Instantons and Self-Dual Gauge Fields

Jeffrey D. Olson*

May 11, 2004

Abstract

Self-dual gauge fields play an important role in four-dimensional Riemannian geometry. In this essay we review some of the basic ideas surrounding these gauge fields. In particular we review the Atiyah-Hitchin-Singer theorem on the dimension of the moduli space of such solutions. We also review the role of self-dual gauge fields in the geometry of Euclidean Yang-Mills theory where these solutions go by the name of instantons. Finally, we work out the simplest example of a self-dual gauge field on a principal SU(2)-bundle over S^4, the BPST instanton.

Contents

1 Introduction **2**

2 Review of Gauge Theory **2**

2.1 Principal bundles . . . 2

2.2 Connection and curvature . . . 4

2.3 Gauge transformations . . . 9

3 Self-Duality **11**

3.1 Yang-Mills theory . . . 12

4 Moduli Space of Self-Dual Connections **12**

5 Instantons on S^4 **16**

5.1 Principal Sp(1)-bundles over S^4 . . . 16

5.2 Self-dual connections for $k = 1$ . . . 17

*Email: jdolson@physics.utexas.edu

1 Introduction

Four-dimensional Riemannian geometry is interesting for a number of reasons. One of the reasons has to do with the fact that the Hodge star operator, $*\colon \Lambda^k \to \Lambda^{4-k}$, which acts on the bundle of differential forms squares to the identity when restricted to the space of 2-forms. This allows one to decompose the bundle of 2-forms into the positive and negative eigenspaces of $*$:

$$\Lambda^2 = \Lambda^2_+ \oplus \Lambda^2_-.$$

The sections of these bundles are called *self-dual* and *anti-self-dual* forms respectively. In Riemannian geometry, as well as in gauge theory, this decomposition is interesting because of the relation of 2-forms to curvature. By applying the Hodge star to the curvature 2-form one can talk about self-dual and anti-self-dual curvatures. These special curvatures turn out to play an important role in four-dimensional geometry and gauge theory.

Physicists, in particular, have found that self-dual gauge fields—which in physics go by the name *instantons*—are solutions to the Euclidean Yang-Mills equations. Indeed, on compact manifolds these solutions are actually absolute minimum of the Yang-Mills action. Instantons are important because they lead to topologically nontrivial vacuum configurations.

The outline of this paper is as follows: In §2, we review some of the necessary differential geometry and gauge theory ideas that will be needed to discuss self-duality in gauge theory. Mostly this consists of a set of definitions and results with almost no proofs. None of the proofs are very difficult and make for good exercises. They can also be found in the references [Nab, BB, Sha]. In §3 we discuss self-dual connections on principal bundles, self-dual curvature in Riemannian geometry, and the relation to Yang-Mills theory. Then in §4 we discuss a theorem by Atiyah, Hitchin, and Singer [AHS] which computes the dimension of the moduli space of self-dual connections on a self-dual Riemannian 4-manifold by an application of the Atiyah-Singer index theorem. In the last section we work out an explicit example of instanton on principal SU(2)-bundle over S^4. This is the so-called BPST instanton first constructed in [BPST] as a solution to the Euclidean Yang-Mills equations on $\mathbb{R}^4$. Much of the material for this section was taken from [Nab].

2 Review of Gauge Theory

Principal bundles—together with connections on them—form the basic geometric setting of gauge theory. For completeness, and to establish some terminology and notation, we review the basic concepts in this section.

2.1 Principal bundles

Definition 2.1. A **principal G-bundle** is a smooth fiber bundle $P \xrightarrow{\pi} M$ together with a Lie group G and a smooth right action $P \times G \to P$ which preserves the fibers of P and acts simply transitively on each fiber. Right multiplication $R_g\colon P \to P$ by $g \in G$ is denoted $R_g(p) = p \cdot g$.

For any principal bundle $P \xrightarrow{\pi} M$ we define the space of **vertical vectors** at $p \in P$ to be $\mathrm{Vert}_p P = \ker \pi_*$ where $\pi_*\colon T_pP \to T_{\pi(p)}M$ is the derivative of the projection map π at $p \in P$. Note that $\dim(\mathrm{Vert}_p P) = \dim G = \dim P - \dim M$. For a fixed $p \in P$ let $\sigma_p\colon G \to P$ be the map given by $\sigma_p(g) = p \cdot g$. The derivative of this map at the identity, $(\sigma_p)_*\colon \mathfrak{g} \to \mathrm{Vert}_p P$, gives an isomorphism between $\mathfrak{g}$ (the Lie algebra of G) and $\mathrm{Vert}_p P$. To every $\xi \in \mathfrak{g}$ we define a **fundamental vertical vector field**, $\hat{\xi}$, on P by $\hat{\xi}_p = (\sigma_p)_*\xi$, or equivalently, $\hat{\xi}_{p\cdot g} = (\sigma_p)_*(L_g)_*\xi$.

Definition 2.2. Let $P \xrightarrow{\pi} M$ be a principal G-bundle, and let F be a smooth left G-space with action $\rho\colon G \to \mathrm{Diff}(F)$. The **associated bundle** $P \times_\rho F$ determined by ρ is the quotient $(P \times F)/G$ where the action of G on $P \times F$ is given by $(p, \xi) \cdot g = (p \cdot g, \rho(g^{-1})\xi)$. With the natural projection $[p, \xi] \mapsto \pi(p)$, the associated bundle is a fiber bundle over M with fiber F and structure group $\rho(G) \cong G/\ker(\rho)$.

If F is a vector space and ρ a linear representation then the associated bundle construction defines a vector bundle over M. An important example of this construction is the **adjoint bundle**, $\mathfrak{g}_P = P \times_{\mathrm{Ad}} \mathfrak{g}$ where $\mathrm{Ad}\colon G \to \mathrm{Aut}(\mathfrak{g})$ is the adjoint representation of G.

Definition 2.3. Let $P \to M$ be a principal G-bundle, and let F be a smooth left G-space with action $\rho\colon G \to \mathrm{Diff}(F)$. A function $f\colon P \to F$ is **right equivariant** if $f(p \cdot g) = \rho(g^{-1})f(p)$.

The importance of this definition is that sections of the associated bundle $E = P \times_\rho F$ are in 1-1 correspondence with right equivariant functions $f\colon P \to F$. Indeed, if f is such a function then $x \mapsto [p, f(p)]$ is a section of E. Here p is any element of $\pi^{-1}(x)$. The equivariance condition ensures that this definition is independent of the choice of p.

We will often be concerned with k-forms on M taking values in some associated vector bundle $E = P \times_\rho V$ over M. It will be convenient to establish a similar relation between these forms and V-valued k-forms on P, which we denote by $\mathcal{A}^k(P, V) = \Gamma(V \otimes \Lambda^k T^*P)$. Let $\alpha \in \mathcal{A}^k(P, V)$, define $\alpha' \in \mathcal{A}^k(M, E)$ by $\alpha'_x(v_1, \cdots, v_k) = [p, \alpha_p(\tilde{v}_1, \cdots, \tilde{v}_k)]$ where $\pi(p) = x$ and $\pi_*\tilde{v}_i = v_i$. In order for this to be independent of the choices made, α must be right equivariant and must vanish on vertical vectors. This leads to the following:

Definition 2.4. Let $P \to M$ be a principal G-bundle and let $P \times_\rho V$ be an associated vector bundle. The space of **basic V-valued k-forms**, denoted $\bar{\mathcal{A}}^k(P, V)$, is defined to be the set of V-valued k-forms α such that

1. $R_g^*\alpha = \rho(g^{-1})\alpha$ for all $g \in G$ (**right equivariance**),
2. $\alpha(X, \cdots) = 0$ for all $X \in \mathrm{Vert}\, P$.

The basic V-valued k-forms are in 1-1 correspondence with k-forms on the base space M taking values in the associated bundle $E = P \times_\rho V$. That is, $\bar{\mathcal{A}}^k(P, V) \cong \mathcal{A}^k(M, E) = \Gamma(E \otimes \Lambda^k T^*M)$. We will occasionally make implicit use of this isomorphism by passing from one space to the other without warning.

