

Topological Aspects of Yang-Mills Theory

M. F. Atiyah¹ and J. D. S. Jones²

¹ Mathematical Institute, University of Oxford, Oxford OX1 3LB, England

² Magdalen College, Oxford OX1 3LB, England

Abstract. The space of maps $S^3 \rightarrow G$ has components which give the topological quantum number of Yang-Mills theory for the group G . Each component of the space has further topological invariants. When $G = \text{SU}(2)$ we show that these invariants (the homology groups) are “captured” by the space of instantons. Using these invariants we show that potentials must exist for which the massless Dirac equation (in Euclidean 4-space) has arbitrarily many independent solutions (for fixed instanton number).

§1. Introduction

In non-abelian 4-dimensional gauge theories it is by now well-known that certain topological aspects play an important role. More specifically the fact that, for a simple non-abelian compact Lie group G , the third homotopy group $\pi_3(G)$ is isomorphic to the integers leads to a “topological quantum number” k . The purpose of this paper is to draw attention to further topological features and to show how these are related to analytical aspects of the gauge theory.

Our basic observation is that the Euclidean Yang-Mills Lagrangian is defined on a function space with many components (labelled by the integers) and that each component has further internal topological invariants. Homotopically the function space is determined by the asymptotic data and so it can be identified with the space $\Omega^3(G)$ of maps $S^3 \rightarrow G$ (normalized to preserve base points). The components of this space give $\pi_3(G)$ and are labelled by integers k , and each component $\Omega_k^3(G)$ is a space with much internal structure extensively studied by topologists.

As usual in order to deal with the asymptotic conditions in R^4 we shall work on the 4-sphere $S^4 = R^4 \cup \infty$ which is the conformal compactification. In this version the relevant function space is the space of connections (potentials) modulo gauge transformations. The space of connections is a linear space, with no topological invariants, but after factoring out by gauge transformations we get a space $\mathcal{C}(G)$ which is homotopically $\Omega^3(G)$. These basic facts are described in §2.

One of the first consequences of the non-triviality of the space $\Omega^3(G)$ or $\mathcal{C}(G)$ is the impossibility of fixing a gauge, i.e. there are topological obstructions to choosing continuously a potential in each gauge-equivalence class. This aspect is briefly discussed in §2 but for a fuller treatment we refer to the paper of Singer [18].

In §3 we make a particular study of $SU(2)$ -instantons, the minimum action solutions of the Euclidean Yang-Mills equations. We show that, for large k , the homology of $\Omega_k^3(SU(2))$ is “contained in” the homology of the space of instantons of degree k [see Theorem (3.1) for a more precise statement]. Thus the complicated homology of the function space is faithfully reflected in the global nature of the instanton space. One noteworthy feature of the homology is that it is all generated by configurations of k single instantons in the manner of Segal [17]. At this point we should perhaps point out that a complete construction for all instantons is now known in terms of linear algebra [6]. However this description, although in principle very explicit, does not immediately give information on the topology of the instanton space. An interesting problem for further investigation will be to connect up our results [Theorem (3.1)] with those of [6].

In §4 we consider the (massless) Dirac equation coupled to instantons (of any degree $k \geq 0$). It is well-known that this equation has a k -dimensional space of solutions [7]. As we move over the instanton space M_k this gives a vector bundle with fibre C^k . We investigate the global topological properties of this bundle and in particular derive results on its Chern classes [Propositions (4.4) and (4.5)]. We then use these results to prove topologically the existence of connections (not solutions of the Yang-Mills equations) for which the coupled Dirac equation has arbitrary large solution space [Theorem (4.6)].

These results for $SU(2)$ can be extended to other groups G and set in a larger context by considering the family of *all* Dirac operators parametrized by the space $\mathcal{C}(G)$ of connection classes. This has a generalized index in the sense of [10], namely the homotopy class of the map $\mathcal{C}(G) \rightarrow \mathcal{F}$ given by the Dirac operators (where \mathcal{F} is the space of Fredholm operators). This generalized index can be identified, using results of [3], and leads to Theorem (4.12). This theorem has cohomological consequences and as a particular application we show [Theorem (4.14)] how to extend Theorem (4.6) to deal with all representations of $SU(2)$.

In order to preserve the flow of the argument in §4 we defer till §5 the proofs of a number of technical results, particularly those involving cohomological calculations.

Finally in §6 we make some brief comparisons with 2-dimensional Yang-Mills and with the 2-dimensional non-linear σ -model. There are some very interesting analogies on the topological level.

Since we shall be dealing with infinite-dimensional function spaces care has to be taken over the topological details. However all our results essentially concern finite-dimensional subspaces of the function-space. For example statements about homology are always of this kind. For this reason we shall not be over-pedantic on this question. In particular we shall use “homotopy equivalence” where it might be technically more accurate to use the weaker notion of homotopy equivalence on compact subsets.

§2. The Space of Connections

We begin by recalling the way in which the “topological quantum number” k arises for a Yang-Mills theory in Euclidean 4-space. We start with a given gauge group G which is assumed to be a compact simple Lie group [e.g. $SU(2)$]. The gauge potential $A(x)$ is a Lie-algebra valued 1-form on R^4 . To get a finite Yang-Mills action we assume that as $x \rightarrow \infty$, $A(x)$ becomes a pure gauge. This means that asymptotically $A(x)$ takes the form

$$A(x) \sim g(x)^{-1} dg(x) \tag{2.1}$$

where $g : S^3 \rightarrow G$ is defined on the “sphere at ∞ ”. If we normalize g by requiring $g(s_0) = 1$, where $s_0 \in S^3$ is some fixed point, then g is uniquely determined by A . The homotopy classes of such maps g form the third homotopy group $\pi_3(G)$ and this is known to be isomorphic to the integers for all G [if $G = SU(2)$ the integer is given by the degree of the map].

Suppose now that we consider not just a single potential $A(x)$ but a continuous family $A(x, y)$ in which y runs over some compact parameter space Y . Then asymptotically, putting a parameter y into (2.1) we have

$$A(x, y) \sim g(x, y)^{-1} dg(x, y) \tag{2.2}$$

(where $d = d_x$ differentiates only the x -variables). In (2.2) g is a map $S^3 \times Y \rightarrow G$ or equivalently a map $Y \rightarrow \Omega^3(G)$, where $\Omega^3(G)$ is the function space of maps $S^3 \rightarrow G$ taking s_0 to 1. The homotopy classes of such maps form an abelian group denoted by $[Y, \Omega^3(G)]$. If Y reduces to a one-point space then we are in our previous situation and the above group gives just the components of $\Omega^3(G)$, in other words $\pi_3(G)$. If the individual components $\Omega_k^3(G)$ were themselves contractible then no further topological invariants would arise from any choice of Y . However it is well known in algebraic topology that each component $\Omega_k^3(G)$ is far from being contractible and hence for suitable Y the groups $[Y, \Omega^3(G)]$ will give non-trivial information going beyond the classification by components. For example if $Y = S^n$ then

$$[Y, \Omega^3(G)] \cong \pi_{n+3}(G).$$

For $G = SU(m)$ we have

$$\pi_4(SU(2)) \cong \pi_5(SU(2)) \cong Z_2 \quad (\text{integers modulo } 2)$$

$$\pi_5(SU(m)) \cong Z \quad (\text{integers}) \quad \text{for } m \geq 3.$$

Thus we see that the space of potentials, subject to the asymptotic condition (2.1), has a rich topological structure.

In order to deal rigorously with this asymptotic condition it is convenient to pass to the 4-sphere $S^4 = R^4 \cup \infty$. This is a sensible procedure for the Yang-Mills equations since they are conformally invariant and S^4 is conformally flat. We shall therefore reformulate our discussion in terms of S^4 , and it is now convenient to use differential-geometric language in which the potential is regarded as defining a connection A in a principle fibre bundle P with group G . The integer k now appears as labelling the isomorphism class of P over S^4 . The asymptotic behaviour of the Euclidean potential has gone into constructing P .

To get further than the integer invariant k we now consider isomorphism classes of bundles P with connection A . For technical reasons, related to the normalization of g imposed after (2.1), we shall actually pick a base point $\infty \in S^4$ and each P will be assumed to have a base point p_∞ over ∞ . Moreover isomorphisms will be required to take base point to base point. With this convention the set of isomorphism classes of all (P, A) will be denoted by $\mathcal{C}(G)$. The topology of A induces a topology on $\mathcal{C}(G)$, so that $\mathcal{C}(G)$ becomes a topological space. This space can also be described more explicitly as follows. First of all $\mathcal{C}(G)$ has a component $\mathcal{C}_k(G)$, for each integer k , corresponding to the class of P . For a given P_k let \mathcal{A}_k denote the space of all connections A on P_k , and let \mathcal{G}_k denote the group of all (base-point preserving) automorphisms of P_k . An element of \mathcal{G}_k is thus a gauge transformation which is the identity at the base point of S^4 , i.e. at ∞ . Clearly \mathcal{G}_k acts on \mathcal{A}_k and $\mathcal{C}_k(G) = \mathcal{A}_k/\mathcal{G}_k$. The topology on $\mathcal{C}_k(G)$ is inherited from that of \mathcal{A}_k and is quite well-behaved in view of the following facts:

- i) \mathcal{A}_k is a linear space,
- ii) \mathcal{G}_k acts freely on \mathcal{A}_k [i.e. $g(A) = A \Rightarrow g = 1$],
- iii) local slices exist for the action of \mathcal{G}_k on \mathcal{A}_k (i.e. subspaces of \mathcal{A}_k meeting \mathcal{G}_k -orbits in one point).

i) is clear. To prove ii) assume $g(A) = A$ for some $g \in \mathcal{G}_k$ and $A \in \mathcal{A}_k$. Then g preserves the parallel transport defined by A and, since g is the identity at ∞ , parallel transport makes it the identity everywhere. The local slices in iii) are essentially given by the transverse gauge. More precisely for $A \in \mathcal{A}_k$ the orbit of \mathcal{G}_k through A is given infinitesimally by all $d_A(\phi)$, where ϕ is a section of the adjoint bundle of P_k (with fibre the Lie algebra of G) vanishing at ∞ , and d_A is the covariant derivative defined by the connection A . An infinitesimal slice is therefore given by solutions of the adjoint equation $d_A^*(\psi) = 0$. Standard analytic arguments involving Sobolev spaces then show that this infinitesimal slice generates a local slice. For further detail on this point we refer to [18].

