin{aligned} \frac{d}{dt}((\mathcal{A}, \psi) \cdot e^{it\theta(x)})|_{t=0} &= \frac{d}{dt}(\mathcal{A} + 2itd\theta, e^{-it\theta(x)}\psi)|_{t=0} \\ &= (2id\theta, -i\theta(x)\psi) \\ &= (2da, -a\psi) \\ &= (2da, -a \cdot \psi). \end{aligned}$$

Again as in Donaldson theory, the “large” moduli space

$$\mathcal{B}(\mathcal{L}) = \mathcal{A}(\mathcal{L})/\mathcal{G}(\mathcal{L})$$

of configurations is a smooth Banach manifold away from the reducible configurations (i.e., away from those $[\mathbf{A}, \psi]$ with $\psi \equiv 0$) and the monopole moduli space $\mathcal{M}(\mathcal{L})$ is a subset of it. Define the **Seiberg-Witten map**

$$F : \mathcal{A}(\mathcal{L}) \longrightarrow \Omega_+^2(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^-(\mathcal{L}))$$

by

$$F(\mathbf{A}, \psi) = (\mathbf{F}_\mathbf{A}^+ - \sigma^+(\psi \otimes \psi^*)_0, \not{D}_\mathbf{A}\psi). \quad (7.117)$$

Then (\mathbf{A}, ψ) satisfies (SW) if and only if $F(\mathbf{A}, \psi) = (0, 0)$. The derivative of F at $(\mathbf{A}, \psi) \in \mathcal{M}(\mathcal{L})$, i.e., the linearization of the Seiberg-Witten equations at (\mathbf{A}, ψ) , is

$$F_{*(\mathbf{A}, \psi)} : \Omega^1(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^+(\mathcal{L})) \longrightarrow \Omega_+^2(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^-(\mathcal{L})) \quad (7.118)$$

$$F_{*(\mathbf{A}, \psi)} = \begin{pmatrix} d^+ & -D_\psi \\ \frac{1}{2}\psi & \not{D}_\mathbf{A} \end{pmatrix}$$

where

$$d^+ : \Omega^1(B, \text{Im}\mathbb{C}) \longrightarrow \Omega_+^2(B, \text{Im}\mathbb{C})$$

is “d” followed by the projection onto the self-dual part,

$$\frac{1}{2}\psi : \Omega^1(B, \text{Im}\mathbb{C}) \longrightarrow \Gamma(\mathcal{S}^-(\mathcal{L}))$$

takes a one-form α to $\alpha \cdot \frac{1}{2}\psi$ (which is understood to mean Clifford multiplication by the vector field dual to the one-form α) and

$$D_\psi : \Gamma(\mathcal{S}^+(\mathcal{L})) \longrightarrow \Omega_+^2(B, \text{Im}\mathbb{C})$$

is given by

$$D_\psi(\eta) = \sigma^+ \left(\psi \otimes \eta^* + \eta \otimes \psi^* - \frac{1}{2}(\langle \psi, \eta \rangle + \overline{\langle \psi, \eta \rangle}) \mathbb{1} \right)$$

(the object inside the parentheses being a section of $\text{End}_0(\mathcal{S}^+(\mathcal{L}))$ which σ^+ identifies with a self-dual two-form on B).

Remark. (7.118) can be verified with a local argument analogous to that for (7.116).

Associated to any solution (\mathbf{A}, ψ) to (SW) is a **fundamental elliptic complex** $\mathcal{E}(\mathbf{A}, \psi)$

$$\begin{aligned} 0 \longrightarrow \Omega^0(B, \text{Im}\mathbb{C}) &\longrightarrow \Omega^1(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^+(\mathcal{L})) \\ &\longrightarrow \Omega_+^2(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^-(\mathcal{L})) \longrightarrow 0 \end{aligned}$$

where the second and third maps are, respectively, the derivative (7.116) of the action of $\mathcal{G}(\mathcal{L})$ on (\mathbf{A}, ψ) , and the derivative (7.118) of the Seiberg-Witten map at (\mathbf{A}, ψ) . This complex has finite-dimensional cohomology groups $H^i(\mathbf{A}, \psi)$, $i = 0, 1, 2$, which admit interpretations analogous to those in Donaldson theory: $H^0(\mathbf{A}, \psi) =$ tangent space to the stabilizer of (\mathbf{A}, ψ) in $\mathcal{G}(\mathcal{L})$ so

$$\begin{aligned} H^0(\mathbf{A}, \psi) = 0 &\iff (\mathbf{A}, \psi) \text{ irreducible} \\ &\iff \psi \neq 0 \end{aligned}$$

$H^1(\mathbf{A}, \psi) =$ formal tangent space to $\mathcal{M}(\mathcal{L})$ at $[\mathbf{A}, \psi]$

$H^2(\mathbf{A}, \psi) =$ obstruction space, i.e.,

$$\begin{aligned} H^2(\mathbf{A}, \psi) = 0 &\iff \text{Implicit Function Theorem} \\ &\text{gives a local manifold} \\ &\text{structure for } F^{-1}(0, 0) \text{ near} \\ &(\mathbf{A}, \psi) \text{ of dimension } \dim H^1(\mathbf{A}, \psi). \end{aligned}$$

If $H^2(\mathbf{A}, \psi) = 0$ and $H^0(\mathbf{A}, \psi) = 0$, then the local manifold structure for $F^{-1}(0, 0)$ near (\mathbf{A}, ψ) projects injectively into the moduli space $\mathcal{M}(\mathcal{L})$ and, near $[\mathbf{A}, \psi]$, $\mathcal{M}(\mathcal{L})$ is a smooth manifold of dimension

$$\begin{aligned} \dim H^1(\mathbf{A}, \psi) &= -\dim H^0(\mathbf{A}, \psi) + \dim H^1(\mathbf{A}, \psi) - \dim H^2(\mathbf{A}, \psi) \\ &= -\text{Index}(\mathcal{E}(\mathbf{A}, \psi)) \\ &= \frac{1}{4}(c_1(L^0)^2 - 2\chi(B) - 3\sigma(B)) \end{aligned}$$

where the last expression comes from the Atiyah-Singer Index Theorem and $c_1(L^0)^2$ means $\int_B c_1(L^0) \wedge c_1(L^0)$.

As in Donaldson theory, $H^0(\mathbf{A}, \psi) = 0$ and $H^2(\mathbf{A}, \psi) = 0$ are the “generic” situation, but this means something slightly different here. First, the part that is the same.

Theorem 7.8 (Generic Metrics Theorem) *Let B denote a compact, connected, simply connected, oriented, smooth four-manifold with $b_2^+(B) > 0$. Then there is a dense G_δ -subset $\text{Gen}(\mathcal{R})$ of the space $\mathcal{R}(B)$ of Riemannian metrics on B with the following property: For any $\mathbf{g} \in \text{Gen}(\mathcal{R})$ and any corresponding spin^c structure \mathcal{L} , any solution (\mathbf{A}, ψ) to the Seiberg-Witten equations is irreducible, i.e., satisfies $H^0(\mathbf{A}, \psi) = 0$. If $b_2^+(B) > 1$, then for any generic path $\mathbf{g}(t)$, $0 \leq t \leq 1$, of Riemannian metrics in $\mathcal{R}(B)$ there are no reducible solutions to the Seiberg-Witten equations for any spin^c structure corresponding to any of the metrics $\mathbf{g}(t)$, $0 \leq t \leq 1$.*

For $H^2(\mathbf{A}, \psi) = 0$ there is no known generic metrics theorem of this sort. In this case one must perturb, not the metric, but the equations themselves. More precisely, we fix the metric \mathbf{g} and the spin^c structure \mathcal{L} . For any fixed $\boldsymbol{\eta} \in \Omega_+^2(B, \text{Im}\mathbb{C})$ we introduce the $(\boldsymbol{\eta})$ **perturbed Seiberg-Witten (PSW) equations**

$$\mathcal{D}_{\mathbf{A}}\psi = 0 \tag{7.119}$$

$$\mathbf{F}_{\mathbf{A}}^+ = \sigma^+(\psi \otimes \psi^*)_0 + \boldsymbol{\eta}. \tag{7.120}$$