To continue our digression on vector-valued forms: note that on any manifold P the exterior derivative $d\colon \mathcal{A}^k(P, V) \to \mathcal{A}^{k+1}(P, V)$ can be defined componentwise relative to any basis for V. That is, given $\omega \in \mathcal{A}^k(P, V)$ define $d\omega = (d\omega^i)e_i$, where $\{e_i\}$ is any basis for V.

The wedge product does not generalize quite so nicely. In general, we must regard it as a map $\wedge\colon \mathcal{A}^p(P,V)\times\mathcal{A}^q(P,V)\to\mathcal{A}^{p+q}(P,V\otimes V)$ given by $\omega\wedge\eta=(\omega^i\wedge\eta^j)(e_i\otimes e_j)$. However, if we are given a multiplication map $\mu\colon V\otimes V\to V$ we can induce a map $\mu^*\colon \mathcal{A}^k(P,V\otimes V)\to\mathcal{A}^k(P,V)$ by setting $(\mu^*\omega)(X^1,\cdots,X^k)=\mu(\omega(X^1,\cdots,X^k))$. We thus obtain a map $\mu^*\circ\wedge\colon \mathcal{A}^p(P,V)\times\mathcal{A}^q(P,V)\to\mathcal{A}^{p+q}(P,V)$.

Our particular interest is the case where $V=\mathfrak{g}$ is a Lie algebra. Here the natural multiplication is given by the Lie bracket $\mu(v\otimes w)=[v,w]$. We shall use the same bracket notation to denote the composition $\mu^*\circ\wedge$. That is, $[\omega,\eta]=\mu^*(\omega\wedge\eta)$. For 1-forms $\omega_1,\omega_2\in\mathcal{A}^1(P,\mathfrak{g})$ this amounts to the following:

$$[\omega_1,\omega_2](X,Y)=[\omega_1(X),\omega_2(Y)]-[\omega_1(Y),\omega_2(X)]. \tag{1}$$

In particular,

$$[\omega,\omega](X,Y)=2[\omega(X),\omega(Y)]. \tag{2}$$

Lemma 2.5. *Let $\omega_p,\omega_q,\omega_r\in\mathcal{A}^\bullet(P,\mathfrak{g})$ be forms of degree p, q, and r respectively. Then*

1. $[\omega_q,\omega_p]=-(-)^{pq}[\omega_p,\omega_q]$
2. $(-)^{rq}[\omega_r,[\omega_p,\omega_q]]+(-)^{qp}[\omega_q,[\omega_r,\omega_p]]+(-)^{pr}[\omega_p,[\omega_q,\omega_r]]=0$
3. $d[\omega_p,\omega_q]=[d\omega_p,\omega_q]+(-)^p[\omega_p,d\omega_q]$

That is, the algebra $\mathcal{A}^\bullet(P,\mathfrak{g})$ is a $\mathbb{Z}_2$-graded Lie algebra with $d\colon \mathcal{A}^\bullet(P,\mathfrak{g})\to\mathcal{A}^\bullet(P,\mathfrak{g})$ a graded derivation.

2.2 Connection and curvature

§ The Maurer-Cartan form

Every Lie group G comes equipped with a natural Lie algebra valued 1-form called the Maurer-Cartan form. This forms plays an important role in gauge theory.

Definition 2.6. Let G be a Lie group and let $\mathfrak{g}=T_eG$ be its Lie algebra. The **Maurer-Cartan form**[1], $\Theta\in\mathcal{A}^1(G,\mathfrak{g})$, is a $\mathfrak{g}$-valued 1-form on G defined by $\Theta_g(X)=(L_{g^{-1}})_*X$. where $X\in T_gG$.

If $\mathfrak{g}$ is regarded as the space of left-invariant vector fields rather than the tangent space to the identity, then Θ sends each vector $X\in T_gG$ to the unique left-invariant vector field extending X.

Theorem 2.7. *Let G be a Lie group and let Θ be its Maurer-Cartan form.*

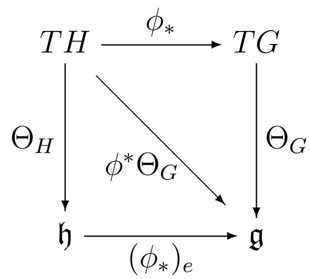
1. $L_g^*\Theta=\Theta$
2. $R_g^*\Theta=\operatorname{Ad}_{g^{-1}}\Theta$

[1]Specifically this is the left-invariant Maurer-Cartan form. Naturally, there is a right-invariant Maurer-Cartan form as well.

3. Let $\{e_1, \cdots, e_n\}$ be a basis for $\mathfrak{g}$ and let $\{\Theta^1, \cdots, \Theta^n\}$ be the unique left-invariant $\mathbb{R}$-valued 1-forms on G which form a dual basis to $\{e_1, \cdots, e_n\}$ at the identity. Then $\Theta = \Theta^i e_i$.

The classical way of writing the Maurer-Cartan form on matrix Lie groups is $\Theta = g^{-1}dg$ where dg is interpreted as a matrix of coordinate differentials. Here the prefactor g^{-1} accounts for the left-translation to the identity.

Theorem 2.8. *Let $\phi\colon H \to G$ be a homomorphism of Lie groups. Then $\phi^*\Theta_G = (\phi_*)_e\Theta_H$. That is, the diagram*



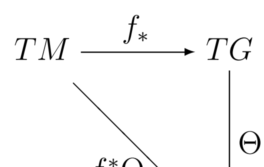
commutes. In particular, if H is a Lie subgroup of G then $\Theta_H = \Theta_G|_H$.

The exterior derivative of the Maurer-Cartan form naturally turns out to be an important quantity on a Lie group. So important, in fact, that it completely determines the local structure of a Lie group. The result is often expressed as follows:

Theorem 2.9 (Structural equation). *Let G be a Lie group and let Θ be its Maurer-Cartan form. Then*

$$d\Theta + \frac{1}{2}[\Theta, \Theta] = 0. \tag{3}$$

In applications to gauge theory we will often encounter the following construction. Let $f\colon M \to G$ be a smooth map from a manifold M into a Lie group G. We define the (left) **Darboux derivative** of f to be pullback $f^*\Theta$ of the Maurer-Cartan form on G to M. Note that this gives a Lie algebra valued 1-form on M:



The pushforward f_* is usually thought of as the derivative of the map f, however, this is not entirely analogous to the notion of a derivative in calculus since f_* has the original map f built into it. The effect of composing f_* with Θ is to *forget* this information by left-translating everything to the identity. Hence, the Darboux derivative is the natural generalization of the derivative to Lie group valued functions.

The interested reader should consult [Sha] for a more thorough discussion of the Maurer-Cartan form and the Darboux derivative (as well as an entertaining discussion of the fundamental theorem of *non-abelian* calculus).

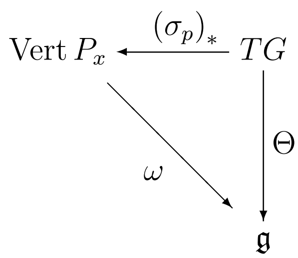
§ Gauge connections

Connections on principal bundles are generalizations of the Maurer-Cartan form on Lie groups.

Definition 2.10. Let $P \to M$ be a principal G-bundle. A **connection**[2] on P is a $\mathfrak{g}$-valued 1-form $\omega \in \mathcal{A}^1(P, \mathfrak{g})$ satisfying two properties:

1. $R_g^* \omega = \mathrm{Ad}_{g^{-1}} \circ \omega$ for all $g \in G$ (**right equivariance**),
2. $\omega(\hat{\xi}) = \xi$ for all $\xi \in \mathfrak{g}$.

Note that ω is *not* a basic 1-form since it does not vanish on vertical vectors. Instead, when restricted to a fiber, ω is essentially just the Maurer-Cartan form on G. Specifically, given the isomorphism $\sigma_p \colon G \to P_x$ for a given p in the fiber over x, the second condition above says that the diagram



commutes. Note, however, that the difference between any two connections *is* a basic 1-form: $\omega' - \omega \in \bar{\mathcal{A}}^1(P, \mathfrak{g})$. That is, the space of connection 1-forms is an affine space modeled on $\bar{\mathcal{A}}^1(P, \mathfrak{g})$.