Properties ii) and iii) imply that $\mathcal{A}_k \rightarrow \mathcal{A}_k/\mathcal{G}_k = \mathcal{C}_k(G)$ is a principal fibre bundle with group \mathcal{G}_k .

The relation of the space $\mathcal{C}(G)$, constructed on S^4 , to our earlier discussion in R^4 , is clarified by the following:

Proposition (2.3). *The space $\mathcal{C}(G)$ is homotopically equivalent to $\Omega^3(G)$.*

The proof of (2.3) is quite routine. We first deform $\mathcal{C}(G)$ into the subspace $\mathcal{C}(G)_\infty$ of connection classes which are flat near $\infty \in S^4$. For any such connection A we pick a flat section α of P with $\alpha(\infty) = p_\infty$ and any section β of P over R^4 which agrees with α on the fixed radial direction s_0 (such β exist because R^4 retracts onto the radial line). On a small 3-sphere around ∞ , α and β now differ by a map $g : S^3 \rightarrow G$ with $g(s_0) = 1$. By assigning g to A we get a map $\mathcal{C}(G)_\infty \rightarrow \Omega^3(G)$ and it is easy to check that this is a homotopy equivalence.

Further insight into (2.3) can be gained if we use “classifying spaces”. We recall that if K is any topological group, it has associated to it a space BK , called its classifying space, which is well-defined up to homotopy. It can be constructed as the base space of any principal K -bundle (called the universal bundle) with contractible total space. It is inverse to the loop space construction in the sense that $B[\Omega(K)] \sim K_0$ and $\Omega(BK) \sim K$. Here $\Omega(K)$, the function space of based maps

$S^1 \rightarrow K$ is itself a topological group using multiplication of values, and K_0 is the identity component of K .

As a first instance of these ideas consider the fibre bundle $\mathcal{A}_k \rightarrow \mathcal{C}_k(G)$ with group \mathcal{G}_k . Since \mathcal{A}_k is a linear space it is contractible and hence $\mathcal{C}_k(G) \sim B\mathcal{G}_k$. For $k=0$, $P_0 = S^4 \times G$, hence $\mathcal{G}_0 = \Omega^4(G) \sim \Omega(\Omega^3(G))$ and so $B\mathcal{G}_0 \sim B\Omega(\Omega^3(G)) \sim \Omega_0^3(G)$. This gives the equivalence $\mathcal{G}_0 \sim \Omega_0^3(G)$ for the $k=0$ components in (2.3). The equivalence for other k is an automatic consequence because the homotopy type of both $\mathcal{C}_k(G)$ and $\Omega_k^3(G)$ is actually independent of k . For $\Omega^3(G)$ this follows at once from the fact that it is a topological group. For $\mathcal{C}(G)$ we can define a composition law which plays the same role for homotopy. This is the ‘‘connected sum’’ operation defined most conveniently on $\mathcal{C}(G)_\infty$ by cutting out small balls at ∞ and gluing the complements together.

If we use the classifying space BG of G we can also describe explicitly the inverse map $\Omega^3(G) \rightarrow \mathcal{C}(G)$. Using the equivalence $\Omega BG \sim G$ and the induced equivalence $\Omega^4(BG) \sim \Omega^3(G)$ it is sufficient to describe a map $\Omega^4(BG) \rightarrow \mathcal{C}(G)$. For this we pick, once and for all, a standard connection ξ for the universal G -bundle QG over BG and a point of QG over the base-point of BG . Then for any (based) map $f : S^4 \rightarrow BG$ we form the pull-back connection $f^*(\xi)$ for the induced G -bundle $f^*(QG)$ over S^4 . Assigning to f the class of $f^*(\xi)$ defines the required map $\Omega^4(BG) \rightarrow \mathcal{C}(G)$.

This map can be thought of in more concrete terms if we introduce a compact parameter space Y and consider the induced homomorphism

$$[Y, \Omega^4(BG)] \rightarrow [Y, \mathcal{C}(G)],$$

where $[Y, \]$ denotes the homotopy classes of maps, which here form abelian groups.

By the ‘‘universal’’ property of BG the group $[Y, \Omega^4(BG)]$ classifies, up to isomorphism, G -bundles Q over $S^4 \times Y$ with a section q over $\infty \times Y$. If we choose any partial connection A (using only the S^4 -directions) for such a Q we get a family (Q_y, A_y) of bundles with connection over S^4 , parametrized by $y \in Y$. The class of this family is the corresponding element of $[Y, \mathcal{C}(G)]$.

Since the Yang-Mills Lagrangian \mathcal{L} depends only on the isomorphism class of the pair (P, A) it is really a function on the space $\mathcal{C}(G)$. In fact the base points we have used so far do not affect \mathcal{L} and so \mathcal{L} , as a function on $\mathcal{C}(G)$ has a further group invariance. Since the base point p can be altered by operating with G we see that G acts on $\mathcal{C}(G)$ and \mathcal{L} is invariant under this action. The reason why we do not proceed to form the quotient space $\mathcal{C}(G)/G$ is that the action of G is not free. First of all the (finite) centre C of G gives an automorphism (not preserving base-points) of any (P, A) . Hence C acts trivially on $\mathcal{C}(G)$ and so the action of G on $\mathcal{C}(G)$ factors through the adjoint group $\text{ad}(G) = G/C = \bar{G}$. More generally if $g \in G$ leaves fixed the class of (P, A) in $\mathcal{C}(G)$ then g centralizes the holonomy group of A at ∞ . Thus \bar{G} acts freely on the subspace of $\mathcal{C}(G)$ representing *irreducible* connections, i.e. connections for which the holonomy group is the whole of G . In particular when $G = \text{SU}(2)$ the reducible connections occur only for $k=0$, hence \bar{G} acts freely on $\mathcal{C}_k(G)$ for all $k \neq 0$. The quotient space $\mathcal{C}_k(G)/\bar{G}$ can then be identified with the classifying space of the group \mathcal{G}'_k of *all* automorphisms of P_k . Note that \mathcal{G}'_k contains \mathcal{G}_k as a normal subgroup with $\text{ad}(G)$ as quotient group. Certainly \mathcal{G}'_k is

not contractible and hence the fibration $\mathcal{A}_k \rightarrow \mathcal{A}_k/\mathcal{G}'_k$ can have no section. In other words we cannot continuously “fix the gauge”. This observation is due to Singer [18] who has extended it to cover $k=0$ and other G . The non-triviality of the earlier fibration $\mathcal{A}_k \rightarrow \mathcal{A}_k/\mathcal{G}_k = \mathcal{C}_k(G)$ can be interpreted as saying that we cannot fix the gauge even up to an ambiguity in G .

§3. SU(2)-Instantons

In §2 we saw that the Yang-Mills Lagrangian \mathcal{L} is essentially a function on the space $\mathcal{C}(G)$ which is in addition G -invariant, and that $\mathcal{C}(G) \sim \Omega^3(G)$ is an interesting topological space. By analogy with other examples in the calculus of variations and with finite-dimensional Morse theory we might expect an intimate relation between the homology of the function space $\mathcal{C}(G)$ and that of the critical point set of \mathcal{L} , i.e. the space of solutions of the Yang-Mills equations. In particular we might expect the absolute minimum of \mathcal{L} to carry a substantial part of the homology of $\mathcal{C}(G)$. The purpose of this section is to prove a precise theorem of this type for the case when $G = \text{SU}(2)$.

We recall that the absolute minimum of \mathcal{L} is given by the *self-dual* solutions of Yang-Mills equations if $k > 0$, and by the *anti-self-dual* solutions if $k < 0$ (see [15] or [8]). We will concentrate on the case $k > 0$. Let $\mathcal{M}_k \subset \mathcal{A}_k$ denote the space of self-dual connections (instantons) for $G = \text{SU}(2)$. Then as shown in [8] the quotient space $M'_k = \mathcal{M}_k/\mathcal{G}'_k$ is a manifold of dimension $8k - 3$: this is the “moduli space of instantons” of degree k . If we divide \mathcal{M}_k by the smaller group \mathcal{G}_k we get a larger manifold M_k which has dimension $8k$ and is fibered over M'_k with fibre $\text{ad}(G) = \text{SU}(2)/\{\pm 1\} \cong \text{SO}(3)$. A point of M_k represents, up to isomorphism, a bundle P_k with self-dual connection and a given point p_∞ over ∞ . In terms of the corresponding vector bundle, with fibre C^2 , the point p_∞ gives a preferred basis at ∞ .