Remark. *The motivation here is easy to understand. Solutions to (SW) are solutions to the equation $F(\mathbf{A}, \psi) = (0, 0)$, where F is the Seiberg-Witten map (7.117). For this to be a manifold, $(0, 0)$ must be a regular value of F . If it is not, the Sard-Smale Theorem suggests that a small perturbation of $(0, 0)$ in $\Omega_+^2(B, \text{Im}\mathbb{C}) \oplus \Gamma(\mathcal{S}^-(\mathcal{L}))$ of the form $(\boldsymbol{\eta}, 0)$ will be a regular value so that $F(\mathbf{A}, \psi) = (\boldsymbol{\eta}, 0)$ will define a manifold of (\mathbf{A}, ψ) . But $F(\mathbf{A}, \psi) = (\boldsymbol{\eta}, 0)$ is just (PSW).*

The linearized complex at any solution to (PSW) is given by the same maps as for (SW) so the cohomology is the same. $\mathcal{G}(\mathcal{L})$ acts on solutions to (PSW) in the same way so there is a moduli space $\mathcal{M}(\mathcal{L}, \boldsymbol{\eta})$ of solutions and everything we have said above for (SW) is also true for (PSW).

Theorem 7.9 (Generic Perturbations Theorem) *Let B denote a compact, connected, simply connected, oriented, smooth four-manifold. Fix a Riemannian metric \mathbf{g} and a spin^c structure \mathcal{L} for B . Then there is a dense G_δ -subset $\text{Gen}(\Omega_+^2)$ in the space $\Omega_+^2(B, \text{Im}\mathbb{C})$ of $\text{Im}\mathbb{C}$ -valued self-dual two-forms on B with the following properties: For $\boldsymbol{\eta} \in \text{Gen}(\Omega_+^2)$, every solution (\mathbf{A}, ψ) to the perturbed Seiberg-Witten equations (7.119) and (7.120) has $H^2(\mathbf{A}, \psi) = 0$. If $b_2^+(B) > 0$ and $\mathbf{g} \in \text{Gen}(\mathcal{R})$, then, for any $\boldsymbol{\eta} \in \text{Gen}(\Omega_+^2)$, the moduli space $\mathcal{M}(\mathcal{L}, \boldsymbol{\eta})$ is a smooth submanifold of $\mathcal{B}(\mathcal{L})$ of dimension $\frac{1}{4}(c_1(L^0)^2 - 2\chi(B) - 3\sigma(B))$.*

Remark. *In particular, if $c_1(L^0)^2 - 2\chi(B) - 3\sigma(B) < 0$, then the moduli space is generically empty.*

Exactly as in the case of Donaldson theory one can show that, for a fixed metric \mathbf{g} and any associated spin^c structure \mathcal{L} , a choice of orientation for the vector space $H_+^2(B, \mathbb{R})$ canonically orients all of the moduli spaces $\mathcal{M}(\mathcal{L}, \boldsymbol{\eta})$. Likewise as in Donaldson theory, when $b_2^+(B) > 1$ there is a cobordism result which roughly says that for a generic one-parameter family $\mathbf{g}(t)$, $0 \leq t \leq 1$, of metrics and a generic one-parameter family $\boldsymbol{\eta}(t)$, $0 \leq t \leq 1$, of perturbations, the moduli spaces parametrized by t fit together to form a smooth manifold with boundary containing no points corresponding to reducible solutions. The boundary is the disjoint union of moduli spaces for $(\mathbf{g}(0), \boldsymbol{\eta}(0))$ and $(\mathbf{g}(1), \boldsymbol{\eta}(1))$. Moreover, selecting an orientation for $H_+^2(B, \mathbb{R})$ orients this parametrized moduli space and the two boundary moduli spaces inherit opposite orientations.

Remark. *There is a technical point which we glossed over here and should mention because it has no analogue in Donaldson theory. Changing the metric \mathbf{g} changes the orthonormal frame bundle and so, one would think, the spin^c structure. It would appear that the discussion above is incomplete without a specification of a spin^c structure for each t . In fact, however, one can show that frame bundles for different metrics are naturally isomorphic and so one can pull back spin^c structures by the isomorphisms, thus effectively “fixing” \mathcal{L} (up to equivalence) regardless of the choice of \mathbf{g} .*

Except for a few minor simplifications and adaptations the story of the Seiberg-Witten moduli space thus far has been virtually indistinguishable from what we had to say about the anti-self-dual moduli space in Section 4. The one aspect of Seiberg-Witten theory that differs significantly from Donaldson theory (and that accounts for its relative simplicity) is that there is no need for an “Uhlenbeck-style compactification”:

For any metric \mathbf{g} , and spin^c structure \mathcal{L} and any perturbation $\boldsymbol{\eta}$, the moduli space $\mathcal{M}(\mathcal{L}, \boldsymbol{\eta})$ is always compact.

The proof of this involves what is called an elliptic “bootstrapping” argument (which we will not describe) based on the crucial fact that the spinor field ψ and curvature F_A for any solution (A, ψ) to (PSW) satisfy **uniform a priori bounds** (this is categorically false for the anti-self-dual equations because these are conformally invariant in dimension four). Because of its significance we will sketch a proof of this but, since the perturbation adds only arithmetic to the argument, we

will do this for the unperturbed equations, written in the form

$$\mathcal{D}_{\mathbf{A}}\psi = 0 \quad (7.121)$$

$$\rho^+(\mathbf{F}_{\mathbf{A}}^+) = (\psi \otimes \psi^*)_0 \quad (7.122)$$

(recall that ρ^+ is the inverse of σ^+). We will appeal to the famous *Weitzenböck formula* from differential geometry which, in our present circumstances, reads

$$\mathcal{D}_{\mathbf{A}}^* \circ \mathcal{D}_{\mathbf{A}}\psi = \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi + \frac{1}{4}\kappa\psi + \rho^+(\mathbf{F}_{\mathbf{A}}^+)\psi \quad (7.123)$$

where $D_{\mathbf{A}}^*$ is the formal adjoint of $\mathcal{D}_{\mathbf{A}} : \Gamma(\mathcal{S}^+(\mathcal{L})) \rightarrow \Gamma(\mathcal{S}^-(\mathcal{L}))$, $\nabla_{\mathbf{A}}^*$ is the formal adjoint of the covariant derivative $\nabla_{\mathbf{A}} : \Gamma(\mathcal{S}^+(\mathcal{L})) \rightarrow \Omega^1(B, \mathcal{S}^+(\mathcal{L}))$ and κ is the scalar curvature of B (for the metric \mathbf{g}). Because (\mathbf{A}, ψ) is a solution to the Seiberg-Witten equations, (7.123) reduces to

$$\nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi + \frac{1}{4}\kappa\psi + (\psi \otimes \psi^*)_0\psi = 0. \quad (7.124)$$

Take the pointwise inner product with ψ to obtain

$$\langle \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi, \psi \rangle + \frac{1}{4}\kappa\|\psi\|^2 + \frac{1}{2}\|\psi\|^4 = 0 \quad (7.125)$$

(for the last term compute $(\psi \otimes \psi^*)_0\psi$ from (7.75) and then take the inner product with ψ).

