GAUGE THEORIES, INSTANTONS AND ALGEBRAIC GEOMETRY

M. FORGER*

Institut für Theoretische Physik, Freie Universität Berlin, Arnimallee 3, 1000 Berlin 33

(*Received December 10, 1977*)

We review the definition of instanton (= pseudoparticle) solutions and their importance in the context of nonabelian gauge (= Yang-Mills) theories, as well as the recent progress, due to Atiyah and Ward, in their construction, using the Penrose twistor transform and methods of algebraic geometry. In particular, we present a proof of the theorem of Atiyah and Ward on the correspondence between SU(2) instanton solutions over the 4-sphere and certain algebraic 2-dimensional complex vector bundles over complex projective 3-space.

1. Introduction

During the last two years, the instanton problem, first formulated by Belavin, Polyakov, Schwarz and Tyupkin [1], [2], has received a great deal of attention both from physicists and mathematicians. The present review will therefore begin with a brief discussion of its position and importance in field theory. Sections 2–3 explain the context of gauge theories—as models of field theory—in which it arises and discuss some of its elementary aspects. Sections 4–7 are devoted to the progress that has been made in its solution so far, and especially to the beautiful recent work of Atiyah, Hitchin, Singer and Ward [3], [4], [5], [6]. There are no new results.

Recall first that in classical field theory, a system can be specified by its action functional S, defined on some set Δ of dynamical variables. The variational principle then states that the possible classical configurations of the system are the extremals of this action, which coincide with the solutions of the corresponding Euler–Lagrange field equations. On the other hand, for the definition of the corresponding quantum field theory, one usually starts from (formally) the same action functional S. Following the procedure of Feynman path integral quantization, one then has to construct a probability measure $d\mu$ on Δ such that the vacuum expectation value of any (observable or nonobservable) functional $\mathcal{O}$ on Δ is given by

$$\langle 0 \mid \mathcal{O}_{\mathrm{quant}} \mid 0 \rangle = \int_{\Delta} \mathcal{O}_{\mathrm{class}}(\phi) \exp\left(-S(\phi)\right) d\mu(\phi). \tag{1.1}$$

Since one is in general unable to perform such a functional integral explicitly, one tries to extract information about the qualitative behaviour of the correlation functions

* Supported by the Studienstiftung des deutschen Volkes.

$\langle 0 \mid \mathcal{O}_{\text{quant}} \mid 0 \rangle$ from a detailed investigation of the classical theory, hoping that eventually one might thus find a key to at least a qualitative understanding of certain phenomena in elementary particle physics, and in particular of the presently fundamental problem of quark confinement. It is certainly not unrealistic to believe that such an extraction should be possible in principle, since the classical theory is in some sense a limit of the quantum theory as $\hbar \to 0$, and the limiting procedure cannot destroy these informations entirely.

It should be mentioned here that whereas field theory admits a direct physical interpretation only in the Minkowski domain (i.e. over a space with a Lorentz metric $+ - \dots -$ or $+ \dots + -$, according to convention), the functional integration technique works—if at all—only in the Euclidean domain (i.e. over a space with a Riemann metric $+ \dots +$). But the models in field theory studied so far are all defined over flat space, and in this case there is a one-to-one correspondence between the theories in the Minkowski domain and in the Euclidean domain, based on complex analysis in a common flat complex space containing Minkowski space and Euclidean space as two different real subspaces. Curiously enough, this correspondence can be made precise on the quantum level (in the axiomatic approach à la Wightman and Osterwalder/Schrader as well as to every order of perturbation theory), but is rather formal on the classical level since it usually does not contain any reasonable relation between Cauchy initial value problems in the Minkowski domain and boundary value problems in the Euclidean domain. This leads us to expect the information on the qualitative behaviour of the correlation functions inherent in the two versions of the classical theory to be different from each other, at least under appropriate initial resp. boundary conditions.

These arguments provide a motivation for investigating the classical Minkowski *and* Euclidean versions of models in field theory. In the Euclidean case—on which we shall concentrate from now on since the instanton problem falls into this class—(1.1) suggests that one should first look for the minima of the action which may be expected to give the leading contributions to the functional integrals: In fact, experience with explicitly soluble models of field theory leads us to believe that in most cases (where the (Euclidean) action S is positive definite), the measure $d\mu$ is sufficiently well-behaved not to cancel the effect of the exponential damping factor $\exp(-S)$. However, there may exist subsidiary conditions such that apart from the absolute minimum, all the minima relative to some fixed value(s) for the functional(s) defining the subsidiary condition(s) give contributions to the functional integrals which are of the same order of magnitude and so cannot be neglected. Polyakov has argued [7] that this actually happens in the case of the instanton problem, where the subsidiary condition is given in terms of the "instanton number" or "winding number" and is of a topological nature.

To be slightly more specific, let us recall that *instantons*—also called *pseudoparticles*—are minima of the action, or equivalently, solutions to a certain natural system of first order nonlinear partial differential equations in classical 4-dimensional Euclidean pure gauge theories. (Compare Sec. 2 for the terminology.) In the context of nonabelian quantum gauge theories, their existence indicates [8], [9] that via a tunnelling effect, the apparent degeneracy of the vacuum is removed, and various forms of symmetry breaking occur, without the generation of additional (unwanted) Goldstone bosons. For more details, we refer the reader to the review papers [10], [11] where the role of classical solutions of the field equations for a nonperturbative analysis of quantum field theories is discussed extensively.

2. Gauge theories and characteristic classes

Since the geometric formulation of classical gauge theories in terms of connections in principal bundles and associated vector bundles, etc., is by now well understood [12], [13], [14], [15], we give only a brief review of the definitions. For the mathematical concepts involved, we refer the reader to the standard literature [16], [17].

To define a classical (Euclidean resp. Minkowski) pure gauge, or Yang–Mills, theory (in d dimensions), we have to be given the following data:

(a) A base space B, which is an oriented (d-dimensional) manifold, equipped with a (Riemannian resp. Lorentz) metric g. Then B also carries a natural volume form ε, defined as taking the value $+1$ on any positive orthonormal frame. Moreover, given any Riemannian real vector bundle E over B, the usual pointwise operations of the Hodge theory can be generalized from ordinary to E-valued differential forms to yield

(i) an exterior product $\wedge_E$ taking an E-valued p-form α and an E-valued q-form β to an ordinary $(p+q)$-form $\alpha\wedge_E\beta$ (using only the Riemannian metric on the fibres of E here),

(ii) an inner product $(\cdot,\cdot)$ taking an E-valued p-form α and an E-valued q-form β to a function (α,β) ($(\alpha,\beta)=0$ if $p\neq q$),

(iii) a star operator $*$ taking an E-valued p-form α to an E-valued $(d-p)$-form $*\alpha$,

satisfying

$$**\alpha=(-1)^{p(d-p)+s}\alpha,\qquad \alpha\wedge_E *\beta=\beta\wedge_E *\alpha=(-1)^s(\alpha,\beta)\varepsilon \tag{2.1}$$

for E-valued p-forms α, β. Here, s denotes the dimension of the maximal subspace on which g is negative definite (i.e. $s=0$ resp. $s=d-1$).

(b) An internal symmetry group or gauge group G, which is a compact Lie group, equipped with a positive definite inner product $(\cdot,\cdot)$ on its Lie algebra $\mathfrak{g}$, invariant under the adjoint representation Ad of G on $\mathfrak{g}$. (If G is also semisimple, we choose $(\cdot,\cdot)$ to be the negative of the Killing form, i.e. $-(X,Y)=\operatorname{tr}\operatorname{ad}(X)\operatorname{ad}(Y)$ for $X,Y\in\mathfrak{g}$.)

The dynamical variable is now an equivalence class $[P,A]$ of pairs (P,A), where P is a principal G-bundle over B and A is a connection form on P, with (P,A) and (P',A') equivalent iff there exists an affine isomorphism of principal G-bundles with connection between (P,A) and (P',A'). We call this equivalence "gauge equivalence" since an isomorphism of this type is known among physicists as a (global) gauge transformation, at least in the case of trivial bundles. The principle of gauge invariance then states that all physically observable quantities given in terms of the pair (P, A) are in fact gauge invariant, i.e. depend only on its gauge equivalence class $[P, A]$. In particular, this applies to the Yang–Mills action S which is defined as follows:

The connection form A, which is an Ad-covariant, $\mathfrak{g}$-valued 1-form on P appropriately normalized on vertical tangent vectors, determines the curvature form F as the horizontal, Ad-covariant $\mathfrak{g}$-valued 2-form

$$F = dA + \tfrac{1}{2}[A, A] \tag{2.2}$$

on P. Now F can also be considered as a 2-form on B, taking its values in the Riemannian real vector bundle $P \times_G \mathfrak{g}$ over B associated to P and the adjoint representation Ad of G on $\mathfrak{g}$, with its Riemannian metric induced from the Ad-invariant positive definite inner product $(\cdot, \cdot)$ on $\mathfrak{g}$. Hence the operations mentioned under (a) above can be applied to define the Yang–Mills action density $(F, F)\varepsilon = (-1)^s F \wedge_{\mathfrak{g}} *F$ as a gauge invariant ordinary d-form on B. ($\wedge_{\mathfrak{g}}$ is an abbreviation for $\wedge_{P \times_G \mathfrak{g}}$.) Up to a normalization factor depending on the specific choice of the gauge group G and the inner product $(\cdot, \cdot)$ on its Lie algebra $\mathfrak{g}$ and involving the coupling constant of the theory, its integral is the Yang–Mills action

$$S = \mathrm{const} \int_B (F, F)\varepsilon = \mathrm{const}(-1)^s \int_B F \wedge_{\mathfrak{g}} *F. \tag{2.3}$$

The extremals of this action are the solutions to the so-called *free Yang–Mills equations*

$$D *F = 0. \tag{2.4}$$

Sometimes, following the example ($G = U(1)$) of Maxwell theory, one also considers the Bianchi identity

$$DF = 0 \tag{2.5}$$

as part of the Yang–Mills equations. Here, D denotes the covariant exterior derivative in the vector bundle $P \times_G \mathfrak{g}$ associated to the connection form A on the principal G-bundle P.

Before proceeding further, we want to explain briefly why the classical gauge theories described above are called "pure", and why the Yang–Mills equations (2.4) are called "free". The reason is that the dynamical variable contains only a multiplet A of gauge fields, but no multiplet ψ of matter fields (and/or Higgs fields), and that the r.h.s. of (2.4), which in the more general case is a source term, vanishes. In fact, if more generally we want to define a classical (Euclidean resp. Minkowski) gauge, or Yang–Mills, theory (in d dimensions), we have to be given as additional data:

(c) A unitary representation of G on a finite-dimensional complex vector space E.

