M. F. ATIYAH

Geometry
of Yang-Mills Fields

PREFACE

These Lectures Notes are an expanded version of the Fermi Lectures which I gave at the Scuola Normale in Pisa in June 1978. They also cover material presented in the spring of 1978 in the Loeb Lectures at Harvard and the Whittemore Lectures at Yale. In all cases I was addressing a mixed audience of mathematicians and physicists and the presentation had to be tailored accordingly. In writing up the lectures I have tried as far as possible to keep this dual audience in mind, and the early chapters in particular attempt to bridge the gap between the two points of view. In the later chapters, where the material becomes more technical, there is a danger of falling between two stools. On the one hand the mathematical jargon may be unintelligible to the physicist, while the presentation may, by mathematical standards, be lacking in rigour. This is a risk I have deliberately taken. The initiated mathematician should be able to fill in most of the gaps by himself or by referring to other published papers. Physicists who have survived the early chapters may derive some benefit by being exposed to new mathematical techniques, applied to problems they are familiar with. With this aim in mind I have throughout presented the mathematical material in a somewhat unorthodox order, following a pattern which I felt would relate the new techniques to familiar ground for physicists.

The main new results presented in the lectures, namely the construction of all multi-instanton solutions of Yang-Mills fields, is the culmination of several years of fruitful interaction between many physicists and mathematicians. The major breakthrough came with the observation [42] by R. S. Ward that the complex methods developed by R. Penrose in his « twistor programme » were ideally suited to the study of the Yang-Mills equations. The instanton problem was then seen [4] to be equivalent to a problem in complex analysis and finally to one in algebraic geometry. Using the powerful methods of modern algebraic geometry and the specific results of G. Horrocks and W. Barth it was not long before the problem was finally solved [2].

The first two chapters provide an introduction to the basic concepts, a statement of the problem and an explicit description of the solution. The

next two chapters are devoted to the Penrose theory and its application to the Yang-Mills equations. In Chapter V I present Horrocks' construction in algebraic geometry which is equivalent via the Penrose theory to the explicit instanton construction of Chapter II. Chapter VI introduces the important mathematical tool of sheaf cohomology and relates it to physically interesting equations. There are a number of digressions which may help to make the material less mysterious and more understandable. Chapter VII is an account of the theorem of Barth [8] which shows that the Horrocks construction of Chapter II yields all relevant bundles and hence that the construction of Chapter II yields all instantons. Finally in Chapter VIII we discuss some other aspects and open problems concerning the Yang-Mills equations.

Although the presentation is somewhat discursive, and includes much background material, it is also reasonably complete from the mathematical point of view. The one point where the proof is only sketched is the identification in Chapter VI of the sheaf cohomology group $H^1(P_3, E(-2))$ with the solution space of an appropriate Laplace operator. A detailed account of this can be found in various forms in [18] [29] [36]. An alternative presentation of the whole instanton theory is contained in the papers of Drinfeld and Manin [16] [17] [18] [19], and mathematicians, particularly if they are proficient in algebraic geometry, may prefer to read these.

My acquaintance with the geometry of Yang-Mills equations arose from lectures given in Oxford in Autumn 1976 by I. M. Singer, and I am very grateful to him for arousing my interest in this aspect of theoretical physics. We have collaborated since on many topics in this area. I have also, over the past few years, greatly benefited from numerous discussions with R. Penrose concerning twistor theory and complex analysis. In developing the mathematical theory of instantons I have throughout worked in close collaboration with N. J. Hitchin, and these lectures embody the results of our joint efforts. I am in addition greatly indebted to my now numerous friends in the physics community who have helped to give me some small understanding of the fascinating mathematical problems facing elementary particle physics.

Finally I should express my thanks to the Accademia Nazionale dei Lincei and to the Scuola Normale for their invitation to deliver the Fermi Lectures and for their hospitality in Pisa.

CONTENTS

CHAPTER I

The Yang-Mills Lagrangian

1. – Physics background.

The aim of quantum field theory is broadly speaking to put all elementary particles on the same footing as photons. Whereas photons appear as the quanta of classical electromagnetic theory other elementary particles should arise by the quantization of appropriate classical field theories. In recent years gauge theories have appeared the most promising candidates, and the Yang-Mills equation is the generalization of Maxwell's equations (in vacuo). The circle group which embodies the phase factor in Maxwell theory is generalized to a non-abelian compact Lie group G such as $SU(2)$ or $SU(3)$, the choice of group being dictated by the empirically observed symmetries of elementary particles. The non-abelian nature of G leads to non-linearity for the Yang-Mills equations. This non-linearity is of course the source of great mathematical difficulties and the quantization of non-abelian gauge theories is still in its infancy.

One recognized way of attempting to develop the quantum theory is to use the Feynman functional integral approach which involves integrating $\exp(iS)$ where S is the action. If we analytically continue to imaginary time, so that Minkowski space gets replaced by Euclidean 4-space, the Euclidean action is a positive multiple of i and so the integrand $\exp(iS)$ becomes a decaying exponential whose maximum value occurs at the minimum of the Euclidean action. It is reasonable therefore to ask for the determination of the classical field configurations in Euclidean space which minimize the action, subject to appropriate asymptotic conditions in 4-space. These classical solutions are the « instantons » of the Yang-Mills theory, and it will be the primary purpose of these lectures to show how to find all instantons. For further explanations of their physical significance, particularly in relation to tunnelling, I refer to [12] or [30]. From a very general point of view one can also say that a thorough understanding of the classical equations is likely to be a pre-requisite for developing the quan-

tum theory, and one may hope that important structural features will appear at the classical level.

If one were to search ab initio for a non-linear generalization of Maxwell's equation to explain elementary particles, there are various symmetry properties one would require. These are

(i)   *external symmetries* under the Lorentz and Poincaré groups and under the conformal group if one is taking the rest-mass to be zero,

(ii) *internal symmetries* under groups like $SU(2)$ or $SU(3)$ to account for the known features of elementary particles,

(iii) *covariance* or the ability to be coupled to gravitation by working on a curved space-time.

Gauge theories satisfy these basic requirements because they are geometric in character. In fact on the mathematical side gauge theory is a well established branch of differential geometry known as the theory of fibre bundles with connection. It has much in common with Riemannian geometry which provided Einstein with the basis for his theory of general relativity. As is well known Einstein spent many years on a fruitless search for a unified field theory, a search which most physicists regarded as a chimera. If the current expectations of Yang-Mills theory are eventually fulfilled, it will in some measure justify Einstein's point of view that the basic laws of physics should all be combined in geometrical form.

Gauge theory first appeared in physics in the early attempt by H. Weyl [43] to unify general relativity and electro-magnetism. Weyl had noticed the conformal invariance of Maxwell's equations and sought to exploit this fact by interpreting the Maxwell field as the distortion of relativistic length produced by moving round a closed path. Weyl's interpretation was disputed by Einstein and never generally accepted. However after the advent of quantum mechanics with its all-important complex wave-functions it became clear that phase rather than scale was the correct concept for Maxwell's equations, or in modern language that the gauge group was the circle rather than the multiplicative numbers. Unfortunately, while scale changes could be fitted into Einstein's theory by replacing the metric with a conformal structure, there was no room for phase to be incorporated into general relativity. Rather the gauge theory had to be superimposed as an additional structure on space-time and the unification sought by Weyl then disappeared.

Non-abelian gauge theories were introduced in 1954 by Yang and Mills [32] and have been increasingly studied by physicists since that time. The relation with the mathematical theory of fibre bundles was either

ignored or considered irrelevant until comparatively recently, when non-perturbative aspects related to instantons have come to the fore. Mathematically this involves global questions of fibre bundle theory incorporating both topology and analysis, as opposed to the purely local theory of classical differential geometry. A great deal of modern geometry of a sophisticated character is involved in dealing with such global problems and the techniques developed by mathematicians are unfamiliar to physicists. One purpose of these lectures is to try to bridge the gap between mathematicians and physicists by explaining the relevant techniques as simply as possible and illustrating how they apply to the determination of instantons. The fact that so many new mathematical tools are naturally involved with this problem may lead to some optimism concerning the ultimate aim of developing the quantized form of gauge theories.

2. – Gauge potentials and fields.

We shall now recall the data of a classical theory as understood by physicists and then reinterpret them in geometrical form.

We begin by fixing a compact Lie group G, typically $SU(2)$ or $SU(3)$, but not excluding yet the abelian group $U(1)$. We consider its Lie algebra $L(G)$, which for $SU(n)$ consists of skew-hermitian $n \times n$ matrices of trace zero. A gauge potential is then a set of functions $A_\mu(x)$ taking values in $L(G)$, where $x = (x_1 \ldots x_4)$ is a point of Minkowski or Euclidean space and $\mu = 1, \ldots, 4$ is a spatial index. Associated with this potential we also consider the operator

$$(2.1) \qquad \nabla_\mu = \partial_\mu + A_\mu$$

where $\partial_\mu = \partial/\partial x_\mu$. This operator acts on a vector function $(f_1(x), \ldots, f_m(x))$ whenever we are given an m-dimensional representation of G: for example when $G = SU(n)$ we can take $m = n$ using the standard representation.

Computing the commutator of ∇_μ and ∇_ν we get the *gauge field* $F_{\mu\nu}$, given by

$$F_{\mu\nu} = [\nabla_\mu, \nabla_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$$

where the commutator $[A_\mu, A_\nu]$ is taken in the Lie algebra of G. The important point to note is that for non-abelian G this commutator does not vanish and so F is a non-linear function of A. For $G = U(1)$ however this term drops out and we get the usual linear relation between the field and vector potential that characterizes Maxwell theory.

The usual non-uniqueness of the potential has its counterpart in the general case in the form of *gauge transformations*. By definition a gauge transformation is a function $g(x)$ taking values in G and transforming the potential A_μ by the formula
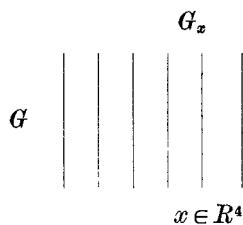
$$A_\mu \to g^{-1} A_\mu g + g^{-1} \partial_\mu g$$

which corresponds to sending ∇_μ into $g^{-1}\nabla_\mu$ (here we consider G as a group of matrices so that $\partial_\mu g$ is simply the differentiated matrix). The gauge field $F_{\mu\nu}$ then transforms by
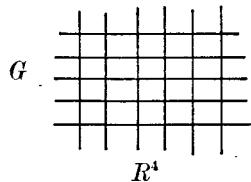
$$F_{\mu\nu} \to g^{-1} F_{\mu\nu} g \ .$$

It is important to observe that the A_μ transform inhomogeneously whereas the $F_{\mu\nu}$ transform homogeneously. In other words $F_{\mu\nu}$ is a vectorial (or tensorial) object whereas A_μ is an affine object (with no preferred zero).

Geometrically or mechanically we can interpret this data as follows. Imagine a structured particle, that is a particle which has a location at a point x of R^4 and an internal structure, or set of states, labelled by elements g of G. We then consider the total space P of all states of such a particle. In general we conceive of the internal spaces G_x and G_y for $x \neq y$ as not being identified and so we draw the picture of P as a collection of « fibres »

In the absence of any external field however we consider that all G_x can be identified to each other so that in addition to the vertical lines or fibres we can also draw horizontal lines (called sections) making the usual Cartesian type of grid

Now we imagine an external field imposed which has the effect of distorting the relative alignment of the fibres so that no coherent identification is possible between the G_x at different points. However we assume that G_x and G_y can still be identified if we choose a definite path in R^4 from x to y. In more physical terms we imagine the particle moving from x to y and carrying its internal space with it. In Minkowski space such a motion would take place along the world line of the particle. This identification of fibres along paths is called « parallel transport ». If we now imagine two different path joining x to y then there is no reason for the two different parallel transports to agree and they are assumed to differ by multiplication with a group element, which could be viewed as a generalized « phase shift ». This phase shift is interpreted as produced by the external field. In geometrical terms it is viewed as the total « curvature » or distortion of the fibre bundle over the region enclosed by the two paths.
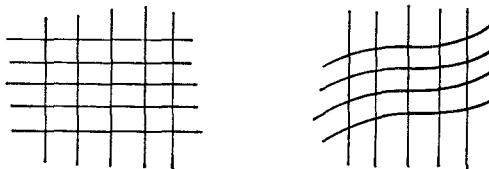
If we now infinitesimalize this picture in Newtonian style we get the infinitesimal parallel transport at a point x in a given direction. This will be an infinitesimal shift A of the fibre G_x into the nearby fibre, and is called a *connection*. The infinitesimal *curvature* F depends on two directions at x and takes values in the Lie algebra of G_x, *i.e.* it is an infinitesimal « phase shift ». As usual the infinitesimal picture, that is the connection, can be integrated up to give the global picture of parallel transport along curves: the two points of view are mathematically equivalent.

If we now compare this picture with the situation where we had no field and all fibres G_x were coherently identified we can view parallel transport as a change of phase in a fixed copy of G, and the connection as an element $A_\mu(x)$ of the Lie algebra depending on the point x and the μ-th direction. Thus we recover the gauge potential of the physicists' language. Similarly the curvature F becomes the gauge field $F_{\mu\nu}(x)$, taking values in the fixed Lie algebra of G. Thus the curvature F can be thought of as the distortion produced by an external field, or it can be *identified* with the field when we think of a field of force as measured by its local effects. This identification of fields with geometrical distortion is of course at the heart of Einstein's theory of gravitation. The difference here is that the distortion is not taking place in the geometry of space-time but in the geometry of some fictitious state-space of internal structure super-imposed on space-time. This difference makes the relevant geometry less obvious and historically, both in physics and in mathematics, the geometry of fibre-bundles came later than the geometry of space. It is significant however that both mathematicians and physicists, each for their own reasons, were led to study these objects which in fact turn up naturally in a great variety of contexts.

Despite its later historical appearance the geometry of fibre bundles of

the type we have been describing is much simpler technically than the Riemann-Einstein geometry of space. This is because the relevant group of our theory was taken as a finite-dimensional group G whereas in Riemannian geometry we have to deal with the group of all coordinate transformations. To clarify this point we shall now return to our fibre bundles and re-examine their relation to gauge theory.

In order to describe our geometrical connection in algebraic terms we compare our parallel transport with the situation in the absence of a field. In this case we used a coherent identification of all the fibres G_x. Now it is important to emphasize that this coherence represents the absence of a field but the particular choice of coherent identification is at our disposal. A particular choice is called picking a gauge and a change from one choice to another is a gauge transformation. Pictorially we imagine two different
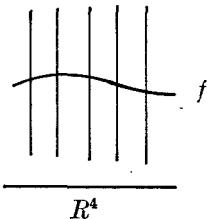

sets of horizontals for our fibre bundle and the change from one to the other is described by a function $g(x)$, taking values in G. No particular choice is regarded as being preferred (despite the appearance of the picture!). Once a gauge has been picked the connection and curvature can be written down in coordinate form. The group of gauge transformations plays a role analogous to that of coordinate transformations in Riemannian geometry. Since it is basically a simpler group the geometry of fibre-bundles is an easier theory: it is, in a definite sense, « less non-linear ».

It should be emphasized that a connection is a definite geometric object and is more primitive than the curvature. As a consequence one should consider the gauge potential as more primitive than the gauge field. This is borne out physically even in electro-magnetism by an experiment which shows that the field may be identically zero but physical effects are still detected due to the fact that parallel transport need not be trivial if the region of space is not simply-connected. The vanishing of curvature only gives information about parallel transport round very small closed paths. In physical terminology parallel transport in general is described by talking about non-integrable phase factors. Non-integrability locally refers to a non-vanishing field, whereas large scale non-integrability is topological in character (going round a wire for example) and may arise even for zero fields (outside the wire). Classically potentials were introduced as a mathematical

device to simplify the field equations and the ambiguity (or gauge freedom) in choice of potential was taken as an indication that the potential had no genuine physical meaning. The geometrical point of view shows that this is too narrow an interpretation. The connection is a geometric object and so the potential should be regarded as physical. The unphysical thing is the choice of gauge in which one chooses to describe the potential, corresponding to the fact that our geometrical fibre bundle where the connection sits has no natural horizontal sections. These remarks about the predominant role of the potential will acquire more substance when we discuss the field equations in the general non-abelian case.

So far we have talked only about fibre bundles in which the fibre was the group G. These are called principal fibre bundles by differential geometers. However for applications one is usually interested in associated bundles in which the fibre is a vector space C^n corresponding to a representation of G. Again the typical case is to take $G = U(n)$. The geometric picture is essentially similar in that we consider a space (the vector bundle) E, fibred over R^4 so that the fibre E_x is thought of as a vector space depending smoothly on x. Parallel transport from x to y is regarded as a unitary transformation from E_x to E_y. Thus parallel transport in the principal bundle gives rise to parallel transport in the vector bundle and the same applies to connection and curvature. In particular a section of the vector bundle, namely a function $f(x)$ defined on R^4 and taking values in the variable vector space E_x is thought of in terms of its graph. A connection enables us to shift

$$R^4$$

this graph infinitesimally in a given direction of R^4. This shift is precisely the covariant derivative $\nabla_\mu f$. This is a geometric notion independent of any choice of gauge. Once we choose a gauge we can describe f algebraically by a n-vector $(f_1(x), ..., f_n(x))$ of ordinary functions and the covariant derivative is then given explicitly by formula (2.1). The curvature $F_{\mu\nu}$ defined as the commutator $[\nabla_\mu, \nabla_\nu]$ is again seen to be geometric in character, appearing now as an (algebraic) operator on sections of the vector bundle.

3. – The field equations.

We come now to the field equations for a gauge theory which will generalize Maxwell's equations. Written in terms of the covariant derivative ∇_μ, and in units where the velocity of light is 1, the equations can be written in terms of commutators

$$(3.1) \qquad [\nabla_\mu, [\nabla_\nu, \nabla_\sigma]] + [\nabla_\nu, [\nabla_\sigma, \nabla_\mu]] + [\nabla_\sigma, [\nabla_\mu, \nabla_\nu]] = 0$$

$$(3.2) \qquad [\nabla_\mu, [\nabla_\mu, \nabla_\nu]] = 0 \ .$$

In (3.2) we sum over μ and in Minkowski space the term corresponding to the time component has a minus sign (whereas the Euclidean analogue has all positive signs).

These two equations involving the potential have a rather different character, in that the first is an identity (the Bianchi identity of differential geometry) and only the second, the *Yang-Mills equation*, imposes a condition on the potential. For $G = U(1)$ these equations written in terms of the field $F_{\mu\nu}$ are Maxwell's equations in vacuo. The first equation is then just the integrability condition on $F_{\mu\nu}$ which asserts that (at least locally) we can introduce a potential A_μ so that

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ .$$

For non-abelian G it is not possible to write these equations in terms of $F_{\mu\nu}$ alone because the covariant derivatives ∇_μ explicitly involve the potential A_μ. This emphasizes once more the primary role of the potential as opposed to the field.