The definition given above—in terms of a Lie algebra valued 1-form—is the most convenient one for gauge theory, however, there is an alternate definition which is more geometrical: a connection is a G-invariant distribution of horizontal subspaces of TP. That is, a connection is a smooth assignment of a subspace $\mathrm{Hor}_p P \subset T_p P$ to each point $p \in P$ such that

1. $T_p P = \mathrm{Hor}_p P \oplus \mathrm{Vert}_p P$,
2. $\mathrm{Hor}_{p \cdot g} P = (R_g)_* \mathrm{Hor}_p P$.

The projection $\pi \colon P \to M$ then induces an isomorphism $\pi_* \colon \mathrm{Hor}_p P \to T_{\pi(p)} M$. The relationship between the two definitions is as follows: Given a connection 1-form $\omega \in \mathcal{A}^1(P, \mathfrak{g})$ define the horizontal subspace at p to be $\mathrm{Hor}_p P = \ker \omega_p$. Conversely, given a G-invariant horizontal distribution $\mathrm{Hor}\, P \subset TP$ define $\omega_p = (\sigma_p)_*^{-1} \circ \mathcal{P}_V$ where $\mathcal{P}_V \colon T_p P \to \mathrm{Vert}_p P$ is the projection onto the vertical subspace and $(\sigma_p)_* \colon \mathfrak{g} \to \mathrm{Vert}_p P$ is the natural isomorphism.

Definition 2.11. Let $P \to M$ be a principal G-bundle, let ω be a connection 1-form on P, and let $\mathcal{A}^k(P, V)$ be the set of k-forms on P taking values in some vector space V. Define the **covariant derivative** $D_\omega \colon \mathcal{A}^k(P, V) \to \mathcal{A}^{k+1}(P, V)$ by

$$D_\omega \alpha = d\alpha \circ \mathcal{P}_H, \tag{4}$$

[2] This is sometimes called an **Ehresmann connection**.

where $\mathcal{P}_H$ projects all arguments onto the horizontal subspace defined by the connection, i.e. $(D_\omega\alpha)(X^1, \cdots, X^k) = d\alpha(X^1_H, \cdots, X^k_H)$.

This definition, while perhaps strange at first sight, turns out to be the appropriate notion of a covariant derivative on a principal bundle. The reason lies in the fact that it allows one to define a traditional covariant derivative on any vector bundle associated to P. This is a consequence of the following:

Theorem 2.12. *Let $P \to M$ be a principal G-bundle with connection ω, and let $E = P \times_\rho V$ be a vector bundle associated to P. Then the covariant derivative restricts to a map $D_\omega \colon \bar{\mathcal{A}}^k(P, V) \to \bar{\mathcal{A}}^{k+1}(P, V)$.*

The established isomorphism between $\bar{\mathcal{A}}^k(P, V)$ and $\mathcal{A}^k(M, E)$ then induces a covariant derivative on $\mathcal{A}^k(M, E)$. In particular, for $k = 0$ we obtain a map $\nabla \colon \Gamma(E) \to \Gamma(E \otimes T^*M)$ which satisfies all the required properties of a covariant derivative. In the case where $E = \mathfrak{g}_P$ is the adjoint bundle of P there is a convenient formula for the covariant derivative restricted to $\bar{\mathcal{A}}^k(P, \mathfrak{g})$:

Lemma 2.13. *Let $P \to M$ be a principal G-bundle with connection ω, and let $\mathfrak{g}_P = P \times_{\mathrm{Ad}} \mathfrak{g}$ be the adjoint bundle. Then for $\alpha \in \bar{\mathcal{A}}^k(P, \mathfrak{g})$*

$$D_\omega\alpha = d\alpha + [\omega, \alpha]. \tag{5}$$

§ Curvature

Having defined the notion of a covariant derivative on vector-valued forms the natural thing to do is to take the covariant derivative of the connection. This gives us the appropriate notion of curvature on a principal bundle.

Definition 2.14. The **curvature** Ω of a connection ω on a principal G-bundle $P \to M$ is the covariant derivative of the connection: $\Omega = D_\omega\omega \in \mathcal{A}^2(P, \mathfrak{g})$.

It is clear from the definition of D_ω that Ω vanishes on vertical vectors. Furthermore, the right equivariance of ω ensures that Ω is also right equivariant. Thus Ω is actually a basic 2-form: $\Omega \in \bar{\mathcal{A}}^2(P, \mathfrak{g})$, and can be regarded as 2-form on M with values in the adjoint bundle: $\Omega \in \mathcal{A}^2(M, \mathfrak{g}_P) = \Gamma(\mathfrak{g}_P \otimes \Lambda^2 T^*M)$.

Theorem 2.15. *Let $P \to M$ be a principal G-bundle with connection ω. Then the curvature of ω is given by*

$$\Omega = d\omega + \frac{1}{2}[\omega, \omega]. \tag{6}$$

Since every Lie group G acts simply transitively on itself from the right, one can regard G as a principal bundle over a one-point set. The Maurer-Cartan form then satisfies the required properties of a connection on G. Theorem 2.9, together with the above, then asserts that the curvature of this connection vanishes.

Theorem 2.16 (Bianchi identity). *Let ω be a connection on a principal bundle and let Ω be the curvature of ω. Then*

$$D_\omega \Omega = D_\omega^2 \omega = 0, \tag{7}$$

or equivalently,

$$d\omega = [\Omega, \omega]. \tag{8}$$

In general, unlike the ordinary exterior derivative, $D_\omega^2 \neq 0$. The connection is special in this regard. When acting on sections of $\bar{\mathcal{A}}^k(P, V)$ it turns out that D_ω^2, while nonvanishing, is purely algebraic in nature. In particular, for $V = \mathfrak{g}$ one can show that

$$D_\omega^2 \alpha = [\Omega, \alpha]. \tag{9}$$

There is a similar expression for more general V.

§ Holonomy

Given a connection on a vector bundle one can define the holonomy group of the connection as the group of transformations obtained by parallel transporting vectors around closed loops in the base space. Not surprisingly, there is a related idea for connections on principal bundles.

Let $P \to M$ be a principal G-bundle with connection ω. A smooth curve $\gamma\colon [0,1] \to P$ is said to be **horizontal** if the tangent vector to the curve at each point is horizontal. Define an equivalence relation on P by saying $p \sim q$ if p and q can be joined by a piecewise-smooth horizontal curve. Now fix a point $p \in P$ and define the **holonomy group** of the connection ω based at the point p to be

$$\mathrm{Hol}_p(\omega) = \{g \in G \mid p \cdot g \sim p\}. \tag{10}$$

Just as with holonomy groups on vector bundles, the group $\mathrm{Hol}_p(\omega)$ depends on the base point p only up to conjugation in G (i.e. the holonomy groups based at different points in P are conjugate, and so isomorphic, subgroups of G).

Definition 2.17. Let $P \to M$ be a principal G-bundle with connection ω. The connection is said to be **irreducible** if $\mathrm{Hol}(\omega) = G$. Conversely, if the holonomy group is a proper subgroup of G then ω is **reducible**.

The point of this definition is that if ω is reducible, with $\mathrm{Hol}_p(\omega) = H \subset G$, then there exits a principal subbundle Q of P, defined by

$$Q = \{q \in P \mid q \sim p\}, \tag{11}$$

with structure group H and an irreducible connection $\omega' = \omega|_Q$.[3] Since one can always reduce in this manner, it makes good sense when studying connections on principal bundles to restrict oneself to irreducible connections.

[3] Technically, one should require H to be a closed Lie subgroup of G in order for Q to be closed submanifold of P.

2.3 Gauge transformations

§ Local gauge transformations

When working with principal bundles it is often convenient or necessary to work in a given local trivialization. To work with connections and curvature in this setting one must know how these objects transform under a change of trivialization. Such transformations are called **local gauge transformations**.