The dynamical variable is then a gauge equivalence class $[P, A, \psi]$ of triples (P, A, ψ), where P, A are as before and ψ is a section of a Hermitian vector bundle E over B, with (P, A, ψ) and (P', A', ψ') gauge equivalent iff there exists an isomorphism between (P, A) and (P', A') as before and such that ψ and ψ' are related by the induced vector bundle isomorphism between E and E'. (E is the associated vector bundle $P \times_G \mathrm{E}$ over B if ψ

is to contain only scalar matter fields, and—assuming that $w_2(B) = 0$ and choosing a spin structure on B—its tensor product with an appropriate spinor bundle over B if ψ is to contain arbitrary spinor matter fields.) Moreover, the action functional contains additional terms involving ψ and its covariant derivative $D\psi$, and its extremals are the solutions to the Yang–Mills equations $D * F = j$, $DF = 0$, where the source term j is the matter field current whose definition also involves ψ and its covariant derivative $D\psi$, plus a partial differential equation for ψ. (This is the covariant Klein–Gordon equation if ψ contains only scalar matter fields, and an appropriate covariant higher spin equation, such as the covariant Dirac equation, if ψ contains arbitrary spinor matter fields.) For more details, see [18].

In the following, according to what has been said in the Introduction, B is assumed to be Riemannian, i.e. $s = 0$.

We now turn to the problem of choosing appropriate boundary conditions. Of course, the action is always finite if B is compact. However, it turns out that even if B is not compact, one can often find a suitable compactification, i.e. a compact base space $\bar{B}$ containing the noncompact base space B as an open dense oriented submanifold, such that the Riemann metric g on B is equal or at least conformally equivalent to the restriction of the Riemann metric $\bar{g}$ on $\bar{B}$ to B. Then for any pair (P, A) which arises from a pair $(\bar{P}, \bar{A})$ by restriction (with P resp. $\bar{P}$ a principal G-bundle over B resp. $\bar{B}$ and A resp. $\bar{A}$ a connection form on P resp. $\bar{P}$), the action is finite; in fact,

$$S = S(P, A) = S(\bar{P}, \bar{A}) < \infty. \tag{2.6}$$

In other words, the existence of an extension $(\bar{P}, \bar{A})$ of (P, A) over the given compactification $\bar{B}$ of B (a geometric boundary condition) implies the finiteness of the action of (P, A) (an integrability requirement and hence an analytic boundary condition). Whether or to what extent the converse is true, is—even in very special cases—still an open problem.

The fact that the geometric boundary condition has turned out to be much easier to handle than the analytic boundary condition constitutes the prime motivation for considering classical Euclidean gauge theories over compact rather than noncompact base spaces. Of course, if the theory is originally defined over a noncompact base space B, additional criteria concerning the choice of the compactification $\bar{B}$ should be given. For example, the instanton problem was originally formulated over $B = \mathbf{R}^4$ [2], but the physicists' arguments and computations all amount to working over the one-point compactification $\bar{B} = S^4$, which in some sense is the "simplest" and "minimal" one. The usual argument here is that this compactification is natural due to the conformal invariance of the problem. Strictly speaking, however, one can only talk about conformal invariance if one is able to lift the action of the conformal group from the base space B to the principal G-bundle P over B (we omit the bars here); this is the typical lifting problem one faces whenever one investigates space-time symmetries in classical gauge theories. Now such a lifting will not exist in general—there is an obstruction which can be expressed in terms of cohomology—but it does of course exist and is unique up to a bundle automorphism, i.e. a (global) gauge transformation, if P is trivial. Actually, as long as one is dealing with equations—such as the free Yang–Mills equations—which involve only the curvature F and not the connection A, it is sufficient to lift to $P \times_G \mathfrak{g}$ rather than P itself, which is certainly possible if $P \times_G \mathfrak{g}$ rather than P itself is trivial. This situation occurs if P admits a reduction of structure group from G to the centre of G, and in particular if G is Abelian, so that in Abelian theories such as Maxwell theory, the problem disappears.

As a consequence of their intrinsically geometric nature, classical gauge theories admit interesting invariants which come from the topology of the bundles involved and are usually expressed in terms of characteristic classes. Following the construction leading to the Weil homomorphism [16], [17], let us recall briefly how these are defined:

Let $\Gamma \in (V^r\mathfrak{g}^*)_I$ be a symmetric invariant of degree r on the Lie algebra $\mathfrak{g}$ of G, considered as an Ad-invariant homogeneous polynomial of degree r or as an Ad-invariant symmetric r-linear functional on $\mathfrak{g}$. (Both interpretations are equivalent by polarizing resp. by restricting to the diagonal.) Then formally inserting the $(P \times_G \mathfrak{g})$-valued curvature 2-form F on B into each slot of Γ, one obtains a gauge invariant ordinary $2r$-form $\Gamma(F, \ldots, F)$ on B, which due to the Bianchi identity (2.5) for F is closed and hence defines a $2r$-dimensional de Rham cohomology class $[\Gamma(F, \ldots, F)] \in H^{2r}(B, R)$ on B, called the *characteristic class* of P associated to Γ. This terminology is justified in view of the theorem of A. Weil which states that the latter is in fact independent of the connection form A and hence a topological invariant of the underlying principal G-bundle P. The resulting correspondence

$$\begin{array}{ccc} (V\mathfrak{g}^*)_I = \bigoplus\limits_{r \geqslant 0} (V^r\mathfrak{g}^*)_I & \to & \bigoplus\limits_{r \geqslant 0} H^{2r}(B, R) = H^{\mathrm{even}}(B, R) \\ \Gamma & \mapsto & [\Gamma(F, \ldots, F)] \end{array} \tag{2.7}$$

is known as the *Weil homomorphism.*

More explicitly, the form $\Gamma(F, \ldots, F)$ is defined as follows: Observe that due to its Ad-invariance, Γ defines a symmetric r-linear functional Γ_b on the fibre of the associated vector bundle $P \times_G \mathfrak{g}$ over b, for any $b \in B$; this family can be viewed as a vector bundle map from the vector bundle $(P \times_G \mathfrak{g}) \otimes \ldots \otimes (P \times_G \mathfrak{g})$ (r factors) to the trivial vector bundle $B \times R$. Then using a generalized exterior product, $\Gamma(F, \ldots, F)$ is given by

$$\begin{aligned} &\Gamma(F, \ldots, F)_b(u_1, \ldots, u_{2r}) \\ &= (\tfrac{1}{2})^r \sum_{\substack{\sigma \text{ permutation} \\ \text{of } \{1,\ldots,2r\}}} (-1)^\sigma \Gamma_b\{F_b(u_{\sigma(1)}, u_{\sigma(2)}), \ldots, F_b(u_{\sigma(2r-1)}, u_{\sigma(2r)})\} \end{aligned} \tag{2.8}$$

$$\text{for} \quad b \in B,\ u_1, \ldots, u_{2r} \in T_bB.$$

In most cases, $G \subset \mathrm{GL}(n, C)$ is a matrix Lie group. Then the adjoint representation of $\mathrm{GL}(n, C)$ on $\mathrm{gl}(n, C)$, restricted to G, leaves $\mathfrak{g}$ invariant and induces the adjoint representation of G on $\mathfrak{g}$. Hence an $\mathrm{Ad}_{\mathrm{GL}(n,C)}$-invariant homogeneous polynomial of degree r on $\mathrm{gl}(n, C)$ restricts to an Ad_G-invariant homogeneous polynomial of degree r on $\mathfrak{g}$. Examples of the former are given by the rth characteristic coefficient C_r and the rth trace coefficient Tr_r:

$$C_0(X) = 1, \quad C_r(X) = \sum_{1\leqslant i_1<\ldots<i_r\leqslant n} \lambda_{i_1} \ldots \lambda_{i_r},$$

$$\mathrm{Tr}_0(X) = n, \quad \mathrm{Tr}_r(X) = \sum_{1\leqslant i\leqslant n} \lambda_i^r \tag{2.9}$$

for $X \in \mathrm{gl}(n, C)$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ (counted according to their multiplicities).

Up to normalization constants, these yield the rth Chern class c_r and the rth component ch_r of the Chern character, explicitly represetned by the differential forms

$$c_r = \left(-\frac{1}{2\pi i}\right)^r C_r(F, \ldots, F) \tag{2.10}$$

and

$$\mathrm{ch}_r = \left(-\frac{1}{2\pi i}\right)^r \frac{1}{r!} \mathrm{Tr}_r(F, \ldots, F), \tag{2.11}$$

respectively; in particular, $\mathrm{ch}_1 = c_1$ and $\mathrm{ch}_2 = -c_2+c_1^2/2$. Moreover, the associated vector bundle $P\times_G \mathfrak{g}$ can be considered as a vector subbundle of the endomorphism bundle $P\times_G \mathrm{gl}(n, C)$ of the associated vector bundle $P\times_G C^n$, i.e. each fibre of $P\times_G \mathfrak{g}$ (resp. $P\times_G \mathrm{gl}(n, C)$) can be considered as a Lie algebra of linear transformations (resp. as the Lie algebra of all linear transformations) on the corresponding fibre of $P\times_G C^n$. Therefore one also has

(iv) an exterior product $\wedge$ taking a $(P\times_G \mathrm{gl}(n, C))$-valued p-form α and a $(P\times_G \mathrm{gl}(n, C))$-valued q-form β to a $(P\times_G \mathrm{gl}(n, C))$-valued $(p+q)$-form $\alpha\wedge\beta$ (using "matrix multiplication"),

which leads to the following form of (2.11):

$$\mathrm{ch}_r = \left(-\frac{1}{2\pi i}\right)^r \frac{1}{r!} \mathrm{tr} F\wedge \ldots \wedge F. \tag{2.12}$$

In particular, the first two Chern classes of the associated vector bundle $P\times_G C^n$ are represented by the differential forms

$$c_1 = -\frac{1}{2\pi i} \mathrm{tr} F \tag{2.13}$$

and

$$c_2 = \frac{1}{8\pi^2} \mathrm{tr} F\wedge F, \quad \text{if } G \subset \mathrm{SL}(n, C), \tag{2.14}$$

since $G \subset \mathrm{SL}(n, C)$ implies $c_1 = 0$. From topology, the Chern classes c_r are actually known to be integral, i.e. to belong to $H^{2r}(B, Z) \subset H^{2r}(B, R)$.

For many physicists, cohomology classes are still a somewhat mysterious kind of invariant; they would prefer numbers. Fortunately, if B is compact, there is a canonical isomorphism

$$\int : H^d(B, R) \to R \qquad [\omega] \to \int_B \omega \tag{2.15}$$

under which $H^d(B, Z)$ corresponds to Z, so that real resp. integral cohomology classes in the top dimension correspond to real numbers resp. integers. Now the interesting values for the dimension of B are $d = 2$ and $d = 4$, since for models in field theory the 2-dimensional version often serves as a good training ground for the 4-dimensional one. Although this is indeed the case here and rather complete results can be obtained [6], we shall—mainly because of the limited amount of space available—turn directly to the 4-dimensional problem.