Now, $\|\psi(x)\|$ is a continuous function on the compact space B so there is an $x_0 \in B$ at which it achieves an absolute maximum. We claim that, at this point, the first term in (7.125) is non-negative, i.e.,

$$\langle \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi(x_0), \psi(x_0) \rangle \geq 0 \quad (7.126)$$

(note that (7.125) implies that $\langle \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi, \psi \rangle$ must be real). The proof depends on the identity

$$\Delta_{\mathbf{g}}\|\psi\|^2 = -2\|\nabla_{\mathbf{A}}\psi\|^2 + 2\langle \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi, \psi \rangle \quad (7.127)$$

where $\Delta_{\mathbf{g}} = d^* \circ d$ is the scalar Laplacian corresponding to \mathbf{g} . This can be verified by writing out $\Delta_{\mathbf{g}}\|\psi\|^2$ in a local orthonormal frame field. Now, at x_0 , (7.127) gives

$$2\langle \nabla_{\mathbf{A}}^* \circ \nabla_{\mathbf{A}}\psi(x_0), \psi(x_0) \rangle = \Delta_{\mathbf{g}}\|\psi\|^2(x_0) + 2\|\nabla_{\mathbf{A}}\psi\|^2(x_0). \quad (7.128)$$

Now, obviously $2\|\nabla_{\mathbf{A}}\psi\|^2(x_0) \geq 0$. Moreover, since $\|\psi\|^2$ achieves a maximum at x_0 , $\Delta_{\mathbf{g}}\|\psi\|^2(x_0) \geq 0$ as well (in local \mathbf{g} -orthonormal coordinates $\Delta_{\mathbf{g}}$ is just

$-\sum_{i=1}^4 \frac{\partial^2}{(\partial x^i)^2}$) so (7.126) is proved. It then follows from (7.125), evaluated at x_0 , that

$$\frac{1}{4}\kappa(x_0)\|\psi(x_0)\|^2 + \frac{1}{2}\|\psi(x_0)\|^4 \leq 0$$

so

$$\|\psi(x_0)\|^4 \leq -\frac{1}{2}\kappa(x_0)\|\psi(x_0)\|^2.$$

There are two possibilities. Either $\|\psi(x_0)\| = 0$, in which case $\psi \equiv 0$ and (\mathbf{A}, ψ) is a reducible solution (and so $\|\psi\|$ is certainly uniformly bounded). Otherwise we have

$$\|\psi(x_0)\|^2 \leq -\frac{1}{2}\kappa(x_0)$$

and therefore

$$\|\psi(x)\|^2 \leq -\frac{1}{2}\kappa(x_0) \tag{7.129}$$

for every x in B . Now, despite appearances, the right-hand side of (7.129) depends on ψ (through x_0) so, to get a bound on the spinor field of every solution (\mathbf{A}, ψ) we define $k(x_0) = \max\{-\frac{1}{2}\kappa(x_0), 0\}$ for each $x_0 \in B$ and

$$k(B) = \max\{k(x_0); x_0 \in B\}$$

and conclude that for any fixed metric and any spin^c structure, any solution (\mathbf{A}, ψ) to the Seiberg-Witten equations has spinor field ψ bounded by the geometrical constant $k(B)$

$$\|\psi(x)\|^2 \leq k(B) \quad \forall x \in B. \tag{7.130}$$

The second Seiberg-Witten equation (7.122) then gives a uniform bound on the self-dual part of the curvature for any solution. A bit more work then gives a bound on the anti-self-dual part of the curvature. From these one can deduce that, for a given \mathfrak{g} (and \mathfrak{h} in the perturbed case) there are at most finitely many (equivalence classes of) spin^c structures for which the moduli space is nonempty (Theorem 5.2.4 of [36]).

The bounds described thus far are not sufficient to prove the compactness of the moduli space. For this one must bound the connection parts \mathbf{A} of solutions (\mathbf{A}, ψ) “up to gauge”. This is generally accomplished by a gauge fixing argument and the bootstrapping referred to above (see Section 5.3 of [36]). In any case, it can be done and the end result is that Seiberg-Witten moduli spaces are *always* compact (and, generically, are smooth manifolds). This compactness simplifies enormously the task of defining and computing “Donaldson-like” invariants in Seiberg-Witten theory (because there is no need, as there is in the anti-self-dual

case, to compactify the moduli space before integrating cohomology classes over it). Even so we intend to consider what appears to be only a very special case. Thus, we fix a generic metric g and perturbation η and suppose that there exists a spin^c structure \mathcal{L} for which the formal dimension of the moduli space is zero, i.e.,

$$c_1(L^0)^2 = 2\chi(B) + 3\sigma(B) \quad (7.131)$$

(such an \mathcal{L} need not exist). Assuming that an orientation for the vector space $H_+^2(B, \mathbb{R})$ has been fixed, the moduli space is a finite set of isolated points each of which is equipped with a sign ± 1 . The sum of these signs is an integer and, as in Donaldson theory, when $b_2^+(B) > 1$ a cobordism argument shows that the integer is independent of the choice of (generic) metric and perturbation. We call this integer the **zero-dimensional Seiberg-Witten invariant of B associated with \mathcal{L}** and denote it

$$SW_0(B, \mathcal{L}).$$

This is, indeed, an invariant in the sense that, if $f : B' \rightarrow B$ is a diffeomorphism for which the induced map $f^* : H_+^2(B, \mathbb{R}) \rightarrow H_+^2(B', \mathbb{R})$ preserves orientation, then the induced spin^c structure $f^*\mathcal{L}$ for B' also satisfies (7.131) and $SW_0(B', f^*\mathcal{L}) = SW_0(B, \mathcal{L})$.

Remark. *When one goes to the trouble of defining the Seiberg-Witten invariant even when (7.131) is not satisfied (which we will not) one obtains a map $SW(B, \cdot)$ on the set of (equivalence classes of) spin^c structures on B which assigns to each such an integer $SW(B, \mathcal{L}) \in \mathbb{Z}$, taken to be zero if the corresponding moduli space is empty. We have already noted that there can be at most finitely many \mathcal{L} for which $SW(B, \mathcal{L}) \neq 0$. The empirical evidence suggests that, when $b_2^+(B) > 1$, $SW(B, \mathcal{L}) \neq 0$ occurs only for those \mathcal{L} satisfying (7.131). We will say that a B with $b_2^+(B) > 1$ is of **SW-simple type** if $SW(B, \mathcal{L}) \neq 0$ implies that \mathcal{L} must satisfy (7.131), i.e., if nonzero Seiberg-Witten invariants occur only for zero-dimensional moduli spaces. It has been conjectured that every B with $b_2^+(B) > 1$ is of SW-simple type. Finally, we shall refer to the elements of $H^2(B, \mathbb{Z})$ which arise as $c_1(L^0)$ for some spin^c structure \mathcal{L} satisfying (7.131) as **Seiberg-Witten (SW-) basic classes**. Thus, SW-basic classes are just the first Chern classes of complex line bundles corresponding to spin^c structures for which the SW moduli space is zero-dimensional.*

It is not our wish here to relate, once again, the story of the immediate and spectacular impact had on the study of smooth four-manifolds by the introduction of the Seiberg-Witten invariants. To hear, from one of the principal players, about the first few intense, exciting weeks one should begin with [48] and then proceed

to [36] for an outline of these and subsequent developments. Our concern here is with the role of Seiberg-Witten theory as a “dual” version of Donaldson theory and the extraordinary conjecture to which Witten was led by this duality.

8. The Witten Conjecture

The duality symmetry of Witten’s topological quantum field theory led not just from the Donaldson to the Seiberg-Witten invariants, but to a very explicit conjecture by Witten regarding the relationship between them. This conjecture, called “Witten’s magical formula” by Taubes, represents the conclusion of our story. Roughly, the conjecture asserts that, for certain four-manifolds, the zero-dimensional Seiberg-Witten invariants contain all of the information available in all of the Donaldson invariants (zero-dimensional or not). Now, of course, we have introduced only the zero-dimensional Donaldson invariant here and, although it is possible (using a result of Kronheimer and Mrowka [30]) to state the conjecture quite precisely without any real understanding of where the higher dimensional Donaldson invariants come from, this seems rather disingenuous. We would therefore like to offer just some crude sense of the enormous amount of labor that goes into merely the definition of these invariants. For this we will present an outline of the construction with absolutely no details and scarcely a word of explanation. At best this will provide only a schematic for those who seek out the details in other sources ([37] contains some of this detail, but leaves enough out that one can still see the forest for the trees, while essentially the whole story is available in [15]).