Definition 2.18. Let $P \xrightarrow{\pi} M$ be a principal bundle with connection ω and curvature Ω. Let $(\{U_i\}, \{s_i\}, \{\phi_i\})$ be a canonical local trivialization ($s_i \colon U_i \to \pi^{-1}(U_i)$ are local sections and $\phi_i \colon \pi^{-1}(U_i) \to U_i \times G$ are the associated trivialization maps). A choice of such a trivialization is called a **local gauge**. On each U_i we define the **local gauge potential**, $A_i = s_i^* \omega$, and **local field strength**, $F_i = s_i^* \Omega$, as the pullbacks of the connection and curvature respectively.

On a nonempty overlap $U_i \cap U_j$ we have two local gauge potentials and field strengths and need to know how they are related. Define the transition function $t_{ij} \colon U_i \cap U_j \to G$ by $s_j(x) = s_i(x) t_{ij}(x)$. It is easy to see that these transition functions satisfy a cocycle condition: $t_{ij} t_{jk} = t_{ik}$, as well as $t_{ij} = t_{ji}^{-1}$ and $t_{ii} = 1$.

Theorem 2.19. *With the above definitions, on the overlap $U_i \cap U_j$ the gauge potentials are related by*

$$A_j = \operatorname{Ad}_{t_{ij}^{-1}} A_i + t_{ij}^* \Theta, \tag{12}$$

where Θ is the Maurer-Cartan form on G. Likewise, the gauge field strengths are given by

$$F_j = \operatorname{Ad}_{t_{ij}^{-1}} F_i. \tag{13}$$

The transformation law for the local field strength follows quite easily from the fact that the curvature Ω may be regarded as 2-form on M with values in the adjoint bundle. The gauge potential on the other hand has an additional term, $t_{ij}^* \Theta$, which is just the Darboux derivative of the transition function t_{ij}.

It is useful to turn this picture around and start with a set of smooth $\mathfrak{g}$-valued 1-forms of M transforming as in (12) and ask if there is a connection on P for which these forms are the local potentials. In fact, there is a unique such connection. We will not need the explicit construction (the details are rather tedious) so we do not work it out here—it is enough to know that it can be done. Hence, we can always work locally once we have specified a local trivialization. Locally, the field strength can be calculated from the potential just as in (6):

$$F_i = dA_i + \frac{1}{2}[A_i, A_i]. \tag{14}$$

§ Global gauge transformations

In standard differential geometry there is a notion of *local* coordinate changes arising from overlapping coordinate charts, as well as *global* coordinate changes arising from a diffeomorphism. Locally, a diffeomorphism looks just like a change of coordinate charts. There is a similar analogy in gauge theory between a local gauge transformation (a choice of section) and a global one. This leads to the following definition:

Definition 2.20. Let $P \to M$ be a principal G-bundle. A **gauge transformation** of P is a bundle automorphism $F\colon P \to P$ which is fiber preserving and commutes with the action of G. That is,

1. $\pi \circ F = \pi$
2. $F(p \cdot g) = F(p) \cdot g$

The group of all gauge transformations of P, denoted $\mathcal{G}_P$, is called the **gauge transformation group**.

There are a few alternative ways of viewing gauge transformations which are sometimes convenient. Consider the action of G on itself by conjugation, $\Psi\colon G \to \mathrm{Aut}(G)$: $\Psi_g(h) = ghg^{-1}$. With this representation one can construct the associated bundle $P \times_\Psi G$.[4] The sections of $P \times_\Psi G$ can be regarded, in the usual way, as functions $f\colon P \to G$ which are right equivariant with respect to conjugation: $f(p \cdot g) = g^{-1} f(p) g$. The set of all such sections forms a group under pointwise multiplication: $(f_1 f_2)(p) = f_1(p) f_2(p)$.

Lemma 2.21. *Let $f \in \Gamma(P \times_\Psi G)$. Define $\Phi(f)\colon P \to P$ by*

$$\Phi(f)(p) = p \cdot f(p).$$

Then $\Phi(f)$ is a gauge transformation and $\Phi\colon \Gamma(P \times_\Psi G) \to \mathcal{G}_P$ is an isomorphism.

Proof. To show that $\Phi(f)$ is a gauge transformation we must show that it preserves fibers and commutes with the action of G. By construction, we see that $\Phi(f)$ clearly preserves fibers. Now, $\Phi(f)(p \cdot g) = (p \cdot g) \cdot f(p \cdot g) = p \cdot g g^{-1} f(p) g = p \cdot f(p) g = \Phi(f)(p) \cdot g$. So $\Phi(f)$ is a gauge transformation. Note Φ is a group homomorphism: $\Phi(f_1 f_2)(p) = p \cdot f_1(p) f_2(p) = \Phi(f_1)(p) \cdot f_2(p) = \Phi(f_1)(p \cdot f_2(g)) = \Phi(f_1)(\Phi(f_2)(p))$. We can explicitly construct an inverse for Φ by setting $\Phi^{-1}(F)(p) = g$ where g is given by $F(p) = p \cdot g$. □

We then have three equivalent ways of viewing a gauge transformation: a bundle automorphism $F\colon P \to P$, a section of $P \times_\Psi G$, or a map $f\colon P \to G$ satisfying $f(p \cdot g) = g^{-1} f(p) g$.

Gauge transformations act on the forms by pullback. That is, the gauge transformation of the connection ω is just $F^*\omega$, and likewise for the curvature.

Theorem 2.22. *Let P be a principal G-bundle with connection ω and curvature Ω. Let $f\colon P \to G$ be a gauge transformation of P. Then*

$$\Phi(f)^*\omega = \mathrm{Ad}_{f^{-1}} \omega + f^*\Theta, \tag{15}$$

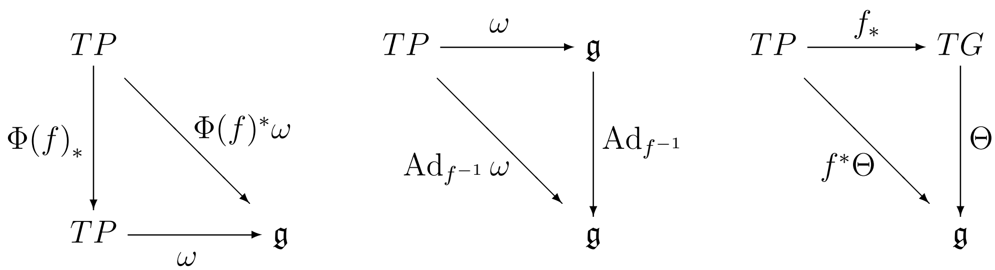
and

$$\Phi(f)^*\Omega = \mathrm{Ad}_{f^{-1}} \Omega. \tag{16}$$

Furthermore, $\Phi(f)^*\omega$ is a connection on P with curvature $\Phi(f)^*\Omega$.

[4] In general, this is neither a vector bundle nor a principal bundle; for even though the fibers are copies of G, the structure group is the inner automorphism group, $\mathrm{Inn}(G)$.

This is just the global version of Theorem 2.19, and provides justification for our definition of a gauge transformation. The meaning of equation (15) is perhaps elucidated by the following diagrams:



Theorem 2.23. *The action of $\mathcal{G}_P$ on $\mathcal{A}^k(P,V)$ restricts to an action on $\bar{\mathcal{A}}^k(P,V)$. That is, for $\alpha \in \bar{\mathcal{A}}^k(P,V)$ and $F \in \mathcal{G}_P$ we have $F^*\alpha \in \bar{\mathcal{A}}^k(P,V)$.*

If we regard the group of gauge transformations, $\mathcal{G}_P$, as sections of $P \times_\Psi G$ then it makes sense to regard sections of the adjoint bundle, $\mathfrak{g}_P = P \times_{\mathrm{Ad}} \mathfrak{g}$, as the "Lie algebra" of the infinite-dimensional "Lie group" $\mathcal{G}_P$. We can make this relation more precise by defining an exponential map $\exp\colon \bar{\mathcal{A}}^0(P,\mathfrak{g}) \to \mathcal{G}_P$ by

$$\exp(\theta)(p) = \exp(\theta(p)), \tag{17}$$

where the exponential on the right hand side is the normal map $\exp\colon \mathfrak{g} \to G$. One can then formulate an infinitesimal gauge transformation:

$$\delta_\theta \alpha \equiv \frac{d}{dt}\left[\Phi(\exp(t\theta))^*\alpha\right]_{t=0}, \tag{18}$$

for $\theta \in \bar{\mathcal{A}}^0(P,\mathfrak{g})$ and $\alpha \in \mathcal{A}^k(P,V)$. The infinitesimal gauge transformation of the connection is then given by the following.