3. Formulation of the instanton problem

In four dimensions, we can integrate the characteristic class of P associated to the symmetric invariant $(\cdot,\cdot)$ of degree 2 on $\mathfrak{g}$, obtaining a topological invariant

$$k = \text{const} \int_B (F, *F)\varepsilon = \text{const} \int_B F \wedge_{\mathfrak{g}} F, \tag{3.1}$$

called the *instanton number* or *winding number*; notice the similarity with the definition of the action

$$S = \text{const} \int_B (F, F)\varepsilon = \text{const} \int_B F \wedge_{\mathfrak{g}} *F, \tag{3.2}$$

also on the level of densities. Since the star operator is an isometry with respect to the inner product $(\cdot,\cdot)$, the inequality

$$0 \leqslant (F \pm *F, F \pm *F) = (F, F) \pm 2(F, *F) + (*F, *F) = 2\big((F, F) \pm (F, *F)\big) \tag{3.3}$$

implies the important relations

$$|(F, *F)| \leqslant (F, F) \quad \text{and} \quad \begin{array}{l} +(F, *F) = (F, F) \Leftrightarrow *F = +F, \\ -(F, *F) = (F, F) \Leftrightarrow *F = -F, \end{array} \tag{3.4}$$

which can be integrated to yield

$$|k| \leqslant S \quad \text{and} \quad \begin{array}{l} +k = S \Leftrightarrow *F = +F, \\ -k = S \Leftrightarrow *F = -F. \end{array} \tag{3.5}$$

Therefore, the minima of the action are the solutions to the so-called *self-dual* or *anti-self-dual equations*

$$*F = F \quad \text{or} \quad *F = -F \tag{3.6}$$

according to whether $k \geqslant 0$ or $k \leqslant 0$, respectively, and these are the instanton solutions (occasionally called *instanton* or *anti-instanton solutions*, respectively) we are interested in.

Evidently, minima of the action are extremals while it is still an unsolved and partially controversial question whether there exist extremals of the action which are not minima.

From the point of view of partial differential equations (for the connection form), the Bianchi identity (2.5) makes it evident that a solution to the first order system (3.6) is also a solution to the second order system (2.4), while it is not surprising that the converse is a nontrivial problem.

To be more specific about the common normalization constant in (3.1) and (3.2), let us assume that G is the simple Lie group $SU(n)$ whose Killing form —which coincides with that of its complexification $SL(n, C)$—is given by

$$-(X, Y) = \operatorname{tr} \operatorname{ad}(X)\operatorname{ad}(Y) = 2n\operatorname{tr}XY \quad \text{for} \quad X, Y \in \mathrm{su}(n). \tag{3.7}$$

This implies the following relation between the products $\wedge_{\mathfrak{g}}$ of Sec. 2, (i) and $\wedge$ of Sec. 2 (iv):

$$\alpha \wedge_{\mathfrak{g}} \beta = -2n \operatorname{tr} \alpha \wedge \beta \tag{3.8}$$

for any $(P \times_{SU(n)} \mathrm{su}(n))$-valued p-form α and any $(P \times_{SU(n)} \mathrm{su}(n))$-valued q-form β. Hence choosing the normalization constant to be $1/16\pi^2 n$, we obtain

$$\begin{aligned} k &= \frac{1}{16\pi^2 n}\int_B (F, *F)\varepsilon = -\frac{1}{8\pi^2}\int_B \operatorname{tr} F \wedge F, \\ S &= \frac{1}{16\pi^2 n}\int_B (F, F)\varepsilon = -\frac{1}{8\pi^2}\int_B \operatorname{tr} F \wedge *F. \end{aligned} \tag{3.9}$$

Thus $-k$ is the integral of the second Chern class $c_2(E)$ of the associated vector bundle $E = P \times_{SU(n)} C^n$; in particular, k is an integer. From now on, we shall identify $c_2(E)$ and $-k$ if not explicitly stated otherwise. Moreover, in this case k is actually the only topological invariant of the theory:

PROPOSITION 1. *If B is connected, the integer k determines the principal* $SU(n)$*-bundle P completely up to an isomorphism.*

Proof: *Step* 1: $n = 2$. Given a principal $SU(2)$-bundle P over B, form the associated complex vector bundle $E = P \times_{SU(2)} C^2$ over B and—considering it as an oriented Riemannian real vector bundle E_R over B—the corresponding oriented sphere bundle S over B; then since $SU(2)$ acts on $S^3 \subset R^4 \cong C^2$ transitively and without fixed points, there is a natural isomorphism of P onto S (as fibre bundles over B). Now S admits a section, and hence P is trivial, iff the Euler class χ_S of S vanishes. But $\chi_S = \mathrm{pf}(E_R) = c_2(E)$.

Step 2: $n > 2$. Since on the level of isomorphism classes, there is a one-to-one correspondence

$$\begin{array}{ccc} \text{principal } SU(n) \text{ bundles} & & n \text{ dimensional complex vector bundles} \\ \text{over } B & \leftrightarrow & \text{over } B \text{ with } c_1 = 0 \\ P & \mapsto & E = P \times_{SU(n)} C^n \end{array},$$

it suffices to show that if $c_2(E) = 0$, E must be trivial. But as B is 4-dimensional, E can always be decomposed in the form $E = F \oplus G$ with G the $(n-2)$-dimensional trivial complex vector bundle over B, and $c_1(F) = c_1(E) = 0$. Therefore $c_2(F) = c_2(E) = 0$, and by Step 1, F is trivial. Hence so is E. ∎

Let us finally state precisely what we mean by "the instanton problem": Given a compact 4-dimensional base space B, a gauge group $G \subset \mathrm{GL}(n, C)$ and an integer k, it is the problem of determining the minima of the action (3.2) under the subsidiary condition (3.1), i.e. the self-dual (if $k \geqslant 0$) resp. anti-self-dual (if $k \leqslant 0$) connections in principal G-bundles P over B with prescribed instanton number k. More explicitly, it can be broken up into a couple of questions of increasing difficulty:

(1) Do solutions exist, i.e. is the minimum attained?

(2) What is the local structure of the space of solutions, i.e. its dimension, and is it locally "regular", i.e. a manifold?

(3) What is the global structure of the space of solutions, e.g. is it connected, simply connected, ...?

(4) Can one give an explicit construction which yields the most general solution?

Here, "solutions" really mean "gauge equivalence classes of solutions", also called "moduli" in the mathematical literature. Notice also that using pull-back via an orientation reversing diffeomorphism on B—such as the antipodal map on S^4, for example—if necessary, one can assume without loss of generality that $k \geqslant 0$ or that $k \leqslant 0$. Of course, this requires B to be reversible; a necessary condition is that the signature of B vanishes (cf. [17], Vol. I, p. 206).

The interesting—and hopefully typical—case where attention has concentrated and results have been obtained is $B = S^4$ (the "simplest" compact 4-dimensional manifold) and $G = \mathrm{SU}(2)$ (the "simplest" compact nonabelian Lie group). The following sections will describe these results and provide at least partial answers to questions (1)–(4) above. However, it seems perhaps useful first to indicate the close analogy with the basic problem in the Hodge theory of harmonic forms, where on the space $\Omega^p(B)$ of ordinary p-forms ω (as the set of dynamical variables) over a compact oriented Riemannian manifold B (as base space), one considers the square of the L^2-norm

$$|\omega|^2 = \int_B (\omega, \omega)\varepsilon = \int_B * \omega \wedge \omega \tag{3.10}$$

(as the action functional) and the de Rham cohomology class

$$[\omega] \in H^p(B, R) \tag{3.11}$$

(as the analogue of the winding number k or rather the corresponding cohomology class), and again asks for the minima of the action (3.10) under the subsidiary condition (3.11), i.e. for the p-forms ω on B with minimal L^2-norm and prescribed cohomology class $[\omega] = c$. As a consequence of the Hodge decomposition theorem [19], it turns out that there is actually a unique minimum, namely the unique harmonic form ω_0 on B with cohomology class $[\omega_0] = c$. For $p = 2$, the analogy becomes even closer if one considers $-i\omega = F$ as the curvature form of a connection form A in a principal $U(1)$-bundle P over B (thus changing the set of dynamical variables from the space $\Omega^2(B)$ to the set of gauge equivalence classes of pairs (P, A)), forcing $[\omega/2\pi]$ to be integral, i.e. to belong to $H^2(B, \mathbf{Z}) \subset H^2(B, \mathbf{R})$, since it is just the first Chern class of the associated complex line bundle $P \times_{U(1)} \mathbf{C}$. Notice, however, that the subsidiary conditions—although they look similar—are by no means identical, so that the resulting partial differential equations ($\Delta\omega = 0$ and $*F = \pm F$, respectively) for a minimum are also different; in particular, the former is of second order, but *linear* (since $U(1)$ is abelian), while the latter is of first order, but *nonlinear* (whenever G is nonabelian). This is the main reason why the instanton problem can be considered as substantially more difficult than a proof of the Hodge decomposition theorem, which already is a deep theorem in harmonic analysis.

4. Explicit solutions and deformation theory (a brief survey)

The first instanton solutions for $|k| = 1$ were given by Belavin, Polyakov, Schwarz and Tyupkin [2]. Later, t'Hooft (unpublished) exhibited an "Ansatz" which was exploited by Jackiw, Nohl and Rebbi [20] to find an $n(k)$-parameter family of instanton solutions for any $k \neq 0$, where

$$n(\pm 1) = 5, \quad n(\pm 2) = 13, \quad n(k) = 5|k|+4 \quad \text{for} \quad |k| \geqslant 3. \tag{4.1}$$

We leave it to the reader to translate their formulae from $\mathbf{R}^4$ to S^4, using the conformal embedding $\mathbf{R}^4 \overset{c}{\hookrightarrow} S^4$ defined by stereographic projection and a suitable transition function for the principal SU(2)-bundle over S^4 classified by the integer $-k$.