Let us begin then by establishing some notation. B will denote a compact, connected, simply connected, oriented, smooth four-manifold with $b_2^+(B) > 1$ and odd (this last condition will be explained shortly). For each $k > 0$ we will write $SU(2) \hookrightarrow P_k \xrightarrow{\pi_k} B$ for the principal $SU(2)$ -bundle over B with Chern class k . $\mathcal{A}(P_k)$ is the affine space of connections on P_k and $\mathcal{G}(P_k)$ is the group of all gauge transformations. The corresponding moduli space is $\mathcal{B}(P_k) = \mathcal{A}(P_k)/\mathcal{G}(P_k)$. The space of irreducible connections will be denoted $\hat{\mathcal{A}}(P_k)$ and the corresponding moduli space is $\hat{\mathcal{B}}(P_k) = \hat{\mathcal{A}}(P_k)/\mathcal{G}(P_k)$.

Remark. *For our present purposes we will take no heed in the notation of the appropriate Sobolev indices.*

We fix once and for all an orientation for the vector space $H_+^2(B, \mathbb{R})$. Now select a generic Riemannian metric \mathbf{g} for B and consider the moduli space $\mathcal{M}(P_k, \mathbf{g})$ of (irreducible) \mathbf{g} -ASD connections on P_k . We have seen that $\mathcal{M}(P_k, \mathbf{g})$ is either

empty or a smooth, orientable manifold of dimension

$$8k - 3(1 + b_2^+(B)). \quad (8.1)$$

Note that our assumption that $b_2^+(B)$ is odd ensures that this is even so we can define

$$d_k = 4k - \frac{3}{2}(1 + b_2^+(B)) \quad (8.2)$$

and be sure that it is an integer.

We have already seen that, when (8.1) is negative the moduli space is generically empty and when it is zero, so that the moduli space is a finite set of signed points, one can define a differential topological invariant of B by adding the signs. We mentioned also that, when (8.1) is positive one would like to define an invariant by, very roughly, “integrating certain carefully chosen cohomology classes over $\mathcal{M}(P_k, \mathbf{g})$.” Viewed from the physical perspective this amounts to computing expectation values for certain carefully chosen observables in Witten’s field theory. The main obstacle to this program is that, when (8.1) is positive, the moduli space is never compact. Thus, the first step in Donaldson’s program of defining new invariants is the construction of a compactification of $\mathcal{M}(P_k, \mathbf{g})$. This is an arduous task and relies heavily on deep analytical results due to Karen Uhlenbeck [49], but, in the end, it can be done and the result is called the **Uhlenbeck compactification** $\bar{\mathcal{M}}(P_k, \mathbf{g})$ of $\mathcal{M}(P_k, \mathbf{g})$. Regrettably, $\bar{\mathcal{M}}(P_k, \mathbf{g})$ is generally not a manifold, but only what is known as a stratified space. As a result, the “integration” of cohomology classes to which we just referred will, in fact, be replaced by the natural pairing of these cohomology classes with a fundamental homology class $[\bar{\mathcal{M}}(P_k, \mathbf{g})]$ of the compactified moduli space (which, unfortunately, exists only under certain restrictions on k that we will describe).

To obtain the “carefully chosen cohomology classes” mentioned above Donaldson constructs a map

$$\mu : H_2(B, \mathbb{Z}) \longrightarrow H^2(\mathcal{M}(P_k, \mathbf{g}), \mathbb{Z}) \quad (8.3)$$

from the second homology of B to the second cohomology of $\mathcal{M}(P_k, \mathbf{g})$. The construction involves a certain auxiliary $\text{SO}(3)$ -bundle over $\hat{B}(P_k) \times B$ and its first Pontryagin class. Donaldson shows that his μ -map (8.3) extends to the Uhlenbeck compactification in the sense that there is a unique map

$$\bar{\mu} : H_2(B, \mathbb{Z}) \longrightarrow H^2(\bar{\mathcal{M}}(P_k, \mathbf{g}), \mathbb{Z}) \quad (8.4)$$

which, when followed by the restriction mapping $H^2(\bar{\mathcal{M}}(P_k, \mathbf{g}), \mathbb{Z}) \rightarrow H^2(\mathcal{M}(P_k, \mathbf{g}), \mathbb{Z})$ is equal to μ . Now, since $2d_k$ is the dimension of $\mathcal{M}(P_k, \mathbf{g})$, if we select

$$(x_1, \dots, x_{d_k}) \in H_2(B, \mathbb{Z}) \times \overset{d_k}{\cdot} \times H_2(B, \mathbb{Z})$$

then $\mu(x_1) \smile \dots \smile \mu(x_{d_k}) \in H^{2d_k}(\mathcal{M}(P_k, \mathfrak{g}), \mathbb{Z})$, where \smile denotes the cup product. Intuitively, one can think of this as a top rank “form” that one might like to integrate over $\mathcal{M}(P_k, \mathfrak{g})$. More precisely, we would like to pair $\bar{\mu}(x_1) \smile \dots \smile \bar{\mu}(x_{d_k})$ with the fundamental class $[\bar{\mathcal{M}}(P_k, \mathfrak{g})]$. From the algebraic topology of stratified spaces, however, one finds that $\bar{\mathcal{M}}(P_k, \mathfrak{g})$ admits a fundamental class only if k is in what is called the **stable range** of B , i.e.,

$$k > \frac{3}{4}(1 + b_2^+(B)) \quad (8.5)$$

or, equivalently,

$$d_k > \frac{3}{2}(1 + b_2^+(B)). \quad (8.6)$$

For these d_k at least we are now in a position to define an invariant. Specifically, for any integer d satisfying

$$d \equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{4} \quad (8.7)$$

and

$$d > \frac{3}{2}(1 + b_2^+(B)) \quad (8.8)$$

we define a map

$$\gamma_d(B) : H_2(B, \mathbb{Z}) \times \overset{d}{.} \times H_2(B, \mathbb{Z}) \longrightarrow \mathbb{Z} \quad (8.9)$$

as follows. By (8.7) we can find a k such that $d = 4k - \frac{3}{2}(1 + b_2^+(B))$ and then consider the principal bundle $SU(2) \hookrightarrow P_k \xrightarrow{\pi_k} B$ with Chern class k . Then $d = d_k$ and (8.6) is satisfied by (8.8). Thus, if $\mathcal{M}(P_k, \mathfrak{g}) \neq \emptyset$, $\bar{\mathcal{M}}(P_k, \mathfrak{g})$ admits a fundamental homology class $[\bar{\mathcal{M}}(P_k, \mathfrak{g})]$ and we may define $\gamma_d(B)$ in (8.9) to be zero if $\mathcal{M}(P_k, \mathfrak{g}) = \emptyset$ and otherwise

$$\gamma_d(B)(x_1, \dots, x_d) = \langle \bar{\mu}(x_1) \smile \dots \smile \bar{\mu}(x_d), [\bar{\mathcal{M}}(P_k, \mathfrak{g})] \rangle \quad (8.10)$$

where $\langle \cdot, \cdot \rangle$ denotes the natural pairing of cohomology and homology. One thinks of the pairing heuristically as the “integration” of the top rank “form” $\bar{\mu}(x_1) \smile \dots \smile \bar{\mu}(x_d)$ over $\bar{\mathcal{M}}(P_k, \mathfrak{g})$ and, under favorable circumstances, (8.10) can actually be computed as such an integral. For these reasons, and to make contact with the notation used in physics, one often sees (8.10) written as

$$\gamma_d(B)(x_1, \dots, x_d) = \int_{\bar{\mathcal{M}}(P_k, \mathfrak{g})} \bar{\mu}(x_1) \wedge \dots \wedge \bar{\mu}(x_d).$$