Our main result is then:

Theorem (3.1). *For $G = \text{SU}(2)$ the inclusion $M_k \subset \mathcal{C}_k$ induces a map in homology*

$$H_q(M_k) \rightarrow H_q(\mathcal{C}_k)$$

which, for $k \geq q$, is a projection onto a direct summand.

In simple terms this theorem asserts that, as k increases, more and more of the homology of the function space \mathcal{C}_k is contained in the space of instantons M_k . Note that $H_q(\mathcal{C}_k)$ is independent of k (since $\mathcal{C}_k \sim \mathcal{C}_0$) whereas $\dim M_k = 8k$ so that M_k is a space which is growing with k and the same is true of $H_q(M_k)$.

Before proceeding to the proof of (3.1) we shall make a number of comments and conjectures. First of all, as mentioned in §1, a complete construction for M_k (for all G) is now known in terms of linear algebra [6]. However this explicit construction does not immediately give geometric or topological information about M_k . For example it is not yet known whether M_k is always connected. Theorem (3.1), which uses an earlier incomplete but easier construction, should be regarded as a partial and provisional result. One might conjecture that a similar result should hold for all G and that $H_q(M_k) \rightarrow H_q(\mathcal{C}_k)$ might actually become an

isomorphism for $k \gg q$. One would also hope for explicit bounds for k as a function of q to replace the asymptotic statement. Finally one might hope for the corresponding (stronger) statements with homotopy replacing homology.

As just indicated the proof of (3.1) uses a construction for $SU(2)$ -instantons due to 't Hooft and others (see [15] or [8]). This construction starts from a configuration of k distinct points in R^4 each with an assigned positive weight. For our purposes the weights play no role and so we shall set them all equal to unity. Then the 't Hooft construction defines a map

$$\theta_k : C_k(R^4) \rightarrow M_k,$$

where $C_k(R^4)$ is the configuration space. By definition $C_k(R^4)$ is the quotient of $\tilde{C}_k(R^4)$, the space of ordered k -tuples of distinct points, by the action of the symmetric group Σ_k . Clearly $\tilde{C}_k(R^4) \subset R^{4k}$ and this defines its topology. The action of Σ_k is free and $C_k(R^4)$ has the quotient topology. Before proceeding further it is perhaps useful to point out that the 't Hooft construction in R^4 , as extended by Jackiw et al. [15] to the conformal compactification S^4 , associates to $(k + 1)$ distinct points of S^4 a self-dual connection on S^4 up to isomorphism. However this construction has no base point at ∞ and so yields only a map

$$C_{k+1}(S^4) \rightarrow M'_k.$$

When one of the $(k + 1)$ points is fixed at ∞ we then get a preferred base point and so recover the 't Hooft construction. In other words the diagram

$$\begin{array}{ccc} C_k(R^4) & \longrightarrow & M_k \\ \downarrow & & \downarrow \\ C_{k+1}(S^4) & \longrightarrow & M'_k \end{array}$$

commutes.

There are various ways to describe the 't Hooft construction (and its conformal extension). It can be described explicitly by giving the Yang-Mills potential in R^4 as a function of k points a_1, \dots, a_k . This description starts from the auxiliary function (or "super-potential")

$$\phi(x) = 1 + \sum_{i=1}^k \frac{1}{|x - a_i|^2}$$

(see [15]). In the algebraic geometric translation of [11] the construction uses $k + 1$ lines of complex projective 3-space. Finally in differential-geometric terms it can be described as follows [8, §6]. We take the $SU(2)$ -connection on R^4 which is the pull-back of the standard $SU(2)$ -connection ξ_k on quaternionic projective k -space $P_k(H)$ by the map

$$u \mapsto (1, (u - a_1)^{-1}, \dots, (u - a_k)^{-1}). \tag{3.2}$$

Here u, a_1, \dots, a_k are all viewed as quaternions. Note that when $u = a_i$ we can eliminate the apparent infinity by multiplying by the quaternionic scalar $(u - a_i)$, a process which is admissible for projective coordinates. The point $u = \infty$ goes into the standard point $(1, 0, \dots, 0)$ and the differential of 3.2 at $u = \infty$ is independent of the a_i , which is why we get our base point at ∞ .

The configuration spaces $C_k(R^4)$ and more generally $C_k(R^n)$ play an interesting and important role in algebraic topology and are intimately related to the function spaces $\Omega_k^n(S^n)$ of (based) mappings $S^n \rightarrow S^n$ of degree k . More specifically there is a map

$$\phi_k^n : C_k(R^n) \rightarrow \Omega_k^n(S^n)$$

which can be defined (up to homotopy) in a variety of ways. The simplest definition is to view k points in R^n as k “electrically charged” particles and to associate to them the corresponding “electric field” E which is a function on R^n taking values in $R^n \cup \infty$. Since $E \rightarrow 0$ at ∞ it extends to a map \bar{E} of $S^n = R^n \cup \infty$ to S^n . It takes the base-point ∞ of the argument S^n into the base-point 0 of the domain S^n , and it is easily checked to have degree k . In this construction it is immaterial (up to homotopy) what law of force or potential one takes as long as it is linear and the field of a single charge has the properties:

- i) $E \rightarrow 0$ at ∞ ,
- ii) $E \rightarrow \infty$ at the source,
- iii) E is spherically symmetric.

Thus we could take the potential to be $\log r$, $\frac{1}{r}$ or $\frac{1}{r^{n-2}}$ (if $n \geq 3$). More drastically we could take a field which has one of these potentials for $r < \varepsilon/2$ and is identically zero for $r > \varepsilon$, where ε is some specified constant. For k charged particles (a_1, \dots, a_k) we could even allow ε to depend continuously on the configuration and be so chosen that the regions of the k fields are disjoint, i.e. taking $2\varepsilon < \sup |a_i - a_j|$.

If $n = 1, 2$ or 4 so that R^n can be identified with one of the basic fields R, C or H , the Segal map can be defined by the function

$$\sum_{i=1}^k \frac{1}{u - a_i},$$

where $(u - a_i)^{-1}$ is the inverse in the appropriate field. This corresponds to the $\log r$ potential, combined with conjugation.

The topological significance of the maps ϕ_k^n is made clear by the following striking theorem of Segal [17].

Theorem (3.3). *The map $\phi_k^n : C_k(R^n) \rightarrow \Omega_k^n(S^n)$ induces isomorphisms in q -dimensional homology provided $k \gg q$.*

In order to apply this theorem we need to relate the Segal map ϕ_k^4 to the 't Hooft map θ_k . First of all we introduce a map

$$\lambda : \Omega^4(S^4) \rightarrow \mathcal{C} \tag{3.4}$$

by assigning to any based map $f : S^4 \rightarrow S^4$ the pull-back $f^*(\xi_1)$ of the standard $SU(2)$ -connection ξ_1 over S^4 (the $k=1$ instanton). This map can also be described as a composition

$$\Omega^4(S^4) \rightarrow \Omega^4(BSU(2)) \rightarrow \mathcal{C}. \tag{3.5}$$

The second map is the one explained in §2 (and is a homotopy equivalence) while the first is induced by the map $j : S^4 \rightarrow BSU(2)$ which classifies the standard bundle on S^4 . More explicitly, using the fact that $SU(2)$ is the group of quaternions of unit norm, we can take $BSU(2) = P_\infty(H)$ and j as the natural inclusion $P_1(H) \rightarrow P_\infty(H)$. Since the standard $SU(2)$ -connection ξ on $P_\infty(H)$ restricts precisely to give ξ_1 on $P_1(H)$, we have $(jf)^*(\xi) = f^*(\xi_1)$ showing that the maps (3.4) and (3.5) do indeed coincide.

Letting λ_k denote the restricting of the map λ to the k -th components, we have the following connection between the Segal and 't Hooft maps:

Lemma (3.6). *We have a homotopy commutative diagram*

$$\begin{array}{ccc}
 & \Omega_k^4(S^4) & \\
 \phi_k^4 \nearrow & & \searrow \lambda_k \\
 C_k(R^4) & \xrightarrow{\theta_k} M_k & \xrightarrow{i_k} \mathcal{C}_k
 \end{array}$$

where i_k is the inclusion map.

Proof. In more concrete terms we have to show the following. If $\alpha = (a_1, \dots, a_k) \in C_k(R^4)$, let $\phi_\alpha : S^4 \rightarrow S^4$ be the Segal map and let $f_\alpha : S^4 \rightarrow P_k(H)$ be the map (3.2) used in the 't Hooft construction. Then we have to show that, up to a standard homotopy, the connections $\phi_\alpha^*(\xi_1)$ and $f_\alpha^*(\xi_k)$ are naturally isomorphic. But this follows at once from the commutative diagram

$$\begin{array}{ccc}
 S^4 & \xrightarrow{f_\alpha} P_k(H) - P_{k-2}(H) \subset P_k(H) & \\
 \searrow \phi_\alpha & & \downarrow \varepsilon \\
 & & S^4
 \end{array}$$

where $P_{k-2}(H)$ is the subspace given by $w_0 = \sum_{i=1}^k w_i = 0$, ε is the homotopy equivalence given by

$$\varepsilon(1, w_1, w_2, \dots, w_k) = \sum w_i$$

and we take ϕ_α in the form

$$\phi_\alpha(u) = \sum_{i=1}^k (u - a_i)^{-1}.$$

Lemma (3.6), combined with the Segal Theorem (3.3), already shows that for $k \geq q$, that part of $H_q(\mathcal{C}_k)$ which is in the image of λ_k comes from $H_q(M_k)$. In order to refine this to get Theorem (3.1) we need a well-known result connecting $\Omega^3(S^3)$ with $\Omega^4(S^4)$, namely

Lemma (3.7). $\Omega^3(S^3)$ is homotopically a direct summand of $\Omega^4(S^4)$. More precisely the suspension map $\Omega^3(S^3) \rightarrow \Omega^4(S^4)$ is homotopically a right inverse to the map $\Omega^4(S^4) \rightarrow \Omega^4(BSU(2)) \sim \Omega^3(S^3)$ of (3.5).