At about the same time, Jackiw and Rebbi [21] and Schwarz [22] proved independently that for any $k \neq 0$, the dimension of the space of instanton solutions (moduli) is actually

$$r(k) = 8|k|-3. \tag{4.2}$$

This value for $r(k)$ had actually been conjectured before using the physical picture that an instanton solution with instanton number k is really a "multiinstanton" suitably composed (in a nonlinear fashion) of k "single instantons", and that each of these is determined by 4 position parameters for its "centre of mass" $+1$ scale parameter for its "size" $+3$ orientation parameters for its "orientation" in the 3-dimensional Lie algebra su(2); finally, 3 parameters have to be subtracted for the "overall gauge freedom" since we are only looking for gauge equivalence classes of solutions. The idea of the proof is to appropriately linearize the field equations (3.6) around some given solution and thus to obtain an elliptic linear differential operator whose analytic index is, due to a vanishing theorem, $r(k)$ and whose topological index is $8|k|-3$. (4.2) then follows from the Atiyah–Singer index theorem. The final step in this direction was taken by Atiyah, Hitchin and Singer [3] who applied an integrability theorem of Kuranishi [23] which justifies the linearization procedure, thus proving the following

THEOREM 1 [3]. *The space of* SU(2)*-instanton solutions (moduli) over S^4 with instanton number $k \neq 0$ is a real analytic manifold of dimension* $8|k|-3$.

This deformation theoretical approach can obviously be generalized to other base spaces than S^4 and other gauge groups than SU(2), and results in this direction have recently been obtained [24], [25]. On the other hand, it tells us nothing about the global structure of the manifold of solutions, or about methods for explicitly constructing the most general solution. There is, however, a different approach due to Atiyah and Ward [5], which goes a long way towards an answer to those global questions, and which is based on the observation that using the Penrose twistor transform, one can translate the instanton problem over S^4 into a problem in algebraic geometry.

5. The Penrose twistor transform

To apply methods of algebraic and/or analytic geometry, it is clearly necessary to introduce some sort of complex structure into the theory. But S^4 does not admit a complex structure, not even an almost complex structure, so that we face the problem of "complexifying" S^4 in some natural and minimal way. One way to motivate the particular procedure we have in mind is the following:

Given a $2n$-dimensional oriented Riemannian manifold X, let $B_{SO(2n)}(X)$ be its positive orthonormal frame bundle, which is a principal $SO(2n)$-bundle over X. According to the general definition of G-structures on manifolds, an almost complex structure on X, compatible with the given orientation and the Riemann metric on X, is a reduction of structure group of $B_{SO(2n)}(X)$ from $SO(2n)$ to $U(n)$, which is well known to exist iff the associated fibre bundle $B_{SO(2n)}(X)\times_{SO(2n)}SO(2n)/U(n)$, with typical fibre the homogeneous space $SO(2n)/U(n)$, admits a section. Observe also that $SO(2n)/U(n)$ is precisely the set of all complex structures on the vector space $\mathbf{R}^{2n}$, compatible with the standard orientation and Riemann metric on $\mathbf{R}^{2n}$. Moreover, if $n = 2$,

$$SO(4)/U(2) \cong (SU(2)\times SU(2))/Z_2/(SU(2)\times U(1))/Z_2 \cong SU(2)/U(1) \cong S^2. \quad (5.1)$$

For $X = S^4$, this associated fibre bundle can be described in several other ways:

As a complex vector space, the (noncommutative) field $\mathbf{H}$ of quaternions is just $\mathbf{C}^2$, i.e. there is an isomorphism

$$\begin{array}{ccc} \mathbf{C}^2 & \to & \mathbf{H} \\ (z_1, z_2) = (x_1+iy_1, x_2+iy_2) & \mapsto & z_1+z_2 j = x_1+iy_1+jx_2+ky_2 = z_1-j\overline{z_2} \end{array} \quad (5.2)$$

inducing a corresponding isomorphism $\mathbf{C}^{2n+2} \xrightarrow{\cong} \mathbf{H}^{n+1}$, for any n. Since $\mathbf{C} \subset \mathbf{H}$, this isomorphism factors to yield a map

$$\pi_n\colon \mathbf{C}P^{2n+1} = \mathbf{C}^{2n+2}/\mathbf{C}-\{0\} \to \mathbf{H}^{n+1}/\mathbf{H}-\{0\} = \mathbf{H}P^n \quad (5.3)$$

of projective spaces which actually turns $\mathbf{C}P^{2n+1}$ into a fibre bundle over $\mathbf{H}P^n$ with typical fibre $\mathbf{C}P^1 = S^2$ and makes the diagram

$$\begin{array}{ccc} & S^{4n+3} & \\ \swarrow & & \searrow \\ \mathbf{C}P^{2n+1} & \xrightarrow{\pi_n} & \mathbf{H}P^n \end{array} \quad (5.4)$$

commutative, where the vertical arrows are the complex and quaternionic Hopf fiberings, respectively. More explicitly, observe that under the canonical projection from $\mathbf{C}^{2n+2}$

to CP^{2n+1}, the set of projective lines, i.e. of lines in CP^{2n+1}, corresponds to the Grassmannian $G_{2,2n+2}$ of 2-planes in C^{2n+2}, and that the antilinear antiinvolution of left multiplication by j in $C^{2n+2} \cong H^{n+1}$ factors to yield an antilinear involution σ on CP^{2n+1}. σ is a "real structure" on CP^{2n+1} different from the ordinary "real structure" on CP^{2n+1} induced by the usual complex conjugation on C^{2n+2}; in terms of Lie groups, these two "real structures" correspond to the two different real forms $SU(2n+2)$ and $SL(2n+2, R)$ of $SL(2n+2, C)$, consisting of those matrices in $SL(2n+2, C)$ which commute with left multiplication by j and with the usual complex conjugation, respectively. Notice that σ has no real (i.e. σ-stable) points since for any nonzero vector $z \in C^{2n+2}$, z and jz are linearly independent and so cannot represent the same point in CP^{2n+1}, but σ does have real (i.e. σ-stable) lines which under the canonical projection from C^{2n+2} to CP^{2n+1} correspond to the 2-planes in C^{2n+2} generated by z and jz, i.e. to the 2-planes Hz, for some nonzero vector $z \in C^{2n+2}$. Hence these real lines are precisely the fibres of π_n and of the form $Hz/C-\{0\}$ for some nonzero vector $z \in C^{2n+2}$, implying that they are isomorphic to $CP^1 \cong S^2$ and that under this isomorphism, σ acts on them as the antipodal map on S^2. Explicitly, we use stereographic projection from the north pole $(0, 1)$ as the second map in the sequence

$$\begin{array}{ccccc} CP^1-\{[0,1]\} & \to & C \cong R^2 & \to & S^2-\{(0,1)\} \\ [z_1, z_2] & \mapsto & z_1^{-1}z_2 = \dfrac{\overline{z_1}z_2}{|z_1|^2} & \mapsto & \left(\dfrac{2\overline{z_1}z_2}{|z|^2}, \dfrac{-|z_1|^2+|z_2|^2}{|z|^2}\right) \end{array} \tag{5.5}$$

which extends to an explicit diffeomorphism $CP^1 \xrightarrow{\cong} S^2$, so that

$$\begin{array}{ccc} CP^1 \to CP^1 & & S^2 \to S^2 \\ [z_1, z_2] \mapsto [-\overline{z_2}, \overline{z_1}] & \text{corresponds to} & a \mapsto -a. \end{array}$$

In particular, if $n = 1$ where $HP^1 \cong S^4$, we see that complex projective 3-space CP^3 is a fibre bundle

$$\pi\colon CP^3 \to S^4 \tag{5.6}$$

over the 4-sphere S^4. Explicitly, we use stereographic projection from the north pole $(0, 1)$ as the second map in the sequence

$$\begin{aligned} HP^1-\{[0,1]\} \to \quad & H = R^4 \\ [z_1+z_2j, z_3+z_4j] \mapsto (z_1+z_2j)^{-1}(z_3+z_4j) &= \frac{\overline{z_1}z_3+z_2\overline{z_4}+(\overline{z_1}z_4-z_2\overline{z_3})j}{|z_1|^2+|z_2|^2} \\ \to \quad & S^4-\{(0,1)\} \\ \mapsto \left(\frac{2\{\overline{z_1}z_3+z_2\overline{z_4}+(\overline{z_1}z_4-z_2\overline{z_3})j\}}{|z_1|^2+|z_2|^2+|z_3|^2+|z_4|^2}\right. &, \left.\frac{-|z_1|^2-|z_2|^2+|z_3|^2+|z_4|^2}{+|z_1|^2+|z_2|^2+|z_3|^2+|z_4|^2}\right) \end{aligned} \tag{5.7}$$

which extends to an explicit diffeomorphism $HP^1 \xrightarrow{\cong} S^4$, so that π is given by the formula

$$\begin{aligned} &\pi([z_1, z_2, z_3, z_4]) \\ &= \left(\frac{2\{\overline{z_1}z_3+z_2\overline{z_4}+(\overline{z_1}z_4-z_2\overline{z_3})j\}}{|z_1|^2+|z_2|^2+|z_3|^2+|z_4|^2}, \frac{-|z_1|^2-|z_2|^2+|z_3|^2+|z_4|^2}{+|z_1|^2+|z_2|^2+|z_3|^2+|z_4|^2}\right). \end{aligned} \tag{5.8}$$

One can show that the fibration π is actually isomorphic to the associated fibre bundle $B_{SO(4)}(S^4) \times_{SO(4)} SO(4)/U(2)$. The fact that S^4 does not admit an almost complex structure can therefore be expressed as the statement that π does not admit a section, and CP^3 can indeed be viewed as a "natural" and "minimal" "complexification" of S^4, in particular as the conformal structure is also preserved:

To see this, we use a different approach which is based on Penrose's concept of twistors, but ultimately leads to the same formula (5.8): Given a 4-dimensional complex vector space T—called *twistor space*—with a basis e_1, e_2, e_3, e_4 and the volume $\omega = e_1 \wedge e_2 \wedge \wedge e_3 \wedge e_4 \in \Lambda^4 T$, consider the 6-dimensional complex vector space $\Lambda^2 T$ with the basis $e_{12}, e_{13}, e_{14}, e_{23}, e_{24}, e_{34}$ $(e_{jk} = e_j \wedge e_k)$ and the nondegenerate symmetric complex bilinear form $\cdot$ given by

$$(u \cdot v)\omega = u \wedge v \quad \text{for} \quad u, v \in \Lambda^2 T. \tag{5.9}$$