Under the assumptions we have made on B , $\gamma_d(B)$ is a d -multilinear, integer-valued function on $H_2(B, \mathbb{Z})$ which does not depend on the choice of the generic metric \mathbf{g} and is, moreover, invariant under orientation preserving diffeomorphisms of B that also preserve the orientation of $H_+^2(B, \mathbb{R})$. Notice that the multilinear function (8.9) is uniquely determined (via polarization) by the corresponding polynomial on $H_2(B, \mathbb{Z})$ for which we use the same symbol:

$$\begin{aligned} \gamma_d(B) : H_2(B, \mathbb{Z}) &\longrightarrow \mathbb{Z} \\ \gamma_d(B)(x) &= \gamma_d(B)(x, \dots, x). \end{aligned} \tag{8.11}$$

Remark. *Just to have one example available we mention the Kummer surface $K3$ (see the Remark just before Donaldson's 1983 Theorem in Section 1). $K3$ has a natural orientation as a complex manifold and $b_2^+(K3) = 3$ (which is greater than 1 and odd). The stable range of $K3$ is therefore given by $k > \frac{3}{4}(1+3)$, i.e., $k > 3$, and $d_k = 4k - \frac{3}{2}(1+3) = 2(2k-3)$. Choosing one of the two possible orientations for $H_+^2(K3, \mathbb{R})$ it has been shown ([42] or [30]) that the Donaldson polynomial of degree $d = d_k = 2n$ with $n (= 2k-3)$ odd and greater than 3 can be described as follows.*

$$\gamma_d(K3)(x) = \gamma_{2n}(K3)(x) = \frac{(2n)!}{2^n n!} (Q_{K3}(x))^n$$

for every $x \in H_2(K3, \mathbb{Z})$, where Q_{K3} is the intersection form of $K3$ and $Q_{K3}(x) = Q_{K3}(x, x)$ is regarded as a quadratic function on $H_2(K3, \mathbb{Z})$.

Our next objectives are to eliminate the stable range assumption (8.8) and relax the restriction (8.7). The first step is to extend the definition (8.9) of $\gamma_d(B)$ to allow for factors of $H_0(B, \mathbb{Z})$ as well. For this one requires another “ μ -map” defined on $H_0(B, \mathbb{Z})$ for which we will use the same symbol

$$\mu : H_0(B, \mathbb{Z}) \longrightarrow H^4(\mathcal{M}(P_k, \mathbf{g}), \mathbb{Z}). \tag{8.12}$$

Since B is assumed connected, $H_0(B, \mathbb{Z}) \cong \mathbb{Z}$ and we will denote a generator simply by ν . As for (8.3), the definition of $\mu(\nu)$ involves the first Pontryagin class of the auxiliary $SO(3)$ -bundle mentioned earlier. Unfortunately, in this case $\mu(\nu)$ does not extend over the entire Uhlenbeck compactification and this introduces technical difficulties. The class $\mu(\nu)$ does, however, admit an extension $\bar{\mu}(\nu)$ to a large enough subset of $\mathcal{M}(P_k, \mathbf{g})$ that, under certain restrictions, one can produce the desired extension of $\gamma_d(B)$ to include this four-dimensional class. Basically, the restriction is that there must be enough two-dimensional classes to satisfy

the stable range condition (8.8) by themselves. Somewhat more precisely, we consider an integer $d > 0$ satisfying (8.7) and written in the form

$$d = a + 2b \quad (8.13)$$

where b is a non-negative integer and a is an integer satisfying

$$a > \frac{3}{2}(1 + b_2^+(B)). \quad (8.14)$$

Then one can define (still using the same symbol)

$$\gamma_d(B) : H_2(B, \mathbb{Z}) \times \dots \times H_2(B, \mathbb{Z}) \times H_0(B, \mathbb{Z}) \times \dots \times H_0(B, \mathbb{Z}) \longrightarrow \mathbb{Z} \quad (8.15)$$

by

$$\begin{aligned} \gamma_d(B)(x_1, \dots, x_a, n_1\nu, \dots, n_b\nu) &= n_1 \dots n_b \langle \bar{\mu}(x_1) \smile \dots \smile \bar{\mu}(x_a) \smile \bar{\mu}(\nu)^b, [\bar{\mathcal{M}}(P_k, \mathbf{g})] \rangle \\ &= n_1 \dots n_b \int_{\bar{\mathcal{M}}(P_k, \mathbf{g})} \bar{\mu}(x_1) \wedge \dots \wedge \bar{\mu}(x_a) \wedge \bar{\mu}(\nu)^b \end{aligned} \quad (8.16)$$

when $\mathcal{M}(P_k, \mathbf{g}) \neq \emptyset$ and $\gamma_d(B) \equiv 0$ when $\mathcal{M}(P_k, \mathbf{g}) = \emptyset$ (of course, k is again the integer for which $d = d_k = 4k - \frac{3}{2}(1 + b_2^+(B))$). For any integer $k > 0$ we will say that a sequence

$$s = (x_1, \dots, x_a, n_1\nu, \dots, n_b\nu)$$

with $x_i \in H_2(B, \mathbb{Z})$, $i = 1, \dots, a$, and $n_j\nu \in H_0(B, \mathbb{Z})$ is **k -stable for B** if $a + 2b = d_k = 4k - \frac{3}{2}(1 + b_2^+(B))$, $b \geq 0$ and $a > \frac{3}{2}(1 + b_2^+(B))$. At this point we have defined $\gamma_d(B)(s)$ whenever s is k -stable for B and $d = a + 2b$.

The key to eliminating the stable range assumption in the definition of the Donaldson invariants is a “blow-up formula” which we now describe. Recall that a blow-up of a four-manifold B is the four-manifold $B \# \overline{\mathbb{C}\mathbb{P}^2}$ obtained by forming the connected sum of B and a copy of $\overline{\mathbb{C}\mathbb{P}^2}$. Recall also that $H_2(\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z}) \cong \mathbb{Z}$, generated by a copy of $\overline{\mathbb{C}\mathbb{P}^1} \cong S^2$ in $\overline{\mathbb{C}\mathbb{P}^2}$, and that, relative to such a basis, $Q_{\overline{\mathbb{C}\mathbb{P}^2}} = (-1)$ so $b_2^+(\overline{\mathbb{C}\mathbb{P}^2}) = 0$. Since the intersection form is additive on connected sums we have

$$b_2^+(B \# \overline{\mathbb{C}\mathbb{P}^2}) = b_2^+(B)$$

so, in particular, the stable ranges of B and $B \# \overline{\mathbb{C}\mathbb{P}^2}$ are the same (see (8.8)). Furthermore, since $Q_{\overline{\mathbb{C}\mathbb{P}^2}}$ is negative definite, $H_2^+(\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{R})$ is trivial and this gives

rise to a natural identification of $H_2^+(B\#\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{R})$ and $H_2^+(B, \mathbb{R})$. In particular, orienting the latter also orients the former and this, in turn, orients all of the moduli spaces of ASD connections on bundles over $B\#\overline{\mathbb{C}\mathbb{P}^2}$. We fix such an orientation once and for all. Now, since $H_2(B\#\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z}) \cong H_2(B, \mathbb{Z}) \oplus H_2(\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z})$ we may identify homology classes in either B or $\overline{\mathbb{C}\mathbb{P}^2}$ with homology classes in $B\#\overline{\mathbb{C}\mathbb{P}^2}$. Fix a generator e for $H_2(\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z}) \subseteq H_2(B\#\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z})$. Donaldson has proved each of the following:

1. If s is k -stable for B (and therefore also for $B\#\overline{\mathbb{C}\mathbb{P}^2}$), then

$$\gamma_{d_k}(B\#\overline{\mathbb{C}\mathbb{P}^2})(s) = \gamma_{d_k}(B)(s).$$

2. Suppose $i = 1, 2,$ or 3 and s is not k -stable for B , but (s, e, \dots, e) is k -stable for $B\#\overline{\mathbb{C}\mathbb{P}^2}$. Then

$$\gamma_{d_k}(B\#\overline{\mathbb{C}\mathbb{P}^2})(s, e, \dots, e) = 0.$$

3. If s is k -stable for B (and therefore also for $B\#\overline{\mathbb{C}\mathbb{P}^2}$), then (s, e, e, e, e) is $(k+1)$ -stable for $B\#\overline{\mathbb{C}\mathbb{P}^2}$ and

$$\gamma_{d_k}(B)(s) = -\frac{1}{2}\gamma_{d_{k+1}}(B\#\overline{\mathbb{C}\mathbb{P}^2})(s, e, e, e, e).$$

From these one obtains the following more general result. Let n be a positive integer and consider the n -fold blow-up

$$B\#n\overline{\mathbb{C}\mathbb{P}^2} = B\#\overline{\mathbb{C}\mathbb{P}^2}_1\#\dots\#\overline{\mathbb{C}\mathbb{P}^2}_n$$

(where each $\overline{\mathbb{C}\mathbb{P}^2}_i$ is a copy of $\overline{\mathbb{C}\mathbb{P}^2}$). For each $i = 1, \dots, n$, let e_i be a generator of $H_2(\overline{\mathbb{C}\mathbb{P}^2}_i, \mathbb{Z}) \subseteq H_2(B\#n\overline{\mathbb{C}\mathbb{P}^2}, \mathbb{Z})$. If s is k -stable for B , then $(s, e_1, e_1, e_1, e_1, \dots, e_n, e_n, e_n, e_n)$ is $(k+n)$ -stable for $B\#n\overline{\mathbb{C}\mathbb{P}^2}$ and

$$\gamma_{d_k}(B)(s) = \left(-\frac{1}{2}\right)^n \gamma_{d_{k+n}}(B\#n\overline{\mathbb{C}\mathbb{P}^2})(s, e_1, e_1, e_1, e_1, \dots, e_n, e_n, e_n, e_n). \quad (8.17)$$

The point to this **blow-up formula** is that even when $s = (x_1, \dots, x_a, n_1\nu, \dots, n_b\nu)$ satisfies $a + 2b = d_k$, but *not* $a > \frac{3}{2}(1 + b_2^+(B))$, so that s is not k -stable for B and $\gamma_{d_k}(B)(s)$ is not yet defined, the right-hand side will be defined provided only that n is sufficiently large (note that $(a + 4n) + 2b = (a + 2b) + 4n = d_k + 4n = 4k + \frac{3}{2}(1 + b_2^+(B)) + 4n = 4(k + n) + \frac{3}{2}(1 + b_2^+(B)) = d_{k+n}$).

Moreover, by #3 above, the right-hand side of (8.17) takes the same value for all such sufficiently large n . Thus, the right-hand side of (8.17) allows us to *define* the left-hand side for any s with $a + 2b = d_k$. Because of the factor $(-\frac{1}{2})^n$, however, $\gamma_{d_k}(B)(s)$ will generally take values $\mathbb{Z}[\frac{1}{2}]$ for such s .

Let us summarize what we have accomplished up to this point. Let d be an integer satisfying $d \equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{4}$. Choose k so that $d = 4k - \frac{3}{2}(1 + b_2^+(B))$ and consider the corresponding principal bundle $SU(2) \hookrightarrow P_k \rightarrow B$ and moduli space $\mathcal{M}(P_k, \mathbf{g})$, where \mathbf{g} is a generic metric. Then the formal dimension of $\mathcal{M}(P_k, \mathbf{g})$ is $2d$. If $d < 0$ the moduli space is empty and $\gamma_d(B)$ is taken to be zero. If $d = 0$ we have produced a numerical invariant $\gamma_0(B)$ in Section 4. If $d > 0$, then for all non-negative integers a and b with $a + 2b = d$ we have defined a multilinear map

$$\gamma_d(B) : H_2(B, \mathbb{Z}) \times \dots \times H_2(B, \mathbb{Z}) \times H_0(B, \mathbb{Z}) \times \dots \times H_0(B, \mathbb{Z}) \longrightarrow \mathbb{Z} \left[\frac{1}{2} \right]$$

(by (8.16) if $a > \frac{3}{2}(1 + b_2^+(B))$ and by the blow-up formula (8.17) with sufficiently large n otherwise). Now, we are primarily interested in the invariants $\gamma_d(B)$ which operate only on two-dimensional homology classes of B , but we need the extra generality ($b \neq 0$) to relax the requirement that d be congruent $\pmod{4}$ to $-\frac{3}{2}(1 + b_2^+(B))$ to

$$d \equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{2} \tag{8.18}$$

and also to formulate the pivotal notion of “simple type” for four-manifolds. First, consider a $d > 0$ that satisfies (8.18), but *not* (8.7). Then $d = 2k - \frac{3}{2}(1 + b_2^+(B))$, where k must be odd. Consequently, $d + 2 = 2(k + 1) - \frac{3}{2}(1 + b_2^+(B))$ and $k + 1$ is even so $d + 2 \equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{4}$. Thus,

$$\gamma_{d+2}(B) : H_2(B, \mathbb{Z}) \times \dots \times H_2(B, \mathbb{Z}) \times H_0(B, \mathbb{Z}) \longrightarrow \mathbb{Z} \left[\frac{1}{2} \right]$$

is defined and so we may define

$$\gamma_d(B) : H_2(B, \mathbb{Z}) \times \dots \times H_2(B, \mathbb{Z}) \longrightarrow \mathbb{Z} \left[\frac{1}{2} \right]$$

by

$$\gamma_d(B)(x_1, \dots, x_d) = \frac{1}{2} \gamma_{d+2}(B)(x_1, \dots, x_d, \nu) \tag{8.19}$$

(keep in mind that this is the definition only when d satisfies (8.18), but not (8.7)). If $d \not\equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{2}$ we take $\gamma_d(B)(x_1, \dots, x_d) = 0$ for all

$x_1, \dots, x_d \in H_2(B, \mathbb{Z})$. Writing

$$\gamma_d(B)(x, \overset{d}{\cdot}, x) = \gamma_d(B)(x)$$

for each $x \in H_2(B, \mathbb{Z})$ we may now introduce a formal power series

$$\mathcal{D}_B(x) = \sum_{d=0}^{\infty} \frac{\gamma_d(B)(x)}{d!} \quad (8.20)$$

on $H_2(B, \mathbb{Z})$ called the **Donaldson series** of B . Notice that the nonzero terms in the series (8.20) are of two types, corresponding to those d which satisfy (8.7) and those which satisfy (8.18), but not (8.7) so that one can write $\mathcal{D}_B(x)$ in more detail as

$$\mathcal{D}_B(x) = \sum \frac{\gamma_d(B)(x, \overset{d}{\cdot}, x)}{d!} + \frac{1}{2} \sum \frac{\gamma_{d+2}(B)(x, \overset{d}{\cdot}, x, \nu)}{d!}.$$

Thus, if one could somehow independently determine the Taylor series of \mathcal{D}_B , then one could extract from it all of the $\gamma_d(B)(x, \overset{d}{\cdot}, x)$ and $\gamma_{d+2}(B)(x, \overset{d}{\cdot}, x, \nu)$ for all $x \in H_2(B, \mathbb{Z})$ and $d \geq 0$. Now let us say that a (compact, connected, simply connected, oriented) smooth four-manifold with $b_2^+(B) > 1$ and odd is of **KM-simple type** if

$$\gamma_{d+4}(B)(x_1, \dots, x_d, \nu, \nu) = 4\gamma_d(B)(x_1, \dots, x_d) \quad (8.21)$$

for all $d > 0$ and all $x_1, \dots, x_d \in H_2(B, \mathbb{Z})$. Although this may seem, at first glance, to be a very special and perhaps even unnatural assumption to make of B , there are, in fact, no known counterexamples to the conjecture that *every* compact, connected, simply connected, oriented, smooth four-manifold with $b_2^+(B) > 1$ and odd is of KM-simple type. Given this assumption one can inductively extract from \mathcal{D}_B all of the

$$\gamma_d(B)(x, \overset{a}{\cdot}, x, \nu, \overset{b}{\cdot}, \nu)$$

where $a + 2b = d$. Polarization in x then gives all of the

$$\gamma_d(B)(x_1, \dots, x_a, \nu, \overset{b}{\cdot}, \nu)$$

and from this and multilinearity in the last b slots one obtains all of the

$$\gamma_d(B)(x_1, \dots, x_a, n_1\nu, \dots, n_b\nu).$$

Thus, for manifolds of KM-simple type, the Donaldson series contains all of the information in all of the Donaldson invariants.