The Yang-Mills equation (3.2) is derived from a Lagrangian $\mathfrak{L}$ by integrating over R^4 a Lagrange density which is an invariantly defined quadratic expression in the curvature. For $G = U(n)$ or $SU(n)$ one puts (up to constant factor)

$$(3.3) \qquad \mathfrak{L} = -\tfrac{1}{2} \int_{R^4} \mathrm{Trace}\,(F_{\mu\nu} F^{\mu\nu})\, dx_1\, dx_2\, dx_3\, dx_4 \ ,$$

where $F^{\mu\nu}$ is obtained in the usual way from $F_{\mu\nu}$, raising indices by the standard metric tensor of Minkowski or Euclidean space and we sum over all μ, ν. Equations (3.2) are the corresponding Euler Lagrange equations. The minus sign was inserted in (3.3) so that in the Euclidean case we get

a positive Lagrangian, the point being that $-\text{Trace}\,(AB)$ is positive definite on the Lie algebra of $U(n)$. For other Lie groups one can either use an embedding in $U(n)$ or more intrinsically one replaces Trace (AB) by the Killing form which is the standard invariant bilinear form on $L(G)$: the two methods give the same answer up to a positive scalar multiple.

In the Euclidean case the Lagrangian can be viewed as the natural L^2-norm of the curvature, that is the integral over R^4 of the sums of the squares of the absolute values of all its components in a standard orthonormal base. More invariantly one can rewrite this as follows. First we recall that a skew tensor $\alpha_{\mu\nu}$ corresponds to an exterior differential 2-form

$$\alpha = \tfrac{1}{2}\sum_{\nu,\mu}\alpha_{\mu\nu}\,dx^\mu\wedge dx^\nu = \sum_{\mu<\nu}\alpha_{\mu\nu}\,dx^\mu\wedge dx^\nu\ .$$

The dual 2-form $*\alpha$ defined relative say to the standard Euclidean metric is given by replacing α_{12} by α_{34} etc. with an appropriate minus sign if the indices involve an odd permutation. It satisfies $*^2 = 1$. The natural L^2-inner product of 2-forms is then defined by

$$\langle\alpha,\beta\rangle = \int_{R^4}\alpha\wedge*\beta\ ,$$

where $\wedge$ denotes the exterior multiplication and gives here an exterior 4-form, *i.e.* a volume form which can therefore be integrated. Passing now to the curvature F which is a 2-form with values in the Lie algebra $L(G_x)$ we define $*F$ in the same way, switching spatial indices and leaving untouched the Lie algebra variables. Then we put

$$(3.4)\qquad \|F\|^2 = \langle F,F\rangle = -\int_{R^4}\text{Trace}\,(F\wedge*F)$$

and this is the invariant way of writing the Lagrangian.

Equation (3.1) asserts that the covariant derivative of F, skew-symmetrized, is zero or symbolically

$$(3.5)\qquad\qquad\qquad \nabla\wedge F = 0\ .$$

Using the duality operator $*$ we see that (3.2) can then be written

$$(3.6)\qquad\qquad\qquad \nabla\wedge*F = 0\ .$$

This way of writing the Yang-Mills Lagrangian and equations exhibits

clearly its invariance and covariance properties. First of all the equations clearly make sense on a curved (Riemannian) 4-space, since the *-operator uses only the infinitesimal duality. Secondly the *-operator on 2-forms in 4-space is conformally invariant in the sense that two metrics ds^2 and $\varrho(x)\,ds^2$ give the same *. Thus the Yang-Mills equation (and the Yang-Mills Lagrangian) depend only on the conformal structure of 4-space. This important property of Maxwell's theory is therefore preserved in the non-abelian case.

The apparent symmetry or duality between (3.5) and (3.6) is delusory as we have explained, although in Maxwell theory it reflects the duality between electricity and magnetism and attempts have been made to understand the non-abelian analogue. This is a deep question and the proper understanding of this duality is likely to be found only at the quantum level [22] [40]. However at the classical level we note an elementary consequence of (3.5) and (3.6), namely that (3.6) follows from the identity (3.5) if the field F satisfies one of the equations

$$(3.7) \qquad\qquad {}^*F = F \qquad \text{(self-duality)}$$

$$(3.8) \qquad\qquad {}^*F = -F \qquad \text{(anti-self-duality)} \,.$$

Thus we have here *first-order* non-linear equations for the potential which imply the second-order Yang-Mills equations. These equations have a particularly simple significance in the Euclidean case as we shall see in the next section. Note that the definition of * involves an orientation of R^4 (an ordering of the coordinates $x_1, \dots, x_4$) and that (3.7) and (3.8) switch when we reverse the orientation. Thus there is no essential mathematical difference between the two cases.

4. – Asymptotic conditions and topology.

We now restrict ourselves to the Euclidean 4-space so that the Yang-Mills Lagrangian $\mathcal{L}$ is positive. It is natural to consider potentials for which the action $\mathcal{L}$ is finite, so that the integral over R^4 converges. To achieve this we assume that the field F decays sufficiently fast as we go to infinity. If we work in a given gauge this simply means that $F_{\mu\nu}(x) \to 0$ sufficiently fast as $|x| \to \infty$. At first sight this might seem to require the gauge potential $A_\mu(x)$ to have similar decay together with its first derivatives. However, because of gauge freedom, all that is necessary is that, for large $|x|$, we can find a gauge transformation $g(x)$ so that the potential in the new

gauge should decay. This means that

$$(4.1) \qquad A_\mu(x) \sim g^{-1}(x)\, \partial_\mu g(x) \qquad \text{as } |x| \to \infty$$

where $\sim$ implies asymptotic behaviour including first derivatives. The important point is that the gauge transformation $g(x)$ need only be defined for large $|x|$. In fact it may be impossible to extend the definition of $g(x)$ continuously to the whole 4-space. To see this consider the restriction of $g(x)$ to a sphere $|x| = R$ of large radius, and take for example $G = SU(2)$. Then g gives a continuous map

$$g: S^3_R \to SU(2)$$

and both sides are topologically 3-spheres. Such a map has a well-defined integer invariant, its degree k, which counts (with appropriate multiplicities and signs) the number of points $x \in S^3_R$ which map to a given general value in $SU(2)$. The function g can be extended continuously to $|x| \leqslant R$ if and only if $k = 0$. The identity map (thinking of both spaces as the standard S^3) has $k = +1$, and $k = -1$ corresponds to an orientation reversal.

An analogous and more easily visualized situation occurs for $G = U(1)$ and R^4 replaced by R^2 in which case g becomes a map of the circle to itself and the degree k is the « winding number ». Note however that if we work on R^4 with $G = U(1)$ there is no topological invariant since every continuous map of S^3 to the circle can be deformed to a constant map.

For any simple non-abelian compact Lie group G a corresponding result holds, namely that continuous maps of S^3 into G have an integer topological invariant which classifies the map up to deformation. This integer arises from the fact that every such G contains copies of $SU(2)$ (or $SO(3)$) as subgroups.

Thus in non-abelian gauge theories on R^4, potentials which are asymptotically flat, *i.e.* have fields asymptotic to zero, fall into distinct families indexed by the corresponding integer k. This is frequently referred to as a topological quantum number even though at this stage we are only dealing with classical fields.

In dealing with asymptotic properties various technical analytical questions arise concerning the precise rate of decay. There is one natural and convenient definition of decay to take arising from conformal invariance. We recall that stereographic projection of a sphere onto flat space is always a conformal map, relating the standard curved metric of the sphere to the flat Euclidean metric. In particular the sphere S^4 (unit sphere in R^5) gets mapped conformally onto R^4.

Alternatively we can say that S^4 is the conformal compactification of R^4 obtained by adding a point at ∞. A potential on R^4 is then said to decay at ∞ if it extends to a potential on S^4. If the integer k is non-zero this means that we cannot describe our potential using a single gauge, we need one gauge in the finite region $|x| \leqslant R$ and another gauge near ∞ i.e. for $|x| \geqslant R$, the two gauges being related on $|x| = R$ by the gauge transformation $g(x)$ of degree k. This means that our fibre bundle P over S^4 is no longer the topological product $S^4 \times G$. In fact the topological theory of fibre bundles over spaces like S^4 which are not contractible tells us that in this case they are precisely classified by the same integer k. Thus the integer k which appeared in the asymptotic description on R^4 is now directly coded into the topology of the space P. For example if $G = SU(2)$ and $k = 1$, P turns out to be topologically the sphere S^7 which is quite different from the product $S^4 \times S^3$: this will be explained in more detail later. The potential or connection in our fibre bundle now has a well-defined curvature F on the whole of S^4 and the particular role of the base point $\infty \in S^4$ can now be ignored. Note that $F \to 0$ at ∞ on R^4 but F need not be zero on S^4 at the point ∞ because when viewed as a differential form one uses different coordinates on R^4 and S^4. Clearly the Lagrangian $\mathcal{L} = \|F\|^2$ computed relative to the curved metric of S^4 is necessarily finite because S^4 is compact (we always assume enough local differentiability so that F is always continuous at least). Moreover because of the conformal invariance of the Yang-Mills functional $\mathcal{L}$ takes the same value whether computed on S^4 or R^4.

Now on closed manifolds like S^4 there are well known theorems of global differential geometry which relate topological invariants to integral expressions in the curvature. One might expect such results because if the curvature is everywhere zero the connection is flat and (on a simply-connected space) one obtains a global gauge implying $k = 0$. In dimension 2 the classical theorem of Gauss expressing the Euler characteristic as the integral of the scalar curvature is the prototype of higher-dimensional generalizations. In our case the formula takes the form

$$(4.2) \qquad\qquad 8\pi^2 k = -\int_{S^4} \mathrm{Trace}\,(F \wedge F)\,,$$

for the group $SU(2)$. For other groups the same result holds but with a different normalization, the integer k being replaced by a suitable multiple: for precise details the reader may refer to [3]. We now have a topological constraint on F and this should be fed into the formula (3.4) for the Lagrangian. To do this it is convenient to decompose F under the action of $*$ into

its self-dual part F^+ and its anti-self-dual part F^-:

$$F = F^+ \oplus F^-$$

so that $*F^+ = F^+$ and $*F^- = -F^-$ (recall $*^2 = 1$ for the positive definite case). Then (3.4) and (4.2) can be written

(3.4)′
$$\mathfrak{L} = \|F^+\|^2 + \|F^-\|^2$$

$$8\pi^2 k = \|F^+\|^2 - \|F^-\|^2$$

from which we deduce that $\mathfrak{L} \geqslant 8\pi^2 |k|$ and equality holds if and only if $*F = (\text{sign } k)F$. Thus the special solutions (3.7) and (3.8) of the Yang-Mills equations correspond to the absolute minimum of the Lagrangian (assuming the value $8\pi^2 |k|$) is attained. This argument is due to Belavin et al. [10] who also showed that for $k = \pm 1$ the minimum is attained, and the term instanton has been coined for such solutions of the Yang-Mills equation. Later other solutions were discovered for all k [14][31] and these were called multi-instantons.

The general problem to which we shall address ourselves is the determination, in as explicit a fashion as possible, of all multi-instantons, not only for $SU(2)$ but for all compact classical groups. As we shall see, this problem admits of a surprisingly simple and complete answer, but the proofs require a great deal of sophisticated mathematical machinery.

At this point we shall merely note that any gauge transform of a multi-instanton is again a multi-instanton, and such solutions will be regarded as equivalent. In geometric terms this means that two fibre bundles over S^4 with connections (satisfying $*F = \pm F$) which are isomorphic will be identified: they cannot be distinguished geometrically.

In the next chapter we shall describe explicitly how to write down the most general multi-instanton for $SU(2)$, and we shall indicate how this has to be generalized to other groups.

CHAPTER II

Description of Instantons

1. – Quaternions.

In R^2 it is well known that many formulae are simpler when written in terms of complex numbers. It was the discovery of Hamilton that for R^4 one can introduce a non-commutative extension of the complex numbers called quaternions, and it was Hamilton's hope that quaternions would turn out to be the natural tool for an algebraic description of the physical world. Although this was an exaggerated hope there is some merit in Hamilton's point of view as we shall see shortly. For the present however, we simply use quaternions as a convenient algebraic formalism that simplifies notation. This is particularly relevant to the group $SU(2)$ which as we shall see can be identified with the group of quaternions of unit norm, in the same way that $U(1)$ is the complex numbers of unit norm.

We begin by briefly recalling the definition and elementary properties of quaternions. Just as the complex numbers C are formed from the real numbers R by adjoining a symbol i with $i^2 = -1$, so the quaternions H (in honour of Hamilton) are formed from R by adjoining three symbols i, j, k satisfying the identities:

(1.1) $i^2 = j^2 = k^2 = -1$

$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j.$$

Thus a general quaternion x is of the form

(1.2) $$x = x_1 + x_2 i + x_3 j + x_4 k$$

where x_1, x_2, x_3, x_4 are real numbers. The conjugate quaternion $\bar{x}$ is defined by

$$\bar{x} = x_1 - x_2 i - x_3 j - x_4 k$$

and conjugation is an anti-involution, i.e. $\overline{(xy)} = \bar{y}\bar{x}$. In virtue of the identity (1.1), one finds

$$x\bar{x} = \bar{x}x = \sum_{\mu=1}^{4} x_\mu^2$$

This quantity is denoted $|x|^2$ and is zero only for $x = 0$. If $x \neq 0$ it has a unique inverse x^{-1} given by

$$x^{-1} = \bar{x}/|x|^2 .$$

The quaternions x with norm 1, *i.e.* $|x| = 1$, form therefore a multiplicative group which is geometrically the 3-sphere

$$\sum_{\mu=1}^{4} x_\mu^2 = 1 .$$

In analogy again with the complex numbers we refer to the component x_1 in (1.2) as the real part of x and the remainder $x - x_1$ as the imaginary part.

If we identify i with the usual complex number we can regard the complex numbers C as contained in H (taking $x_3 = x_4 = 0$). Moreover every quaternion x as in (1.2) has a unique expression.

$$x = z_1 + z_2 j , \quad \text{where } z_1 = x_1 + x_2 i \text{ and } z_2 = x_3 + x_4 i .$$

This identifies H with C^2. Now consider the quaternion multiplication $x \to xg$ where $g = g_1 + g_2 j$, with $g_1, g_2 \in C$. Computing we find

$$xg = (z_1 + z_2 j)(g_1 + g_2 j) = z_1 g_1 - z_2 \bar{g}_2 + (z_1 g_2 + z_2 \bar{g}_1) j$$

so that the vector (z_1, z_2) is multiplied on the right by the 2×2 complex matrix

$$\begin{pmatrix} g_1 & g_2 \\ -\bar{g}_2 & \bar{g}_1 \end{pmatrix} .$$

Thus, if we wish, we can identify the algebra of quaternions as a sub-algebra of the 2×2 complex matrices in which i, j, k are the matrices

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} .$$

In particular the group $Sp(1)$ of quaternions of unit norm gets identified

with $SU(2)$ and its Lie algebra can be viewed as the pure imaginary quaternions with natural basis i, j, k.

We now identify R^4 with H via (1.2) and an $SU(2)$-potential will be given by functions $A_\mu(x)$ whose values are imaginary quaternions. It will be convenient if we go further and write

$$A(x) = \sum_{\mu=1}^{4} A_\mu(x)\,dx^\mu$$

so that $A(x)$ is a differential form with values in $\mathrm{Im}\,(H)$. Finally we shall consider the quaternion differential

$$dx = dx^1 + dx^2\,i + dx^3\,j + dx^4\,k$$

and its conjugate

$$d\overline{x} = dx^1 - dx^2\,i - dx^3\,j - dx^4\,k$$

just as in complex variable theory one uses $dz = dx + i\,dy$ and $d\overline{z} = dx - i\,dy$. If $f(x)$ is any function of the quaternion variable x with quaternion values the expression

(1.3) $\qquad\qquad A(x) = \mathrm{Im}\,\{f(x)\,dx\} = \tfrac{1}{2}\{f(x)\,dx - d\overline{x}\,\overline{f(x)}\}$ $\qquad$.

will represent an $SU(2)$-potential. Here $f(x)\,dx$ is computed formally, written as $\sum a_\mu\,dx^\mu$ with $a_\mu \in H$ and then $A_\mu = \mathrm{Im}\,(a_\mu)$. Note that, before taking the imaginary part we have a potential for the group H^* of all non-zero quaternions which is $SU(2)$ times a scale factor.

We shall also write the curvature F as an exterior 2-form

$$F = \sum_{\mu<\nu} F_{\mu\nu}\,dx^\mu \wedge dx^\nu = \tfrac{1}{2}\sum_{\mu,\nu} F_{\mu\nu}\,dx^\mu \wedge dx^\nu \ .$$

Then F can be computed from A by

$$F = dA + A \wedge A$$

where

$$dA = \sum_\mu dA_\mu \wedge dx^\mu = \tfrac{1}{2}\sum_{\mu,\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu)\,dx^\mu \wedge dx^\nu$$

and

$$A \wedge A = \sum_{\mu,\nu} A_\mu A_\nu\,dx^\mu \wedge dx^\nu = \tfrac{1}{2}[A_\mu,\,A_\nu]\,dx^\mu \wedge dx^\nu \ .$$

When A is written in the quaternionic form (1.3) we get a similar formula

$$(1.4) \qquad F = \mathrm{Im}\,\{df \wedge dx + f\,dx \wedge f\,dx\}\,.$$

Taking imaginary parts commutes with formation of curvature because of our remark above concerning the larger group H^*.

The use of the quaternion differentials dx and $d\bar{x}$ is also convenient in connection with the study of self-duality. To see this let us compute $dx \wedge d\bar{x}$. We get

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

The coefficients of i, j, k in this expression are precisely a basis for the self-dual 2-forms, i.e. 2-forms ω with $*\omega = \omega$. Hence $dx \wedge d\bar{x}$ is a 2-form, with values in the Lie algebra of $SU(2)$, which is self-dual. A similar computation shows that $d\bar{x} \wedge dx$ is anti-self-dual.

2. – The basic instanton.

Using the quaternionic notation of the preceding section we shall now exhibit the basic instanton with $k = +1$ (and the anti-instanton with $k = -1$).

Consider the $SU(2)$-potential A defined by

$$(2.1) \qquad A(x) = \mathrm{Im}\left\{\frac{\bar{x}\,dx}{1 + |x|^2}\right\} = \frac{1}{2}\left\{\frac{\bar{x}\,dx - d\bar{x}\,x}{1 + |x|^2}\right\}\,.$$

The explicit components $A_\mu(x)$ can of course be read off from this succinct formula, for example

$$A_1(x) = \frac{-x_2 i - x_3 j - x_4 k}{1 + |x|^2}\,, \qquad A_2(x) = \frac{x_1 i - x_4 j + x_3 k}{1 + |x|^2}\,.$$

Computing the curvature F of this potential as in (1.4) we get

$$(2.2) \qquad F = \mathrm{Im}\left\{\frac{d\bar{x} \wedge dx}{1 + |x|^2} + \bar{x}\,d(1 + |x|^3)^{-1} \wedge dx + \frac{\bar{x}\,dx \wedge \bar{x}\,dx}{(1 + |x|^2)^2}\right\}\,.$$

Writing $|x|^2 = x\bar{x}$ the middle term in this expression gives

$$-\frac{\bar{x}\,dx\wedge\bar{x}\,dx}{(1+|x|^2)^2}-\frac{\bar{x}x\,d\bar{x}\wedge dx}{(1+|x|^2)^2}.$$

Substituting this in (2.2) and simplifying we get the purely imaginary expression:

$$(2.3)\qquad\qquad F=\frac{d\bar{x}\wedge dx}{(1+|x|^2)^2}$$

so that, as explained in §1, F is anti-self-dual, i.e. $*F = -F$. As $|x|\to\infty$ we see from (2.1) that

$$(2.4)\qquad\qquad A(x)\sim\operatorname{Im}(x^{-1}dx)=\varphi(x)^{-1}d\varphi(x)$$

where $\varphi(x)=x/|x|$. This shows that A is asymptotically the gauge transform of 0 by the gauge transformation $g(x)=\varphi(x)$, or equivalently that if we apply the inverse gauge transformation $\varphi(x)^{-1}$ to A we get 0 asymptotically. On the unit sphere $|x|=1$ in quaternion space we have $\varphi(x)^{-1}=\bar{x}$ and the map $x\to\bar{x}$ of S^3 to itself has degree -1. Thus (2.1) describes an anti-instanton.