A 6-dimensional real subspace R_σ (resp. $R_\perp$) of $\Lambda^2 T$ on which $\cdot$ induces a nondegenerate symmetric real bilinear form of signature $-----+$ (resp. $-+--+-$) is then defined as the eigenspace for eigenvalue 1 of the antilinear involution σ (resp. $\perp$) on $\Lambda^2 T$ that arises from left multiplication by j on $T \cong C^4 \cong H^2$ via the wedge product (resp. from a nondegenerate Hermitian sesquilinear form of signature $+-+-$ on T via the wedge product and a complex analogue of the definition of the star operator). In this subspace we have the projective space $PL_{1,5}$ (resp. $PL_{2,4}$) of null lines, i.e. of lines on the light cone $L_{1,5} = \{u \in R_\sigma / \ u \cdot u = 0\}$ (resp. $L_{2,4} = \{u \in R_\perp / \ u \cdot u = 0\}$), which is known to be the conformal compactification of Euclidean (resp. Minkowski) 4-space R^4, and there are isomorphisms

$$\begin{array}{rcl} S^4 & \to & PL_{1,5} \\ a & \mapsto & [a, 1] \end{array} \quad \left(\text{resp.} \begin{array}{rcl} S^1 \times S^3 & \to & PL_{2,4} \\ (e^{i\phi}, a) & \mapsto & [a_1, \cos\phi, a_2, a_3, \sin\phi, a_4] \end{array}\right). \tag{5.10}$$

Points in this conformal compactification can be uniquely represented by 2-planes in $T \cong C^4$ which are spanned by z and jz for some nonzero vector $z \in C^4$ (resp. which are null with respect to the Hermitian form on C^4), or equivalently, which are spanned by vectors $x, y \in C^4$ such that $x \wedge y \in R_\sigma$ (resp. $R_\perp$); in fact, this defines a diffeomorphism—the so-called *Penrose twistor transform*—from a real submanifold of the Grassmannian $G_{2,4}$ of 2-planes in C^4 onto the conformal compactification $S^4 \cong PL_{1,5}$ (resp. $S^1 \times S^3 \cong PL_{2,4}$) of R^4. (For more details on the Minkowski version of all these statements, see e.g. [26].) Now projecting any point of CP^3 onto the unique real line in CP^3 (with respect to σ) passing through it, and applying the Penrose twistor transform to the resulting 2-plane in C^4, one recaptures (5.8).

It is also useful to observe that the Penrose twistor transform really extends into the complex domain: In fact, $PL_{1,5}$ and $PL_{2,4}$ are both real submanifolds of the complex manifold PL_6^C of null lines in $\Lambda^2 T$ (with respect to $\cdot$), which in turn is a 4-dimensional complex quadric in the 5-dimensional complex manifold CP^5 of all lines in $\Lambda^2 T$, and the Penrose twistor transform is actually a biholomorphic diffeomorphism from $G_{2,4}$ onto PL_6^C.

6. Instantons and vector bundles over CP^3

The fibration π: $CP^3 \to S^4$ described in the previous section is the crucial tool in the proof of the following

THEOREM 2 [5]. *There is a one-to-one correspondence between*

(i) SU(2)-*antiinstanton solutions* (*moduli*) *over* S^4 *with instanton number* $k < 0$, *i.e. gauge equivalence classes of anti-self-dual connections* A *in a principal* SU(2)-*bundle over* S^4, *with second Chern class* $-k > 0$,

and

(ii) *isomorphism classes of holomorphic 2-dimensional complex vector bundles* E *over* CP^3 *with first Chern class* $c_1(E) = 0$, *second Chern class* $c_2(E) = -k > 0$, *and subject to the following two conditions*:

(1) *Triviality Condition*: *The restriction of* E *to any real line in* CP^3 (*with respect to* σ), *i.e. to the fibres of* π, *is* (*holomorphically*) *trivial.*

(2) *Unitarity Condition*: E *has a* (*holomorphic*) "*symplectic structure*" $\hat{\sigma}$: $E \to \sigma^*\overline{E}$.

Remark 1. Given a holomorphic 2-dimensional complex vector bundle E over CP^3 with first Chern class $c_1(E) = 0$, its restriction to any line l in CP^3 decomposes into the direct sum of two holomorphic complex line bundles $L_1(l)$ and $L_2(l)$ over l; moreover, the $L_i(l)$ are essentially unique and determined by an integer $k_i(l)$, corresponding to the first Chern class $c_1(L_i(l))$ [27]. (Observe that $l \cong CP^1 \cong S^2$, whether l is real or not.) While $k_1(l)+k_2(l)$, corresponding to the first Chern class $c_1(E|l)$, is a topological invariant and hence invariant under deformations of l, actually $k_1(l)+k_2(l) = 0$ for all l since $c_1(E) = 0$, $k_1(l)-k_2(l)$ may jump under deformations of l, and we call

$$\mathscr{J} = \{l \text{ line in } CP^3/\ k_1(l)-k_2(l) \neq 0\} \tag{6.1}$$

the *set of jumping lines* of E. It is then obvious that

$$E|l \text{ is (holomorphically) trivial} \Leftrightarrow l \notin \mathscr{J}, \tag{6.2}$$

so that the triviality condition is equivalent to the nonexistence of real jumping lines. Now apply the Penrose twistor transform, which takes lines in CP^3 to points in the complex quadric $PL_6^{\mathbb{C}} \subset CP^5$ and the real lines in CP^3 to points in the real submanifold $S^4 \cong PL_{1,5}$ of $PL_6^{\mathbb{C}}$, as described in Sec. 5. Then intuitively speaking, one could say that it takes the bundle E—whether it satisfies the triviality condition or not—to some sort of "meromorphic connection" A with "curvature" F over $PL_6^{\mathbb{C}}$, and the jumping lines of E to the poles of A and F, so that the triviality condition keeps these singularities away from the real domain.

Remark 2. Given a 2-dimensional complex vector bundle E over B, a symplectic structure resp. real structure on E can be defined as an antilinear automorphism $\hat{\sigma}$: $E \to E$ such that $\hat{\sigma}^2 = -1$ resp. $\hat{\sigma}^2 = +1$. This terminology is based on the fact that $\hat{\sigma}$ can be viewed as a reduction of structure group of the principal GL(2, C)-bundle L of linear frames of E from GL(2, C) to its "symplectic" real form GL(1, H) resp. to its "real" real form GL(2, R). We shall only encounter the situation where E comes with a natural volume ω, i.e. a nowhere vanishing section ω of $\Lambda^2 E$, and where $\hat{\sigma}$ is compatible with ω in the sense that $(\hat{\sigma}\wedge\hat{\sigma})(\omega) = \omega$. Then ω defines a reduction of structure group P of L from GL(2, $\boldsymbol{C}$) to SL(2, $\boldsymbol{C}$), and $\hat{\sigma}$ can be viewed as a further reduction of structure group Q^s resp. Q^r of the principal SL(2, $\boldsymbol{C}$)-bundle P from SL(2, $\boldsymbol{C}$) to its "symplectic" real form Sp(1) resp. to its "real" real form SL(2, $\boldsymbol{R}$). In fact, using $\hat{\sigma}$ to transform frames yields an involutive automorphism $\check{\sigma}$: $P \to P$ which is $\varkappa$-covariant, i.e. satisfies $\check{\sigma}(p\cdot g) = \check{\sigma}(p)\cdot\varkappa(g)$ for $p \in P$, $g \in \mathrm{SL}(2, \boldsymbol{C})$, where $\varkappa$ is the homomorphism

$$\begin{array}{rcl} \varkappa\colon \mathrm{SL}(2,\boldsymbol{C}) & \to & \mathrm{SL}(2,\boldsymbol{C}) \\ g & \mapsto & g^{*-1} \end{array} \quad \text{resp.} \quad \begin{array}{rcl} \varkappa\colon \mathrm{SL}(2,\boldsymbol{C}) & \to & \mathrm{SL}(2,\boldsymbol{C}) \\ g & \mapsto & \bar{g} \end{array} \tag{6.3}$$

and this establishes a one-to-one correspondence between the $\hat{\sigma}$'s and the $\check{\sigma}$'s. Moreover, Q^s resp. Q^r is precisely the set of fixed points of $\check{\sigma}$, and conversely, Q^s resp. Q^r determines $\check{\sigma}$ completely. Finally, in the symplectic case, one knows that Sp(1) $\cong$ SU(2), and indeed symplectic structures $\hat{\sigma}$ are in one-to-one correspondence with Hermitian structures $\langle\cdot,\cdot\rangle$ via the formula

$$\hat{\sigma}u \wedge v = \langle u, v\rangle \quad \text{for} \quad u, v \in E_b,\ b \in B. \tag{6.4}$$

Remark 3. For holomorphic vector bundles over complex manifolds, symplectic or real structures as defined in Remark 2 are incompatible with analyticity: they are antiholomorphic vertically ($\hat{\sigma}$ is fibrewise antilinear, and $\check{\sigma}$ is $\varkappa$-covariant, with $\varkappa$ antiholomorphic), but holomorphic horizontally ($\hat{\sigma}$ and $\check{\sigma}$ induce the identity on the base). Therefore, given a holomorphic 2-dimensional complex vector bundle E over CP^3, with a holomorphic volume ω turning P into a holomorphic principal SL(2, $\boldsymbol{C}$)-bundle over CP^3, we use pull-back via the antiholomorphic involution σ on CP^3 and define a holomorphic "symplectic structure" resp. "real structure" on E to be a holomorphic isomorphism $\hat{\sigma}$: $E \to \sigma^*\bar{E}$ such that $\hat{\sigma}^2 = -1$ resp. $\hat{\sigma}^2 = +1$, or using $\hat{\sigma}$ to transform frames, to be an involutive holomorphic isomorphism $\check{\sigma}$: $P \to \sigma^*\bar{P}^\varkappa$. (Observe that if E resp. P are holomorphic bundles over CP^3, the conjugate manifolds $\bar{E}$ resp. $\bar{P}^\varkappa$ are holomorphic bundles over the conjugate manifold $\overline{CP}^3$; the superscript $\varkappa$ indicates that in order to obtain a holomorphic action of SL(2, $\boldsymbol{C}$), one defines right translation by g on $\bar{P}^\varkappa$ to be right translation by $\varkappa(g)$ on P. Now σ being antiholomorphic, $\sigma^*\bar{E}$ resp. $\sigma^*\bar{P}^\varkappa$ are holomorphic bundles over CP^3.) Concretely, for $z \in CP^3$, $\hat{\sigma}_z$ resp. $\check{\sigma}_z$ is an isomorphism $\hat{\sigma}_z$: $E_z \to \bar{E}_{\sigma(z)}$ resp. $\check{\sigma}_z$: $P_z \to \bar{P}^\varkappa_{\sigma(z)}$, i.e. an antilinear isomorphism $\hat{\sigma}_z$: $E_z \to E_{\sigma(z)}$ resp. a $\varkappa$-covariant isomorphism $\check{\sigma}_z$: $P_z \to P_{\sigma(z)}$, depending holomorphically on z. Of course, this definition gives up the direct interpretation of Remark 2 in terms of reductions of structure group—hence the quotation marks—, but as will become clear below, it comes in again through the back door via the triviality condition of Theorem 2.

Proof of Theorem 2: We shall now explain the ideas involved in the proof of Theorem 2. For convenience, symbols carrying a $\sim$ will refer to bundles over S^4, while other symbols will refer to bundles over CP^3.