Remark. *Let us continue the example of the $K3$ surface (see the Remark following (8.11)) by writing out its Donaldson series. At this point we know the following. $b_2^+(K3) = 3$ implies that*

$$d_k = 4k - \frac{3}{2}(1 + b_2^+(B)) = 2(2k - 3) = 2n, \quad n \equiv 1 \pmod{2}. \quad (8.22)$$

To be in the stable range we must also have

$$n > 3. \quad (8.23)$$

For such n we have invariants $\gamma_{2n}(K3) : H_2(K3, \mathbb{Z}) \rightarrow \mathbb{Z}$ given by

$$\gamma_{2n}(K3)(x) = \frac{(2n)!}{2^n n!} (Q_{K3}(x))^n. \quad (8.24)$$

To describe the remaining invariants we notice first that there is no zero-dimensional moduli space in this case ($d_k = 0$ implies $k = \frac{3}{2}$, but k must be an integer, i.e., Chern number). Next, when $k = 1$, $d_k = -2 < 0$ and the moduli space is generically empty so the corresponding invariant is zero. On the other hand, $k = 1$ and $k = 2$ give $d_k = 2(n = 1)$ and $d_k = 6(n = 3)$, respectively, which are positive, but not in the stable range. The corresponding (potentially) nonzero invariants must therefore be obtained from the blow-up formula (8.17). The Donaldson series also contains terms corresponding to $d \equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{2} \equiv 0 \pmod{2}$, but $d \not\equiv -\frac{3}{2}(1 + b_2^+(B)) \pmod{4} \equiv 2 \pmod{4}$, i.e., d of the form

$$\begin{aligned} d &= 2n \\ n &\equiv 0 \pmod{2}. \end{aligned} \quad (8.25)$$

These are all obtained from (8.19) which, in this case, becomes

$$\gamma_{2n}(K3)(x) = \gamma_{2n}(K3)(x, \cdot^{2n}, x) = \frac{1}{2} \gamma_{2n+2}(K3)(x, \cdot^{2n}, x, \nu). \quad (8.26)$$

Now, all of this has been calculated (see, for example, Proposition 7.13 of [30]) and we will say only that in all cases $\gamma_{2n}(K3)(x)$ is given by the same formula that appears in (8.24). Thus, the Donaldson series is given by

$$\begin{aligned} \mathcal{D}_{K3}(x) &= \sum_{d=0}^{\infty} \frac{\gamma_d(K3)(x)}{d!} = \sum_{n=0}^{\infty} \frac{\gamma_{2n}(K3)(x)}{(2n)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \frac{(2n)!}{2^n n!} (Q_{K3}(x))^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} (Q_{K3}(x)/2)^n \end{aligned}$$

and so it is customary to write

$$\mathcal{D}_{K3}(x) = \exp(Q_{K3}(x)/2). \quad (8.27)$$

The appearance of the exponential function in the expression (8.27) for the Donaldson series of $K3$ is just a special case of a remarkable structure theorem proved by Kronheimer and Mrowka [30] which plays a central role in the conjecture of Witten that is our focus here.

Theorem 8.1. *If B is of KM -simple type, then there exist cohomology classes $K_1, \dots, K_S \in H^2(B, \mathbb{Z})$, called **KM -basic classes**, and rational numbers $a_1, \dots, a_S \in \mathbb{Q}$, called **coefficients**, such that*

$$\mathcal{D}_B(x) = \exp(Q_B(x)/2) \sum_{r=1}^S a_r \exp(K_r(x)) \quad (8.28)$$

for every $x \in H_2(B, \mathbb{Z})$. Moreover, each KM -basic class K_r is an integral lift of the second Stiefel-Whitney class $w_2(B)$ of B , i.e., $w_2(B) \equiv K_r \pmod{2}$.

Thus, for manifolds of KM -simple type, all of the information contained in the (infinitely many) Donaldson invariants of B is contained in a finite set of cohomology classes and rational numbers (for $K3$ the only KM -basic class is $0 \in H^2(K3, \mathbb{Z})$ and the only coefficient is 1). The appearance of Theorem 8.1 in 1994 represented a profound breakthrough in Donaldson theory. Enormously complicated calculations of an apparently infinite set of invariants were suddenly replaced by the (certainly not easy, but at least finite) problem of determining the KM -basic classes and coefficients. As fate would have it, however, the Fall of 1994 witnessed another event which rendered this triumph of Kronheimer and Mrowka moot. Edward Witten, at the end of his now famous lecture at M.I.T. (described in [48]) made the conjecture which, within weeks, brought about the demise of Donaldson theory and initiated the entirely new approach to the study of smooth four-manifolds that has come to be known as Seiberg-Witten theory. We know that Witten was led to his conjecture by physics, not mathematics and we are now, at long last, prepared to state the conjecture precisely.

Conjecture 8.2. Witten's Conjecture: *Let B be a compact, connected, simply connected, oriented, smooth 4-manifold with $b_2^+(B) > 1$ and odd. Then*

1. B is of SW -simple type if and only if it is of KM -simple type and, in this case,

2. *SW-basic classes coincide with KM-basic classes and*

$$\mathcal{D}_B(x) = \exp(Q_B(x)/2) \sum 2^{m(B)} SW_0(B, \mathcal{L}) \exp(c_1(L^0)(x)) \quad (8.29)$$

where $m(B) = 2 + \frac{1}{4}(7\chi(B) + \mathbb{1}\sigma(B))$, $\chi(B)$ is the Euler characteristic of B , $\sigma(B)$ is the signature of B and the sum is over all (equivalence classes of) spin^c structures \mathcal{L} for B satisfying (7.131), i.e., $c_1(L^0)^2 = 2\chi(B) + 3\sigma(B)$.

The content of the conjecture is that, for manifolds of simple type, the basic classes are just those elements of $H^2(B, \mathbb{Z})$ corresponding to spin^c structures with zero-dimensional SW moduli spaces and the coefficients are, up to the topological factor $2^{m(B)}$, just the corresponding zero-dimensional SW invariants.

There are a number of attitudes one might adopt toward a conjecture of this sort. One might, of course, try to prove it and, although this has not been the principal focus of work in this area, much deep and interesting mathematics has been directed toward this end. A strategy suggested by Pidstrigatch and Tyurin [44] for relating the two sets of invariants by viewing the Donaldson and Seiberg-Witten moduli spaces as singular submanifolds of a larger moduli space of “nonabelian monopoles” has been taken up in earnest by Feehan and Leness in a long series of difficult and technical papers (see [17] for a synopsis). Another approach, due to Vajiac [50], is much closer to the spirit of the view from physics, but remains formal in the sense that the result is derived by applying as yet unproved infinite dimensional versions of the equivariant localization theorems. By far the more prevalent attitude has been that, even if a proof is hard to come by, the conjecture has been checked in every case in which all of the invariants are known and has survived so that it would seem to make good practical sense for the topologist to (at least provisionally) abandon the Donaldson invariants for the much more tractable Seiberg-Witten invariants. Perhaps the most enlightened attitude, however, and one which has been emphasized by Atiyah, is that if physics is truly capable of casting such a penetrating light upon mathematics at the very deepest levels, then mathematicians will want to take heed and turn their attention once again to their historical roots in physics.