Clearly if we replace x by $\bar{x}$ throughout we will obtain an instanton with potential and field given by

$$(2.5)\qquad\qquad A=\operatorname{Im}\left\{\frac{x\,d\bar{x}}{1+|x|^2}\right\},\qquad F=\frac{dx\wedge d\bar{x}}{(1+|x|^2)^2}.$$

These formulae are of course just those of Belavin et al. [10] written in quaternionic notation.

To examine more carefully the behaviour of the anti-instanton (2.1) as $|x|\to\infty$ we change gauge by $\varphi(x)^{-1}$, corresponding to (2.4), and introduce the quaternion coordinate $y=x^{-1}$ around the point at ∞, regarded now as a point of S^4. Taking the imaginary part of the identity

$$x\left(\frac{\bar{x}\,dx}{1+|x|^2}\right)x^{-1}+x\,dx^{-1}=\frac{\bar{y}\,dy}{1+|y|^2}$$

shows that the anti-instanton extends to S^4 and has precisely the same form at ∞ as it has near 0. Similar calculations hold naturally for the instanton (2.5).

If we simply put $y = x^{-1}$ in (2.1) we get

$$(2.6) \qquad A(y) = - \operatorname{Im} \left\{ \frac{dy \, y^{-1}}{1 + |y|^2} \right\}$$

and this describes the anti-instanton in the «singular» or «asymptotic» gauge, namely the gauge in which $A \to 0$ as $|y| \to \infty$, but which is badly behaved at $y = 0$, where $A(y)$ is singular. As we have seen this singularity can be removed by the appropriate gauge transformation, but the singular form (2.6) is useful in practice.

It is perhaps worth pointing out that all the above formulae hold for complex numbers instead of quaternions, in which case we obtain $U(1)$-gauge potentials and fields on R^2 or S^2. Self-duality no longer makes sense in dimension two but the 2-form F given by (2.3) is a multiple of the invariant spherical area. In other words F is invariant under $SU(2)$ acting by fractional linear transformations $x \to (ax + b)(cx + d)$ of the complex variable x.

Since the equations $*F = \pm F$ are conformally invariant it follows that any conformal transformation of S^4 into itself will convert the basic instanton into some other instanton. We recall now that, just as the proper (*i.e.* orientation preserving) conformal group of S^2 is $SL(2, C)/\{\pm 1\}$ acting via fractional linear transformations of a complex variable, so the proper conformal group of S^4 is $SL(2, H)/\{\pm 1\}$ acting similarly on a quaternionic variable. Thus the transformations $x \to axb$ with x, a, b all quaternions and $a \neq 0$, $b \neq 0$ generate the rotation group $SO(4)$ together with scale change; $x \to x + c$ gives translations while $x \to 1/x = \bar{x}/|x|^2$ gives a proper inversion (i.e. inversion together with a compensating reflection $x \to \bar{x}$ to restore orientation). Because of the non-commutativity of the quaternions every proper conformal transformation of S^4 can be written either using left multiplication or using right multiplication, i.e.

$$x \to (ax + b)(cx + d)^{-1}$$

or

$$x \to (x\gamma + \delta)^{-1}(x\alpha + \beta) \,.$$

Taking quaternionic conjugates interchanges these two ways of representing the conformal transformations. More precisely let S, T denote the two transformations above with $(\alpha, \beta, \gamma, \delta) = (\bar{a}, \bar{b}, \bar{c}, \bar{d})$ and let C denote conjugation $x \to \bar{x}$, then $T = CSC$. Note that C, being a reflection, is an improper conformal transformation.

We return now to the basic anti-instanton (2.1) and we apply conformal transformations to it. Recall that, up to a gauge transformation, (2.1) is

preserved by the inversion $x \to x^{-1}$. It is evidently unchanged by $x \to ax$ with $|a| = 1$ and $x \to xa$ produces only a constant gauge transformation. Thus it is essentially invariant by $SO(4)$. In fact it is invariant up to gauge transformations by the larger group $SO(5)$ which may be viewed here as $Sp(2)/\{\pm 1\}$ where $Sp(2) \subset SL(2, H)$ is the compact subgroup leaving norms fixed. This verification is best left till later when the proper geometric interpretation of (2.1) will make this invariance evident. To get new anti-instantons therefore we should use elements representing $SL(2, H)$ modulo $Sp(2)$. Such elements are naturally given by the transformations

$$(2.7) \qquad\qquad x \to \mu(x - b)$$

where μ is a positive real scalar and b is a quaternion.

These parameters can be regarded as parametrizing $SL(2, H)/Sp(2)$, the space of quaternion norms on H^2 with volume 1, by associating to (μ, b) the positive self-adjoint matrix

$$\begin{pmatrix} \mu & b \\ b^* & v \end{pmatrix}$$

with $\mu v - |b|^2 = 1$.

As μ, b vary the transformation (2.7) applied to (2.1) generates a 5-parameter family of anti-instantons with « centre » b and « scale » μ. From (2.3) we see that the field density is a maximum at the centre and its strength there is μ^2. This shows that no two members of our family can be gauge equivalent. The more difficult result, which will emerge much later, is that every anti-instanton (i.e. $k = -1$) is gauge equivalent to one of our family.

It will be convenient to apply inversion to (2.7) (and a sign change) to (2.7) to get

$$(2.8) \qquad\qquad x \to \lambda(b - x)^{-1} \;.$$

This transformation applied to (2.3) gives us the general anti-instanton in an asymptotic gauge as explained before. We can also apply it to (2.5) to generate the family of all instantons in an asymptotic gauge.

We now come to the more difficult question of constructing multi-instantons for larger values of k. For this purpose we introduce the space H^k consisting of column vectors u with quaternion components u_α ($\alpha = 1, ..., k$), and we define an $SU(2)$-potential on the space $H^k = R^{4k}$ by a formula quite similar to (2.1), namely

$$(2.9) \qquad\qquad A(u) = \operatorname{Im}\left\{ \frac{u^* \, du}{1 + |u|^2} \right\}$$

where u^* stands for the transposed conjugate of the column vector u

and $u^* du$ stands for the matrix product, so that

$$u^* du = \sum_{\alpha=1}^{k} \overline{u}_\alpha du_\alpha$$

and $|u|^2 = u^* u = \sum_\alpha |u_\alpha|^2$ is the Euclidean norm.

Note that (2.9) restricts to (2.1) on each coordinate axis (all $u_\alpha = 0$ except $\alpha = \beta$) and is unchanged by the group $\mathrm{Sp}\,(k)$ acting on H^k. Thus it restricts to (2.1) on any one-dimensional H-subspace of H^k. It has therefore a high degree of symmetry and we shall in due course explain its geometrical significance. For the present we regard (2.9) as simply an auxiliary formula used to construct potentials on $H = R^4$ by using suitable functions $u = f(x)$, i.e. maps $f: H \to H^k$. Given any such f we substitute in (2.9) and obtain the potential

$$(2.10) \qquad A_f(x) = \mathrm{Im}\left\{\frac{f^*(x)\,df(x)}{1 + |f(x)|^2}\right\} = \mathrm{Im}\left\{\sum_\alpha \frac{\overline{f_\alpha}(x)\,df_\alpha(x)}{1 + |f(x)|^2}\right\}.$$

For our function $f(x)$ we now take the matrix analogue of (2.8), combined with a conjugation (to give instantons rather than anti-instantons), namely

$$(2.11) \qquad u(x) = [\lambda(B - x)^{-1}]^* .$$

Here B is a symmetric $k \times k$ matrix of quaternions, λ is a row vector $(\lambda_1, ..., \lambda_k)$ of quaternions and x stands for the scalar quaternion xI where I is the unit $k \times k$ matrix. For $k = 1$ the parameters were arbitrary except that λ had to be invertible. In the general case however the parameters λ, B will have to satisfy algebraic constraints as follows:

(I) $B^* B + \lambda^* \lambda$ is a *real* $k \times k$ matrix

(II) For every $x \in H$ the equations

$$(B - x)\xi = 0 , \qquad \lambda\xi = 0 \text{ with } \xi \in H^k \text{ imply } \xi = 0 .$$

Condition (I) asserts that the coefficients of i, j, k in the $k \times k$ quaternion matrix $B^* B + \lambda^* \lambda$ all vanish. This gives a system of quadratic relations on the coefficients of B and λ.

Condition (II) is a non-degeneracy or open condition which can also be formulated as saying that the $(k + 1) \times k$ matrix $\begin{pmatrix} \lambda \\ B - x \end{pmatrix}$ has maximal rank k for all $x \in H$. Condition (I) is the crucial algebraic one which will

ensure that the potential $A_{\lambda,B}(x)$ defined by substituting (2.11) in (2.10) is self-dual. Condition (II) will ensure the solution is non-degenerate and in particular that the points x for which $(B-x)$ is singular give singularities of the potential which can be removed by a gauge transformation.

In principle, given gauge potentials $A_{\lambda,B}(x)$ defined as above it is a direct matter of computation to verify that the resulting field F satisfies $*F = -F$. However the computations can be carried out more elegantly once we have explained the geometrical meaning of some of the formulae. This will be done in the next section. Much more difficult is the proof that our construction gives all self-dual fields. This requires the introduction of quite new ideas and techniques which will be explained in subsequent chapters.

The matrix transformation (2.11) includes, as a specially simple case, the obvious choice of a diagonal matrix B (with diagonal entries $b_1, ..., b_k \in H$) and real positive scalars $\lambda_1, ..., \lambda_k$. Provided the b_i are all distinct, conditions (I) and (II) will be satisfied. The resulting k-instanton therefore looks like a superposition of k instantons with scales λ_i and centres b_i. These special solutions were discovered by 't Hooft and others [14] [31]. The general solution cannot however be put into this form (even after a conformal transformation).

If we replace the vector $\lambda = (\lambda_1, ..., \lambda_k)$ in (2.11) by $q\lambda$ where q is a quaternion of unit norm then the resulting potential A given by (2.9) is unaltered. Similarly if we replace λ by λT and B by $T^{-1}BT$, where T is a (real) orthogonal $k \times k$ matrix, then the potential gets conjugated by the constant matrix T which simply gives a gauge transformation. Note that both these alterations of (λ, B) preserve the conditions (I) and (II) above. We shall in due course prove that no other transformations of the parameters (λ, B), except those just described, give gauge equivalent potentials. Thus the main theorem to be proved can be stated as follows [2] [18] [19]

THEOREM. *Every k-instanton for $SU(2)$ arises from parameters (λ, B) satisfying* (I) *and* (II), *the potential being given in an asymptotic gauge by formula* (2.9) *where $u(x)$ is defined by* (2.11). *The potentials defined by (λ, B) and (λ', B') are gauge-equivalent if and only if $\lambda' = q\lambda T$, $B' = T^{-1}BT$ with $q \in Sp(1)$ and $T \in O(k)$.*

It is a simple matter to count the number of effective parameters involved in our construction. The initial data of a pair (λ, B) involves

$$4k + 4 \cdot \tfrac{1}{2}k(k+1) = 2k^2 + 6k$$

real parameters. The number of real equations involved in (I) is $3k(k-1)/2$ while the groups $Sp(1)$ and $O(k)$ have dimensions 3 and $\tfrac{1}{2}k(k-1)$ respect-

ively. Computing naively we find

$$2k^2 + 6k - \frac{3k(k-1)}{2} - 3 - \tfrac{1}{2}k(k-1) = 8k - 3$$

as the number of effective parameters. This checks with the calculation by infinitesimal variation methods [3] [37]. If we start from a 't Hooft solution (B diagonal, λ real) and consider small perturbations of the solution these parameters can be interpreted as follows. Each b_i has 4 parameters and each $\lambda_i \in H$ has 4 parameters, giving $8k$ in all: we subtract 3 because of the action of $Sp(1)$. However we cannot satisfy condition (I) with B diagonal and the λ_i not real. What happens is that condition (I) requires non-diagonal terms to enter in B so as to cancel the imaginary contributions from λ. Near a 't Hooft solution one can actually solve by power series expansion for these off-diagonal terms so that $(b_1 \ldots b_k, \lambda_1 \ldots \lambda_k)$ modulo $Sp(1)$ give local parameters for the space of k-instantons. However these are not global parameters and the global structure (even topologically) of this « moduli » space is quite complicated (see [5]).

For the group $Sp(n)$ (consisting of $n \times n$ quaternion matrices which also preserve the norm of H^n) there is an entirely analogous solution of the multi-instanton problem. We replace (λ, B) by (Λ, B) where Λ is an $n \times k$-matrix satisfying the analogues of (I) and (II), namely

(I)$_n$ $B^*B + \Lambda^*\Lambda$ is real

(II)$_n$ $(B - x)\xi = 0$ and $\Lambda\xi = 0$ imply $\xi = 0$.

We then define the $k \times n$-matrix U as a function of $x \in H$ by

$$(2.12) \qquad U(x) = [\Lambda(B - x)^{-1}]^* .$$

Finally the potential $A(x)$ is defined in terms of $U(x)$ by a formula generalizing (2.9):

$$(2.13) \qquad A(x) = \sigma U^* dU \sigma + \sigma^{-1} d\sigma$$

where $\sigma = (1 + U^*U)^{-\frac{1}{2}}$ is now a self-adjoint $n \times n$-matrix. Note that, when $n = 1$, σ is a real scalar and (2.13) coincides with the alternative form (2.9).

3. – Geometrical interpretation.

There is a geometrical way to construct potentials which is very simple but appears not to be familiar to physicists. Roughly speaking the con-

struction is analogous to the way in which Riemannian metrics on a manifold can be constructed by embedding the manifold in a Euclidean space and considering the induced metric. Historically this is of course the way Riemannian metrics first arose (e.g. surfaces in 3-space).

We recall that a vector bundle E over a space X is a family of vector spaces E_x parametrized (continuously) by $x \in X$: for example if X is an n-dimensional manifold its tangent spaces form such a vector bundle with $E_x \cong R^n$. We say that E is embedded in the trivial bundle $X \times R^N$ if each E_x is embedded in R^N, the embedding varying continuously (or differentiably) with x. For example if X is embedded as a manifold in R^N its tangent spaces (translated to the origin) get identified with subspaces of R^N. When E is embedded in $X \times R^N$ a section of E, namely a function $f(x)$ taking its values in E_x can be viewed as a function with values in R^N. We can then form its partial derivatives $\partial_\mu f$, but these need no longer take values in the sub-spaces E_x. However if we let P_x be any linear transformation on R^N varying smoothly with x and projecting onto E_x (i.e. $P_x^2 = P_x$) we can put

$$(3.1) \qquad\qquad \nabla_\mu f = P \partial_\mu f$$

and we get a covariant derivative defined on E. If E is the tangent bundle of a manifold X and P is orthogonal projection ∇_μ is just the usual covariant derivative of Riemannian geometry (given by the Levi-Civita connection).

In general (3.1) just corresponds to a $GL(n, R)$-connection or potential, but if we impose additional structures, preserved by P, we can get potentials for appropriate subgroups such as $O(n)$, $U(m)$ or $Sp(l)$ ($n = 2m$ or $n = 4l$ respectively). Thus to get $Sp(1)$ potentials we would consider quaternionic lines in H^k and use orthogonal projection.

Choosing a gauge for the bundle E will give rise to linear maps $u_x : R^n \to R^N$ whose image is just $E_x \subset R^N$. If inner products are fixed throughout so that u is an orthogonal gauge then the orthogonal projection P_x onto E_x is given by $P = uu^*$, while $u^*u = 1$. To compute the covariant derivative ∇ in the gauge u we put $f = ug$ where g is now a function on X with values in R^n and find

$$\nabla(ug) = uu^* d(ug) = u\{dg + u^*(du)g\}$$

showing that the gauge potential A is given by

$$(3.2) \qquad\qquad A = u^* du \quad \text{or} \quad A_\mu = u^* \partial_\mu u .$$

Note that u is here an $(N \times n)$ matrix of functions so that A_μ is an $(n \times n)$ matrix. If $N = n$ then (3.2) asserts that A is gauge equivalent to zero, corresponding to the fact that $E_x = R^N$ does not really depend on x. However for $N > n$ (3.2) gives interesting potentials. The formula for the gauge field is

$$(3.3) \qquad F = du^* \, du + u^* \, du \wedge u^* \, du$$

or

$$F_{\mu\nu} = \partial_\mu u^* \, \partial_\nu u - \partial_\nu u^* \, \partial_\mu u + [u^* \, \partial_\mu u, \, u^* \, \partial_\nu u] \, .$$

For many purposes it is unnecessary to pick a gauge since we can always work directly inside the larger R^N space which has a natural basis. We shall illustrate this by deriving an alternative expression for the field, expressed directly in terms of the projection operator P and the complementary projection $Q = 1 - P$. Let us take the $GL(N, R)$-potential B defined by

$$(3.4) \qquad B = Q \, dQ \, .$$

Computing the covariant derivative $\nabla_B = d + B$ on functions f which lie in E i.e. satisfying $Pf = f$ or $Qf = 0$, we see that

$$\nabla_B f = df + Q(dQ)f = df - Q^2 \, df = P \, df = \nabla f$$

where ∇ is the covariant derivative on E defined by (3.1). Thus ∇_B extends ∇ to all R^N-valued functions.

Now differentiating $Q^2 = Q$ we get $Q \, dQ + dQ \, Q = dQ$ and so

$$B^2 = Q \, dQ \wedge Q \, dQ = Q(dQ - Q \, dQ) \wedge dQ = 0 \, .$$

Hence the field F_B of ∇_B is given simply by

$$(3.5) \qquad F_B = dQ \wedge dQ \, .$$

Restricting this to E we see that the field F corresponding to the covariant derivative (3.1) is given by

$$(3.6) \qquad F = P \, dQ \wedge dQ P = P \, dP \wedge dP P \, .$$

The components $F_{\mu\nu}$ of F are here linear transformations on the image of P. If we choose an orthogonal gauge u and take $P = uu^*$ then (3.6) becomes

$$u\{u^* \, d(uu^*) \wedge d(uu^*) u\} u^*$$

and the term in brackets gives

$$du^* \wedge du + u^* du \wedge u^* du + u^* du \wedge du^* u + du^* u \wedge du^* u \,.$$

The last two terms cancel, since $u^* du = - du^* u$, and so we find again formula (3.3).

The bundle E when embedded in R^N has a complementary bundle $E^\perp$ given by the image of Q. The relationship between E and $E^\perp$ is symmetrical with the roles of P and Q being interchanged. If $v \colon R^{N-n} \to R^N$ is an orthogonal gauge for $E^\perp$ so that $Q = vv^*$ we can compute the field F of E from (3.6) by

$$(3.7) \qquad F = P\, d(vv^*) \wedge d(vv^*)\, P$$

$$= P\, dv \wedge dv^*\, P$$

the other terms dropping out since $Pv = 0$.