First, given a principal SU(2)-bundle $\tilde{Q}$ over S^4, we can pull it back to a principal SU(2)-bundle Q over CP^3 via π. Complexifying the group, i.e. extending the structure group of both bundles from SU(2) to SL(2, C), we obtain principal SL(2, C)-bundles $\tilde{P}$ over S^4 and P over CP^3 together with embeddings $\tilde{Q} \hookrightarrow \tilde{P}$, $Q \hookrightarrow P$ and a commutative diagram

$$\begin{tikzcd}
P \arrow[r, "\tilde{\pi}"] \arrow[dd, bend right=60, "\varrho"'] & \tilde{P} \arrow[dd, bend left=60, "\tilde{\varrho}"] \\
Q \arrow[u, hook] \arrow[r, "\tilde{\pi}"] \arrow[d, "\varrho"] & \tilde{Q} \arrow[u, hook] \arrow[d, "\tilde{\varrho}"] \\
CP^3 \arrow[r, "\pi"] & S^4
\end{tikzcd} \tag{6.5}$$

We also have the associated 2-dimensional complex vector bundles

$$\tilde{E} = \tilde{P} \times_{\mathrm{SL}(2,C)} C^2 = \tilde{Q} \times_{\mathrm{SU}(2)} C^2 \quad \text{over} \quad S^4 \tag{6.6}$$

and

$$E = P \times_{\mathrm{SL}(2,C)} C^2 = Q \times_{\mathrm{SU}(2)} C^2 \quad \text{over} \quad CP^3 \tag{6.7}$$

with the commutative diagram

$$\begin{array}{ccc} E & \xrightarrow{\hat{\pi}} & \tilde{E} \\ \downarrow & & \downarrow \\ CP^3 & \xrightarrow{\pi} & S^4 \end{array} . \tag{6.8}$$

By construction, the bundles over CP^3 are (smoothly) trivial over the fibres of π. $\tilde{E}$ and E carry natural volumes $\tilde{\omega}$ and ω, i.e. nowhere vanishing smooth sections of $\Lambda^2\tilde{E}$ and $\Lambda^2 E$, defining the reduction $\tilde{P}$ and P of structure group of the bundle of linear frames of $\tilde{E}$ and E from GL(2, C) to SL(2, C), respectively, and $\omega = \pi^*\tilde{\omega}$. Finally, writing

$$\begin{aligned} Q = \pi^*\tilde{Q} &= \{(z, \tilde{q}) \in CP^3 \times \tilde{Q} / \ \pi(z) = \tilde{\varrho}(\tilde{q})\}, \\ P = \pi^*\tilde{P} &= \{(z, \tilde{p}) \in CP^3 \times \tilde{P} / \ \pi(z) = \tilde{\varrho}(\tilde{p})\}, \end{aligned} \tag{6.9}$$

the reduction $\tilde{Q}$ of structure group of $\tilde{P}$ from SL(2, C) to SU(2) defines a "symplectic structure"

$$\begin{array}{rccc} \check{\sigma}: & P & \to & P^{\varkappa} \\ & (z, \tilde{q}\cdot g) & \mapsto & (\sigma(z), \tilde{q}\cdot\varkappa(g)) \end{array} \tag{6.10}$$

on P in the sense of Remark 3. (Observe $\pi \circ \sigma = \pi$, $P = \pi^*\tilde{P} \Rightarrow \sigma^*P^{\varkappa} \cong P^{\varkappa}$; the bar is superfluous since we have not yet introduced any complex structure.)

Conversely, start from a holomorphic 2-dimensional complex vector bundle E over CP^3 which is (holomorphically) trivial over the fibres of π. For $x \in S^4$, consider the space

$\tilde{E}_x$ of holomorphic sections of $E|\pi^{-1}(x)$; it is a 2-dimensional complex vector space since $\pi^{-1}(x) \cong S^2$ and by Liouville's theorem, such a section is uniquely determined by its value at a single point $z \in \pi^{-1}(x)$. This defines a smooth 2-dimensional complex vector bundle $\tilde{E}$ over S^4 such that $E \cong \pi^*\tilde{E}$; the canonical isomorphism here is given by evaluation of a section at a point. Now if $c_1(E) = 0$, the holomorphic complex line bundle $\Lambda^2 E$ over CP^3 is (holomorphically) trivial and hence admits a nowhere vanishing holomorphic section ω, uniquely determined up to a nonzero scalar factor. In particular, for $x \in S^4$, $\omega|_{\pi^{-1}(x)}$ is a nowhere vanishing holomorphic section of $\Lambda^2 E|\pi^{-1}(x)$ and can be identified with a nonzero element $\tilde{\omega}_x$ of $\Lambda^2\tilde{E}_x$. This defines a nowhere vanishing smooth section $\tilde{\omega}$ of the complex line bundle $\Lambda^2\tilde{E}$ over S^4, uniquely determined up to a nonzero scalar factor. Finally, for $x \in S^4$ the "symplectic structure" $\hat{\sigma}$ on E induces a holomorphic isomorphism $\hat{\sigma}_x$: $E|\pi^{-1}(x) \to \sigma^*\bar{E}|\pi^{-1}(x)$, and hence an antilinear isomorphism $\tilde{\hat{\sigma}}_x$: $\tilde{E}_x \to \tilde{E}_x$ such that $\tilde{\hat{\sigma}}_x^2 = -1$ by

$$(\tilde{\hat{\sigma}}_x u)(z) = \hat{\sigma}_x\big(u(\sigma z)\big) \quad \text{for} \quad u \in \tilde{E}_x,\ z \in \pi^{-1}(x). \tag{6.11}$$

This defines a smooth antilinear automorphism $\tilde{\hat{\sigma}}$: $\tilde{E} \to \tilde{E}$ such that $\tilde{\hat{\sigma}}^2 = -1$. Now using the volume $\tilde{\omega}$ and the symplectic structure $\tilde{\hat{\sigma}}$ on $\tilde{E}$ to reduce structure groups as described in Remark 2, we are—up to canonical isomorphisms—back to the previous situation.

In the following, we write $\tilde{A}$ resp. A for a connection 1-form in $\tilde{Q}$ resp. Q as well as in $\tilde{P}$ resp. P, with $\tilde{F}$ resp. F the corresponding curvature 2-form, and $\tilde{A}/\tilde{F}$ on $\tilde{P}$ resp. A/F on P is given by extending $\tilde{A}/\tilde{F}$ on $\tilde{Q}$ resp. A/F on Q from $\tilde{Q}$ to $\tilde{P}$ resp. from Q to P.

With these rather technical preliminaries out of the way, the proof of Theorem 2 is based on the following sequence of propositions:

PROPOSITION 2. *There is a one-to-one correspondence between connections A in Q and almost complex structures J on the total space P which are compatible with the bundle structure, given by the requirement that the horizontal subspaces of A in P are complex subspaces of the tangent spaces to P, i.e. stable under J, or equivalently, that the connection 1-form A on P is of type* $(1, 0)$.

For the idea of the proof, cf. e.g. [16], Vol. II, pp. 178–180.

Remark 4. Compatibility of an almost complex structure J: $TP \to TP$ on P with the bundle structure means that

(i) J is $\mathrm{SL}(2, C)$-covariant;

(ii) J turns the projection ϱ: $P \to CP^3$ into an almost complex map;

(iii) J maps the vertical bundle $\mathrm{Ver}\,P$ of P onto itself and makes the diagram

$$\begin{array}{ccc} \mathrm{Ver}\,P & \xrightarrow{J|_{\mathrm{Ver}\,P}} & \mathrm{Ver}\,P \\ \uparrow & & \uparrow \\ P \times \mathrm{sl}(2, C) & \xrightarrow[\mathrm{id}_P \times J_0]{} & P \times \mathrm{sl}(2, C) \end{array}$$

commutative, where the vertical maps are the usual trivialisation

$$\begin{array}{ccc} P\times \mathrm{sl}(2, C) & \to & \mathrm{Ver}\, P \\ (p, X) & \mapsto & \dfrac{d}{dt}(p\cdot \exp tX)|_{t=0} \end{array}$$

of Ver P [16], [17], and J_0 is multiplication by i in the complex Lie algebra sl$(2, C)$.

The theorem of Newlander–Nirenberg [16] states that the almost complex structure J on P is integrable—and hence turns P into a holomorphic principal SL$(2, C)$-bundle over CP^3—iff its torsion $N = N(J)$ vanishes, i.e. iff

$$N(X, Y) = 2\{[JX, JY]-[X, Y]-J[X, JY]-J[JX, Y]\} = 0 \tag{6.12}$$

for all vector fields X, Y on P or—which is sufficient—for all invariant vector fields X, Y on P. Due to Remark 4, (iii), (6.12) is automatically satisfied on vertical vector fields, and due to Remark 4, (ii) and the fact that the curvature F of A determines the vertical component of the Lie bracket between two invariant horizontal vector fields, we obtain

PROPOSITION 3. *With the notation of Proposition* 2, J *has no torsion iff* F *on* P *is of type* $(1, 1)$, *i.e.* (6.12) *holds for all vector fields* X, Y *on* P *iff*

$$F(JX, JY) = F(X, Y) \tag{6.13}$$

for all vector fields X, Y *on* P.

For a more detailed proof of a similar statement, cf. [16], pp. 180–181.

Remark 5. Propositions 2 and 3 of course apply to a more general situation: In fact, one can replace CP^3 by any complex manifold and the structure groups SU(2) of Q and SL$(2, C)$ of P by any real Lie group K and complex Lie group G, respectively, such that $K \subset G$ and the Lie algebra of K is a real form of the Lie algebra of G.

The crucial observation, based on the specific form of the fibration π: $CP^3 \to S^4$ and originally due to Ward, is now

PROPOSITION 4. *With the notation as before, assume that* A *on* Q (*and* P) *is the pull-back of a connection* 1-*form* $\tilde{A}$ *on* $\tilde{Q}$ (*and* $\tilde{P}$) *via* $\tilde{\pi}$, *so that* F *on* Q (*and* P) *is the pull-back of the curvature* 2-*form* $\tilde{F}$ *on* $\tilde{Q}$ (*and* $\tilde{P}$) *via* $\tilde{\pi}$. *Then* $\tilde{F}$ *satisfies the anti-self-dual equation* $*\tilde{F} = -\tilde{F}$ *on* $\tilde{Q}$, *or equivalently, on* $\tilde{P}$, *iff* F *on* P *is of type* $(1, 1)$, *i.e. iff* (6.13) *holds for all vector fields* X, Y *on* P.