References

- [1] Alvarez-Gaumé L. and Zamora F., *Duality in Quantum Field Theory (and String Theory)*, arXiv:hep-th/9709180.
- [2] Atiyah M. and Bott R., *The Moment Map and Equivariant Cohomology*, *Topology* **23** (1984) 1-28.
- [3] Atiyah M., Hitchin N. and Singer I., *Self-Duality in Four-Dimensional Riemannian Geometry*, *Proc. Roy. Soc. Lond. A* **362** (1978) 425-461.
- [4] Atiyah M. and Jeffrey L., *Topological Lagrangians and Cohomology*, *J. Geom. Phys.* **7** (1990) 119-136.
- [5] Belavin A., Polyakov A., Schwarz A. and Tyupkin Y., *Pseudoparticle Solutions of the Yang-Mills Equations*, *Phys. Lett. B* **59** (1975) 85-87.
- [6] Berline N., Getzler E. and Vergne M., *Heat Kernels and Dirac Operators*, Springer, New York, 1996.
- [7] Blau M., *The Mathai-Quillen Formalism and Topological Field Theory*, *J. Geom. Phys.* **11** (1993) 95-127.
- [8] Blau M. and Thompson G., *Localization and Diagonalization*, *J. Math. Phys.* **36** (1995) 2192-2236.
- [9] Bleecker D., *Gauge Theory and Variational Principles*, Addison-Wesley, Reading, MA, 1981.
- [10] Cordes S., Moore G. and Ramgoolam S., *Lectures on 2D Yang-Mills Theory, Equivariant Cohomology and Topological Field Theory*, arXiv: hep-th/9402107.
- [11] Deligne P. *et al* (Eds.), *Quantum Fields and Strings: A Course for Mathematicians*, vols. 1&2, AMS, Providence, RI, 1999.
- [12] Dirac P.A.M., *Quantised Singularities in the Electromagnetic Field*, *Proc. Roy. Soc. A* **133** (1931), 60-72.
- [13] Donaldson S., *An Application of Gauge Theory to Four-Dimensional Topology*, *J. Diff. Geom.* **18** (1983) 279-315.
- [14] Donaldson S., *Polynomial Invariants for Smooth 4-Manifolds*, *Topology* **29** (1990), 257-315.
- [15] Donaldson S. and Kronheimer P., *The Geometry of Four-Manifolds*, Oxford University Press, Oxford, 1990.
- [16] 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.

- [17] Feehan P. and Leness T., *On Donaldson and Seiberg-Witten Invariants*, arXiv: math.DG/0106221.
- [18] Freed D. and Uhlenbeck K., *Instantons and Four-Manifolds*, Springer, New York, Berlin, 1984.
- [19] Freedman M., *The Topology of Four-Dimensional Manifolds*, J. Diff. Geom. **17** (1982) 357-454.
- [20] Friedman R. and Morgan J., *Smooth Four-Manifolds and Complex Surfaces*, Springer, New York, 1994.
- [21] Goddard P. and Olive D., *Magnetic Monopoles in Gauge Field Theories*, Rep. Progr. Phys. **41** (1978) 1357-1437.
- [22] Guillemin V. and Sternberg S., *Geometric Asymptotics*, Amer. Math. Soc., Providence, RI, 1977.
- [23] Guillemin V. and Sternberg S., *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, 1984.
- [24] Guillemin V. and Sternberg S., *Supersymmetry and Equivariant De Rham Theory*, Springer, New York, 1999.
- [25] 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 pp. 1-48.
- [26] Jaffe A. and Taubes C., *Vortices and Monopoles*, Birkhauser, Boston, MA, 1980.
- [27] Kalkman J., *BRST Model for Equivariant Cohomology and Representatives for the Equivariant Thom Class*, Comm. Math. Phys. **153** (1993) 447-463.
- [28] Kobayashi S. and Nomizu K., *Foundations of Differential Geometry, vols. I & II*, Wiley Interscience, 1963 and 1969.
- [29] Kronheimer P. and Mrowka T., *Recurrence Relations and Asymptotics for Four-Manifold Invariants*, Bull. Amer. Math. Soc. **30** (1994) 215-221.
- [30] Kronheimer P. and Mrowka T., *Embedded Surfaces and the Structure of Donaldson's Polynomial Invariants*, J. Diff. Geom. **3** (1995) 573-734.
- [31] Labastida J. and Marino M., *A Topological Lagrangian for Monopoles on Four-Manifolds*, Phys. Lett. B **351** (1995) 146-152.
- [32] Labastida J. and Lozano C., *Lectures on Topological Quantum Field Theory*, arXiv: hep-th/9709192.
- [33] Lang S., *Introduction to Differentiable Manifolds*, John Wiley & Sons, New York, London, 1962.
- [34] Lawson H., *The Theory of Gauge Fields in Four Dimensions*, Amer. Math. Soc., Providence, RI, 1985.

-
- [35] Lawson H. and Michelsohn M-L., *Spin Geometry*, Princeton University Press, Princeton, NJ, 1989.
- [36] Morgan J., *The Seiberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds*, Princeton University Press, Princeton, NJ, 1996.
- [37] Morgan J., *An Introduction to Gauge Theory*, in *Gauge Theory and the Topology of Four-Manifolds*, Amer. Math. Soc., Providence RI, 1998.
- [38] Mathai V. and Quillen D., *Superconnections, Thom Classes and Equivariant Differential Forms*, *Topology* **25** (1986) 85-110.
- [39] Naber G., *Topological Methods in Euclidean Spaces*, Dover Publications, Mineola, NY, 2000.
- [40] Naber G., *Topology, Geometry, and Gauge Fields: Foundations*, Springer, New York, 1997.
- [41] Naber G., *Topology, Geometry, and Gauge Fields: Interactions*, Springer, New York, 2000.
- [42] O'Grady K., *Donaldson's Polynomials for K3 Surfaces*, *J. Diff. Geom.* **35** (1992) 415-427.
- [43] Parker T., *Gauge Theories on Four-Dimensional Riemannian Manifolds*, *Comm. Math. Phys.* **85** (1982) 563-602.
- [44] Pidstrigach V. and Tyurin A., *Localization of Donaldson's Invariants along Seiberg-Witten Classes*, arXiv: dg-ga/9507004.
- [45] Seiberg N. and Witten E., *Monopoles, Duality and Chiral Symmetry Breaking in $N = 2$ Supersymmetric QCD*, *Nucl. Phys. B* **431** (1994) 581-640.
- [46] Szabo R., *Equivariant Localization of Path Integrals*, arXiv: hep-th/9608068.
- [47] Taubes C., *Self-Dual Connections on Non-Self-Dual 4-Manifolds*, *J. Diff. Geom.* **17** (1982) 139-170.
- [48] Taubes C., *The Seiberg-Witten Invariants* (Videotape), AMS Selected Lectures in Mathematics, Amer. Math. Soc., Providence, RI, 1995.
- [49] Uhlenbeck K., *Removable Singularities in Yang-Mills Fields*, *Comm. Math. Phys.* **83** (1982) 11-30.
- [50] Vajiac A., *A Generalization of Witten's Conjecture Relating Donaldson and Seiberg-Witten Invariants*, arXiv: hep-th/0003214.
- [51] Warner F., *Foundations of Differentiable Manifolds and Lie Groups*, Springer, New York, 1983.
- [52] Witten E., *Topological Quantum Field Theory*, *Comm. Math. Phys.* **117** (1988) 353-386.

- [53] Witten E., *Two Dimensional Gauge Theories Revisited*, J. Geom. Phys. **9** (1992) 303-368.
- [54] Witten E., *Monopoles and Four-Manifolds*, Math. Res. Lett. **1** (1994) 769-796.
- [55] 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.
E-mail: gregory.l.naber@drexel.edu