Lemma 2.24. *Let P be a principal G-bundle with connection ω and let $\theta \in \bar{\mathcal{A}}^0(P,\mathfrak{g})$. Then*

$$\delta_\theta\omega = D_\omega\theta = d\theta + [\omega,\theta]. \tag{19}$$

Proof. Let $f = \exp(t\theta)$, then $\delta_\theta\omega = \frac{d}{dt}[\Phi(f)^*\omega]$ which by (15) is

$$\delta_\theta\omega = \frac{d}{dt}[\mathrm{Ad}_{f^{-1}}\,\omega] + \frac{d}{dt}[f^*\Theta].$$

Now $\frac{d}{dt}[\mathrm{Ad}_{f^{-1}}\,\omega] = -\operatorname{ad}_\theta\omega = [\omega,\theta]$ and $\frac{d}{dt}[f^*\Theta] = \frac{d}{dt}[e^{-t\theta}de^{t\theta}] = d\theta$. The result then follows from Lemma 2.13. □

3 Self-Duality

We can easily extend the Hodge star operator on a Riemannian manifold M to act on k-forms with values in some vector bundle $E \to M$. For $s \otimes \alpha \in \Gamma(E \otimes \Lambda^k T^*M)$ define

$$*(s \otimes \alpha) = s \otimes (*\alpha) \tag{20}$$

This allows us to extend the notion of (anti-)self-duality to vector-valued forms on a Riemannian manifold.

Definition 3.1. Let $P \to M$ be a principal G-bundle over a Riemannian 4-manifold M. A connection ω on P is said to be **self-dual** if its curvature (considered as a 2-form on M with values in the adjoint bundle) is. That is, if $*\Omega = \Omega$. Likewise, a connection is said to be **anti-self-dual** if $*\Omega = -\Omega$. These connections also go by the name **instantons** and **anti-instantons** respectively.

3.1 Yang-Mills theory

Let $P \to M$ be principal G-bundle over a compact Riemannian four-manifold M with connection ω and curvature Ω. This defines a Euclidean Yang-Mills theory with gauge group G. The action for this theory, considered as a functional of the connection ω, is given by

$$S = -\frac{1}{4}\int_M \operatorname{tr}(\Omega \wedge *\Omega) \tag{21}$$

where the trace is taken to mean the trace in the adjoint representation of the Lie algebra, i.e. given a basis $\{e_1, \cdots, e_n\}$ for the Lie algebra $\mathfrak{g}$ define $\operatorname{tr}(\Omega \wedge *\Omega) = \operatorname{tr}(\operatorname{ad}_{e_i} \circ \operatorname{ad}_{e_j})\Omega^i \wedge (*\Omega)^j$. This is essentially just the Killing form on the Lie algebra $\mathfrak{g}$.

The interesting thing to note about this action is that its absolute minima are given by self-dual connections. To see this note that we can always decompose the curvature 2-form into a self-dual and an anti-self-dual piece: $\Omega = \Omega_+ + \Omega_-$. The Yang-Mills action is then given by

$$S = -\frac{1}{4}\int_M \operatorname{tr}(\Omega \wedge \Omega_+) + \frac{1}{4}\int_M \operatorname{tr}(\Omega \wedge \Omega_-)$$

Now compare this to the first Pontrjagin class of the adjoint bundle:

$$\begin{aligned} p_1(\mathfrak{g}_P)[M] &= -\frac{1}{8\pi^2}\int_M \operatorname{tr}(\Omega \wedge \Omega) \\ &= -\frac{1}{8\pi^2}\int_M \operatorname{tr}(\Omega \wedge \Omega_+) - \frac{1}{8\pi^2}\int_M \operatorname{tr}(\Omega \wedge \Omega_-) \end{aligned}$$

We see that $S/(2\pi^2) \geq p_1(\mathfrak{g}_P)[M]$ with equality iff $\Omega_- = 0$. Since $p_1(\mathfrak{g}_P)[M]$ is a topological invariant we can conclude that the minima of the Yang-Mills action are exactly the self-dual connections on P.

4 Moduli Space of Self-Dual Connections

Given a principal G-bundle $P \to M$ we can ask the question: "How many different connections are there on P?" It is useful to restrict this question in two ways. First, we restrict ourselves to *irreducible* connections (see Defn. 2.17). Second, when we say *different* connections we should really restrict ourselves to connections which are not related by a gauge transformation. For if we are given one connection on P we can form (infinitely many) other connections by pulling back under gauge transformations. However, we would like to regard all of these as equivalent.

Definition 4.1. Let $P \to M$ be a principal G-bundle. Let $\mathcal{C}(P)$ denote the space of all connections on P and let $\mathcal{C}(P)_0$ be the subspace of all irreducible connections. Define the **moduli space** of irreducible connections on P as the quotient $\mathcal{M} = \mathcal{C}(P)_0/\mathcal{G}_P$, where $\mathcal{G}_P$ is the group of gauge transformations of P.

If M is a Riemannian four-manifold let $\mathcal{C}^{\pm}(P)$ be the space of (anti)self-dual connections on P and define the moduli space of irreducible (anti)self-dual connections on P as the quotient $\mathcal{M}^{\pm} = \mathcal{C}^{\pm}(P)_0/\mathcal{G}_P$.

It turns out that, under suitable conditions, the moduli space $\mathcal{M}^{\pm}$, if non-empty, is actually a finite-dimensional smooth manifold with a computable dimension. This was first shown in 1978 by Atiyah, Hitchin, and Singer [AHS] who used the index theorem to compute the dimension. The precise statement of their result is as follows[5,6]:

Theorem 4.2 (Atiyah, Hitchin, and Singer). *Let $P \to M$ be a principal G-bundle over a compact self-dual Riemannian 4-manifold M with positive scalar curvature where G is a compact semisimple Lie group. Then the moduli space $\mathcal{M}^+$ of irreducible, self-dual connections on P is either empty or a manifold of dimension*

$$2p_1(\mathfrak{g}_P)[M] - \frac{1}{2}(\chi - \tau) \dim G,$$

where $p_1(\mathfrak{g}_P)$ is the first Pontrjagin class of the adjoint bundle, and χ and τ are the Euler characteristic and Hirzebruch signature of M respectively.

The proof of Theorem 4.2 is in three steps:

1. *Infinitesimal.* Construct the tangent space to $\mathcal{M}^+$ and compute its dimension using the Atiyah-Singer index theorem together with a vanishing theorem.

2. *Local.* Show that the infinitesimal deformations can be integrated to obtain a local moduli space.

3. *Global.* Show that these local moduli spaces patch together to form a Hausdorff manifold.

We will not prove the theorem here, but rather focus on the construction in step 1. Even here we shall leave out some details. The complete proof can be found in [AHS] as well as in [BB] (in greater detail).

Recall that $\mathcal{C}(P)$, the space of connection 1-forms on P, is an affine space modeled on $\bar{A}^1(P, \mathfrak{g})$. That is, any two connections $\omega, \omega' \in \mathcal{C}(P)$ differ by an element of $\bar{A}^1(P, \mathfrak{g})$. Let us assume that there exists at least one self-dual connection ω_0 on P. Consider a one-parameter family of connections through ω_0 given by $\omega_t = \omega_0 + \alpha(t)$ where $\alpha(t)$ is a curve in $\bar{A}^1(P, \mathfrak{g})$ with $\alpha(0) = 0$. Differentiating this curve at $t = 0$ gives an element $\dot{\alpha} = \frac{d}{dt}\alpha|_{t=0}$ of the tangent

[5] We should remark that a number of the conditions in Theorem 4.2 can be weakened in many circumstances (e.g. "mildly-irreducible" instead of irreducible, etc.), but we will not trouble ourselves with these details here.