Proof. For any G there is a natural map of its suspension SG into BG and the composition

$$G \rightarrow \Omega(SG) \rightarrow \Omega(BG)$$

gives the natural homotopy equivalence $G \sim \Omega BG$. Taking $G = \text{SU}(2) = S^3$ gives

$$S^3 \rightarrow \Omega(S^4) \rightarrow \Omega(\text{BSU}(2)).$$

Now apply Ω^3 , i.e. take maps of S^3 into the sequence, and the Lemma follows.

Remark. As the proof shows this Lemma uses the accidental fact that the sphere S^3 coincides with the Lie group $\text{SU}(2)$. It is therefore closely tied to the particular dimensionality of our spaces, and the fact that $R^4 = H$ is the field of quaternions.

As we have already noted the homotopy equivalence

$$\Omega^4(BG) \sim \Omega^3(G) \sim \mathcal{C}(G)$$

of (2.3) coincides with the second map in (3.5). Hence (3.7) tells us that $\Omega^3(S^3) \rightarrow \Omega^4(S^4)$ is a right homotopy inverse to $\lambda: \Omega^4(S^4) \rightarrow \mathcal{C}$. Now it is an easy property of the Segal maps ϕ_k^n that they commute with suspension, i.e. that we have homotopy commutative diagrams

$$\begin{array}{ccc} C_k(R^n) & \xrightarrow{\phi_k^n} & \Omega_k^n(S^n) \\ \downarrow i & & \downarrow \sigma \\ C_k(R^{n+1}) & \xrightarrow{\phi_k^{n+1}} & \Omega_k^{n+1}(S^{n+1}) \end{array}$$

where σ is suspension and i is induced by the inclusion $R^n \subset R^{n+1}$. Combining this diagram for $n=3$ with that of (3.6) and the result of (3.7) we get finally a homotopy commutative diagram :

$$\begin{array}{ccccc} C_k(R^3) & \xrightarrow{\hspace{2cm}} & \Omega_k^3(S^3) & & \\ & \searrow & \nearrow & & \\ & & M_k & \xrightarrow{\mathcal{C}_k} & \\ & \nearrow & \searrow & & \\ C_k(R^4) & \xrightarrow{\hspace{2cm}} & \Omega_k^4(S^4) & & \end{array}$$

In particular, the top part of the diagram shows that the map $\phi_k^3: C_k(R^3) \rightarrow \Omega_k^3(S^3)$ factors homotopically through M_k . By Segal's theorem (3.3), ϕ_k^3 induces isomorphisms in H_q provided $k \gg q$. It follows that

$$H_q(M_k) \rightarrow H_q(\Omega_k^3(S^3)) \cong H_q(\mathcal{C}_k)$$

is projection onto a direct summand, which proves Theorem (3.1).

Segal's theorem (3.3) asserts that we have isomorphisms in H_q for $k \geq k_0(q)$. The same lower bound $k_0(q)$ then works in Theorem (3.1). For $q=1$, H_1 is the abelianized fundamental group π_1 and

$$\pi_1(C_k(R^3)) \approx \Sigma_k$$

since the covering space $\tilde{C}_k(R^3)$ of ordered k -tuples is easily seen to be simply-connected. Hence, for $k \geq 2$, H_1 is of order 2 and so $k_0(1)=2$, i.e.

$$H_1(M_k) \rightarrow H_1(\mathcal{C}_k) \cong Z_2$$

is projection onto a direct summand for $k \geq 2$. Similar explicit results for H_q ($q \geq 1$) depend on calculating the homology of the configuration space $C_k(R^3)$. Some results in this direction will be given in the next section.

A significant feature of Segal's theorem (3.3) is that we cannot replace homology by homotopy. This is clear by looking at fundamental groups, since (for $n \geq 3$)

$$\pi_1(C_k(R^n)) \cong \Sigma_k, \quad \pi_1(\Omega_k^n(S^n)) \cong Z_2.$$

In fact the whole point of theorems of this type is precisely that the fundamental group is essentially killed but that homology is preserved. Now we have shown that M_k lies between $C_k(R^3)$ and $\Omega_k^3(S^3)$, but at present we do not know which space it resembles most. It seems more likely that M_k has a small fundamental group and resembles $\Omega_k^3(S^3)$. This is compatible with the slight evidence so far available since dimension counts show that the 't Hooft solutions fail to fill up M_k for $k \geq 3$, whereas Hartshorne [14] has shown that they give all of M_k for $k \leq 2$. Note that $k = 3$ is precisely the value at which Σ_k becomes larger than Z_2 .

§ 4. The Dirac Operator

In this section we shall study the (massless) Dirac operator on S^4 , coupled to a gauge field, and we shall show that the analysis of this operator is intimately related to the topology of the space $\mathcal{C}(G)$ of (classes of) connections studied in § 2. We shall deal in detail with the case when $G = \text{SU}(2)$, using the results of § 3, but the general discussion will apply to any G .

We recall first that the Dirac operator on S^4 acts on spinor fields interchanging positive and negative (helicity). We denote by D the Dirac operator on positive spinors: its adjoint D^* is then the Dirac operator on negative spinors. More generally, if P is a principal G -bundle on S^4 with connection A and if $\rho: G \rightarrow U(n)$ is a unitary representation of G , we can extend D to an operator D_A acting on positive V -valued spinors, where V is the C^n -bundle associated to P by ρ . This is D "coupled to A ".

We shall be particularly interested in the null space $\mathcal{N}_A = \mathcal{N}(D_A)$, i.e. the space of solutions of the equation $D_A u = 0$. Since D_A is elliptic and S^4 is compact the space \mathcal{N}_A has a finite dimension say r_A . Replacing D_A by its adjoint D_A^* gives similarly a space $\mathcal{N}(D_A^*)$ with finite dimension say s_A . We now recall that by definition, the *index* of the elliptic operator D_A is given by

$$\text{index } D_A = \dim \mathcal{N}(D_A) - \dim \mathcal{N}(D_A^*) = r_A - s_A. \tag{4.1}$$

The significance of the index is that it is invariant under perturbation. Thus $\text{index } D_A$ depends only on the isomorphism class of the bundle P and hence on the integer k , which classifies P as in § 2. In fact the index theorem of [9] shows that

$$\text{index } D_A = \varrho(k), \tag{4.2}$$

where $\varrho(k)$ is a certain multiple of k depending on the homomorphism ρ : for details see [8].

If A is a Yang-Mills minimum, i.e. $A \in \mathcal{M}_k(G)$ in the notation of §2, then as shown in [8] we have a “vanishing theorem”, namely for $k \geq 0$.

$$\mathcal{N}(D_A^*) = 0. \tag{4.3}$$

Thus $s_A = 0$ and so $r_A = \varrho(k)$ is constant for all $A \in \mathcal{M}_k(G)$. Dividing by the group \mathcal{G}_k of gauge equivalences we then get a vector bundle N over the quotient space $M_k(G)$, with fibre $C^{\varrho(k)}$. As we shall see shortly, for $G = \text{SU}(2)$, this vector bundle N is topologically non-trivial. In other words, not only does the topological invariant k describe the number of independent solutions of the Dirac equation coupled to instantons, but the global variation of these solutions over the space of instantons is related to other topological invariants of the space $\mathcal{G}_k(G)$.

We now take $G = \text{SU}(2)$ and $\varrho: \text{SU}(2) \rightarrow U(2)$ the inclusion. In this case $\varrho(k) = k$ in (4.2) (see [8]). Moreover if we use only the 't Hooft connections defined by configurations of k points a_1, \dots, a_k of \mathbb{R}^4 one can show explicitly that the solutions of the corresponding Dirac equation have a natural basis e_1, \dots, e_k , each e_i being correlated with a point a_i (a proof will be given in §5). In terms of the bundle N this can be reformulated as follows:

Proposition (4.4). *Let $C_k(\mathbb{R}^3) \rightarrow M_k$ be the map given by the 't Hooft construction as in §2. Then the pull-back of N to $C_k(\mathbb{R}^3)$ is isomorphic to the vector bundle associated to the standard k -dimensional representation σ_k of $\pi_1(C_k(\mathbb{R}^3)) \cong \Sigma_k$.*

Note. In (4.4) N is naturally (the complexification of) a real vector bundle and the isomorphism is over the reals. The real structure of N arises because spinors with values in C^2 have a real structure, corresponding to the fact that the representation $C^2 \otimes C^2$ of $\text{SU}(2) \times \text{SU}(2)$ comes from the real representation $\text{SU}(2) \times \text{SU}(2) \rightarrow \text{SO}(4)$ on \mathbb{R}^4 .