If v is simply a linear, but not orthogonal, gauge these formulae get modified slightly. We take the polar decomposition $v = \omega \varrho$ where $\varrho^2 = v^* v$ and $Q = \omega \omega^* = v \varrho^{-2} v^*$. Substituting in (3.6) for Q we get

$$(3.8) \qquad F = P\, dv \varrho^{-2}\, dv^*\, P$$

All these formulae hold unchanged if we replace the real numbers by complex numbers or quaternions. In the quaternion case if we want to consider all our matrix operators as left operators then we should regard H^N as a right vector space, i.e. a scalar quaternion q acts on a quaternion vector ξ by $\xi \to \xi \bar{q}$.

We now apply this general construction with $X = S^4 = P_1(H)$, the quaternion projective line. The calculations which follow are similar to those in [13] and [15]. We consider a point of $P_1(H)$ as given by homogeneous coordinates (x, y) with $x, y \in H$ and scalar multiplication on the right so that (x, y) and (xq, yq) denote the same point. Now let

$$(3.9) \qquad v(x, y) = Cx + Dy$$

be a $(k + n) \times k$ matrix of quaternions, C and D being constant matrices (independent of the scalar quaternion variables x, y). We now assume

$$(3.10) \qquad v(x, y) \text{ has maximal rank for all } (x, y) \neq (0, 0).$$

The columns of $v(x, y)$ then span a subspace of H^{n+k} having dimension k

and depending only on the ratio xy^{-1}, i.e. on the point of S^4. The orthogonal complement is then a subspace $E_{(x,y)}$ of dimension n. We now give this vector bundle E over S^4 the covariant derivative induced from H^{n+k} by orthogonal projection as in (3.1). The field or curvature can then be computed by one of the above formulae. If we restrict to $R^4 \subset S^4$ where $y \neq 0$ then we can take affine coordinates $(x, 1)$ and $v(x) = v(x, 1)$ then gives a linear gauge for $E^\perp$. Substituting for v in (3.8) gives the following expression for the field F:

$$(3.11) \qquad F = PC\, dx \varrho^{-2}\, d\bar{x} C^* P$$

where $\varrho^2 = v^* v = (\bar{x}C^* + D^*)(Cx + D)$.

If we now assume that

$$(3.12) \qquad (\bar{x}C^* + D^*)(Cx + D)$$

is a *real* matrix for all $x \in H$ then the term ϱ^{-2} in (3.11) commutes with the scalar quaternion dx and shows that F involves only the self-dual expression $dx\, d\bar{x}$. Hence we have verified that (3.9), subject to conditions (3.10) and (3.12), gives rise to a multi-instanton for the group $Sp(n)$. Note that the vector bundle E on which the self-dual potential is defined has the image of v as its orthogonal complement $E^\perp$.

It is now easy to verify that the instanton number, i.e. the topological invariant, of E is precisely k. Since this invariant is additive for direct sums it is equivalent to check that $E^\perp$ has invariant $- k$. But $E^\perp$ is, by definition, a direct sum of k quaternion line-bundles corresponding to the k basic vectors of H^k (i.e. the columns to v). Each of these line-bundles can be identified with the standard line-bundle over $S^4 = P_1(H)$ which associates to (x, y) the one-dimensional subspace of H^2 consisting of scalar multiples of (x, y). This has invariant ± 1 depending on conventions. In our case the sign must be $- 1$ so that the invariant of the self-dual bundle E becomes the positive integer k.

If we want to write down explicitly a gauge potential A corresponding to the matrix $v(x)$ we must first pick an orthogonal gauge for the bundle E, i.e. a $(k + n) \times n$ matrix $u(x)$ such that

$$(3.13) \qquad u^* v = 0 \qquad u^* u = 1\,.$$

Then $A = u^*\, du$ as in (3.2). Note that conditions (3.13) are preserved if we replace u by ug where $g(x) \in Sp(n)$ and this produces a general gauge transformation on A.

The matrices u, v can be put into a sort of normal form which gives more explicit formulae although this will introduce « apparent singularities ». To obtain these normal forms we first decompose $v(x, y)$ into blocks as follows

$$v(x,\, y) = \begin{pmatrix} C_0 x + D_0 y \\ C_1 x + D_1 y \end{pmatrix}$$

where C_0, D_0 are $n \times k$ and C_1, D_1 are $k \times k$ matrices. Since v is assumed by (3.10) to have maximal rank we may, after a change of the (x, y) variables (a conformal transformation of S^4), assume C_1 non-singular. Replacing v by RvS where S is a real $k \times k$ matrix and $R \in Sp(n + k)$ we can then take $C_1 = -I$ (I the unit $k \times k$) matrix and $C_0 = 0$.

Now putting $y = 1$ we get $v(x)$ in the form

(3.14) $$v(x) = \begin{pmatrix} A \\ B - xI \end{pmatrix}$$

where A is an $n \times k$ matrix and B is a $k \times k$ matrix, both have quaternion entries but are independent of the quaternion variable x. Condition (3.12) is equivalent to requiring both

(3.15) $A^*A + B^*B$ is real

(3.16) $B^*x + \bar{x}B$ is real for all $x \in H$.

But (3.16) is easily seen to be equivalent to requiring

(3.17) B is a symmetric matrix .

We now take u in the form $u = \begin{pmatrix} -I \\ U \end{pmatrix} \sigma$ where I is the unit $n \times n$-matrix, U is an $k \times n$-matrix and σ is a self-adjoint $n \times n$-matrix. Equations (3.13) become

(3.18) $-A + U^*(B - xI) = 0 \qquad \sigma^{-2} = I + U^*U$.

Except for singularities at points x where $B - xI$ is singular we can solve for U:

(3.19) $U^* = A(B - xI)^{-1}$.

Substituting $u = \begin{pmatrix} -I \\ U \end{pmatrix} \sigma$ in (3.2) we find the gauge potential given in

terms of U and $\sigma = (1 + U^*U)^{-\frac{1}{2}}$ by

$$(3.20) \qquad\qquad A = \sigma U^* dU \sigma + \sigma^{-1} d\sigma \; .$$

Equations (3.19) and (3.20) are precisely those given in section 2. We have thus verified that the formulae of section 2 do indeed give k-instantons. We have also explained why the singularities of section 2 are only « apparent », and due to the particular choice of gauge.

Geometrically our construction of multi-instantons can also be formulated as follows. On the Grassmannian $G_{k,n}(H)$ of n-dimensional subspaces of H^{n+k} the standard vector bundle with fibre H^n has a standard connection (induced by orthogonal projection) and our instanton connections on S^4 are induced by suitable maps $f: S^4 \to G_{k,n}(H)$. Equation (3.19) describes f explicitly in terms of appropriate coordinates in the two manifolds. The standard connection on $G_{k,n}(H)$ is automatically invariant under the group $Sp(n+k)$. In particular for $n = k = 1$ this shows that the basic instanton on S^4 is invariant under $Sp(2) \cong \mathrm{Spin}(5)$.

CHAPTER III

The Penrose Twistor Space

1. – Complex projective 3-space.

Our attack on the instanton problem will rest on complex analytic methods which are part of the general twistor theory of R. Penrose [35]. Very roughly Penrose's programme consists in re-interpreting physical space-time data in terms of corresponding data in a space of 3 complex variables. This space is the (projective) twistor space and the transformation of data can be called the twistor transform. The Penrose theory when continued to Euclidean 4-space can be developed in a slightly different way. In this section we shall describe the basic geometrical picture and comment on it from a variety of viewpoints.

As in previous sections we shall use the quaternions H and we shall identify S^4 with $P_1(H)$, the projective line over the quaternions. We shall use left scalars so that (q_1, q_2) and $(\lambda q_1, \lambda q_2)$ represent the same point of $P_1(H)$. Taking conjugates (which reverses the orientation) would convert to right scalars.

We identify the complex numbers C with the subfield of H generated by 1 and i, and H then becomes identified with C^2 by writing quaternions in the form $z_1 + z_2 j$ with $z_1, z_2 \in C$. Similarly H^2 gets identified with C^4. Now consider the complex projective 3-space $P_3(C)$, the space parametrizing complex lines (through 0) in C^4. If to each complex line we associate the quaternion line it generates, we get a map

(1.1)
$$P_3(C) \to P_1(H) .$$

In terms of homogeneous coordinates this map is given by

$$(z_1, z_2, z_3, z_4) \to (z_1 + z_2 j, z_3 + z_4 j) .$$

If we fix a quaternion line (a copy of C^2) then all complex lines in it form a copy of $P_1(C) = S^2$. Thus (1.1) is a fibre bundle with fibre $P_1(C)$.

Left multiplication by j induces a transformation σ on $P_3(C)$ which is anti-linear (i.e. anti-holomorphic in local complex coordinates) and satisfies $\sigma^2 = 1$. In homogeneous coordinates

j provides a real structure

$$(1.2) \qquad \sigma(z_1, z_2, z_3, z_4) = (-\bar{z}_2, \bar{z}_1, -\bar{z}_4, \bar{z}_3) \ .$$

Clearly σ preserves the fibration (1.1), acting trivially on $P_1(H)$, and acting as the anti-podal map on each fibre (on S^2). We shall consider σ as defining a « real structure » for $P_3(C)$ which is different from the usual real structure given by just conjugating all coordinates. This terminology is standard in algebraic geometry and means that, for some suitable algebraic embedding of $P_3(C)$ in $P_N(C)$, σ will be given by conjugating, in the usual way the co-ordinates of $P_N(C)$. As a lower-dimensional example consider a single $P_1(C)$-fibre with its anti-podal σ. This can be embedded in $P_2(C)$ as a conic with equation $\omega_1^2 + \omega_2^2 + \omega_3^2 = 0$. Note that this has no real points, corresponding to the fact that σ has no fixed points.

Although σ on $P_3(C)$ has no fixed points it does have fixed lines; these are precisely the fibres of (1.1). We call these the *real lines*. Thus S^4 appears as the parameter space of all the real lines.

Now we recall the famous Klein representation of all lines of $P_3(C)$. Given two distinct points (z_α), (ω_α) of $P_3(C)$ we introduce the Plücker coordinates

$$p_{\alpha\beta} = z_\alpha \omega_\beta - z_\beta \omega_\alpha$$

which give a skew-symmetric matrix characterizing the line joining (z) and (ω). The six homogeneous variables $p_{12}, p_{13}, p_{14}, p_{23}, p_{34}, p_{42}$ satisfy one quadratic identity

$$(1.3) \qquad p_{12}p_{34} + p_{13}p_{42} + p_{14}p_{23} = 0$$

(this expression is the square root of $\det(p_{\alpha\beta})$). Thus the parameter space of all lines in $P_3(C)$ is the complex 4-dimensional quadric $Q_4 \subset P_5(C)$ given by equation (1.3). A real structure on $P_3(C)$ induces a real structure on Q_4. For the standard real structure on $P_3(C)$ the real structure of Q_4 is given by conjugating the $p_{\alpha\beta}$ and so Q_4 has (1.3) as its real equation: this is a quadratic form of type $(3, 3)$, i.e. having 3 plus signs and 3 minus signs in its diagonalization. For our real structure given by σ on $P_3(C)$ we get a different real form of Q_4 corresponding to the signature $(5, 1)$. To verify this note that (1.2) has the following effect on the $p_{\alpha\beta}$

$$p_{12} \to -\overline{p_{12}}, \qquad p_{34} \to -\overline{p_{34}}, \qquad p_{13} \to \overline{p_{24}}, \qquad p_{14} \to -\overline{p_{23}} \ .$$

Hence the six quantities

$$X_1 = ip_{12}, \qquad X_2 = ip_{34}, \qquad X_3 = p_{13} + p_{24},$$

$$X_4 = i(p_{13} - p_{24}), \qquad X_5 = i(p_{14} + p_{23}), \qquad X_6 = p_{14} - p_{23}$$

are conjugated normally by (1.2). Rewriting equation (1.3) in terms of the coordinates $X_1, ..., X_6$ we get

(1.4) $$4X_1X_2 + X_3^2 + X_4^2 + X_5^2 + X_6^2 = 0$$

confirming that the signature is $(5, 1)$. Moreover the real points of (1.4), representing S^4, are indeed given by the affine equation (taking $X_1 - X_2 = 1$)

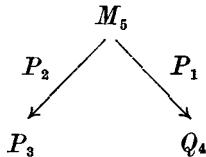$$X_3^2 + X_4^2 + X_5^2 + X_6^2 + (X_1 + X_2)^2 = 1 .$$

Reverting to the complex geometry of the Klein representation we recall that there are two families of projective planes lying on Q_4. If the equation of Q_4 is written

$$Y_1^2 + Y_2^2 + Y_3^2 = Z_1^2 + Z_2^2 + Z_3^2$$

then the planes are given by the equations $Y = AZ$ where $A \in O(3, C)$. The two families depend on the sign of $\det A$. One family corresponds to points of $P_3(C)$ and the other to planes of $P_3(C)$ or equivalently to points of the dual $P_3(C)$. This means that as a line l in $P_3(C)$ varies through a given point A its representative point $L \in Q_4$ varies in a plane α on Q_4. Similarly if l varies in a plane B its representative L varies in a plane β on Q_4, β being of the opposite family to α.

With our real structure σ every α-plane contains a unique real point, namely the intersection $\alpha \cap \sigma(\alpha)$. This corresponds to the real line in $P_3(C)$ joining A to $\sigma(A)$, and so describes once more the map $P_3(C) \to S^4$.

All this can be conveniently summarized by introducing the correspondence space $M_5 \subset P_3 \times Q_4$ consisting of « incident » pairs (A, L), i.e. A lies on the line l (or equivalently L lies in the plane α). We then have two fibre maps

$$
\begin{array}{ccc}
 & M_5 & \\
P_2 \swarrow & & \searrow P_1 \\
P_3 & & Q_4
\end{array}
$$

with fibres P_2, P_1 as indicated. All structures here are complex algebraic. If we now pick our real structure σ then this defines

(i) the real subspace S^4 of Q_4

(ii) a section $s: P_3 \to M_5$ which picks out the unique point in the P_2-fibre which is real (regarded as a subspace of Q_4).

Thus the inverse image of S^4 in M_5 gets identified via (ii) with $P_3(C)$ and so we recover our basic fibration $P_3(C) \to S^4$. Note that whereas the projection $M_5 \to P_3$ is complex analytic the section $s: P_3 \to M_5$ is not.

2. – Lie groups.

The Klein representation can also be looked at from the point of view of Lie groups. The group $SL(4, C)$, of complex 4×4 matrices of determinant 1, acts on $P_3(C)$ by projective transformations, and hence it acts on the space of lines i.e. the quadric Q_4. But the group of projective transformations in P_5 keeping Q_4 fixed is just the complex orthogonal group $O(6, C)$ modulo $\{\pm 1\}$ and so we get a homomorphism

$$(2.1) \qquad\qquad SL(4, C) \to O(6, C)/\{\pm 1\}$$

with kernel the fourth roots of unity. Since these two groups have the same dimension it follows that (2.1) is a local isomorphism or that

$$(2.2) \qquad\qquad SL(4, C) \simeq \text{Spin } (6, C) \ .$$

This is one of the coincidences that happen in low dimensions between the various classical groups.

If we start from the group Spin $(6, C)$ then the isomorphism (2.2) implies that Spin $(6, C)$ has a representation on C^4. This is one of the half-spin representations. The other half-spin representation is on the dual C^4.

These complex Lie groups have many real forms, and the local isomorphism (2.1) leads in particular to the following local isomorphisms of real Lie groups:

(i) $SL(4, R) \sim SO(3, 3)$

(ii) $SU(2, 2) \sim SO(4, 2)$

(iii) $SL(2, H) \sim SO(5, 1)$

(iv) $SU(4) \sim SO(6)$.

Case (ii) arises from Minkowski space and is the one mainly studied in the Penrose theory. Case (iii) arises from Euclidean space and is the one that concerns us. The action of $SL(2, H)$ on $P_3(C)$ preserves the fibration (1.1) and induces the conformal group action on S^4. In other words the conformal group of S^4 acts naturally on the fibration (1.1).

Taking maximal compact subgroups of (iii) we get the local isomorphism

$$Sp(2) \sim SO(5) \ .$$

Thus $Sp(2)$ is the group of automorphisms of the fibration (1.1) which preserves in addition the natural metrics on $P_3(C)$ and on S^4. Since $Sp(2)$ acts transitively on $P_3(C)$ we can express $P_3(C)$ as a homogeneous (coset) space:

$$P_3(C) = \frac{Sp(2)}{U(1) \times Sp(1)} \ .$$

Since

$$S^4 = \frac{SO(5)}{SO(4)} = \frac{Sp(2)}{Sp(1) \times Sp(1)}$$

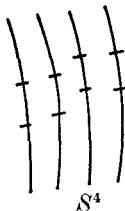we see that the fibration (1.1), expressed in terms of compact Lie groups becomes

$$(2.3) \qquad \frac{Sp(2)}{U(1) \times Sp(1)} \rightarrow \frac{Sp(2)}{Sp(1) \times Sp(1)}$$

with fibre $Sp(1)/U(1) = S^2$. This description shows clearly that, starting from S^4 we have essentially two different choices to produce $P_3(C)$ depending on which $Sp(1)$ factor we pick. These two choices switch when we reverse orientation on S^4 and lead to dualizing $P_3(C)$.

The fibration (2.3) is also related to the basic $Sp(1)$-instanton on S^4. As explained in Chapter II the basic instanton can be thought of as the quaternion line-bundle over $P_1(H)$ with its connection induced from the fixed space H^2. An equivalent description is to say that the principal $Sp(1)$-bundle of the instanton is the fibration

$$(2.4) \qquad \frac{Sp(2)}{Sp(1)} \rightarrow \frac{Sp(2)}{Sp(1) \times Sp(1)} \ .$$

The connection can now be described in terms of horizontals to the fibres

If we give all spaces their natural metrics, inherited from the bi-invariant metric of the compact group $Sp(2)$, we can choose the orthogonals to the fibres as horizontals. This gives a connection admitting $Sp(2) \sim SO(5)$ as symmetry group. Moreover this is the *unique* choice that has this property because any invariant choice of horizontals must, at each point, be invariant under the action of the isotropy group $Sp(1)$, which acts by inequivalent representations in the vertical (fibre) and horizontal directions (vertically we have the adjoint representation of $Sp(1)$ and horizontally we have the C^2 representation of $Sp(1) \cong SU(2)$). Since the instanton connection admits $SO(5)$ as symmetry group it follows that our connection must be the instanton or anti-instanton, depending on which $Sp(1)$ factor we choose on the left hand side of (2.4).

Thus $P_3(C)$ is naturally the quotient of the principal bundle of the $Sp(1)$-instanton by the action of $U(1)$. Equivalently $P_3(C)$ can be obtained from the instanton, considered as $H = C^2$ bundle over S^4, by replacing each fibre by the corresponding projective space $P_1(C)$. This way of obtaining $P_3(C)$ from S^4 can be generalized to other 4-manifolds as explained in [3].

3. – Complex coordinates in R^4.