[6] Note that our normalization of p_1 differs from [AHS] by a factor of 2.

space to $\mathcal{C}(P)$ at ω_0. (We use the term *tangent space* rather loosely here since $\mathcal{C}(P)$ is, in general, infinite-dimensional). In order to construct the tangent space to $\mathcal{M}^+$ at ω_0 we need to answer two questions: (1) When is $\dot{\alpha}$ tangent to a slice through self-dual connections, and (2) When is $\dot{\alpha}$ tangent to a slice though gauge equivalent connections?

1. Let Ω_t and Ω_0 be the curvatures of ω_t and ω_0. Then

$$\begin{aligned}\Omega_t &= d\omega_t + \frac{1}{2}[\omega_t, \omega_t] \\ &= d\omega_0 + d\alpha + \frac{1}{2}[\omega_0, \omega_0] + [\omega_0, \alpha] + \frac{1}{2}[\alpha, \alpha] \\ &= \Omega_0 + D_{\omega_0}\alpha + \frac{1}{2}[\alpha, \alpha].\end{aligned}$$

   So that $\Omega_t - \Omega_0 = D_{\omega_0}\alpha + \frac{1}{2}[\alpha, \alpha]$. Now ω_t is a curve through self-dual connections iff $p_-\Omega_t = 0$ where $p_- = \frac{1}{2}(1 - *)$ is the projection[7] onto the anti-self-dual 2-forms. This is true iff

$$p_-\left(D_{\omega_0}\alpha + \frac{1}{2}[\alpha, \alpha]\right) = 0.$$

   Taking the derivative at $t = 0$ gives

$$\begin{aligned}\frac{d}{dt}p_-\left(\Omega_t - \Omega_0\right)|_{t=0} &= p_-\left(D_{\omega_0}\dot{\alpha} + \frac{1}{2}[\dot{\alpha}, \alpha] + \frac{1}{2}[\alpha, \dot{\alpha}]\right)\Bigg|_{t=0} \\ &= p_- D_{\omega_0}\dot{\alpha}\end{aligned}$$

   since $\alpha(0) = 0$. So $\dot{\alpha}$ is tangent to a slice through self-dual connections iff it lies in the kernel of $p_-D_{\omega_0}$

$$\dot{\alpha} \in \ker(p_-D_{\omega_0}).$$

2. Now suppose that ω_t is a slice through gauge equivalent connections so that $\omega_t = F_t^*\omega_0$ for some $F_t \in \mathcal{G}_P$. By Lemma 2.24 we see that $\dot{\alpha} = D_{\omega_0}\theta$ for some $\theta \in \bar{\mathcal{A}}^0(P, \mathfrak{g})$. So $\dot{\alpha}$ is tangent to a slice through self-dual connections iff it lies in the image of D_{ω_0}.

Thus $\dot{\alpha}$ is tangent to a curve through inequivalent self-dual connections precisely when[8] $\dot{\alpha} \in \ker(p_-D_\omega)/\operatorname{im}(D_\omega)$. We should note that for $\theta \in \bar{\mathcal{A}}^0(P, \mathfrak{g})$ we have $p_-D_\omega(D_\omega\theta) = p_-D_\omega^2\theta = p_-[\Omega, \theta] = 0$ since Ω is self-dual by assumption, so $\operatorname{im}(D_\omega) \subseteq \ker(p_-D_\omega)$. The tangent space to $\mathcal{M}^+$ can then be identified with the first cohomology group of the following elliptic complex:

$$0 \to \bar{\mathcal{A}}^0(P, \mathfrak{g}) \xrightarrow{D_\omega} \bar{\mathcal{A}}^1(P, \mathfrak{g}) \xrightarrow{p_-D_\omega} \bar{\mathcal{A}}^2_-(P, \mathfrak{g}) \to 0. \tag{22}$$

The meaning of this is perhaps more clear if we use the isomorphism $\bar{\mathcal{A}}^k(P, \mathfrak{g}) \cong \mathcal{A}^k(M, \mathfrak{g}_P)$ to write this as

$$0 \to \mathcal{A}^0(M, \mathfrak{g}_P) \xrightarrow{d_\nabla} \mathcal{A}^1(M, \mathfrak{g}_P) \xrightarrow{p_-d_\nabla} \mathcal{A}^2_-(M, \mathfrak{g}_P) \to 0, \tag{23}$$

[7] Here $*$ is computed on $\bar{\mathcal{A}}^2(P, \mathfrak{g})$ using the established isomorphism $\bar{\mathcal{A}}^2(P, \mathfrak{g}) \cong \mathcal{A}^2(M, \mathfrak{g}_P)$

[8] We will drop the 0 on ω_0 from here on.

where d_∇ are the induced covariant derivatives on $\mathcal{A}^k(M, \mathfrak{g}_P)$. This complex is sometimes called the *twisted anti-self-dual complex* (twisted because the forms take values in $\mathfrak{g}_P$ rather than $\mathbb{R}$). If we write the dimensions of the cohomology groups of this complex as h_0, h_1, h_2 then by the above arguments $\dim \mathcal{M}^+ = h_1$. The goal then is compute the index, $h_0 - h_1 + h_2$, of this complex using the index theorem and then show that $h_0 = h_2 = 0$ under the assumptions mentioned in Theorem 4.2.

We shall not argue here as to why h_0 and h_2 vanish other than to mention that the vanishing of h_0 makes use of the irreducibility of ω and the semi-simplicity of G. The vanishing of h_2 follows from the fact that M is self-dual with positive scalar curvature. The reader should consult [AHS] or [BB] for the details.

Before considering the index of the complex (23) we should try and calculate the index of the untwisted anti-self-dual complex.

Lemma 4.3. *Let M be a Riemannian four-manifold. The index of the anti-self-dual complex,*

$$0 \to \mathcal{A}^0(M) \xrightarrow{d} \mathcal{A}^1(M) \xrightarrow{p_- d} \mathcal{A}^2_-(M) \to 0, \tag{24}$$

is $\frac{1}{2}(\chi - \tau)$ where χ is the Euler characteristic of M and τ is the Hirzebruch signature.

Proof. Using Poincaré duality we can write $\chi = 2b^0 + 2b^1 + b^2_+ + b^2_-$ and $\tau = b^2_+ - b^2_-$. Thus

$$\frac{1}{2}(\chi - \tau) = b^0 - b^1 + b^2_-.$$

What remains to show is that b^0, b^1, and b^2_- are the dimensions of the cohomology groups in (24). We leave this as an exercise. □

The idea now is compute the index of the twisted anti-self-dual complex using the Atiyah-Singer index theorem. One proceeds by folding (23) into the two term complex[9]

$$d^*_\nabla + p_- d_\nabla \colon \mathcal{A}^1(M, \mathfrak{g}_P) \to \mathcal{A}^0(M, \mathfrak{g}_P) \oplus \mathcal{A}^2_-(M, \mathfrak{g}_P). \tag{25}$$

The Dirac operator for this complex is $\mathcal{D} \colon \Gamma(S^+ \otimes S^- \otimes \mathfrak{g}_P) \to \Gamma(S^- \otimes S^- \otimes \mathfrak{g}_P)$, where $S^\pm$ are the bundles of positive and negative chirality spinors on M. The index theorem then gives

$$\begin{aligned}
\operatorname{index} \mathcal{D} &= \int_M \hat{A}(M) \operatorname{ch}(S^- \otimes \mathfrak{g}_P) \\
&= \int_M \hat{A}(M) \operatorname{ch}(S^-) \operatorname{ch}(\mathfrak{g}_P) \\
&= \dim(\mathfrak{g}_P) \int_M \hat{A}(M) \operatorname{ch}(S^-) + \dim(S^-) \int_M p_1(\mathfrak{g}_P) \\
&= (\operatorname{index} \mathcal{D}_0) \dim G + 2p_1(\mathfrak{g}_P)[M]
\end{aligned}$$

where $\mathcal{D}_0 \colon \Gamma(S^+ \otimes S^-) \to \Gamma(S^- \otimes S^-)$ is (minus) the Dirac operator associated with the untwisted anti-self-dual complex (24). By Lemma 4.3 we then have

$$\dim \mathcal{M}^+ = \operatorname{index} \mathcal{D} = 2p_1(\mathfrak{g}_P)[M] - \frac{1}{2}(\chi - \tau) \dim G \tag{26}$$

as desired.