The standard topological invariants of a vector bundle are its characteristic classes, namely the Chern classes $c_i \in H^{2i}(\quad, Z)$ in integral cohomology and (for real bundles) the Stiefel-Whitney classes $w_i \in H^i(\quad, Z_2)$ in mod 2 cohomology. Using (4.4) the characteristic classes of N can be calculated algebraically from the cohomology of the symmetric group. The simplest to calculate is w_1 . Since a class in $H^1(\quad, Z_2)$ is given by a homomorphism $\pi_1 \rightarrow Z_2$ it follows that $w_1(\sigma_k)$ is given by the sign homomorphism $\Sigma_k \rightarrow \{\pm 1\}$ (or Z_2). It follows from (4.4) that $w_1(N)$, pulled back to $C_k(\mathbb{R}^3)$ gives the generator of $H^1(C_k(\mathbb{R}^3), Z_2)$ for $k \geq 2$. For higher-dimensional classes we find in particular the following results which will be proved in §5:

Proposition (4.5). *Let p be an odd prime, then the bundle on $C_k(\mathbb{R}^3)$ associated to σ_k has the following non-zero classes mod p :*

- i) $c_{p-1} \not\equiv 0 \pmod p$ in $C_p(\mathbb{R}^3)$, and more generally,
- ii) $c_{n(p-1)} \not\equiv 0 \pmod p$ in $C_{np}(\mathbb{R}^3)$ for all $n \geq 1$,
- iii) $c_{p-1}^n \not\equiv 0 \pmod p$ in $C_{np}(\mathbb{R}^3)$ $1 \leq n < p$.

On the other hand if $t < p - 1$ we have

- iv) $c_t \equiv 0 \pmod p$ in $C_{np}(\mathbb{R}^3)$ for $1 \leq n < p$.

Propositions (4.4) and (4.5) show that the bundle N of null-spaces of the Dirac operator is highly non-trivial as asserted earlier.

So far we have only used instanton connections, since for these we have the vanishing theorem (4.3). In fact (4.3) does not hold for all connections. More precisely we shall prove

Theorem (4.6). *In each component $\mathcal{C}_k(\text{SU}(2))$ there exist connections A for which the space of solutions of the Dirac equation $D_A u = 0$ (and its adjoint) has arbitrarily large dimension.*

It may well be possible, given sufficient ingenuity, to exhibit explicit connections for which the Dirac equation has many solutions. Our proof will be quite different in that topological considerations alone force the existence of such connections. Rather surprisingly we shall use the topological non-triviality of the bundles N (over M_k) to prove Theorem (4.6). Thus the global behaviour of the solutions of the Dirac equation coupled to instantons will force the existence of other connections A which are very far removed from being instantons in that the null-spaces $\mathcal{N}(D_A^*)$ are large instead of being zero. The explanation of this apparent paradox is that to prove (4.6) for a given integer k we make use of the bundles N over $M_l \subset \mathcal{C}_l$ for larger values l . Since $\mathcal{C}_k \sim \mathcal{C}_l$ we can homotopically transport M_l to \mathcal{C}_k , where it will *not* be related to M_k . Essentially the connections A of (4.6) will be found on this transported copy M_l^* of M_l . If we carry out the transportation carefully then the Yang-Mills action on M_l^* will be approximately $2l - k$ (when normalized so that the action of a basic instanton is equal to 1). Thus to obtain the connections A of (4.6) we have to consider arbitrarily large values of the action.

As an example if $k = 0$, so that M_0 is one point, we can take $l = 2$. Transporting to \mathcal{C}_0 the closed path on M_2 which arises from interchanging the two points in a configuration of $C_2(\mathbb{R}^3)$ we can use the non-vanishing of w_1 to deduce that (4.3) must fail for some A on our path.

In dealing with the family of all D_A we have therefore to face the fact that the null-spaces $\mathcal{N}(D_A)$ have dimensions which jump. We cannot therefore make these null-spaces into a vector bundle to obtain cohomological invariants. There is however a standard way to get round this problem which we now digress to explain.

First of all in dealing with differential operators (which are unbounded operators on the Hilbert space L^2) it is standard to turn them into bounded operators by altering the norms (i.e. using Sobolev spaces). In this way elliptic differential operators become bounded *Fredholm* operators T on a Hilbert space H , i.e. they satisfy

- i) $\dim \mathcal{N}(T) < \infty$,
- ii) $\dim \mathcal{N}(T^*) < \infty$,
- iii) $T(H)$ is closed in H .

The space of all Fredholm operators, denoted by \mathcal{F} , has a metric topology given by the operator norm. For any $T \in \mathcal{F}$ its index is [by i) and ii)] well-defined and one can show that \mathcal{F} has one component \mathcal{F}_k for each value of the index. More generally the homotopy type of \mathcal{F} is completely known. There is a natural equivalence

$$\mathcal{F} \sim Z \times BU, \tag{4.7}$$

where $BU = \varinjlim_n BU(n)$ is the limit of the classifying spaces of the unitary groups $U(n)$. This result is (in a slightly different version) proved in the Appendix to [2]. As a consequence of (4.7) the cohomology of each component of \mathcal{F} can be identified with the polynomial ring in the universal Chern classes c_1, c_2, \dots . Hence if T_y is a continuous family of elliptic differential operators parametrized by $y \in Y$, we get a map $Y \rightarrow \mathcal{F}$ and hence we can define Chern classes in the cohomology of Y . If we have $\mathcal{N}(T_y^*) = 0$ for all $y \in Y$ then these Chern classes coincide with those of the vector bundle formed by the null spaces $\mathcal{N}(T_y)$. Similar results hold for real operators, $O(n)$ replacing $U(n)$ and the mod 2 classes w_i replace the c_i .

In our case after factoring out by the gauge equivalences we have the Dirac family parametrized by \mathcal{C} , hence a map $\mathcal{C} \rightarrow \mathcal{F}$ and classes w_i, c_i in the cohomology of each component \mathcal{C}_k . Moreover on restricting to $M_k \subset \mathcal{C}_k$ we recover the classes arising earlier from the bundle N . The classes are essentially independent of k in view of the following easy Lemma (proved in §5).

Lemma (4.8). *The map $\mathcal{C}_k \rightarrow \mathcal{F}_k$ given by the family of Dirac operators is (up to homotopy) independent of k , when we identify $\mathcal{C}_k \sim \mathcal{C}_0$ and $\mathcal{F}_k \sim \mathcal{F}_0$.*

Having extended our characteristic classes from M_k to the whole of \mathcal{C}_k we shall now use them to prove Theorem (4.6). As a simple illustration of the method we shall first prove the partial result indicated earlier, namely the existence of connections in \mathcal{C}_0 for which the Dirac equation has non-zero solutions. Assume the contrary then $\mathcal{N}(D_A) = 0$ for all $A \in \mathcal{C}_0$. Hence the (real) Dirac family $\mathcal{C}_0 \rightarrow \mathcal{F}$ has the characteristic classes of the 0-dimensional vector bundle and so in particular $w_1 = 0$. But in $M_2 \subset \mathcal{C}_2$ we have already seen that $w_1 \neq 0$, giving the required contradiction. The proof moreover shows that any circle of connections in \mathcal{C}_0 , on which the generator of $H^1(\mathcal{C}_0, \mathbb{Z}_2)$ is non-zero, must contain a connection A for which the Dirac equation has non-trivial solutions.

For the more general case we need to consider the closed subspace $\mathcal{F}^{r,s}$ of \mathcal{F}_k consisting of operators T with $\dim \mathcal{N}(T) \geq r$ (and hence $\dim \mathcal{N}(T^*) \geq s = r - k$). We suppose $k \geq 0$, then a result of Koschorke [16] asserts that $\mathcal{F}^{r,s}$ has codimension $2rs$ and represents the cohomology class $\mathcal{X}^{r,s}$ given by the determinant

$$\mathcal{X}^{r,s} = \begin{vmatrix} c_r & c_{r+1} & \dots & c_{r+s-1} \\ c_{r-1} & c_r & & \vdots \\ \vdots & & \ddots & c_r \\ c_{r-s+1} & & & c_r \end{vmatrix}.$$

If now T_y is a family of elliptic operators of index k with $\dim \mathcal{N}(T_y) < r$ for all $y \in Y$, the image of the map $Y \rightarrow \mathcal{F}_k$ does not meet $\mathcal{F}^{r,s}$ and so the class $\mathcal{X}^{r,s}$ goes to zero in the cohomology of Y . If on the other hand we can compute the Chern classes for this family and show that $\mathcal{X}^{r,s} \neq 0$ we can then argue that the image of Y must meet $\mathcal{F}^{r,s}$, i.e. $\dim \mathcal{N}(T_y) \geq r$ for some $y \in Y$.

The strategy of proof of Theorem (4.6) is now clear. We shall use Propositions (4.4) and (4.5) to prove that $\mathcal{X}^{r,s} \neq 0$ for suitable values of r, s . Given any $k \geq 0$ and any $r > k$ we pick a prime $p > r$ and consider the class $\mathcal{X}^{p-1, p-1-k}$ for the family of Dirac operators. By Lemma (4.8) it is immaterial whether we consider this in the cohomology of \mathcal{C}_k or in the cohomology of any other component \mathcal{C}_i . We

take $l=np$ with $n=p-1-k$, so that $1 \leq n < p$. We now calculate the pull-back of $\mathcal{X}^{p-1, p-1-k}$ to $C_{np}(R^3)$ using (4.4) and (4.5). Reducing modulo p and using iv) we get

$$\mathcal{X}^{p-1, p-1-k} \equiv (c_{p-1})^{p-1-k} \not\equiv 0 \pmod p \text{ by iii).}$$

This completes the proof of Theorem (4.6) for $k \geq 0$. For $k < 0$ we replace D by D^* and apply the same argument.