The introduction of the Penrose space $P_3(C)$ as a tool to study problems on S^4 (or R^4) can be motivated in the following way. It is very well known that many classical problems in 2 real variables (x, y) are best solved by introducing the single complex variable $z = x + iy$, and then using the powerful methods of holomorphic function theory. When we pass from R^2 to R^4 one might naively try a similar device by introducing two complex variables

$$z_1 = x_1 + ix_2 , \qquad z_2 = x_3 + ix_4 .$$

There is one immediate objection to such a method, namely that it is too closely tied to the particular coordinate system and therefore is unlikely

to give significant results. For example a permutation of the four coordinates x_1, x_2, x_3, x_4 would lead to new complex variables not well related to the first choice. In R^2 on the other hand provided we have fixed the metric and orientation the complex structure is unambiguous. The complex number i is given by rotation through $\pi/2$ in the positive sense.
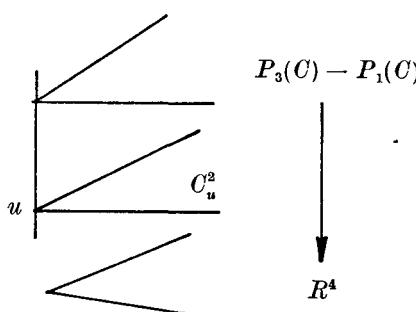
Since there is no natural choice of complex structure on R^4 we can try to consider simultaneously *all* choices which are compatible with the metric and orientation. Effectively we have to define i as a proper orthogonal transformation with $i^2 = -1$. If we have made one such choice then transforming (conjugating) by elements of $SO(4)$ will produce all other choices. Moreover transforming by elements of $U(2)$ leaves the first choice unaltered. Hence the set of all complex structures is naturally parametrized by the coset space

$$\frac{SO(4)}{U(2)} = \frac{SU(2) \times SU(2)}{U(1) \times SU(2)} = S^2 .$$

Thus to each $u \in S^2$ we have a complex structure on R^4, namely an isomorphism $R^4 \cong C^2$.

If we now want to introduce complex variable methods in R^4 we need 3 complex variables (u, z_1^u, z_2^u): the first variable u tells us which complex structure to use and the next two are the complex coordinates themselves. Since the coordinates z_1, z_2 depend on u the situation is a little delicate. The Penrose picture clarifies this geometrically as we shall now explain.

For simplicity we shall now work only over $R^4 \subset S^4$ and so we remove the « point at ∞ » in S^4 and correspondingly we remove the projective « line at ∞ » in P_3 that lies over it in the fibration (1.1). Then we get a fibration $P_3(C) - P_1(C) \to R^4$. Projective planes in $P_3(C)$ which meet in the « line at ∞ » become « parallel » affine planes in $P_3(C) - P_1(C)$. Thus we have the following picture for our fibration:

Over a given point (say the origin) of R^4 we have the fibre $P_1(C) = S^2$ parametrized by u. The plane of our parallel system through u is a copy of C^2 and under the projection gets identified with R^4. This is the way in which R^4 gets the complex structure corresponding to u. The fact that this complex structure is changing with u means that the vertical identification between the different C_u^2 does not preserve the complex structure. This can be expressed differently by noting that the «horizontal» projection $P_3(C) - P_1(C) \rightarrow$ $\rightarrow P_1(C)$ mapping C_u^2 into u is that of a complex vector bundle which is not isomorphic to the product $P_1(C) \times C^2$, although the underlying real vector bundle is isomorphic to $P_1(C) \times R^4$. This can happen because bundles over $S^2 = P_1(C)$ are topologically classified by maps of the equatorial S^1 into the group of the bundle and the fundamental groups π_1 in our case are:

$$\pi_1(U(2)) \cong Z \text{ (integers)} \quad \pi_1(SO(4)) \cong Z_2 \text{ (integers mod 2)}$$

and our bundle corresponds to the integer 2 which gives zero in $\pi_1(SO(4))$.

This topological fact therefore lies behind the « linkage » between the complex coordinates z_1^u, z_2^u and the complex parameter u. Working however in the space $P_3(C) - P_1(C)$ we can locally introduce three independent (« unlinked ») complex variables and use these instead of (u, z_1^u, z_2^u). Thus the Penrose space enables us to introduce 3 complex variables in a natural way to study problems in R^4.

The same picture applies infinitesimally for the complete fibration $P_3(C) \rightarrow S^4$. At each point $u \in P_3(C)$ we can consider the tangent space T_u and the subspace L_u of tangents to the fibre through u. The quotient space T_u/L_u is then a complex vector space which by projection gets naturally identified with the real tangent space to S^4 at the point below u.

Note in particular that, since complex manifolds have a natural orientation, S^4 inherits a natural orientation from its description as the base of the fibration $P_3(C) \rightarrow S^4$. This is the orientation which we fix (the opposite convention is adopted in [3]).

CHAPTER IV

Holomorphic Bundles

1. – Holomorphic and unitary gauges.

In Chapter I we explained that gauge theories had a direct differential-geometric interpretation in terms of fibre bundles with connection. In this chapter we shall encounter holomorphic fibre bundles and since these are unfamiliar to physicists and have somewhat different features we begin with some elementary remarks designed to explain the essential points.

Let us begin by considering gauge theory for the non-compact group $GL(n, C)$. Geometrically it will be convenient to consider the representation on C^n and discuss complex vector bundles. Thus over our base space X (e.g. S^4) we consider a vector bundle E with fibre C^n, the fibre at $x \in X$ being denoted by E_x. Intuitively we consider E_x as a vector space varying continuously with x. A linear gauge for E means a choice of basis of E_x varying continuously with x. A gauge transformation is then the function $g(x) \in GL(n, C)$ which provides the change of basis at each point x.

In the definition of a vector bundle it is always assumed that local gauges always exist, so that E is locally isomorphic to the product $X \times C^n$. A global gauge does not necessarily exist since E need not be globally isomorphic to the product. The instanton bundles already provide examples of this situation with $X = S^4$ and $n = 2$.

If we assume that each fibre E_x has a positive inner product, varying continuously with x, then we can consider unitary gauges in which the basis at each x consists of an orthonormal base. A change from one unitary gauge to another is then described by a gauge transformation $g(x) \in U(n)$. Local unitary gauges always exist and global unitary gauges exist if and only if a global linear gauge exists, since the topological properties of $GL(n, C)$ are essentially carried by $U(n)$.
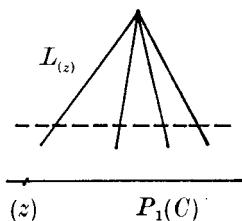
Instead of fixing the inner product and defining unitary gauges in terms of it we can reverse the procedure by choosing one linear gauge, decreeing

that this is orthonormal, and considering all gauges obtained by applying unitary gauge transformations.

If X is a differentiable manifold then it is natural to require that the word continuous be replaced throughout by (sufficiently) differentiable. This introduces no essential differences.

If X is replaced by a complex analytic manifold Z, for example $P_3(C)$, then we can introduce the notion of a holomorphic, or complex analytic, vector bundle. Naively we now think of the fibres E_z as varying holomorphically with $z \in Z$. A holomorphic gauge is then a basis in each E_z varying holomorphically with z and a holomorphic gauge transformation is given by a holomorphic function $g(z)$ with values in $GL(n, C)$. The holomorphic structure can be defined by fixing one gauge, decreeing this to be holomorphic, and then allowing as new holomorphic gauges only those obtained by applying holomorphic gauge transformations.

The analogy and difference between the holomorphic and unitary case is fairly clear. The important point to note is that unitary gauge transformations are defined by a point-wise restriction on values whereas holomorphic gauge transformations cannot be defined this way. Note that the two types of gauge transformations are diametrically opposed since a transformation which is both unitary and holomorphic is necessarily constant. To illustrate these basic ideas let us consider a simple example. Take as base space $P_1(C)$ and associate to each point $(z) \in P_1(C)$ the complex line $L_{(z)} \subset C^2$ which that point parametrizes, namely all muntiples λz.

It is clear intuitively that $L_{(z)}$ varies holomorphically with (z). A natural holomorphic gauge is given by intersecting $L_{(z)}$ with any affine line in C^2 (not through 0) (see dotted line in figure): for example the line $z_2 = 1$. This gauge is well-defined except at the point $(1, 0) \in P_1(C)$. Similarly $z_1 = 1$ gives a gauge except at $(0, 1)$. The gauge transformation between these two is given by the function z_1/z_2 and this is of course a holomorphic function on $P_1(C)$ outside the two points $(1, 0)$ and $(0, 1)$.

If we give C^2 its natural inner product each fibre $L_{(z)}$ inherits an inner

product and we can introduce unitary gauges. Note again that the linear holomorphic gauges above are definitely not unitary.

If we associate to each (z) the orthogonal complement $L^{\perp}_{(z)}$ we again get a vector bundle with unitary structure. However the process of taking orthogonal complements involves complex conjugation and so the vector spaces $L^{\perp}_{(z)}$ do *not* vary holomorphically with (z). Thus L is a holomorphic vector bundle but $L^{\perp}$ is not. This is a major difference between the unitary and holomorphic theories.

Although taking orthogonal complements is not a holomorphic process we can instead form the quotient space $N_{(z)} = C^2/L_{(z)}$. This does again form a holomorphic line-bundle N, a holomorphic gauge coming from any holomorphic function on $P_1(C)$ with values in C^2. If we take linear duals the bundle N' appears as a sub-bundle of $P_1(C) \times (C^2)'$. In general if E is any holomorphic vector bundle, its linear dual E' is again a holomorphic vector bundle.

If we replace $P_1(C)$ by $P_n(C)$ then again we get a holomorphic line bundle L. The quotient C^{n+1}/L is now a holomorphic C^n-bundle over $P_n(C)$.

So far we have just holomorphic or unitary bundles with no further structure. Now we come to the question of connections. In the case of a linear vector bundle E with group $GL(n, C)$ a connection can be given by the covariant derivative ∇. The components ∇_μ, relative to coordinates x_μ in the base manifold, act on sections of the vector bundle, i.e. functions $f(x) \in E_x$. In a given linear gauge

$$\nabla f = df + Af ,$$

where $A = \sum A_\mu dx^\mu$ is the gauge potential.

If E has a unitary structure and we require ∇ to be compatible with unitarity (i.e. the corresponding parallel transport preserves length) then in any unitary gauge $A^* = -A$.

Before proceeding to the holomorphic case let us review a few elementary definitions used in complex manifold theory. If $(z_1 \ldots z_n)$ are local complex coordinates on Z we introduce the formal differentials dz_α, $d\bar{z}_\alpha$ defined by

$$dz_\alpha = dx_\alpha + i\,dy_\alpha , \quad d\bar{z}_\alpha = dx_\alpha - i\,dy_\alpha$$

where $z_\alpha = x_\alpha + iy_\alpha$. The total differential

$$df = \sum_{\alpha=1}^{n} \left(\frac{\partial f}{\partial x_\alpha}\, dx_\alpha + \frac{\partial f}{\partial y_\alpha}\, dy_\alpha \right)$$

can be decomposed into two parts

$$df = d'f + d''f$$

where $d'f$ involves only the dz_α and $d''f$ involves only the $d\bar{z}_\alpha$. This decomposition is independent of the choice of local holomorphic coordinates and the equation $d''f = 0$ is the Cauchy-Riemann equation, characterizing holomorphic functions.

If now E is a holomorphic vector bundle over Z and ∇ is a covariant derivative, we can write

$$\nabla = \nabla' + \nabla''.$$

If we put

$$\nabla = \sum (\nabla_\alpha dx_\alpha + \widetilde{\nabla}_\alpha dy_\alpha)$$

then $\nabla' = \sum \nabla'_\alpha dz_\alpha$, $\nabla'' = \sum \nabla''_\alpha d\bar{z}_\alpha$ where

$$\nabla'_\alpha = \tfrac{1}{2}(\nabla_\alpha - i\widetilde{\nabla}_\alpha)$$
$$\nabla''_\alpha = \tfrac{1}{2}(\nabla_\alpha + i\widetilde{\nabla}_\alpha).$$

We shall say that ∇ is compatible with the holomorphic structure of E if $\nabla''f = 0$ for every holomorphic section f of E. This is equivalent to requiring that, in any holomorphic gauge, the gauge potential A, when decomposed as $A = A' + A''$, has $A'' = 0$. Such a potential is said to be of type $(1, 0)$. In general a differential form is said to be of type (p, q) if it involves p of the dz_α and q of the $d\bar{z}_\alpha$.

The link between connections on unitary and holomorphic bundles is provided by the following simple and well-known result:

PROPOSITION (1.1). *Let E be a holomorphic vector bundle with a unitary structure. Then there is a unique connection compatible with both structures, i.e. such that*

(i) *in every unitary gauge the gauge potential A satisfies $A^* = -A$,*

(ii) *in every holomorphic gauge $A'' = 0$.*

The curvature F of this unique connection is of type $(1, 1)$.

PROOF. – First pick a holomorphic gauge and choose $A'' = 0$ to satisfy (ii). Now transform to a unitary gauge by the gauge transformation g. If the transformed potential is $B = B' + B''$, then $B'' = d''g \cdot g^{-1} + A'' = d''g \cdot g^{-1}$

is determined by g. Finally to satisfy (i) we must take $B' = -(B'')^*$. This uniquely determines B and hence fixes the potential. Computing the field F in the holomorphic gauge we have

$$F = dA' + [A', A'] = \{d'A' + [A', A']\} + d''A'$$

exhibiting only terms of type $(2, 0)$ and $(1, 1)$. But unitarity implies that, in any gauge, $F^* = -F$, and so the component of type $(2, 0)$ is also zero. Thus F is of type $(1, 1)$.

Proposition (1.1) has an important converse which can be stated as follows:

THEOREM (1.2). *Let E be a complex vector bundle, with a unitary structure, over a complex manifold. Let E have a unitary connection whose curvature is of type $(1, 1)$. Then there is a unique holomorphic structure on E such that the connection is that given by* (1.1).

This theorem is essentially a consequence of the Newlander-Nirenberg integrability theorem for complex structures [33]. The basic idea is that the holomorphic structure of E is defined by taking the solutions of the equation $\nabla''f = 0$ as its holomorphic sections. Here ∇'' denotes the $(0, 1)$ component of the covariant derivative of the connection. The condition on the curvature implies that $[\nabla''_\alpha, \nabla''_\beta] = 0$, where $\nabla'' = \sum_\alpha \nabla''_\alpha d\bar{z}_\alpha$, and the z_α are local complex coordinates on our manifold. The integrability theorem of [33] then implies that the equation $\nabla''f = 0$ has locally enough solutions to provide a basis for E. For further details we refer to [3]: see also [25].

2. – Twistor interpretation of instantons.

In this section we shall show how to interpret the self-duality equations for a Yang-Mills field on S^4 in terms of complex analysis on the twistor space $P_3(C)$.

As an algebraic preliminary we shall need to understand the significance of the equations $^*\omega = \pm \omega$ for a 2-form ω on R^4 in terms of complex coordinates. We recall that once we have introduced complex coordinates, identifying R^4 with C^2, a 2-form ω can be expressed in terms of its type decomposition:

(2.1) $$\omega = \omega^{2,0} + \omega^{1,1} + \omega^{0,2}$$

We ask how this decomposition corresponds to the *-decomposition

$$(2.2) \qquad \omega = \omega^+ + \omega^-$$

according to the eigenvalues ± 1 of *. This is just a question of linear algebra and it can be viewed in terms of group representations. Equation (2.2) corresponds to decomposing a representation of $SO(4)$ into two irreducible pieces (each of dimension 3), while (2.1) corresponds to decomposing this representation under the subgroup $U(2)$. The $(2, 0)$ and $(0, 2)$ components have dimension one while the 4-dimensional representation in type $(1, 1)$ has a further decomposition:

$$(2.3) \qquad \omega^{1,1} = \omega_0^{1,1} + \alpha \ .$$

Here α is a multiple of the $(1, 1)$ form μ corresponding to the hermitian metric and $\omega_0^{1,1}$ is the « primitive » part, orthogonal to μ. The 3-dimensional representation of $U(2)$ on this primitive part is easily seen to be irreducible and so this must coincide with one of the two irreducible pieces of (2.2). But the form μ corresponding to the metric is self-dual. Hence we must have

$$(2.4) \qquad \omega^- = \omega_0^{1,1} \ .$$

In particular this shows that the space Ω^- of ω with $*\omega = - \omega$ is of type $(1, 1)$ for all complex structures (compatible with metric and orientation). The converse is also true, because the space

$$V = \bigcap_u \Omega_u^{1,1}$$

of 2-forms on R^4 which are of type $(1, 1)$ for all complex structures u is invariant under $SO(4)$, contains Ω^- but is not the whole space: it must therefor coincide with Ω^-.

Thus we have established the following algebraic lemma.

LEMMA (2.5). *A 2-form ω on R^4 is anti-self-dual if and only if it is of type $(1, 1)$ for all compatible complex structures.*

We shall now apply this lemma to a 2-form ω on S^4 which we lift to obtain a 2-form $\tilde{\omega}$ on $P_3(C)$. This form $\tilde{\omega}$ is purely horizontal, i.e. $\tilde{\omega}_{\alpha\beta} = 0$ if α or β is a fibre direction. To compute the horizontal part of $\tilde{\omega}$ at a point u we note the interpretation, explained in the preceding section, according

to which u parametrizes (infinitesimally) complex structures on S^4 at the point below. From this and our Lemma we deduce

PROPOSITION (2.6). *A 2-form ω on S^4 is anti-self-dual if and only if its lift $\tilde{\omega}$ to $P_3(C)$ is of type $(1,1)$.*

Note that this proposition is a purely local one, valid for any open set U in S^4 and its counterpart $\tilde{U}$ in $P_3(C)$.

Finally we consider a complex vector bundle E on S^4 with a unitary structure and connection. Let F be its curvature. If we lift E to give a bundle $\tilde{E}$ with connection on $P_3(C)$ its curvature is just the lift $\tilde{F}$ of F. Applying (2.6) to the matrix coefficients of F we deduce

PROPOSITION (2.7). *A vector bundle E on S^4 with unitary structure and connection has anti-self-dual curvature if and only if the lifted bundle $\tilde{E}$ on $P_3(C)$ with the lifted connection has curvature of type $(1,1)$.*

Using (2.7) and theorem (1.2) we deduce the important result

$$(2.8) \qquad\qquad E \text{ anti-self-dual} \to \tilde{E} \text{ holomorphic} .$$

We plan now to make this statement more precise by specifying which holomorphic bundles on $P_3(C)$ arise in this way and how one gets back from $\tilde{E}$ to E.

Note first of all that, restricted to any fibre P_x of $P_3(C) \to S^4$, $\tilde{E}$ is holomorphically trivial, a basis of E_x giving rise to a holomorphic basis or gauge of $\tilde{E}|P_x$. Conversely this shows that E_x can be uniquely defined as the space of holomorphic sections of $\tilde{E}|P_x$.

We turn next to consider the unitary structures on E and $\tilde{E}$. The unitary structure on E can be given by an anti-linear isomorphism $\tau \colon E \to E^*$ such that $(u, \tau v)$ is a positive hermitian form (E^* denotes the dual of E). Passing to $\tilde{E}$ we use τ to define a lifting $\tilde{\tau}$ of the conjugation σ on $P_3(C)$, namely we define a commutative diagram

$$\begin{array}{ccc} \tilde{E} & \xrightarrow{\tilde{\tau}} & \tilde{E}* \\ \downarrow & & \downarrow \\ P & \xrightarrow{\sigma} & P \end{array} .$$

The map $\tilde{\tau}$ is anti-holomorphic, i.e. if we give $\tilde{E}*$ its opposite complex structures it becomes a holomorphic isomorphism. This follows from the fact that $\tilde{\tau}$ preserves the unitary structure and connection and hence (by the uniqueness part of Theorem (1.2)) it preserves the holomorphic structures.