Quite independently of this result, the assumption occurring at the beginning of Proposition 4 can also be translated. In fact, any point $\tilde{q}_0 \in \tilde{Q} \subset \tilde{P}$ defines a section

$$\begin{array}{cccc} s_0\colon \pi^{-1}(x) & \to & Q|\pi^{-1}(x) \subset P|\pi^{-1}(x) \\ z & \mapsto & (z, q_0) \end{array} \tag{6.14}$$

of $Q|\pi^{-1}(x) = \pi^{-1}(x)\times \tilde{Q}_x \subset P|\pi^{-1}(x) = \pi^{-1}(x)\times \tilde{P}_x$, where $x = \tilde{\varrho}(\tilde{q}_0) \in S^4$, and we have

PROPOSITION 5. *With the notation as before, A on Q, or equivalently, on P, is the pull-back of a connection 1-form $\tilde{A}$ on $\tilde{Q}$, or equivalently, on $\tilde{P}$, via $\hat{\pi}$ iff the almost complex structure J on P is trivial over the fibres of π. More explicitly,*

$$s_0 \text{ is almost complex} \Leftrightarrow s_0^* J = (J \text{ on } \pi^{-1}(x)) \Leftrightarrow s_0^* A = 0$$
$$\text{for any section } s_0 \text{ of } P|\pi^{-1}(x) \text{ as above}$$

is equivalent to

$$A = \hat{\pi}^* \tilde{A} \quad \text{for some connection 1-form } \tilde{A}.$$

Hence starting from an anti-self-dual connection $\tilde{A}$ on the principal SU(2)-bundle $\tilde{Q}$ over S^4, one ends up with a complex structure on the principal SL(2, C)-bundle P over CP^3 which is trivial over the fibres of π, so that E becomes a holomorphic 2-dimensional complex vector bundle over CP^3, (holomorphically) trivial over the fibres of π and equipped with a natural holomorphic volume ω (thus in particular, $c_1(E) = 0$). Moreover, the "symplectic structure" $\hat{\sigma}$: $E \to \sigma^* \bar{E}$ on E is holomorphic since by construction, $\check{\sigma}$: $P \to \sigma^* \bar{P}^{\varkappa}$ preserves the horizontal spaces of A on P and hence is holomorphic. The converse direction of the proof is now also clear, and the second Chern classes involved yield the same integer since Chern classes are natural under pull-back. This completes the proof of Theorem 2. ∎

Remark 6. As has become clear from the proof, the central idea behind Theorem 2 is to use fibration π: $CP^3 \to S^4$ for coding connections over S^4 into almost complex structures on the total spaces of 2-dimensional complex vector bundles over CP^3, compatible with the bundle structure. The reason why this idea has something to do with instantons is Ward's observation that the anti-self-dual equation (3.6) for a connection, which (in local coordinates) is a system of *nonlinear* first order partial differential equations for its components, is precisely the *integrability condition* (6.12) for an almost complex structure, which (in local coordinates) is a system of *linear* first order partial differential equations for its components; the nonlinearity has been "eaten up" by the Penrose twistor transform. In other words, this transformation is precisely "the right" *Gelfand integral transform*[1] for the SU(2) instanton problem over S^4, transforming it into a problem in algebraic geometry which we shall discuss briefly in the next section. In fact, this is true more generally for the SU(n) instanton problem over S^4 with $n \geqslant 2$ (one only has to change the "symplectic structure" to a "Hermitian structure" providing for the reduction of structure group from SL(n, C) to SU(n)), but the amount of knowledge about the corresponding problem in algebraic geometry depends drastically on n: there is quite a lot of information if $n = 2$, but almost nothing is known if $n \geqslant 3$.

[1] I am indebted to V. Guillemin, D. Kazhdan and S. Sternberg for introducing me to the concept of Gelfand integral transforms as the general picture behind practically all the presently known techniques of explicitly solving nonlinear partial differential equations, in particular behind the inverse scattering method (*à la* Gelfand–Levitan–Marchenko/Lax). If chosen suitably—which is part of the problem and unfortunately has to be done case by case—it transforms the original problem into a simpler one which can then eventually be solved.

7. Instantons and algebraic curves in CP^3

Let us mention first that Serre's basic theorems on the relation between analytic and algebraic geometry [28] imply that in the statement of Theorem 2, (ii), the term "holomorphic" may be replaced by the term "algebraic"; thus Theorem 2 indeed provides a translation of the SU(2) instanton problem over S^4 into a classification problem in algebraic geometry. Lack of space and the incompetence of the author prevent us from going into a detailed discussion of the progress that algebraic geometers have recently made in the area [29], [30], [31], [32], so let us just mention that

(1) The manifold of solutions for $k = 1$ [2] is the 5-dimensional unit ball, i.e. the interior of the 4-sphere in R^5. In particular, it is connected, simply connected, and a single 5-dimensional orbit under the action of the conformal group SO(5, 1).

(2) The manifold of solutions for $k = 2$ [2] is a 13-dimensional manifold which is connected, but doubly rather than simply connected (i.e. the fundamental group is Z_2), and a one-parameter family of 12-dimensional orbits under the action of the conformal group SO(5, 1) [32].

However, we do want to sketch the role of algebraic curves in this context and their interpretation in terms of "instanton physics".

Algebraic curves enter the picture as the sets of zeros of suitable algebraic sections. More precisely, let E be an algebraic 2-dimensional complex vector bundle over CP^3 satisfying the conditions of Theorem 2, (ii), let H be the canonical line bundle over CP^3 determined by any hyperplane section, and let $H^{-1} = H^*$ be its dual, so that H^{-1} is the universal line bundle over CP^3, i.e.

$$\text{point in } H^{-1} = (\text{line in } C^4, \text{ vector on that line}),$$

and $c_1(H) = 1$. For $l \in Z$, set

$$E(l) = E \otimes H^l = \begin{cases} E \otimes H \otimes \dots \otimes H & (|l| \text{ times}) \quad \text{for} \quad l > 0, \\ E & \qquad\qquad\qquad\quad \text{for} \quad l = 0, \\ E \otimes H^{-1} \otimes \dots \otimes H^{-1} & (|l| \text{ times}) \quad \text{for} \quad l < 0. \end{cases} \tag{7.1}$$

Then for $l < 0$, $E(l)$ has no nonzero algebraic[3] sections. (Indeed, given an algebraic[3] section s of $E(l)$, then restricting to any real line $L \cong CP^1 \cong S^2$ in CP^3, $E(l)|L \cong H^l|L \oplus \oplus H^l|L$ since $E|L$ is trivial, so that $s|_L$ corresponds to two algebraic[3] sections of $H^l|L$ which must vanish identically if $l < 0$.) If $k > 0$, one can even show that E itself has no nonzero algebraic[3] sections either, implying that E is a stable bundle, which is equivalent to the statement that End(E) does not admit any algebraic[3] sections, i.e. that E does not admit any algebraic[3] vector bundle maps $E \to E$, except constant multiples of the identity [31]. On the contrary, for a suitable integer $l > 0$, $E(l)$ does have nonzero algebraic[3] sections s vanishing along smooth algebraic curves Γ in CP^3. (The idea is that E always

[2] In this section, the integer k will stand for minus the instanton number, so that $c_2(E) = k > 0$.

[3] Recall that in all these cases, algebraic = holomorphic and rational = meromorphic [28].

admits rational[3] sections, and that tensoring with a sufficiently high power of H, one can remove the singularities; recall that the algebraic[3] sections of H over CP^3 correspond to the homogeneous polynomials of degree l on C^4.)

The curves Γ arising in this way are not arbitrary. Indeed, identify CP^3 with the zero section in $E(l)$, then over $CP^3 \subset E(l)$, the tangent bundle $TE(l)$ of the total space $E(l)$ decomposes naturally into the vertical and a horizontal part:

$$TE(l)|CP^3 = \mathrm{Ver}\,E(l)|CP^3 \oplus \mathrm{Hor}\,E(l)|CP^3 \cong E(l) \oplus T(CP^3). \tag{7.2}$$

But Γ is precisely the set of $z \in CP^3$ such that $s(z) \in CP^3$, so that we obtain a commutative diagram

$$\begin{array}{ccc} T(CP^3)|\Gamma & \xrightarrow{Ts|\Gamma} & TE(l)|\Gamma \\ \downarrow & & \downarrow {\scriptstyle \text{vertical projection in (7.2)}}, \\ N(\Gamma) = T(CP^3)|\Gamma/T\Gamma & \xrightarrow{\cong} & E(l)|\Gamma \end{array} \tag{7.3}$$

where $N(\Gamma)$ is the normal bundle of Γ, showing that the derivative Ts of the section s, restricted to Γ, induces an isomorphism

$$N(\Gamma) \cong E(l)|\Gamma. \tag{7.4}$$

In particular, taking wedge products and observing that $\Lambda^2 E$ is trivial since $c_1(E) = 0$, we obtain isomorphisms

$$\Lambda^2 N(\Gamma) = \Lambda^2 E(l)|\Gamma = (\Lambda^2 E \otimes H^{2l})|\Gamma = H^{2l}|\Gamma. \tag{7.5}$$

Notice also that writing Γ as the disjoint union of its connected components

$$\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_r, \tag{7.6}$$

any other isomorphism between $\Lambda^2 N(\Gamma)$ and $H^{2l}|\Gamma$ is then uniquely determined by r nonzero scalars $\lambda_1, \ldots, \lambda_r \in C$. Hence a necessary condition on the algebraic curve Γ to arise as the set of zeros of an algebraic section s as above is that the second exterior power $\Lambda^2 N(\Gamma)$ of its normal bundle $N(\Gamma)$ is the restriction to Γ of an even power of the canonical line bundle H over CP^3.

The restrictions on the possible curves Γ can also be expressed numerically in terms of the formulae

$$(l-2)\,d_i = g_i - 1, \qquad \sum_{i=1}^{r} d_i = k + l^2 \tag{7.7}$$

for the degree d_i and the genus g_i of the component curve Γ_i. (In particular, the first of these is a direct consequence of the isomorphisms (7.5) and

$$\Lambda^2 N(\Gamma) \cong \Lambda^3 \left(T(CP^3)|\Gamma\right)/\Lambda^1 T\Gamma \cong H^4|\Gamma \otimes T^*\Gamma = H^4|\Gamma \otimes K_\Gamma, \tag{7.8}$$

where K_Γ is the canonical line bundle (divisor class) over Γ, by taking the first Chern

[3] Recall that in all these cases, algebraic = holomorphic and rational = meromorphic [28].

class.) Now let us fix a certain value for the instanton number, i.e. for the integer k. Then we obtain some value for the integer l, which is conveniently chosen to be as small as possible, i.e. l is to be the smallest positive integer such that for any algebraic 2-dimensional complex vector bundle E over CP^3 satisfying the conditions of Theorem 2, (ii), $E(l)$ admits nonzero algebraic sections s vanishing along smooth algebraic curves Γ. An estimate for l in terms of k is given by the following

CONJECTURE ([32], proved for $k \leqslant 9$). It suffices to take l to be the smallest positive integer such that $l < \sqrt{3k+1}-2$ (see also (7.11) below).