In view of Theorem (4.6) it is clear that we should consider the stratification of the space \mathcal{C}_k by the dimension of the null-space of the Dirac operator. Denoting by $\mathcal{C}^{r,s}$ the subspace of A of index $r-s=k$ for which $\dim \mathcal{N}(D_A) \geq r$ we have

$$\mathcal{C}_k = \mathcal{C}^{k,0} \supset \mathcal{C}^{k+1,1} \supset \mathcal{C}^{k+2,2} \supset \dots$$

Theorem (4.6) tells us that this descending chain never terminates. Moreover the method of proof clearly shows that this stratification has topological significance.

We return now to the case of a general group G , and a representation $\varrho: G \rightarrow U(n)$. Again the family of Dirac operators gives a map

$$\mathcal{C}(G) \rightarrow \mathcal{F} \tag{4.9}$$

which gives us Chern classes in the cohomology of $\mathcal{C}(G)$. For $G = \text{SU}(2)$ we exploited the connection with configuration spaces in order to study the map (4.9) and evaluate its Chern classes. In general this elementary approach is not available. However, determining the homotopy class of (4.9) is a special case of the generalized index theorem of [10]. In fact the case of the Dirac operators is very basic and was used in [3] in relation to the Bott periodicity theorems. We recall that for the unitary groups Bott's theorem gives a homotopy equivalence

$$\Omega(U) \sim Z \times BU. \tag{4.10}$$

Applying Ω^2 then gives the further equivalence

$$\Omega^3(U) \sim Z \times BU. \tag{4.11}$$

The main result of [3] can now be formulated as follows:

Theorem (4.12). *The map $\mathcal{C}(G) \rightarrow \mathcal{F}$ given by the family of Dirac operators is homotopically equivalent to the map $\Omega^3(G) \rightarrow \Omega^3(U)$ induced by the map $\varrho: G \rightarrow U(n) \rightarrow U$, where we identify $\Omega^3(G)$ with $\mathcal{C}(G)$ as in (2.2) and $\Omega^3(U)$ with \mathcal{F} by (4.11) and (4.7).*

If for example $G = U(n)$ with n large then $\Omega^3(G)$ will approximate $\Omega^3(U)$ in homotopy and homology (up to a suitable dimension) and so the low Chern classes of the Dirac family will be non-zero even in the real cohomology of $\mathcal{C}(G)$. They are then comparatively easy to detect as differential forms. The situation of $\text{SU}(2)$ was more difficult and subtle because there the Chern classes were of finite order and had to be computed in the cohomology mod p .

As an application of Theorem (4.12) let us take $G = \text{SU}(2)$ again but now we take $\varrho: \text{SU}(2) \rightarrow U(m+1)$ to be the m -th symmetric power representation σ^m . Cohomological calculations using (4.12), which will be carried out in § 5, lead to the following:

Proposition (4.13). *Let $f_m(t) = 1 + \sum_{i=1}^{\infty} c_i t^i$ be the formal power series giving the Chern classes of the Dirac family for $G = \text{SU}(2)$ and $q = \sigma^m$. For $m=1$ put $f_1(t) = f(t)$.*

Then f_m is determined from f by the formula:

$$f_m(t) = (f(mt))^m (f((m-1)t))^{2(m-1)} (f((m-2)t))^{3(m-2)} \dots (f(t))^m.$$

If we pull back to the configuration space C_{np} and work with mod p cohomology then (4.4) and (4.5) tell us that the first non-vanishing coefficient in $f(t)$ (excluding the constant term 1) is c_{p-1} . From (4.13) it then follows that the first possible non-vanishing coefficient in $f_m(t)$ occurs in the same dimension and is λ_m times the corresponding coefficient of $f(t)$ where

$$\begin{aligned} \lambda_m &= m^p + 2(m-1)^p + 3(m-2)^p + \dots + m^p \\ &\equiv m + 2(m-1) + 3(m-2) + \dots + m \pmod{p} \\ &\not\equiv 0 \pmod{p}, \text{ if } p \gg m. \end{aligned}$$

It follows that, for $p \gg m$, the Dirac family for $q = \sigma^m$ gives rise to Chern classes which satisfy i) and iv) of (4.5). As will be shown in § 5 these then imply ii) and iii) of 4.5. The same argument as before then proves:

Theorem (4.14). *Theorem (4.6) also holds for $G = \text{SU}(2)$ and all representations $q = \sigma^m$.*

Finally we should point out that the results in [2] also enable us to refine (4.12) by its real analogue. For this we need $\varrho: G \rightarrow \text{Sp}(n)$ to be symplectic [e.g. if $G = \text{SU}(2)$, $q = \sigma^m$ with m odd] and we use the real version of (4.11), namely

$$\Omega^3(\text{Sp}) \sim Z \times BO,$$

where Sp and O denote the unions of the finite-dimensional symplectic and orthogonal groups under inclusion.

§5. Deferred Proofs

In this section we give the technical proofs of results stated and used in the previous section. We begin by discussing the solutions of the Dirac equation for the $\text{SU}(2)$ -connection associated to a configuration of k points a_1, \dots, a_k of R^4 . We have to show that there is a natural basis e_1, \dots, e_k of the solutions with e_i corresponding to a_i . This can be verified explicitly in a variety of ways. In the algebro-geometric interpretation given in [11], instantons correspond to certain algebraic vector bundles E over complex projective 3-space P_3 . The instantons associated to points a_1, \dots, a_k (and $a_{k+1} = \infty$) give a bundle E , such that the twist $E(1)$ has a section s vanishing on the lines A_1, \dots, A_{k+1} defined by the a_i . Moreover the solutions of the Dirac equation correspond to the sheaf cohomology group $H^1(P_3, E(-1))$. This cohomology group can be computed from the exact sequences

$$0 \rightarrow \mathcal{O}(-2) \xrightarrow{s} E(-1) \rightarrow \mathcal{F}_r \rightarrow 0$$

$$0 \rightarrow \mathcal{F}_r \rightarrow \mathcal{O} \rightarrow \mathcal{O}_r \rightarrow 0,$$

where \mathcal{O} is the sheaf of holomorphic functions, $\Gamma = A_1 \cup A_2 \cup \dots \cup A_{k+1}$ is the curve made up of the projective lines A_i , \mathcal{I}_Γ is its ideal sheaf and s is defined by the section s . Taking cohomology and using the fact that all the sheaves $\mathcal{O}(k)$ on P_3 have zero H^1 and H^2 , we find

$$H^1(P_3, E(-1)) \cong H^1(P_3, \mathcal{I}_\Gamma) \cong \text{Coker} \{H^0(P_3, \mathcal{O}) \rightarrow H^0(\Gamma, \mathcal{O}_\Gamma)\}.$$

This shows that $H^1(P_3, E(-1))$ is naturally the quotient of the $(k+1)$ -dimensional vector space having one basis vector e_i for each line A_i by the one-dimensional space spanned by $\sum e_i$. If we fix A_{k+1} (corresponding to $a_{k+1} = \infty$) this shows that e_1, \dots, e_k give a natural basis. In the first instance all vector spaces are complex but E has a real structure which makes all the above isomorphisms compatible with complex conjugation. Thus e_1, \dots, e_k also give a basis for the real solutions of the Dirac equation.

Note that in these calculations we have for simplicity (and in conformity with § 3) normalized out the “weights” $\lambda_1, \dots, \lambda_k$ attached to the points a_1, \dots, a_k .

We come next to the proof of Proposition 4.5, which involves the cohomology of the configuration space $C_k = C_k(R^3)$. Since this space is the quotient of the space $\tilde{C}_k = \tilde{C}_k(R^3)$ of ordered configurations by the symmetric group Σ_k we first need information about the integral cohomology of \tilde{C}_k as a module over Σ_k . Results along these lines can be found in [12] but we shall give direct proofs.

Lemma (5.1). i) $H^* \tilde{C}_k$ is torsion free.

ii) The Poincaré polynomial of $H^* \tilde{C}_k$ is $(1+t^2)(1+2t^2) \dots (1+(k-1)t^2)$.

(Recall the Poincaré polynomial of a graded abelian group A is $\sum_{i \geq 0} a_i t^i$ where $a_i = \text{rank } A_i$.)

Proof. Following Arnol’d [1] consider the map $\tilde{C}_k \rightarrow \tilde{C}_{k-1}$ defined by taking the first $k-1$ points of the ordered configuration. This is a fibration, where the fibre over the point (x_1, \dots, x_{k-1}) is $R^3 - \{x_1, \dots, x_{k-1}\}$. This fibre, F , has the homotopy type of a bouquet of $k-1$ two-dimensional spheres. It follows by induction that $H^{2i-1} \tilde{C}_{k-1} = 0$ and so the spectral sequence of this fibration is concentrated in even dimensions and therefore collapses. We deduce that $H^* \tilde{C}_k$ is isomorphic as a group to $H^* F \otimes H^* \tilde{C}_{k-1}$. By induction it follows that $H^* \tilde{C}_k$ is torsion free. If we write $p_F(t)$ and $p_i(t)$ for the Poincaré polynomials of $H^* F$ and $H^* \tilde{C}_i$ it also follows that $p_k(t) = p_F(t) p_{k-1}(t)$. Now $p_F(t) = (1+(k-1)t^2)$ and the proof of ii) is completed by induction.