Restricting to the fibre P_x our map $\tilde{\tau}$ induces a similar map on holomorphic sections of $\tilde{E}$ and in this way we recover τ_x. Lifting back this gives the unitary structure of $\tilde{E}$ which together with the holomorphic structure yields, by (1.1), a unique connection of type $(1, 1)$. Our final step is to show that this connection descends to a connection on E: by (2.7) this will necessarily be anti-self-dual. The condition that the connection on $\tilde{E}$ descend from $P_3(C)$ to S^4 is that the curvature should be purely horizontal, i.e. that $F_{\alpha\beta} = 0$ if α is a vertical direction (i.e. along the fibres). If α and β are both vertical this is clear because, restricted to a fibre, $\tilde{E}$ is trivial holomorphically and unitarily. The stronger statement, when only α is vertical, follows from the triviality of $\tilde{E}$ restricted to the *first formal neighbourhood* of a fibre. More concretely this triviality means that we can pick a gauge of $\tilde{E}$, near a fixed fibre P_x, which is holomorphic and unitary on P_x and up to first derivatives normal to P_x. The verification of this triviality property of $\tilde{E}$ is best postponed until later (see Chapter VI, § 3) since it involves complex analytical machinery. We note in passing that $\tilde{E}$ is not trivial up to second derivatives unless the whole curvature vanishes.

To summarize our results in a convenient form we shall make the following definitions. Let V be a holomorphic vector bundle over $P_3(C)$, then an anti-linear isomorphism $p\colon V \to V^*$ covering σ on $P_3(C)$ such that

$$(u, pv) = \overline{(v, pu)} \qquad v \in V_{(z)}, \quad u \in V_{\sigma(z)}$$

will be called a *real form* on V. If V is further assumed to be trivial on all real lines of $P_3(C)$ then p induces a non-degenerate hermitian form on the space of holomorphic sections of V restricted to any real line. If this hermitian form is positive we say that our real form is *positive*. Two vector bundles V, W with real forms are called isomorphic if there is a complex analytic isomorphism from V to W commuting with p.

Our conclusion can now be stated as follows:

THEOREM (2.9). *There is a natural* (1-1) *correspondence between*

(i) *anti-self-dual* $U(n)$-*potentials over* S^4 *(up to gauge equivalence) and*

(ii) *holomorphic vector bundles with fibre* C^n *over* $P_3(C)$ *with a positive real form (up to isomorphism).*

REMARKS.

1) This theorem is purely local in character. It holds for any open set U of S^4 and its counterpart $\tilde{U}$ on $P_3(C)$.

2) In a purely complex form this theorem is originally due to R. S Ward [42]. We shall later discuss Ward's proof. The present version is stated briefly in [4] and elaborated in [3].

3) It should be emphasized that in (ii) most of the information is already contained in the holomorphic vector bundle. The positive real form when it exists is generally unique.

4) Both in (i) and (ii) there is an integer topological invariant. In (i) it is the anti-instanton number, while in (ii) it is the second Chern class. These integers are equal (cf. [3]).

Theorem (2.9) can easily be generalized to the orthogonal and symplectic groups. We shall consider the symplectic case in detail, the orthogonal case is quite similar except for sign changes.

Recall that the compact symplectic group $Sp(n)$ is the group of norm-preserving automorphisms of the quaternion vector space H^n. It can be identified with the subgroup of $U(2n)$ commuting with the action of the quaternion j on $H^n = C^{2n}$. Alternatively it is the sugbroup of $U(2n)$ commuting with the skew bilinear form (,) on C^{2n} defined by $(u, jv) = \langle u, v \rangle$ where $\langle , \rangle$ is the hermitian inner product. Hence an $Sp(n)$-potential can be represented geometrically by a vector bundle E with fibre C^{2n} having a unitary connection, together with an isomorphism $\alpha \colon E \to E^*$ which is skew i.e. $(u, \alpha v) = - (v, \alpha u)$ and preserves connection (E^* being endowed with the connection inherited from E by duality).

Suppose now we have an $Sp(n)$-potential on S^4 which is anti-self-dual. Then using (2.9) for $U(2n)$ we see first of all that $\tilde{E}$ on $P_3(C)$ is holomorphic. Moreover the isomorphism $\alpha \colon E \to E^*$ induces a holomorphic isomorphism $\tilde{\alpha} \colon \tilde{E} \to \tilde{E}^*$ which is also skew. Combining $\tilde{\alpha}$ with the anti-linear isomorphism $\sigma \colon \tilde{E} \to \tilde{E}^*$ given by (2.9) we obtain an anti-linear isomorphism $\tilde{E} \to \tilde{E}$ covering σ on $P_3(C)$. We shall denote this map on $\tilde{E}$ also by σ: it satisfies $\sigma^2 = - 1$, and is compatible with $\tilde{\alpha}$ or equivalently with the skew form on $\tilde{E}$ defined by $\tilde{\alpha}$. Thus an anti-self-dual $Sp(n)$-potential on S^4 corresponds to a holomorphic vector bundle $\tilde{E}$, with fibre C^{2n}, over $P_3(C)$ which has two further structures:

(i) a holomorphic non-degenerate skew form on $\tilde{E}$

(ii) an anti-linear map $\sigma \colon \tilde{E} \to \tilde{E}$ lifting σ on $P_3(C)$ such that $\sigma^2 = - 1$ and compatible with the skew form i.e. $(\sigma u, \sigma v) = \overline{(u, v)}$.

Moreover the bundle $\tilde{E}$ is holomorphically trivial on all real lines of $P_3(C)$ and the hermitian form induced by $(u, \sigma v)$ on sections of $\tilde{E}$, restricted to a real line, is positive definite.

In the special case of $Sp(1) \cong SU(2)$ condition (i) reduces to a topological constraint, namely that the first Chern class $c_1(\tilde{E})$ should vanish. The non-degenerate skew form is then unique up to a constant factor. Moreover the anti-linear map $\sigma: \tilde{E} \to \tilde{E}$ is unique (unless $k = 0$) because two such σ differ by a holomorphic automorphism of $\tilde{E}$ and (as we shall see later) $\tilde{E}$ has no automorphisms except scalars: the condition $\sigma^2 = -1$ and positivity of the hermitian form then uniquely fix σ. We thus recover the version of the $SU(2)$ theorem given in [4].

3. – Bundles over $P_1(C)$.

In order to familiarize ourselves with holomorphic vector bundles we shall now describe the simplest case, namely over the complex projective line.

The basic example of a complex line bundle (fibre C^1) is, as we have seen, given by the family $L_{(z)}$ of complex lines in C^2, parametrized by the corresponding point $(z) \in P_1(C)$. This standard bundle is topologically non-trivial, and can be described by two holomorphic gauges near $z = 0$ and $z = \infty$ on P_1 (z being a non-homogeneous coordinate), with the gauge transformation from one to the other being multiplication by z. More generally the bundle $L^n = L \otimes \ldots \otimes L$ (n times) is similarly given by the gauge transformation z^n. Taking $L^{-1} = L^*$ (the dual) we can also allow n to be a negative integer.

One can now ask if there are any further holomorphic line bundles not isomorphic to some L^n. Suppose for example we construct a bundle by using as a gauge transformation any function $f(z)$ which is holomorphic and non-zero near the equator $|z| = 1$. First of all we can define the topological invariant

$$k = \frac{1}{2\pi i} \int \frac{f'(z)}{f(z)} \, dz$$

the integral taken over $|z| = 1$. Suppose first $k = 0$, then we have a well-defined holomorphic function

$$g(z) = \log f(z)$$

in some annulus $r < |z| < R$. Take the Laurent expansion

$$g(z) = \sum_{-\infty}^{\infty} a_n z^n$$

and write it in the form

$$g(z) = g_0(z) - g_\infty(z)$$

where g_0 is holomorphic in $|z| < R$ and g_∞ is holomorphic in $r < |z|$, i.e. g_0 involves terms with $n \leqslant 0$ and g_∞ those with $n \geqslant 0$. This decomposition is therefore unique up to additive constants. Exponentiating we get

$$f(z) = f_0(z) f_\infty(z)^{-1}$$

which we can interpret as saying that the holomorphic bundle given by $f(z)$ is holomorphically trivial. The terms $f_0(z)$ and $f_\infty(z)$ allow us to change our original holomorphic gauges so that they now coincide near the equator and so provide a global holomorphic gauge. Thus if $k = 0$ the bundle is holomorphically trivial and a similar argument tells us in general that

$$(3.1) \qquad\qquad f(z) = f_0(z) z^k f_\infty(z)^{-1}$$

and so the bundle is holomorphically equivalent to L^k. A slight generalization of this argument, allowing more than two local gauges, leads to the conclusion that every holomorphic line-bundle on $P_1(C)$ is isomorphic to some L^k.

Passing now to vector bundles with fibre C^n we can construct obvious examples as direct sums:

$$(3.2) \qquad\qquad E = L^{k_1} \oplus L^{k_2} \oplus ... \oplus L^{k_n} .$$

A much subtler theorem, proved in various versions over the years by Hilbert, G. D. Birkhoff, Grothendieck [26] and others asserts that every holomorphic vector bundle over $P_1(C)$ is isomorphic to such a direct sum of L^{k_i} and the integers $(k_1, ..., k_n)$ are unique up to permutation. In more concrete terms this implies in particular a matrix version of (3.1), namely if $f(z)$ is a holomorphic function defined in the annulus $r < |z| < R$, taking values in $GL(n, C)$, then we can write $f(z)$ in the form

$$(3.3) \qquad\qquad f(z) = f_0(z) \Lambda(z) f_\infty(z)^{-1}$$

where $\Lambda(z)$ is the diagonal matrix with entries z^{k_i} and $f_0(z), f_\infty(z)$ are holomorphic functions with values in $GL(n, C)$ defined for $|z| < R$ and $r < |z|$ respectively.

It is important to note that although the exponents k_i in (3.2) are holomorphic invariants of the bundle E they are *not* topological invariants, only

the sum $k = \sum k_i$ is a topological invariant. Again, in more concrete terms, if in (3.3) $f(z)$ is holomorphic in z and depends continuously on some other parameter t, then the factorization (3.3) cannot in general be made continuous in t. The integers k_i will depend on t in a semi-continuous fashion. Namely as $t \to t_0$ the differences $|k_i(t) - k_j(t)|$ can suddenly increase. In particular this implies that the trivial bundle, with all $k_i = 0$, is stable under small deformations.

A simple geometrical example will illustrate the way in which the integers k_i can suddenly jump. Consider the case $n = 2$, and assume that the topological invariant $k_1 + k_2 = 0$. Instead of vector bundles with fibre C^2 we can equally well work with projective bundles with fibre $P_1(C)$. The trivial bundle is $P_1 \times P_1$, while an example of a non-trivial bundle corresponding to integers $(1, -1)$ is provided by the family of generators of a quadric cone in P_3 (we regard the vertex of the cone as a different point on each generator). If we now take a general quadric in P_3 it can be identified with $P_1 \times P_1$ (cf. a real hyperboloid). By continuously altering the coefficients of the quadratic form we can end up with a singular quadric i.e. a cone.

On $P_1(C)$ therefore all holomorphic vector bundles are known. In particular they are all algebraic, i.e. in suitable gauges the gauge transformations are rational functions of z. Moreover this algebraic structure is essentially unique. This is a special case of a general theorem of Serre [39] which applies to all complex algebraic varieties in projective space of any dimension, and in particular to projective spaces themselves. However the simple and complete classification available over $P_1(C)$ does not extend to other algebraic varieties.

One way to study holomorphic (or algebraic) vector bundles on a projective space $P_m(C)$ (e.g. $m = 3$ which is our case) is to consider the restriction of the bundle to all the projective lines in $P_m(C)$. The semi-continuity property mentioned above tells us that over the « general line » we get a set of integers $(k_1, ..., k_n)$ but there will in general be some exceptional or jumping lines where the differences $|k_i - k_j|$ jump upwards. If for some line all k_i are 0 (i.e. the restricted bundle is trivial) then this must be true for the general line. According to Theorem (2.9) we see that the holomorphic vector bundles on $P_3(C)$ arising from anti-instantons on S^4 have this property, since every real line has all $k_i = 0$. The jumping lines are not real and correspond to points in the complexification of S^4. Serre's theorem assures us that, in a suitable gauge, our anti-self-dual gauge potential on S^4 will be given by *rational functions* in the coordinates. The complex poles of these rational functions account for the jumping lines.

If E is a holomorphic vector bundle on $P_3(C)$ trivial on general lines then, up to a constant matrix, E has a unique global holomorphic gauge along

any general line (since holomorphic functions on P_1 are constant). However if we take a triangle formed by three coplanar general lines there is no reason why the three holomorphic gauges on these lines should be consistent. One can think of a trivial holomorphic bundle on P_1 as having a distinguished connection or parallel transport. Then, going round the triangle formed by three coplanar lines, parallel transport may not return to the identity. Thus we get a version of curvature for E purely out of the holomorphic structure and the compactness of the projective lines. This « curvature » is a global notion, associated with each general triangle, but we can obviously infinitesimalize it to derive something more like the differential-geometric curvature.

From this point of view it is now natural to introduce the space which parametrizes all lines in $P_3(C)$. As explained in Chapter III this space is the Klein quadric Q_4 in $P_5(C)$. The « general » lines of P_3, i.e. lines for which E is trivial, correspond to an open set U of Q_4. We can now construct a holomorphic vector bundle $\mathcal{E}$ over U by defining the fibre $\mathcal{E}_L$ to be the space of holomorphic sections of $E|l$, where $L \in U$ represents the line $l \subset P_3$. All the lines l through a point A correspond to points L of a plane α in Q_4 and the vector spaces $\mathcal{E}_L$ for $l \in \alpha \cap U$ can all be naturally identified with E_A. This gives $\mathcal{E}$ a flat connection along all α-planes through L. These α-planes generate the tangent cone to Q_4 at L and so we get components for a connection along all these directions in the tangent cone. But any holomorphic function on this cone which is homogeneous of degree one on each generating line automatically extends (uniquely) to a linear function on the whole tangent space. This is an elementary geometric property of quadric cones (in any number of dimensions); cf. Chapter VI, § 3. Hence $\mathcal{E}$ has a holomorphic connection, flat along the α-planes. On the other hand on the other system of planes on Q_4, corresponding to planes of P_3, the connection need not be flat. This corresponds to the other notion of curvature on P_3 described above. A connection for a bundle on Q_4 which is flat along α-planes precisely corresponds to anti-self-duality for the curvature. This is essentially Ward's approach, passing from holomorphic bundles on P_3 to holomorphic bundles with holomorphic anti-self-dual connection on an open set of Q_4.

If we now want to consider unitary structures and work on $S^4 \subset Q_4$ we would impose appropriate conjugations throughout. The only drawback of this approach is that, in order to pass from S^4 to a neighbourhood in Q_4 and apply the above argument, we have to assume that our bundle on S^4 is *real-analytic*. Our treatment in Section 2, based on the Newlander-Nirenberg integrability theorem, had the advantage of requiring only differentiability. Analyticity, and eventually rationality of the solutions, is then an automatic consequence.

As mentioned in Section 2 (and will be proved in Chapter VI) a holomorphic bundle E on $P_3(C)$ which is trivial on a given line l is also trivial on the first neighbourhood of l. Thus a basis of E along l can be extended formally up to first derivatives normal to l. This extension essentially is just the connection on $\mathcal{E}$ at the point L defined by Ward's approach. In fact the property of quadric cones used in Ward's construction is closely related to the extension property of E: both can be formulated in terms of the vanishing of a certain sheaf cohomology group (cf. Chapter VI, § 3).

CHAPTER V

Construction of Algebraic Bundles

1. – The linear complex.

In this chapter we shall give an algebraic construction for vector bundles on $P_3(C)$ which, by the results of Chapter IV, will correspond to anti-instantons. We begin in this section by considering in some detail the basic anti-instanton (for $k = -1$ and $G = SU(2)$). The algebraic geometry in this case is classical and is, in traditional language, the geometry of a « linear complex » of lines in $P_3(C)$. It will turn out that this special case is an illuminating preliminary of the general construction and so it repays careful study.

If we start again from our fibration $P_3(C) \to S^4$ and if we use the standard metrics of S^4 and of $P_3(C)$ we can obviously construct a vector bundle over $P_3(C)$ with fibre C^2 by taking the horizontal tangent vectors on $P_3(C)$, i.e. vectors orthogonal to the fibre direction. Now we recall that taking orthogonal complements is not a holomorphic process so that we would not expect this horizontal vector bundle to be holomorphic. However the vertical bundle, from which we start, is itself not holomorphic since its definition involves the « real structure » of $P_3(C)$. In fact it turns out that the horizontal vectors *do* form a holomorphic bundle (while the complementary verticals do not). The restriction of this bundle to any real line is clearly the normal bundle of the line in $P_3(C)$. This is not trivial but becomes so after tensoring with the standard line bundle L on $P_3(C)$. We then end up with a holomorphic vector bundle which satisfies the conditions of Theorem (2.9) of Chapter IV and corresponds to the basic anti-instanton on S^4.

The preceding construction used the metric and real structure of $P_3(C)$ or equivalently the metric and quaternion structure of $C^4 = H^2$. From these structures we can also extract the natural skew-form and this is essentially what is needed to produce the algebraic vector bundle. If $L_{(z)} \subset C^4$ is the line corresponding to the point $(z) \in P_3(C)$ we consider its annihilating or

polar space $L^{\circ}_{(z)}$ with respect to the (non-degenerate) skew-form. The dimension of $L^{\circ}_{(z)}$ is 3 and, since the form is skew-symmetric, $L_{(z)} \subset L^{\circ}_{(z)}$. Hence the quotient space

$$E_{(z)} = L^{\circ}_{(z)}/L_{(z)}$$

is a 2-dimensional vector space depending algebraically on the point $(z) \in P_3(C)$. In other words E is an algebraic vector bundle with fibre C^2 over $P_3(C)$. Moreover the skew-form induces a non-degenerate skew-form on E so that its structure group reduces to $SL(2, C)$.

In projective terms $L^{\circ}_{(z)}$ corresponds to a projective plane in $P_3(C)$ passing through the point (z). Such a point-plane correspondence is classically called a null-correlation and the lines that lie in such a plane and pass through its corresponding point form the associated « linear complex ». These projective lines correspond in C^4 to the 2-dimensional isotropic subspaces of the skew-form, *i.e.* those $C^2 \subset C^4$ on which the skew-form is identically zero. If the skew-form in C^4 is given by the skew-symmetric 4×4 matrix a_{ij} the line $(x)\,(y)$ belongs to the linear complex provided $\sum y_\alpha a_{\alpha\beta} x_\beta = 0$ or equivalently

$$\sum a_{\alpha\beta} p_{\alpha\beta} = 0$$

where $p_{\alpha\beta} = x_\alpha y_\beta - x_\beta y_\alpha$ are the Plücker coordinates of the line.

We shall now show that the lines of this linear complex are precisely the jumping lines of the bundle E. First assume that the line $(x)\,(y)$ does not belong to the linear complex. This means that $L_{(x)}$ does not lie in $L^{\circ}_{(y)}$ and hence that $L^{\circ}_{(x)} \cap L^{\circ}_{(y)} = R$ is a 2-dimensional subspace of C^4 which, for all (z) on the line $(x)\,(y)$, lies in $L^{\circ}_{(z)}$ but does not contain $L_{(z)}$. This shows that E, restricted to the line $(x)\,(y)$ is a trivial bundle. Conversely we next show that E is not trivial when restricted to a line l of the linear complex. The line l corresponds to an isotropic 2-dimensional subspace $W \subset C^4$, and so for all $(z) \in l$ we have the inclusions

$$(1.1) \qquad\qquad L_{(z)} \subset W \subset L^{\circ}_{(z)} \subset C^4 \,.$$

Hence E_z contains $W/L_{(z)}$, i.e. $E|l$ contains the line-bundle W/L (here W denotes the trivial bundle with fixed fibre W). Topological considerations then show that $W/L \cong L^*$. Since L^* has holomorphic sections with zeros it cannot be a sub-bundle of a trivial bundle and so $E|l$ is not trivial: in fact $E|l$ corresponds to the integers $(1, -1)$ in the general classification.