Given k and l, equation (7.7) puts upper bounds on the degrees and genera of the curves that can appear. The simplest case is $l = 1$, where due to $g_i \geqslant 0$, $d_i \geqslant 1$

$$g_i-1 = -d_i \Rightarrow g_i = 0,\ d_i = 1 \text{ and } r = k+1. \tag{7.9}$$

(Observe that conversely, $g_i = 0$ for any one i implies $d_i = 1/2-l$, hence $l = 1$, and (7.9) holds.) Therefore, Γ consists of $k+1$ lines in CP^3. If $l \geqslant 2$, (7.7) imposes no restriction on the d_i which would prevent the g_i and d_i from assuming their maximal value g and d, respectively, when Γ is connected, i.e. when $r = 1$, and

$$g = (l-2)(k+l^2)+1, \quad d = k+l^2, \tag{7.10}$$

so that we obtain the following table:

k	1	2	3	4	5	6	7	8	9
l	1	1	2	2	3	3	3	4	4
g	0	0	1	1	15	16	17	49	51
d	1	1	7	8	14	15	16	24	25

(7.11)

Of course, the procedure of associating curves with bundles described above is highly redundant since the same bundle E gives rise to many nonzero algebraic sections s in $E(l)$ (their number increases with l), and hence to many curves Γ. This redundancy is perhaps most transparent in the statement that E is trivial iff Γ is the complete intersection of two algebraic hypersurfaces in CP^3. Conversely, however, one can associate bundles with curves in an essentially unique way: In fact, given a smooth algebraic curve Γ in CP^3, cover Γ by open sets in CP^3 such that any one of them meets at most one connected component of Γ. Moreover, assume that on any one of them, say U, the ideal of regular functions vanishing along $U\cap\Gamma$ is generated by two relatively prime homogeneous polynomials f_1, f_2 (restricted to U); then if on any other one, say V, the ideal of regular functions vanishing along $V\cap\Gamma$ is generated by two relatively prime homogeneous polynomials g_1, g_2 (restricted to V), one obtains a regular GL(2, C)-valued transition function h_{UV} on $U\cap V$:

$$\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} (h_{UV})_{11} & (h_{UV})_{12} \\ (h_{UV})_{21} & (h_{UV})_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \quad \text{on} \quad U\cap V. \tag{7.12}$$

These transition functions define an algebraic 2-dimensional complex vector bundle E' over an open neighbourhood of Γ in CP^3, with $\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ on U, $\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ on V, etc., fitting together to yield an algebraic section s of E' such that $s^{-1}(0) = \Gamma$. Now if Γ satisfies conditions on its normal bundle of the type mentioned above, E' and s can be extended to all of CP^3, and the nonzero scalars $\lambda_1, \ldots, \lambda_r \in C$ mentioned above appear again (up to a common factor), guaranteeing uniqueness (up to an isomorphism). Finally, tensoring with an appropriate power of H, one kills the first Chern class of E' and recovers E.

Returning to the situation where Γ is given in terms of E, l and s as before, observe that restricting to $CP^3 - \Gamma$, s generates a trivial line subbundle θ of $E(l)$, and there is an exact sequence

$$0 \to \theta \xrightarrow{s} E(l) \to H^{2l} \to 0 \tag{7.13}$$

of algebraic complex vector bundles over $CP^3 - \Gamma$, in which the second map is given by taking the wedge product with s (recall $\Lambda^2 E(l) \cong H^{2l}$ since $\Lambda^2 E$ is trivial). Hence over $CP^3 - \Gamma$, $E(l)$ is an extension of θ by H^{2l}; such extensions are classified by elements of the first sheaf cohomology group $H^1(CP^3 - \Gamma, H^{-2l})$. From the Penrose theory of twistors one knows that these elements in turn correspond to certain solutions of the Maxwell equations for spin $(l-1)$ fields if $l \geqslant 2$, and to scalar densities satisfying the Laplace equation if $l = 1$. These solutions ϕ have singularities on the surface $\tilde{\Gamma}$ in S^4 which corresponds to Γ. The exact type of the singularities is not yet known, but very probably ϕ has to be a distributional solution of the corresponding inhomogeneous equation with the δ-function of $\tilde{\Gamma}$ as a source term.

The algebraic geometry thus leads to the series of Ansätze A_l of Atiyah and Ward [5]. Unfortunately, when $l \geqslant 3$, the curves Γ will be of high genus (cf. (7.11)), so that the surfaces of singularity $\tilde{\Gamma}$ will become rather complicated; moreover, in general there seems to be no natural way to incorporate the reality conditions imposed in Theorem 2, (ii). For $l = 1$, the Ansatz A_1 coincides with the one of t'Hooft: In this case, according to what has been said above, Γ is a set of $k+1$ lines in CP^3, and we also have to fix $k+1$ nonzero scalars $\lambda_1, \ldots, \lambda_{k+1} \in C$ to determine the solution; the reality conditions then amount to the lines being real with respect to σ —so that $\tilde{\Gamma}$ is just a set of $k+1$ points in S^4 —and to $\lambda_1, \ldots, \lambda_{k+1}$ being real and positive. These are precisely the "position" and "scale" parameters in t'Hooft's Ansatz [20].

8. Further comments

Although the results of Atiyah and Ward have opened up a new approach to the instanton problem through methods of algebraic geometry, many questions remain open. For example, algebraic geometry can be expected to shed new light on the striking analogy between self-dual or anti-self-dual gauge fields in four dimensions (compactified to S^4) and the nonlinear σ-model in two dimensions (compactified to S^2). Both seem to be excellent candidates for completely integrable systems (in the Euclidean sense), for which as yet no examples in four dimensions are known. A first step in this direction is the recent discovery of Bäcklund transformations [33], [34] for the Yang–Mills fields, based on the Atiyah–Ward Ansätze. New conserved currents—another characteristic feature of complete integrability—have not yet been found. These questions would of course become even more important should one be able to prove that any solution to the free Yang–Mills equations is actually self-dual or anti-self-dual.

Acknowledgement

This review grew out of a talk given at the meeting on "Methods of Differential Geometry in Mathematical Physics", Clausthal–Zellerfeld, June 30–July 2, 1977. It is based on notes taken from two lectures at the Mathematische Arbeitstagung 1977 in Bonn by M. F. Atiyah, to whom I wish to express my special gratitude. I would also like to acknowledge fruitful discussions with H. Hironaka, N. Hitchin, D. Kazhdan, E. Miller, C. Rebbi, B. Schroer and E. Witten.

Note. While preparing the final draft of the manuscript, I received a paper by R. Stora [35] which is written in a similar spirit.

REFERENCES

[1] A. M. Polyakov: *Phys. Lett.* **59B** (1975), 82.
[2] A. A. Belavin, A. M. Polyakov, A. S. Schwarz and Y. S. Tyupkin: ibid. **59B** (1975), 85.
[3] M. F. Atiyah, N. J. Hitchin and I. M. Singer: *Proc. Nat. Acad. Sci. USA* **74** (1977), 2662.
[4] R. S. Ward: *Phys. Lett.* **61A** (1977), 81.
[5] M. F. Atiyah and R. S. Ward: *Commun. Math. Phys.* **55** (1977), 117.
[6] M. F. Atiyah: *Geometry of Yang–Mills fields*, Preprint, Oxford 1977.
[7] A. M. Polyakov: *Quark confinement and topology of gauge groups*, Lectures at NORDITA, Copenhagen, 1976.
[8] R. Jackiw and C. Rebbi: *Phys. Rev. Lett.* **37** (1976), 172.
[9] C. Callan, R. Dashen and D. Gross: *Phys. Lett.* **63B** (1976), 334.
[10] R. Jackiw: *Semi-classical analysis of quantum field-theory*, Lectures at the XVI Escuela Latino-Americana de Fisica, Caracas, 1976.
[11] —: *Rev. Mod. Phys.* **49** (1977), 681.
[12] A. Trautman: *Rep. Math. Phys.* **1** (1970), 29.
[13] D. A. Popov: *Theor. & Math. Phys.* **24** (1975), 347 (Engl. transl. (1976), 879).
[14] T. T. Wu and C. N. Yang: *Phys. Rev.* **D12** (1975), 3843, 3845.
[15] R. Stora: *Continuum gauge theories*, Lectures at the Summer Institute for Theoretical Physics, Cargèse, 1976.
[16] S. Kobayashi and K. Nomizu: *Foundations of differential geometry*, Vol. I (1963) and Vol. II (1969), Interscience.
[17] W. Greub, S. Halperin and R. Vanstone: *Connections, curvature and cohomology*, Vol. I (1972) and Vol. II (1973), Academic Press.
[18] M. Forger: Forthcoming Thesis.
[19] F. Warner: *Foundations of differentiable manifolds and Lie groups*, Scott, Foresman & Co., 1971.
[20] R. Jackiw, C. Nohl and C. Rebbi: *Phys. Rev.* **D15** (1977), 1642.
[21] R. Jackiw and C. Rebbi: *Phys. Lett.* **67B** (1977), 189.

[22] A. S. Schwarz: ibid. **67B** (1977), 172.
[23] M. Kuranishi: *Existence of locally complete families of complex structures*, in: Proc. Conf. Compl. Anal., Minneapolis 1964.
[24] H. Römer: private communication through B. Schroer, and CERN preprint.
[25] N. J. Hitchin: private communication, and work in preparation.
[26] W. Kopczyński and L. S. Woronowicz: *Rep. Math. Phys.* **2** (1971), 35.
[27] A. Grothendieck: *Amer. J. Math.* **79** (1957), 121.
[28] J. P. Serre: *Ann. Inst. Four.* **6** (1956), 1.
[29] M. Maruyama: *Nagoya Math. J.* **58** (1975), 25.
[30] H. Grauert and G. Mülich: *Manuscr. Math.* **16** (1975), 75.
[31] W. Barth: *Math. Ann.* **226** (1977), 125.
[32] R. Hartshorne: private communication through M. F. Atiyah.
[33] A. A. Belavin and V. E. Zakharov: *Yang–Mills equations as inverse scattering problem*, Preprint, Moscow, 1977.
[34] E. Corrigan, D. B. Fairlie, P. Goddard and R. G. Yates: *Bäcklund transformations and the construction of the Atiyah-Ward Ansätze for self-dual* SU(2) *gauge fields*, Preprint, Durham/Cambridge, 1977.
[35] R. Stora: *Yang–Mills instantons, geometrical aspects*, Lectures at the International School of Mathematical Physics, Erice, 1977.