We will describe the structure of $H^* \tilde{C}_k$ as a Σ_k -module in terms of a basic Σ_k -module M_k . By definition M_k is the highest cohomology group of \tilde{C}_k , $M_k = H^{2(k-1)} \tilde{C}_k$.

Let α be a partition of the integer k , that is $\alpha = (k_1, \dots, k_r)$ where the k_i are integers such that $k_1 + \dots + k_r = k$. Then define M_α to be $M_{k_1} \otimes \dots \otimes M_{k_r}$, Σ_α to be $\Sigma_{k_1} \times \dots \times \Sigma_{k_r}$ and $i_\alpha: \Sigma_\alpha \rightarrow \Sigma_k$ to be the inclusion. Then M_α is a Σ_α -module and we may form the induced Σ_k -module $i_{\alpha*} M_\alpha$.

Lemma (5.2). As a Σ_k -module $H^* \tilde{C}_k \cong \bigoplus_\alpha i_{\alpha*} M_\alpha$ where α runs through all partitions (k_1, \dots, k_r) of k . The grading $|\alpha|$ of the term $i_{\alpha*} M_\alpha$ is $2(k-r)$.

Proof. Let K denote the set $\{1, 2, \dots, k\}$ and, for any subset S of K , let $E(S)$ denote the space of embeddings $S \rightarrow R^3$. Then $E(K) = \tilde{C}_k$. For any partition σ of K into disjoint subsets (K_1, \dots, K_r) put $E(\sigma) = \prod E(K_i)$. Then there is an obvious map $\lambda_\sigma: E(K) \rightarrow E(\sigma)$. If we pick an ordering of the K_i we can define a map $\mu_\sigma: E(\sigma) \rightarrow E(K)$ by identifying the i -th copy of R^3 , occurring in $E(\sigma)$, with the strip $i-1 < y_1 < i$ in the R^3 occurring in $E(K)$ (y_1, y_2, y_3 being the standard coordinates). The composition $\lambda_\sigma \mu_\sigma$ is homotopic to the identity and so the cohomology of $E(\sigma)$ appears as a direct summand of that of $E(K)$. If $\tau = (J_1, \dots, J_s)$ is another partition of K the non-empty intersections $K_i \cap J_j$ define a partition $\sigma\tau$. There is a map $\lambda_{\sigma, \tau}: E(\sigma) \rightarrow E(\sigma\tau)$ defined on each factor $E(K_i)$ using the partition of K_i given by the non-empty intersections $K_i \cap J_j$. Similarly, if we pick an ordering of the K_i there is a map $\mu_{\sigma, \tau}: E(\sigma\tau) \rightarrow E(\tau)$ whose component in the factor $E(J_j)$ is defined using the partition of J_j given by the non-empty intersections $K_i \cap J_j$, with the induced ordering. The commutative diagram

$$\begin{array}{ccc}
 E(\sigma) & \xrightarrow{\mu_\sigma} & E(K) \\
 \lambda_{\sigma, \tau} \downarrow & & \downarrow \lambda_\tau \\
 E(\sigma\tau) & \xrightarrow{\mu_{\sigma, \tau}} & E(\tau)
 \end{array}$$

shows that $\lambda_\tau \mu_\sigma$ factors through $E(\sigma\tau)$. Consider now the case $r = s$. Then by 5.1 the top-dimensional cohomology of $E(\sigma)$ and $E(\tau)$ is in dimension $|\sigma| = 2(k - r)$, while that of $E(\sigma\tau)$ is in dimension $|\sigma\tau|$. If $\sigma \neq \tau$ then $\sigma\tau$ is a strict refinement of σ and so $|\sigma\tau| < |\sigma|$. Hence $\lambda_\tau \mu_\sigma$ induces the zero homomorphism on $H^{|\sigma|}$. It follows that

$$\sum_\sigma \lambda_\sigma^*: \bigoplus_\sigma H^{|\sigma|}(E(\sigma)) \rightarrow H^*(E(K)) \tag{5.3}$$

is an embedding onto direct a summand. But from (5.1), we can compute the ranks of these groups:

$$\begin{aligned}
 \text{rank } H^*(E(K)) &= k! \\
 \text{rank } H^{|\sigma|}(E(\sigma)) &= (k_1 - 1)! (k_2 - 1)! \dots (k_r - 1)! .
 \end{aligned}$$

The fact that every permutation has a unique expression as a product of cycles leads to the identity:

$$k! = \sum_\sigma (k_1 - 1)! (k_2 - 1)! \dots (k_r - 1)!$$

from which we deduce that (5.3) is an isomorphism. Finally we consider the action of the symmetric group Σ_k on the terms in (5.3). Σ_k acts on the partitions σ of K with orbits corresponding to the partitions α of the integer k , and isotropy groups Σ_α . Thus the left-hand side of (5.3) is just $\bigoplus_\alpha i_{\alpha*}(M_\alpha)$ and so the proof of our Lemma is complete.

Lemma (5.2) enables us to compute the cohomology of Σ_k with $H^*(\tilde{C}_k)$ as coefficient module. Recall first that

$$H^s(\Sigma_k, i_{\alpha*}(M_\alpha)) \cong H^s(\Sigma_\alpha, M_\alpha)$$

and hence, for $s > 0$, it is annihilated by a power of the order of Σ_α . In particular if $k = p$ is an odd prime, and if α is not the trivial partition k , the order of Σ_α is prime to p and so

$$H^s(\Sigma_k, i_{\alpha_*}(M_\alpha))_p = 0 \tag{5.4}$$

where $()_p$ denotes localization at p . From (5.2) we see that only the top-dimensional cohomology group $H^{2(p-1)}(\tilde{C}_p)$ corresponds to the trivial partition and hence

$$H^s(\Sigma_p, H^t(\tilde{C}_p))_p = 0 \text{ for } 0 < t < 2p - 2, s > 0. \tag{5.5}$$

For $s = 0$ the groups $H^s(\Sigma_p, H^t(\tilde{C}_p))$ vanish for odd t and are torsion-free for even t . The Hochschild-Serre spectral sequence for the covering $\tilde{C}_p \rightarrow C_p$ then shows that we have an exact sequence

$$0 \rightarrow H^{2p-2}(\Sigma_p)_p \rightarrow H^{2p-2}(C_p)_p \rightarrow H^{2p-2}(\tilde{C}_p)_p$$

and hence

$$H^{2p-2}(\Sigma_p)_p \text{ maps isomorphically to the torsion subgroup of } H^{2p-2}(C_p)_p. \tag{5.6}$$

Now Z_p is the p -Sylow subgroup of Σ_p and hence detects its p -primary cohomology. In fact $H^*(Z_p)$ is generated freely by a class u of order p in H^2 and $H^*(\Sigma_p)_p$ can be identified with the subring generated by u^{p-1} . Moreover the total Chern class of the standard representation σ_p , restricted to Z_p , becomes

$$\begin{aligned} c(\sigma_p) &= (1 + u)(1 + 2u) \dots (1 + (p - 1)u) \\ &= (p - 1)! u^{p-1} = -u^{p-1}. \end{aligned}$$

Combined with (5.6) this shows

$$c_t(\sigma_p) = 0 \text{ in } (H^{2t}(\Sigma_p))_p \text{ for } t < p - 1, \tag{5.7}$$

$$c_{p-1}(\sigma_p) \text{ gives a generator of the torsion subgroup } (Z_p) \text{ of } H^{2p-2}(C_p(R^3))_p. \tag{5.8}$$

Passing from integral to mod p cohomology (5.7) and (5.8) give the $n = 1$ case of (4.5). To deal with larger values of n we use the map

$$(C_p)^n \rightarrow C_{np}$$

corresponding to the map λ_σ (for ordered sets) where σ is the partition of $\{1, \dots, np\}$ given by n blocks of p integers. On fundamental groups this induces the inclusion

$$(\Sigma_p)^n \rightarrow \Sigma_{np}.$$

Since the standard representation σ_{np} restricts to $\sigma_p \oplus \dots \oplus \sigma_p$ its total Chern class $c(\sigma_{np})$ restricts to $c(\sigma_p) \otimes \dots \otimes c(\sigma_p)$ in $H^*((\Sigma_p)^n)$. For the corresponding bundles V_{np}, V_p on C_{np}, C_p the same formula holds. Hence $c_{n(p-1)}(V_{np})$ restricted to $(C_p)^n$ contains the class $c_{p-1}(V_p) \otimes \dots \otimes c_{p-1}(V_p)$.

Reducing modulo p this shows that i) implies ii) in (4.5). Similarly $c_{p-1}^n(V_{np})$ restricted to $(C_p)^n$ contains the term $n! c_{p-1}(V_p) \otimes \dots \otimes c_{p-1}(V_p)$ and so for $n < p$ this is non-zero mod p , proving iii).

It remains to prove iv) for $1 < n < p$. But for $n < p$, Z_p^n is a p -Sylow subgroup of Σ_{np} and hence the p -primary part of $c(\sigma_{np})$ is determined by its restriction to Z_p^n , namely by $c(\sigma_p) \otimes \dots \otimes c(\sigma_p)$ in $H^*((Z_p)^n)$. Since $c(\sigma_p) = 1 - u^{p-1}$ the first non-zero class in the tensor product is c_{p-1} and so iv) is established.