We come now to consider real structures so that, by the results of Chapter IV, we can construct anti-instantons on S^4.

If we take the standard skew-form $(,)$ on $C^4 = H^2$ this is related to j and the positive inner product $\langle , \rangle$ by

$$(1.2) \qquad\qquad \langle u, v \rangle = (u, jv)$$

and in particular j preserves the skew-form in the sense that

$$(1.3) \qquad\qquad (ju, jv) = \overline{(u, v)} .$$

(1.2) implies that $L^\circ_{(z)} = L^\perp_{(jz)}$ where $\perp$ denotes the orthogonal space with respect to the inner product, and so for any $(z) \in P_3$ we have an orthogonal decomposition:

$$(1.4) \qquad\qquad C^4 = L_{(z)} \oplus R_x \oplus L_{(jz)}$$

where $R_x = L^\circ_{(z)} \cap L^\circ_{(jz)}$ and x represents the real line $l_x = \{(z)(jz)\}$.

This shows first that a real line l_x for $x \in S^4$ is never a jumping line for the algebraic bundle $E = L^\circ/L$ on $P_3(C)$. It also shows that j induces on E a real structure, i.e. an anti-linear map $\sigma \colon E \to E$ covering σ on $P_3(C)$ satisfying $\sigma^2 = -1$. Finally this real structure is positive, meaning that $(u, \sigma v)$ induces a positive inner product on the sections of E restricted to any real line. This last statement follows from the fact that, restricted to the real line $(z)(jz)$, the sections of E can be identified with R and the inner product is just that coming, as in (1.2), from the inner product of C^4.

Thus the bundle E on $P_3(C)$ together with its real structure σ gives precisely the data which, according to Chapter IV, corresponds to an anti-instanton on S^4. The bundle R on S^4 has as fibre at the point x the vector space R_x in the decomposition (1.4) with its natural inner product. It remains to specify the connection it inherits. According to Chapter IV the connection is uniquely determined by the condition that, on $P_3(C)$, it should be compatible with both the unitary and holomorphic structures. Now quite generally if $V \subset W$ are holomorphic vector bundles with compatible unitary structures then the canonical connection on V is induced by orthogonal projection from that of W. This is clear because a holomorphic section f of V is also holomorphic in W and so satisfies $\nabla''_W f = 0$ (∇_W being the canonical covariant derivative of W), hence $P\nabla''_W f = 0$ where P is orthogonal projection from W to V. Since this property plus unitarity characterizes the canonical connection we must have $P\nabla_W = \nabla_V$. Dualizing, the same applies to holomorphic quotient bundles. Applying these observations first to the sub-bundle $L^\circ \subset P_3 \times C^4$ and then to the quotient bundle E of L° we see that the canonical connection on E coincides with the connection induced on $L^\circ \cap L^\perp$ from the trivial connection of $P_3 \times C^4$.

From (1.4) we see that $L^\circ_{(z)} \cap L^\perp_{(z)} = R_x$ and so the connection on R is that induced from $S^4 \times C^4$ by orthogonal projection.

In quaternionic notation the decomposition (1.4) can also be written

$$C^4 = R_x \oplus R_x^\perp$$

where $x \in S^4 = P_1(H)$ and $R_x^\perp$ is the quaternionic line in H^2 parametrized by x. Thus R and $R^\perp$ are the standard quaternionic line-bundles over S^4 discussed in Chapter II and the connection we have given R is the natural one explained in Chapter II. With our present sign conventions R is an anti-instanton while $R^\perp$ is an instanton. Both are clearly acted on by the compact symplectic group $Sp(2) \cong \mathrm{Spin}\,(5)$.

If we fix the metric on S^4 as we have done here then we get a unique anti-instanton bundle R. However by applying a conformal transformation we will get new anti-instantons. The space of moduli is then $SL(2, H)/Sp(2)$ and parametrizes the metrics on S^4 in its standard conformal structure. Alternatively the moduli space can be viewed as the interior of the unit ball in R^5 (the hyperbolic 5-space). This unit ball can be thought of as the interior component of the real part of $P_5(C) - Q_4$, the complex moduli space. Note that the exterior component corresponds to a gauge field on S^4 with real singularities on an equatorial 3-sphere (the polar section of the external point).

2. – The Horrocks construction.

We shall now give an algebraic construction for $SL(2, C)$ bundles over $P_3(C)$ which generalizes the linear complex bundle described in Section 1. This construction is due to G. Horrocks and has been further studied by W. Barth and K. Hulek [8]. We shall return to it again in Chapter VII.

The basic idea can be expressed as follows. In section 1 our bundle E arose as the quotient L°/L. We now look for a generalization in which L is allowed to be k-dimensional. The details are as follows.

We fix complex vector spaces V, W of respective dimensions $2k + 2$, k and we suppose V endowed with a non-degenerate skew-form. Now let

(2.1) $$A(z) \colon W \to V$$

be a linear map depending linearly on $z = (z_1, z_2, z_3, z_4)$. Thus $A(z) = \sum_{i=1}^{4} A_i z_i$ where each $A_i \colon W \to V$ is a constant linear map. We let

$U_z = A(z) W \subset V$ be the image space and we assume:

(2.2)   for all $z \neq 0$, the space U_z is k-dimensional and isotropic for the skew-form on V.

With this assumption we have $U_z \subset U_z^\circ$, the polar space, and hence $E_z = U_z^\circ / U_z$ is a vector space depending algebraically on $(z) \in P_3(C)$. Since $\dim U_z = k$, $\dim U_z^\circ = 2k + 2 - k = k + 2$ and so $\dim E_z = 2$. Moreover E_z inherits a non-degenerate skew-form and so E is an algebraic vector bundle over $P_3(C)$ with group $SL(2, C)$.

Note that, when $k = 1$, U_z is one-dimensional and automatically isotropic. For $k > 1$ however the isotropic condition, which can be written in matrix form as

(2.3) $A(z)^t J A(z) \equiv 0$

where A^t is the transpose and J is the matrix of the skew-form, gives a system of quadratic equations on the coefficients of the 4 matrices A_1, A_2, A_3, A_4. The number of coefficients considerably exceeds the number of equations so that we certainly get solutions.

The assumption that $\dim U_z = k$ for all $z \neq 0$ implies that for any two distinct points (x) and (y) in P_3

(2.4) $U_x \cap U_y = 0$.

To see this we note first that the family of all U_z gives a k-dimensional vector bundle over P_3 isomorphic to the sum of k copies of L, each copy arising from a basis vector of W. Now any non-zero vector in $U_x \cap U_y$ would give rise to a holomorphic section (not identically zero) of U restricted to the line l joining (x), (y) in P_3. But $U|l$ is isomorphic to k copies of $L|l$ and this bundle has no holomorphic section besides zero (recall that, over any $P_m(C)$, a holomorphic section of L^* is a linear form in the coordinates, a section of L^{-n} is a homogeneous polynomial of degree n, and therefore L has only the zero section).

Now let us try to identify the jumping lines of the bundle E on P_3. Consider first a line l joining points (x) (y) such that

(2.5) $U_x^\circ \cap U_y = 0$.

This implies that

(2.6) $R = U_x^\circ \cap U_y^\circ$

is a 2-dimensional complement to U_z in U_z° for all points (z) on the line l. This shows that $E|l$ is trivial and so l is not a jumping line. Suppose now

that (2.5) is false then the space R given by (2.6) has non-zero intersection with U_y. Let v be a non-zero vector in $R \cap U_y$, then according to (2.4) $v \notin U_x$ and so v defines an algebraic section of $E|l$ which is zero at y but non-zero at x. This shows that $E|l$ is not trivial and so l is a jumping line. Thus the jumping lines are precisely the lines for which (2.5) fails.

To introduce reality conditions we now assume that we have an anti-linear map σ acting on W, V with $\sigma^2 = +1$ on W and $\sigma^2 = -1$ on V. Moreover we require that σ preserve the skew-form on V and be such that the corresponding hermitian form

$$(2.7) \qquad \langle u, v \rangle = (u, \sigma v)$$

is positive. Our linear transformation $A(z)$ is now assumed to be compatible with σ, i.e.

$$(2.8) \qquad \sigma\{A(z)\omega\} = A(\sigma z)(\sigma \omega)$$

where $\sigma(z)$ is, as before, multiplication by the quaternion i.

Condition (2.7) implies that

$$(2.9) \qquad U_{\sigma z} = \sigma\{U_z\}$$

and (2.7) implies that

$$(2.10) \qquad U_z^\circ = U_{\sigma z}^\perp .$$

Hence we get an orthogonal decomposition:

$$(2.11) \qquad V = U_z \oplus R_x \oplus U_{\sigma z}$$

where $R_x = U_z^\circ \cap U_{\sigma z}^\circ$ depends only on the point $x \in S^4$ parametrizing the real line $(z)(\sigma z)$ of P_3. This shows in particular that the real line l_x is not a jumping line for E and that E inherits a real structure from σ with the properties required, as in Chapter IV, to give an anti-instanton on S^4. Exactly as for the case $k = 1$ we see that the anti-instanton bundle on S^4 is R with the connection induced by orthogonal projection from $P_3 \times V$.

The instanton number of the bundle R over S^4 is $-k$. This follows from the fact that its orthogonal complement is topologically equivalent to the sum of k copies of the basic 1-instanton bundle.

To generalize from $Sp(1)$ to $Sp(n)$ all we have to do in the preceding construction is to take V to have dimension $2k + 2n$. For the group $SO(n)$ we have to take V of dimension $2k + n$ and we have to alter the signs. Thus we require V to have a non-degenerate symmetric bilinear form and $\sigma^2 = +1$ on V and $\sigma^2 = -1$ on W. In all cases the instanton number is $-k$.

We can deal with $U(n)$ by regarding it in the standard way as the subgroup of $SO(2n)$ commuting with J where $J^2 = -1$. We therefore require that V, W should both possess a complex linear automorphism J with $J^2 = -1$, $\sigma J = -J\sigma$. Moreover J is assumed to preserve the inner product on V or equivalently $(Ju, Jv) = -(u, v)$. Finally the linear transformation $A(z): W \to V$ is required to commute with J. Then the decomposition (2.10) is J-invariant. Hence the bundle R has an action of J and (since J preserves inner products) the canonical connection of R is preserved by J. Hence the connection of R, which is anti-self-dual, reduces from $SO(2n)$ to $U(n)$.

Notice that we could proceed one stage further and regard $Sp(n) \subset SO(4n)$ and so describe $Sp(n)$ solutions as $SO(4n)$ solutions with extra structure. However the more direct approach to $Sp(n)$ used earlier is more economical.

Returning now to the symplectic case we see that two triples (W, V, A) (W', V', A') which are isomorphic, in the obvious sense that we have isomorphisms $W \cong W'$, $V \cong V'$ commuting with σ and the skew-forms on V, V', and taking A to A', will give rise to an isomorphism between the bundles R and R' preserving connections. Thus isomorphic triples in the linear algebra sense yield gauge equivalent solutions of the anti-instanton problem. As we shall prove later the converse is also true.

3. – Quaternionic formulae.

In this section we shall show how the construction of the preceding section, for $Sp(n)$, can be formulated in terms of quaternions. We then obtain the explicit formulae given in Chapter II.

We recall that we have identified the space C^4 with H^2 in such a way that the map σ on C^4 is given by left multiplication on H^2 by the quaternion j. Similarly the vector space V of dimension $2k + 2n$ can be viewed as a left quaternion vector space of dimension $k + n$ with j given by the anti-linear map σ.

The vector space W, of complex dimension k, has σ with $\sigma^2 = 1$ and so can be viewed as the complexification of the real vector space W_R left fixed by σ. Then $C^4 \otimes_C W \cong H^2 \otimes_R W_R$ and σ corresponds again to multiplication by j on the quaternion vector space $H^2 \otimes_R W_R$.

The linear map $A(z): W \to V$ can now be viewed as a map

$$A: H^2 \otimes_R W_R \to V$$

which is quaternion linear. This corresponds to the compatibility of $A(z)$ with σ.

If we take a real basis of W_R and an orthogonal H-basis of V, so that V gets identified with H^{k+n}, then A is described by two matrices C, D of quaternions. The row-vectors of C are the image under A of $(1, 0) \otimes$ basis vectors of W_R and D is similarly defined replacing $(1, 0)$ by $(0, 1)$ in H^2. We use matrices as right multipliers here since our scalars act on the left, so that C, D are $k \times (k + n)$ matrices. Regarded as a matrix function of the pair (x, y) of quaternions in H^2 we then have

$$(3.1) \qquad A(x, y) = xC + yD \ .$$

The non-degeneracy condition on A is just that this matrix have maximal rank for all $(x, y) \neq (0, 0)$. Finally the isotropy condition (2.2) is equivalent, using (2.10), to saying that for any two row-vectors u, v of the matrix (3.1) u and iu are orthogonal to jv and $j(iv)$. This means that

$$\mathrm{Re}\,(uv^* j) = \mathrm{Re}\,(uv^* k) = \mathrm{Re}\,(uv^* i) = 0$$

or equivalently that the quaternion uv^* is real. Hence (2.2) is equivalent to

$$(3.2) \qquad \text{for all } (x, y) \in H^2, \quad \text{the matrix } (xC + yD)(C^* \bar{x} + D^* \bar{y})$$

is real.

We thus recover the description given in Chapter II, Section 3 except that we have transposed our matrices since we are now considering left action by scalars. For this reason we now get anti-instantons instead of the instantons of Chapter II.

CHAPTER VI

Linear Field Equations

1. – Bundles and sheaf cohomology.

The theory of holomorphic bundles is intimately related to the cohomology theory of sheaves. Although this cohomology theory arises in various contexts, and can be developed independently of bundle theory, it is rather natural from our point of view to explain it in terms of bundles. The main thing to emphasize just now is that cohomology is a linear or abelian theory and it first arises from abelian gauge groups and, at a later stage, from solvable groups.

We shall begin, as in Chapter IV, § 3, by considering holomorphic bundles over $P_1(C)$. This simple case provides a good illustration of the general theory and ties up with the classical function theory of one complex variable. In addition this special case occupies a key role in our applications, since $P_1(C)$ appears as the fibre of our basic fibration $P_3(C) \to S^4$.

In Chapter IV, § 3, we explained how a holomorphic line-bundle over $P_1(C)$ given by the holomorphic gauge-transformation $f(z)$, near the equator, could be reduced, as in (3.1), to the standard form z^k. In particular when $k = 0$ we could take $g(z) = \log f(z)$ and using the Laurent expansion

$$(1.1) \qquad g(z) = \sum_{-\infty}^{\infty} a_n z^n \qquad r < |z| < R$$

we could write $g(z)$ in the form

$$(1.2) \qquad g(z) = g_0(z) - g_\infty(z)$$

with $g_0(z)$ holomorphic for $|z| < R$ and $g_\infty(z)$ holomorphic for $r < |z|$. Suppose now we ask instead whether we can write $g(z)$ in the form

$$(1.3) \qquad g(z) = z^m g_0(z) - g_\infty(z)$$

where m is some fixed integer, and g_0, g_∞ are as before holomorphic near $0, \infty$ respectively. Clearly if $m \leqslant 1$ this is possible while for $m \geqslant 2$ it is possible if and only if the coefficients of (1.1) satisfy:

$$a_1 = \ldots = a_{m-1} = 0 \ .$$

In other words the functions $z, \ldots, z^{m-1}$ in general provide obstructions to the solubility of (1.2). Alternatively the space of all holomorphic functions $g(z)$, modulo those which can be written in the form (1.3), has dimension $m-1$ and a basis is represented by $z, \ldots, z^{m-1}$. This is our first encounter with a *sheaf cohomology group*: it is usually denoted by $H^1(P_1, L^m)$ or $H^1(P_1, \mathcal{O}(-m))$. The reasons for these notations will be explained shortly.

Equation (1.2) asserts that a holomorphic bundle over $P_1(C)$ with group C (the additive group) is holomorphically trivial: by exponentiation this corresponds to bundles with group C^* (the multiplicative group) for which the topological invariant $k = 0$. We can also consider C as the matrix subgroup of translations in $SL(2, C)$, i.e. matrices of the form

$$\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}.$$

Then (1.2) says that holomorphic bundles for this group over $P_1(C)$ are trivial. Geometrically a bundle for this matrix group corresponds to a 2-dimensional holomorphic vector bundle E over $P_1(C)$ having a sub line-bundle N such that N and the quotient E/N are both trivial. The solubility of (1.2) then asserts that N always possess a holomorphic complement. Recall that, unlike unitary bundles, holomorphic complements do not necessarily exist. In fact (1.3) corresponds to a slightly more general situation of this type. To see this note that eq. (1.3) is equivalent to the matrix equation

(1.4) $$\begin{pmatrix} 1 & g \\ 0 & z^m \end{pmatrix} = \begin{pmatrix} 1 & g_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & z^m \end{pmatrix} \begin{pmatrix} 1 & g_\infty \\ 0 & 1 \end{pmatrix}^{-1}.$$

If we regard the left-hand side as the gauge transformation, near $|z| = 1$, defining a 2-dimensional vector bundle E over $P_1(C)$, then E has a trivial sub-bundle N with $E/N \cong L^{-m}$, and solubility of (1.4) is equivalent to the existence of a holomorphic complement to N. Thus for $m \geqslant 2$ such a complement does not generally exist. More precisely any such bundle defines an element of the vector space $H^1(P_1, L^m)$ and a complement exists only if this element is zero. More generally $H^1(P_1, L^m)$ classifies the extensions E of the trivial bundle by L^{-m}, two such extensions being regarded as equivalent

if there is an isomorphism between the bundles which is the identity on the sub-bundle and quotient bundle.

After this introduction we shall now give the precise definition of sheaf cohomology groups on any compact complex manifold X, e.g. $P_3(C)$. We begin with the simplest case, corresponding to (1.2). Instead of just the 2 open sets $|z| < R$ and $r < |z|$ which cover $P_1(C)$ we now have to allow a finite number of open sets $\{U_\alpha\}$ to cover X: for $P_3(C)$ we could take the 4 sets given by $z_\alpha \neq 0$ $(\alpha = 1, 2, 3, 4)$. A holomorphic 1-cochain is then a collection of holomorphic functions $g_{\alpha\beta}$ defined in $U_\alpha \cap U_\beta$ (conventionally $g_{\beta\alpha} = -g_{\alpha\beta}$). Such a 1-cochain is called a 1-cocycle if it satisfies the transitivity condition

$$(1.5) \qquad g_{\alpha\beta} = g_{\alpha\gamma} + g_{\gamma\beta} \quad \text{in } U_\alpha \cap U_\beta \cap U_\gamma.$$

It is called a 1-coboundary if there exist holomorphic functions h_α defined in U_α so that

$$(1.6) \qquad g_{\alpha\beta} = h_\alpha - h_\beta \quad \text{in } U_\alpha \cap U_\beta.$$

Every 1-coboundary is evidently a 1-cocycle but the converse need not be true. We measure the effect by considering the quotient space H^1 of 1-cocycles modulo 1-coboundaries. If the open sets U_α are sufficiently small and well-chosen (technically all finite intersections should be domains of holomorphy) H^1 is independent of the covering. It is called the first cohomology group of H with coefficients in the sheaf of holomorphic functions and is denoted by $H^1(X, \mathcal{O})$.