[9] We've also reversed the sign.

5 Instantons on S^4

Instantons were first studied in [BPST] as self-dual solutions to the Yang-Mills equations on $\mathbb{R}^4$. In order for the Yang-Mills action to be finite on this noncompact space, the gauge field strength F (or curvature) must approach zero at infinity. It is sufficient, but not necessary, to demand that the gauge potential itself approach zero at infinity. One need only demand the potential asymptotically approach something *gauge equivalent* to zero. A potential which is gauge equivalent to zero is said to be **pure gauge**. The field strength of a pure gauge potential is zero. For compact simple Lie groups it turns out that there are always nontrivial solutions with this asymptotic behavior. It was later realized that many of these solutions could be extended smoothly to S^4 by adding the point at infinity. These solutions then turn out to be self-dual gauge fields on principal G-bundles over S^4.

In the next few sections we work out the simplest nontrivial example (studied in [BPST]) of an instanton on a principal SU(2)-bundle over S^4. Actually, we will find it more convenient to work with the group of unit quaternions, Sp(1), which is isomorphic to SU(2).

5.1 Principal Sp(1)-bundles over S^4

Our atlas for S^4 consists of two charts U_N, U_S describing the northern and southern hemispheres respectively. We take the coordinate functions on these patches to be

$$x = \phi_N(u_0, \cdots, u_4) = \frac{1}{1+u_0}(u_1 + iu_2 + ju_3 + ku_4) \tag{27}$$

$$y = \phi_S(u_0, \cdots, u_4) = \frac{1}{1-u_0}(u_1 - iu_2 - ju_3 - ku_4) \tag{28}$$

where $S^4 = \{u \in \mathbb{R}^5 \mid \sum u_i^2 = 1\}$ and we have identified $\mathbb{R}^4$ with the quaternions $\mathbb{H}$. On the overlap $U_N \cap U_S$—which includes everything but the north pole ($x = 0$) and the south pole ($y = 0$)—the change of coordinates is given by $y = x^{-1} = \bar{x}/|x|^2$.

In terms of transition functions it is rather easy to describe the principal Sp(1)-bundles over S^4. Since any principal bundle over $\mathbb{R}^n$ must be trivial, our coordinate charts $U_{N,S}$ will always be trivializing neighborhoods for the bundle, with essentially one transition function between them. This transition function will determine the bundle up to equivalence. When restricted to the equator of S^4 the transition function must be a smooth map $t\colon S^3 \to \mathrm{Sp}(1)$. Since Sp(1) is topologically a 3-sphere, t is determined up to homotopy by its degree $k \in \mathbb{Z}$. For concreteness we take the transition functions to be

$$t_{NS}(x) = \frac{x^k}{|x|^k} = \frac{\bar{y}^k}{|y|^k} = t_{NS}(y) \tag{29}$$

$$t_{SN}(x) = \frac{\bar{x}^k}{|x|^k} = \frac{y^k}{|y|^k} = t_{SN}(y) \tag{30}$$

The principal Sp(1)-bundles are then in 1-1 correspondence with the integers $\mathbb{Z}$. In the next section we will concentrate on the simplest nontrivial case: $k = 1$. Though we will not need it, we should mention that this bundle is essentially the quaternionic Hopf bundle $\mathrm{Sp}(1) \hookrightarrow S^7 \to S^4$.

For each of these principal Sp(1)-bundles we can use Theorem 4.2 to compute the dimension of instanton moduli space. Using the fact that $\chi(S^4) = 2$, $\tau(S^4) = 0$, and $\dim(\mathrm{Sp}(1)) = 3$ we see that the dimension of the moduli space, if non-empty, is given by $2p_1(\mathfrak{g}_P)[S^4] - 3$. Now it turns out that the first Pontrjagin class of the adjoint bundle associated to the kth Sp(1)-principal bundle is given by

$$p_1(\mathfrak{g}_P)[S^4] = -c_2(\mathfrak{g}_P \otimes \mathbb{C})[S^4] = 4k. \tag{31}$$

It also turns out that for $k > 0$ there always exist self-dual solutions so that the dimension of the moduli space is $\dim \mathcal{M}^+ = 8k - 3$. The integer k is called the **instanton number** of the bundle. The $k < 0$ principal bundles turn out to admit an $8k - 3$ parameter family of anti-self-dual solutions.

5.2 Self-dual connections for $k = 1$

In order to find self-dual connections on principal bundles over S^4 it is useful to think of S^4 as the one-point compactification of $\mathbb{R}^4 \cong U_N$ and look for gauge potentials which minimize the Yang-Mills action on $\mathbb{R}^4$. As discussed above, if the action is even to be finite we must demand that the gauge potential asymptotically approaches pure gauge at infinity (the south pole). That is, we will look for a local gauge potential of the form

$$A_N = f(r) t_{SN}^* \Theta \tag{32}$$

where $\lim_{r\to\infty} f(r) = 1$. We should also demand $f(0) = 0$ if our potential is to be smooth at the origin. Assuming we can find such a potential, we can then determine the potential on U_S by the gauge transformation law

$$A_S = \mathrm{Ad}_{t_{NS}^{-1}} A_N + t_{NS}^* \Theta. \tag{33}$$

§ Pure gauge potentials

To pursue this line of attack we must first describe the pure gauge potential $t_{SN}^*\Theta$. This is the just the pullback of the Maurer-Cartan form on Sp(1) to $\mathbb{H}^\times = \mathbb{H} \setminus \{0\}$ via the projection map $t_{SN} \colon \mathbb{H}^\times \to \mathrm{Sp}(1)$ given by $t_{SN}(x) = \bar{x}/|x|$ (we are taking $k = 1$). We compute this in two steps by taking $t_{SN} = \iota \circ \mathrm{proj}$ where $\mathrm{proj} \colon x \mapsto x/|x|$ is the standard projection onto the unit quaternions and ι is inversion in $\mathbb{H}^\times$.

1. There is a theorem (Proposition 3.4.10 of [Sha]) that says the pullback of the the Maurer-Cartan form of any Lie group under inversion is given by

   $$\iota^*\Theta_g = -\mathrm{Ad}_g \Theta_g. \tag{34}$$

   For a matrix Lie group, if we write the Maurer-Cartan form as $\Theta = g^{-1}dg$ then we have $\iota^*\Theta = -dgg^{-1}$.

2. According to the Theorem 2.8 the pullback of $\mathrm{proj} \colon \mathbb{H}^\times \to \mathrm{Sp}(1)$ is given by

   $$\mathrm{proj}^* \Theta_{\mathrm{Sp}(1)} = (\mathrm{proj}_*)_e \Theta_{\mathbb{H}^\times} = \mathrm{Im}(\Theta_{\mathbb{H}^\times}) = \mathrm{Im}(x^{-1}dx). \tag{35}$$

3. We calculate the pullback of the composition $(\iota \circ \mathrm{proj})$ by another application of Theorem 2.8 this time taking the reverse multiplication law on $\mathbb{H}^\times$. This result is

$$(\iota \circ \mathrm{proj})^*\Theta = (\iota \circ \mathrm{proj})_{*e}\Theta'_{\mathbb{H}^\times} = -(\iota \circ \mathrm{proj})_{*e}\,\mathrm{Ad}\,\Theta_{\mathbb{H}^\times} = -\,\mathrm{Im}(dx x^{-1}). \tag{36}$$

For $x \in \mathbb{H}^\times$ the inverse is given by $x^{-1} = \bar{x}/|x|^2$ and we may write

$$-\,\mathrm{Im}(dx x^{-1}) = -\frac{1}{|x|^2}\,\mathrm{Im}(dx\bar{x}) = \frac{1}{|x|^2}\,\mathrm{Im}(x d\bar{x}). \tag{37}$$

Thus on $U_N \cap U_S \cong \mathbb{H}^\times$ we may describe the pure gauge potential by

$$t_{SN}^*\Theta = \frac{1}{|x|^2}\,\mathrm{Im}(x d\bar{x}). \tag{38}$$

Note that this is not well-defined at the origin, so it cannot be smoothly extended to the whole of U_N. Indeed, we cannot consistently define a pure gauge potential on all of S^4 with the given bundle transition functions (This is possible only for the trivial bundle with $k = 0$).