Next we come to the proof of Lemma (4.8). Since both \mathcal{C} and \mathcal{F} have homotopy composition laws it is enough to prove that the map $\mathcal{C} \rightarrow \mathcal{F}$ given by the Dirac family is compatible with these compositions. Actually it is convenient to use the direct sum operation which clearly gives a commutative diagram

$$\begin{array}{ccc} \mathcal{C}(\text{SU}(2)) \times \mathcal{C}(\text{SU}(2)) & \rightarrow & \overline{\mathcal{F}} \times \overline{\mathcal{F}} \\ \downarrow \alpha & & \downarrow \\ \mathcal{C}(\text{SU}(4)) & \longrightarrow & \mathcal{F} \end{array}$$

In addition we have a homotopy commutative diagram

$$\begin{array}{ccc} \mathcal{C}(\text{SU}(2)) \times \mathcal{C}(\text{SU}(2)) & & \\ \beta \downarrow & \searrow \alpha & \\ \mathcal{C}(\text{SU}(2)) & \xrightarrow{\gamma} & \mathcal{C}(\text{SU}(4)) \end{array}$$

where α is again the direct sum, β is composition and γ is induced by the inclusion $\text{SU}(2) \rightarrow \text{SU}(4)$. If we replace $\mathcal{C}(G)$ by its homotopy equivalent $\Omega^3(G)$ the homotopy $\alpha \sim \gamma\beta$ arises from the well-known (rotation) homotopy between $A \oplus B$ and $AB \oplus 1$ as maps from $\text{SU}(2) \times \text{SU}(2)$ into $\text{SU}(4)$.

Putting together our two diagrams we conclude that $\mathcal{C}(\text{SU}(2)) \rightarrow \mathcal{F}$ is compatible with composition.

Finally we come to the proof of Proposition (4.13). Using Theorem (4.12) we see that we have to compare the two maps

$$\Omega^3(\text{SU}(2)) \rightarrow \Omega^3(U) \sim Z \times BU$$

given by σ^m and σ^1 . Taking Y to be any compact subspace of $\Omega^3(\text{SU}(2))$ it is convenient first to relate the two elements u_m and u in $K(Y)$ given by the maps $Y \rightarrow Z \times BU$. The element u is related to the 2-dimensional bundle ξ on $S^4 \times Y$ [defined by $Y \rightarrow \Omega^3(\text{SU}(2)) \sim \Omega^4(B\text{SU}(2))$] by

$$\beta^2(u) = \xi - 2,$$

where β^2 is the (iterated) Bott periodicity isomorphism

$$K(Y) \rightarrow K(S^4 \times Y, \infty \times Y).$$

Similarly u_m is given by

$$\beta^2(u_m) = \sigma^m(\xi) - (m + 1).$$

Using these two formulae we propose now to express u_m directly in terms of u . For this we introduce the Adams operations ψ^m (see [2]) which are expressible in terms of the σ^m by a polynomial formula:

$$\psi^m = m\sigma^m + \text{composite terms.}$$

Since composite terms vanish on a suspension this gives

$$\psi^m \beta = m \sigma^m \beta.$$

But $\psi^m \beta = m \beta \psi^m$ and since the algebra of operations in K -theory is torsion-free [2] we can cancel by m and deduce

$$\beta \psi^m = \sigma^m \beta.$$

Iterating this once gives

$$\sigma^m \beta^2 = m \beta^2 \psi^m \quad \text{or} \quad \beta^{-2} \sigma^m \beta^2 = m \psi^m.$$

Returning now to our formula for u_m we put $\eta = \xi - 2 = \beta^2(u)$ and expand

$$\sigma^m(\xi) = \sigma^m(\eta + 2) = \sigma^m(\eta) + 2\sigma^{m-1}(\eta) + 3\sigma^{m-2}(\eta) + \dots + (m + 1).$$

Thus

$$\begin{aligned} u_m &= \beta^{-2} \{ \sigma^m(\eta) + 2\sigma^{m-1}(\eta) + \dots + m\sigma^1(\eta) \} \\ &= m\psi^m(u) + 2(m-1)\psi^{m-1}(u) + \dots + m\psi^1(u). \end{aligned}$$

Taking Chern classes now gives the required formula for $f_m(t)$ in terms of $f(t)$ in (4.13): we need only recall that the effect of ψ^m is to replace t by mt and that the Chern polynomial takes sums into products.

§6. Comments and Comparisons

The main theme of this paper has been to demonstrate the intimate relation between the analysis of the Yang-Mills theory (in 4 dimensions) and the topology of the associated function space. It may be helpful, therefore, to compare our situation with that in some other Lagrangian theories.

The best understood case is that of Yang-Mills theory in 2 dimensions. This has been studied in detail, for all closed surfaces, from the point of view of Morse theory [5]. In particular when the base manifold is the 2-sphere we get (see [4]) a Morse picture which is quite analogous to the classical case of geodesics on G . Moreover the Dirac operator can now be interpreted in terms of the Cauchy-Riemann operator $\bar{\partial}$ and the stratification by the dimension of its null-space is closely related to the picture of the Morse flow. The analogy with geodesics is tied to the fact that both function spaces have here the same homotopy type, namely $\Omega(G)$. In both cases the function space is therefore connected (for simple G), unlike Yang-Mills over S^4 . Also critical points of arbitrary Morse index exist, i.e. the Euler equations have solutions which are not minima of the Lagrangian.

Another case which has been extensively studied is the “non-linear σ model” in which the Lagrangian is the “Energy” of a map $f : S^2 \rightarrow S^2$. Solutions of the Euler equations are called harmonic maps and it is known [13] that the only solutions are absolute minima of the Lagrangian and are given by holomorphic maps f (if degree f is positive) or anti-holomorphic maps (if degree f is negative). The function space [if we normalize $f(\infty) = 0$ say] is $\Omega^2(S^2)$ and has components given by the integer degree as in 4-dimensional Yang-Mills. Moreover we have the Segal map

$$C_k(\mathbb{R}^2) \rightarrow M_k$$

which assigns to a configuration of points $a_1, \dots, a_k \in R^2 = C$ the rational function

$$f(z) = \sum \frac{1}{z - a_i}.$$

Here M_k denotes the space of all rational functions f which satisfy $f(\infty) = 0$ and have precisely k poles (counting multiplicities). Arguing precisely as in §3 we deduce that, for $k \geq q$, $M_k \rightarrow \Omega_k^2(S^2)$ induces epimorphism in q -dimensional homology. Note that, even though M_k is a very simple space to describe, its actual homology groups do not appear to be known. The analogy with our topological discussion of Yang-Mills over S^4 is quite striking. This analogy has also suggested that higher critical points of the Yang-Mills functional on S^4 may not exist, i.e. that all solutions of the Yang-Mills equations are either self-dual (if $k \geq 0$) or anti-self-dual (if $k \leq 0$). As yet this question remains open.

For this non-linear σ model the only analogue of the Dirac operator is associated to a $U(1)$ -gauge theory and has no topological features analogous to §4.

References

1. Arnol'd, V.I.: The cohomology ring of the coloured braid group. *Math. Notes Acad. Sci. USSR* **5**, 138—140 (1969)
2. Atiyah, M.F.: *K-theory*. New York: Benjamin 1967
3. Atiyah, M.F.: Bott periodicity and the index of elliptic operators. *Quart. J. Math. (Oxford)* **19**, 113—140 (1968)
4. Atiyah, M.F.: Geometry of Yang-Mills fields. *Proc. Intern. Conf. Math. Physics, Rome 1977*
5. Atiyah, M.F., Bott, R.: (to appear)
6. Atiyah, M.F., Drinfeld, V.G., Hitchin, N.J., Manin, Yu.I.: Construction of instantons. *Phys. Lett.* **65A**, 185—187 (1978)
7. Atiyah, M.F., Hitchin, N., Singer, I.M.: Deformations of instantons. *Proc. Nat. Acad. Sci.* **74**, 2662—2663 (1977)
8. Atiyah, M.F., Hitchin, N., Singer, I.M.: Self-duality in four-dimensional Riemannian geometry. *Proc. Roy. Soc.* (to appear)
9. Atiyah, M.F., Singer, I.M.: The index of elliptic operators. I. *Ann. Math.* **87**, 484—530 (1968)
10. Atiyah, M.F., Singer, I.M.: The index of elliptic operators. IV. *Ann. Math.* **93**, 119—138 (1971)
11. Atiyah, I.M., Ward, R.S.: Instantons and algebraic geometry. *Commun. math. Phys.* **55**, 117—124 (1977)
12. Cohen, F.R., Lada, T.J., May, J.P.: The homology of iterated loop spaces. *Lecture notes in mathematics*, Vol. 533. Berlin-Heidelberg-New York: Springer 1976
13. Eells, J., Wood, J.C.: Restrictions on harmonic maps of surfaces. *Topology* **15**, 263—266 (1976)
14. Hartshorne, R.: Stable vector bundles and instantons. *Commun. math. Phys.* **59**, 1—15 (1978)
15. Jackiw, R., Rebbi, C., Nohl, C.: Conformal properties of pseudo-particle configurations. *Phys. Rev. D* **15**, 1642—1646 (1977)
16. Koschorke, U.: Infinite dimensional K-theory and characteristic classes of Fredholm bundle maps. *Proc. Symp. Pure Math.* **15**, 95—133 (1970)
17. Segal, G.: Configuration spaces and iterated loop spaces. *Inventiones math.* **21**, 213—221 (1973)
18. Singer, I.M.: Some remarks on the Gribov ambiguity. *Commun. math. Phys.* **60**, 7—12 (1978)

Communicated by A. Jaffe

Received March 7, 1978