§ Curvature of a pure gauge potential

Since a pure gauge potential is, by definition, gauge equivalent to the zero potential we know that the local field strength of any such potential vanishes. Computing this for the pure gauge potential given above gives us a useful relation.

Let $\alpha = x d\bar{x}$ be the $\mathbb{H}$-valued 1-form appearing in the potential $A = \frac{1}{r^2}\,\mathrm{Im}(\alpha)$ where $r = |x|$. The field strength is then given by

$$F = dA + \frac{1}{2}[A, A] = 0. \tag{39}$$

We leave it as an exercise for the reader to show that

$$d\,\mathrm{Im}(\alpha) = \mathrm{Im}(d\alpha) = d\alpha = dx \wedge d\bar{x} \tag{40}$$

and

$$\frac{1}{2}[\mathrm{Im}(\alpha), \mathrm{Im}(\alpha)] = \mathrm{Im}(\alpha \wedge \alpha) = \alpha \wedge \alpha \tag{41}$$

where the wedges in these expressions are computed with respect to quaternionic multiplication. Using these we can write

$$\begin{aligned} dA &= \frac{1}{r^2} d(\mathrm{Im}\,\alpha) + d\left(\frac{1}{r^2}\right) \wedge \mathrm{Im}(\alpha) \\ &= \frac{1}{r^2} d\alpha - \frac{2}{r^4}\,\mathrm{Re}(\alpha) \wedge \mathrm{Im}(\alpha) \end{aligned}$$

and

$$\frac{1}{2}[A, A] = \frac{1}{r^4}\alpha \wedge \alpha.$$

Equation (39) then gives

$$\mathrm{Re}(\alpha) \wedge \mathrm{Im}(\alpha) - \frac{1}{2}(\alpha \wedge \alpha) = \frac{r^2}{2} d\alpha. \tag{42}$$

This can be verified directly, although the calculation is somewhat tedious. The interesting thing about the $\mathrm{Im}\,\mathbb{H}$-valued 2-form $d\alpha$ is that it is self-dual! That is, $*d\alpha = d\alpha$. In standard coordinates it can be written

$$\begin{aligned} d\alpha = dx \wedge d\bar{x} = &-2(dx^1 \wedge dx^2 + dx^3 \wedge dx^4)i \\ &-2(dx^1 \wedge dx^3 + dx^4 \wedge dx^2)j \\ &-2(dx^1 \wedge dx^4 + dx^2 \wedge dx^3)k. \end{aligned} \tag{43}$$

In the next section we will compute a field strength proportional to this form.

§ Self-dual equations

We now proceed to look for a potential which is asymptotically pure gauge and whose field strength is self-dual. That is, we look for potential of the form

$$A = \frac{f(r)}{r^2}\,\mathrm{Im}\,\alpha$$

To determine the field strength we compute

$$\begin{aligned} dA &= f d\left(\frac{1}{r^2}\,\mathrm{Im}\,\alpha\right) + df \wedge \frac{1}{r^2}\,\mathrm{Im}\,\alpha \\ &= \frac{f}{r^2} d\alpha - \frac{1}{r^4}\left(2f - r\frac{\partial f}{\partial r}\right)\mathrm{Re}(\alpha) \wedge \mathrm{Im}(\alpha) \end{aligned}$$

and

$$\frac{1}{2}[A, A] = \frac{f^2}{r^4} \alpha \wedge \alpha.$$

We see from (42) that if

$$2f - r\frac{\partial f}{\partial r} = 2f^2 \tag{44}$$

the entire curvature will be proportional to $d\alpha$ and so self-dual. The solution to this equation is

$$f(r) = \frac{r^2}{r^2 + \lambda^2} \tag{45}$$

which we see satisfies our requirements of vanishing at the origin and approaching 1 at ∞. Here λ is an arbitrary constant describing the strength of the instanton. With this one-parameter family of solutions, the gauge potential and self-dual field strength are given by

$$A = \frac{1}{r^2 + \lambda^2}\,\mathrm{Im}(x d\bar{x}) \tag{46}$$

$$F = \frac{\lambda^2}{(r^2 + \lambda^2)^2}\, dx \wedge d\bar{x}. \tag{47}$$

§ The connection on the principal bundle

In order to verify if our potential comes from a connection on the $k = 1$ principal bundle over S^4 we need to gauge transform to the southern hemisphere and verify that everything patches together nicely. In the U_N coordinates x our gauge potential is given by

$$A_N = \frac{|x|^2}{|x|^2 + \lambda^2}\, t_{SN}^* \Theta. \tag{48}$$

In the southern hemisphere we have

$$\begin{aligned} A_S &= \mathrm{Ad}_{t_{NS}^{-1}} A_N + t_{NS}^* \Theta \\ &= \frac{|x|^2}{|x|^2 + \lambda^2}\, \mathrm{Ad}_{t_{NS}^{-1}} t_{SN}^* \Theta + t_{NS}^* \Theta \\ &= -\frac{|x|^2}{|x|^2 + \lambda^2}\, t_{NS}^* \Theta + t_{NS}^* \Theta \\ &= \frac{\lambda^2}{|x|^2 + \lambda^2}\, t_{NS}^* \Theta \\ &= \frac{|y|^2}{|y|^2 + \lambda^{-2}}\, t_{NS}^* \Theta. \end{aligned}$$

In the third step above we have made use of the identity $\mathrm{Ad}_f(f^*\Theta) = -(f^{-1})^*\Theta$ applicable to any smooth map $f\colon M \to G$. We see that in the U_S coordinates y the local gauge potential has precisely the same form as in the northern hemisphere but with the inverse value of λ. In particular, the local gauge potentials $A_{N,S}$ are smooth (and vanishing) at the north and south poles respectively. Since these local gauge potentials are related be a smooth gauge transformation on the overlap they must come from a unique connection on the principal bundle. This connection is self-dual since its curvature in local coordinates on S^4 is given by

$$F = \frac{\lambda^2}{(|x|^2 + \lambda^2)^2}\, dx \wedge d\bar{x}. \tag{49}$$

We have then found a one-parameter family of self-dual connections on the $k = 1$ principal bundle over S^4. In fact, it is easy to modify these equations to obtain the full 5 parameter family of solutions by taking the origin of the instanton to be at an arbitrary point x_0 on U_N (rather than at the north pole):

$$A_N = \mathrm{Im}\left(\frac{x - x_0}{|x - x_0|^2 + \lambda^2}\, d\bar{x} \right) \tag{50}$$

$$F = \frac{\lambda^2}{(|x - x_0|^2 + \lambda^2)^2}\, dx \wedge d\bar{x}. \tag{51}$$

The Atiyah-Hitchin-Singer theorem guarantees us that we have found them all.

References

[AHS] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London **A362** (1978), 425–461.

[BB] B. Booss and D. D. Bleecker, *Topology and analysis: The Atiyah-Singer index formula and gauge-theoretic physics*, Universitext, Springer-Verlag, New York, 1985.

[BPST] A. Belavin, A. Polyakov, A. Schwartz, and Y. Tyupkin, *Pseudoparticle solutions of the Yang-Mills equations*, Phys. Lett. **59B** (1975), 85–87.

[FU] D. S. Freed and K. K. Uhlenbeck, *Instantons and four-manifolds*, 2nd ed., Springer-Verlag, New York, 1991.

[Nab] G. L. Naber, *Topology, geometry, and gauge fields: Foundations*, Texts in Applied Mathematics, Springer-Verlag, New York, 1997.

[Sha] R. W. Sharpe, *Differential geometry: Cartan's generalization of Klein's Erlangen program*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1997.