We cannot pursue cohomology theory in any detail, for which we refer to standard texts, and we confine ourselves for the present to a few further comments.

1) For $P_1(C)$ we used just 2 open sets, hence there was only one $g_{\alpha\beta}$ and every 1-cochain was automatically a 1-cocycle.

2) $H^1(X, \mathcal{O})$ always classifies holomorphic bundles over X with group C or equivalently extensions of the trivial line-bundle by itself.

3) For every integer $q \geqslant 0$ one can define $H^q(X, \mathcal{O})$: $H^0(X, \mathcal{O})$ consists of global holomorphic functions on X (which are constant for compact X). The cohomology groups are linked together by exact sequences.

4) If X is not compact similar definitions work but one must allow infinite coverings which are locally finite.

5) For X compact algebraic, as for $P_m(C)$ for any m, cohomology groups can be computed by using only rational functions. This is one of the main results of Serre [38].

It is implicit in our notation that other coefficients than the holomorphic functions $\mathcal{O}$ can be used to form cohomology groups. Thus in the case of $P_1(C)$ the cohomology group we denoted by $H^1(P_1, L^m)$ uses local holomorphic sections of the holomorphic line-bundle L^m. Clearly the definition above of H^1 goes through unaltered for any X and any holomorphic line-bundle L or even vector bundle E over X. The cohomology group $H^1(X, L)$ can be viewed as classifying extensions of the line-bundle L by the trivial line-bundle. More generally holomorphic vector bundles E of any dimension which have a given sub-bundle E_1 and quotient $E/E_1 = E_2$ are classified by $H^1(X, E_2^* \otimes E_1)$ where E_2^* is the dual bundle of E_2.

If X is compact all H^q of the type considered here are *finite-dimensional vector spaces*. This ceases to be true for non-compact X. For example $H^0(X, \mathcal{O})$ consists of all holomorphic functions on X. Examples with H^1 being infinite-dimensional will occur naturally in our applications as we shall see in the next section.

2. – Linear aspects of the Penrose transform.

The main result of Chapter III was that solutions of the anti-self-dual Yang-Mills equations on S^4 converted, via the Penrose transformation, to holomorphic bundles on $P_3(C)$. There is a parallel twistor interpretation for solutions of certain important *linear* differential equations. We shall show in this section that they correspond to elements of appropriate sheaf cohomology groups.

We begin by considering the case of U(1) Yang-Mills theory. Although there are no global anti-self-dual solutions on S^4, solutions certainly exist in R^4 corresponding to solutions of the Euclidean Maxwell equations for a 2-form ω: $d\omega = 0$, $*\omega = -\omega$. From Chapter III we know that these correspond to holomorphic line-bundles on $P_3(C) - P_1(C)$. If we ignore unitarity and work with complex-valued 2-forms the holomorphic line-bundle is unrestricted, except that it must be topologically trivial. As explained in Section 1 such a line-bundle is then described by an element of the cohomology group $H^1(P_3 - P_1, \mathcal{O})$. Thus this group corresponds under the Penrose transformation to the solutions of the anti-self-dual Maxwell equations on R^4. This is the type of correspondence which we propose to extend. As we shall see all the cohomology groups $H^1(P_3 - P_1, \mathcal{O}(-m))$ correspond in a similar way to solutions of other linear differential equations on R^4.

Before leaving the case of Maxwell fields we should comment on one special topological feature, related to the bundle interpretation, which does not generalize to the other cases. If we work with some open set $U \subset R^4$ and the corresponding open set $\tilde{U} \subset P_3(C)$ then holomorphic line-bundles on $\tilde{U}$ which are trivial on the fibres of $P_3(C) \to S^4$ correspond to anti-self-dual C^*-potentials on U. If U is R^4 or a contractible open set then the homology of $\tilde{U}$ is generated by the fibres and the line-bundle on $\tilde{U}$ is then topologically trivial. However if U has 2-dimensional homology, e.g. if $U = R^4 - R^1$ then the bundle on $\tilde{U}$ need not be topologically trivial and the corresponding bundle (with connection) on U is then also nontrivial. We cannot in such a case take the logarithm of the gauge transformations to reduce from C^* to C. Thus the geometrical correspondence between bundles given in Chapter III contains more information than the correspondence described above between $H^1(\tilde{U}, \mathcal{O})$ and solutions of the anti-self-dual Maxwell equations on U.

The cohomology groups $H^1(\tilde{U}, \mathcal{O}(-m))$ fall naturally into two families according as $m \geqslant 2$ or $m < 2$. The case $m = 2$ is in a sense the most basic and the situation is essentially symmetric about this case, as we shall see. For example $H^1(\tilde{U}, \mathcal{O}(-4))$ will turn out to correspond to solutions of the self-dual Maxwell equations on U.

We begin therefore with the case $m = 2$ and we take any element $\Phi \in H^1(\tilde{U}, \mathcal{O}(-2))$. If P_x is the fibre of $P_3(C) \to S^4$ over the point $x \in U \subset S^4$ we can restrict Φ to P_x to get an element

$$\varphi_x \in H^1(P_x, \mathcal{O}(-2)).$$

As explained in Section 1

$$\dim H^1(P_x, \mathcal{O}(-2)) = 1$$

so that if we fixed, in some standard way, a basis of this space φ_x would become a scalar function of x, defined for $x \in U$. If $U \subset R^4$ then $\tilde{U} \subset P_3(C) - P_1(C)$ has a natural map to $P_1(C)$, projecting along parallel planes, and this identifies all P_x with $x \in U$. Thus in this case φ_x gives a scalar function φ defined in U. Analogy with the Maxwell fields suggests that φ should satisfy a differential equation and the only candidate with the necessary invariance properties is the Laplace operator of R^4. In fact the correspondence $\Phi \leftrightarrow \varphi$ establishes an isomorphism between $H^1(\tilde{U}, \mathcal{O}(-2))$ and the space of solutions of the Laplace equation in U. When $U = R^4$ both spaces are acted on by the 11-parameter group of conformal motions of R^4 (the Euclidean motions together with scalar magnification about an origin) and the isomorphism is compatible with these actions. In particular the subspace

of $H^1(P_3(C) - P_1(C), \mathcal{O}(-2))$ which is homogeneous of degree n, under scale change, corresponds to homogeneous polynomials of degree n on R^4 which satisfy the Laplace equation. We shall indicate how this statement can be verified.

Once we have fixed an origin in R^4, $P_3(C) - P_1(C)$ can be viewed as the normal bundle N of the corresponding projective line. The bundle $L \otimes N$ is trivial or equivalently $N \cong L^* \otimes S^-$, where $S^- \cong C^2$ is a fixed vector space. As explained in [3] (where a different notation is used), S^- can be naturally identified with one of the half-spin spaces of R^4. Holomorphic functions on N which are homogeneous, in each fibre, of degree n form a sheaf over P_1 isomorphic to the sections of $L^n \otimes \mathfrak{F}^n(S^-)$, where $\mathfrak{F}^n$ denotes the polynomials of degree n (so $\mathfrak{F}^n(S^-) \cong C^{n+1}$). Since we want to compute $H^1(N, L^2) = H^1(N, \mathcal{O}(-2))$ we tensor with L^2 and deduce a map

$$(2.1) \qquad H^1(P_1, L^{n+2}) \otimes \mathfrak{F}^n(S^-) \to H^1(P_3 - P_1, \mathcal{O}(-2))_n$$

where the suffix n on the right denotes the part of the cohomology which is homogeneous of degree n. By general theorems one can deduce that (2.1) is actually an isomorphism. Now we saw in Section 1 that $H^1(P_1, L^{n+2})$ was of dimension $n+1$. With a little more care (see section 4) one can naturally identify it with $\mathfrak{F}^n(S^+)$, where S^+ is the other half-spin space of R^4. Now $S^+ \otimes S^- \cong R^4 \otimes C$, the complexification of R^4. Hence (2.1) shows that the space of polynomials φ corresponding to elements $\Phi \in H^1(P_3 - P_1, \mathcal{O}(-2))_n$ consists precisely of the image of

$$(2.2) \qquad \mathfrak{F}^n(S^+) \otimes \mathfrak{F}^n(S^-) \to \mathfrak{F}^n(C^4) .$$

The classical invariant theory of $SO(4)$ then shows that the image of (2.2) is just the space of harmonic polynomials; for example when $n = 2$ the left side of (2.2) has dimension 9 while the right side has dimension 10, containing the extra polynomial $r^2 = \sum_1^4 x_i^2$.

The preceding piece of algebra can be used as a basis for the general correspondence $\Phi \leftrightarrow \varphi$ by associating to each of Φ, φ a power series expansion in homogeneous terms. Appropriate analytic questions of convergence have then to be examined but these can be dealt with by standard methods. Note that such convergence questions are irrelevant when we work on the whole of S^4 since all spaces are then finite-dimensional.

If we work on the whole of S^4 the different fibres P_x cannot all be identified (the bundle $P_3(C) \to S^4$ being topologically non-trivial) and so more care has to be taken in the interpretation of the correspondence $\Phi \leftrightarrow \varphi$.

We must now view φ as a section of the line-bundle W over S^4 whose fibre W_x is the one-dimensional space $H^1(P_x, \mathcal{O}(-2))$. One can show that W^4 is the volume of density bundle on S^4: thus φ is the fourth root of a density. The Laplace operator has a conformally invariant counterpart [34] which acts on such fourth-roots of densities and $\Phi \leftrightarrow \varphi$ is an isomorphism between $H^1(P_3, \mathcal{O}(-2))$ and the space of solutions of this conformally invariant Laplace equation on S^4. Written in terms of the standard metric of S^4 this operator is $\Delta + R/6$, where Δ is the Laplace-Beltrami operator $d^* d$ of S^4 and R is the scalar curvature [34], which vanishes on R^4 but is positive on S^4.

Since $\Delta \geqslant 0$ as an operator, $\Delta + R/6 > 0$ and so the equation $(\Delta + R/6)\varphi = 0$ has no global solutions on S^4 except $\varphi = 0$. This checks with the well-known result of algebraic geometry that $H^1(P_3, \mathcal{O}(-2)) = 0$: in fact the same is true for all $\mathcal{O}(-m)$.

In a later section we shall have more to say about the Laplace operator in relation to the Penrose transform. Returning for the present to the interpretation of other sheaf cohomology groups, we get a very similar picture for the groups $H^1(\tilde{U}, \mathcal{O}(-m))$ $m > 2$. Any element Φ of this cohomology group gives rise to an element $\varphi_x \in H^1(P_x, \mathcal{O}(-m))$ for all $x \in U$. As shown in section 1 this vector space (in which φ_x takes its values) has dimension $m - 1$. Thus, for $U \subset R^4$, φ is an $(m-1)$-component function and it satisfies a first-order differential equation. For $m = 3$ this is the (mass-less) Dirac equation and for $m = 4$ it is the (self-dual) Maxwell equation. All these can be exhibited in conformally invariant form.

For $m < 2$ we have already met the case $m = 0$ in a different context. For other values one has to proceed slightly differently. Restricting Φ to any P_x now gives zero since $H^1(P_x, \mathcal{O}(-m)) = 0$ for $m \leqslant 1$. However the isomorphism (2.1) still holds with 2 replaced by m. In particular taking $n = 2 - m$ we see that

$$(2.3) \qquad \mathfrak{F}^{2-m}(S^-) \cong H^1(P_3 - P_1, \mathcal{O}(-m))_{2-m} .$$

Based on this isomorphism we can now assign to any $\Phi \in H^1(P_3 - P_1, \mathcal{O}(-m))$ a k-component function φ on R^4, where $k = 3 - m$. This is done as follows. Given $x \in R^4$ we take this as origin for scalar magnification and then pick out the component Φ_{2-m} of Φ. By (2.3) we get a vector in the k-dimensional space $\mathfrak{F}^{2-m}(S^-)$. An alternative description is to consider the ν-th formal neighbourhoods $P_x^{(\nu)}$ of the line P_x: this means that we include not only values on P_x but also normal derivatives up to order ν. Taking $\nu = 2 - m$ we restrict Φ to give an element

$$\varphi_x \in H^1(P_x^{(\nu)}, \mathcal{O}(-m)) .$$

It is not hard to show that this is equivalent to the previous definition.

Again one shows that the functions φ arising this way are precisely the solutions of the appropriate first-order differential equation. For $m = 1$ we again get the (mass-less) Dirac equation but this time for the opposite type of spinor from the case $m = 3$. In general the equations for m and $4 - m$ are essentially the same, except that they are of opposite type i.e. switch when we reverse orientation.

3. – Linear equations in a Yang-Mills background.

As we saw in Chapter IV a vector bundle E on S^4 with anti-self-dual connection lifts to give a bundle $\tilde{E}$ on $P_3(C)$ which has a natural holomorphic structure. As in the preceding section we can then form the sheaf cohomology groups $H^1(P_3, \tilde{E}(-m))$ or their counterparts over an open set. It is natural to look for an interpretation of these vector spaces in terms of the original bundle E on S^4. When E is a trivial bundle the results of section 2 tell us that we get spaces of solutions of certain standard linear differential equations. These differential equations have obvious counterparts for a non-trivial bundle in which the derivatives are replaced by appropriate covariant derivatives built from the connection. It is then extremely plausible that our sheaf cohomology groups will correspond to these covariant differential equations, and in this section we shall show that this is indeed the case.

As mentioned in Chapter IV the vector bundle $\tilde{E}$ is trivial along each line P_x, for $x \in S^4$, and furthermore is trivial along the first formal neighbourhood $P_x^{(1)}$. We shall first give the proof of this statement. In fact if V is any holomorphic vector bundle on P_3, defined in a neighbourhood of a fixed projective line P_1, triviality of V restricted to P_1 implies triviality of V restricted to $P_1^{(1)}$. To see this we consider first the exact sequence of sheaves

$$(3.1) \qquad\qquad 0 \to J^2/J \to \mathcal{O}/J^2 \to \mathcal{O}/J \to 0$$

where $\mathcal{O}$ is the sheaf of holomorphic functions and J is the ideal sheaf of P_1 (consisting of functions vanishing on P_1). The three sheaves in (3.1) can then be equivalently described as follows:

$\mathcal{O}/J = \mathcal{O}(P_1)$, holomorphic functions on P_1

$\mathcal{O}/J^2 = \mathcal{O}(P_1^{(1)})$, holomorphic « functions » on $P_1^{(1)}$

$J^2/J =$ sheaf of sections of the co-normal bundle N^* of P_1 in P_3.

If $f \in \mathcal{O}/J^2$ then it has a value on P_1, namely its image in $\mathcal{O}/J$, and extra components corresponding to the first-order terms in the Taylor expansion normal to P_1. These first-order terms lie in J^2/J.

Since the normal bundle $N \cong L^{-1} \oplus L^{-1}$ we have $N^* \cong L \oplus L$ and so in particular

$$H^0(P_1, N^*) = H^1(P_1, N^*) = 0$$

(see §1). From the cohomology exact sequence of (3.1) this implies that the global sections of $\mathcal{O}/J^2$ map isomorphically onto those of $\mathcal{O}/J$: in other words a global function on $P_1^{(1)}$ is constant, as we should expect.

Suppose now we introduce the vector bundle V, assumed trivial on P_1, and tensor (3.1) with V. We get an exact sequence:

$$(3.2) \qquad 0 \to V^{(0)} \otimes N^* \to V^{(1)} \to V^{(0)} \to 0$$

where $V^{(0)}$, $V^{(1)}$ denote the restriction of V to P_1 and $P_1^{(1)}$ respectively. Since $V^{(0)}$ is trivial we again have

$$H^0(P_1, V^{(0)} \otimes N^*) = H^1(P_1, V^{(0)} \otimes N^*) = 0$$

and (3.2) then implies as before that global sections of $V^{(1)}$ map isomorphically onto global sections of $V^{(0)}$. This means that a basis of (constant) sections of the trivial bundle $V^{(0)}$ has a unique extension to $V^{(1)}$, showing in particular that $V^{(1)}$ is trivial as claimed.

When $V = \tilde{E}$ arises from a bundle E on S^4 as in Chapter IV the fibre E_x can be identified with the space of global sections of $V_x^{(0)}$ (i.e. the restriction of V to P_x). Global sections of $V_x^{(1)}$ give vectors in $E_x^{(1)}$, the first jet space of sections of E at x (i.e. including first derivatives). The fact that every global section of $V_x^{(0)}$ extends uniquely to a global section of $V_x^{(1)}$ means that we have a unique way of extending vectors from E_x to $E_x^{(1)}$, i.e. we have a connection on E. This is essentially Ward's definition of the connection as explained in Chapter IV. Note that the property of quadric cones used in Chapter IV, §3 for Ward's construction is equivalent to the vanishing of $H^1(P_3, \mathcal{O}(-1))$ as one sees by an exact sequence argument; a more elementary argument can however be used based on the fact that a quadric surface factorizes as $P_1 \times P_1$.

We turn now to the question of interpreting the sheaf cohomology groups $H^1(\tilde{U}, \tilde{E}(-m))$ where $\tilde{U} \subset P_3(C)$ corresponds to an open set $U \subset S^4$. We begin by considering the case $m > 2$. Then for any $x \in U$, an element

$\Phi \in H^1(\tilde{U}, \tilde{E}(-m))$ defines

$$\varphi_x \in H^1(P_x, \tilde{E}_x^{(0)}(-m)) \cong E_x \otimes H^1(P_x, \mathcal{O}(-m))$$

(since $\tilde{E}$ is trivial on each P_x). In other words φ is a section of $E \otimes W_m$ where W_m is the vector-bundle (with fibre C^{m-1}) which we met in the previous section. For $m > 2$ there is a first-order differential equation for sections of W_m whose solutions correspond to elements of $H^1(\tilde{U}, \mathcal{O}(-m))$. This differential equation may be identified by computing in the first formal neighbourhood $P_x^{(1)}$. Repeating this computation with $\tilde{E}(-m)$, using the triviality of $\tilde{E}_x^{(1)}$, and the way in which this triviality defines the connection on E, it follows that φ satisfies the obvious differential equation for sections of $E \otimes W_m$. For example if $m = 3$, W_m is the 2-component spin bundle, the equation is the (mass-less) Dirac equation and the equation for $E \otimes W_3$ is the Dirac equation extended to E using its connection. In more physical terms this is the Dirac equation in a Yang-Mills background field.

For $m < 2$, the equations are again first-order, and a similar but slightly more complicated argument leads to the same conclusion.

Finally we come to the case $m = 2$ which is both the most interesting and technically the most complicated since it involves a second-order operator, the Laplacian. General considerations, and computations of the above type, lead to the conclusion that elements $\Phi \in H^1(\tilde{U}, \tilde{E}(-2))$ correspond to sections of $E \otimes W_2$ over U (where W_2 is the line-bundle of fourth-roots of densities as in Section 2) which satisfy a second-order differential equation. Using the triviality of $\tilde{E}$ on $P_x^{(1)}$ identifies the first-order and second-order parts of this equation but leaves unidentified the zero-order part. A final calculation is therefore necessary to determine this zero-order part. The most elegant and informative way of performing this calculation is explained in detail in [29]. The idea is to exploit the inter-relation between the different cohomology groups $H^1(\tilde{U}, \tilde{E}(-m))$ for various m, arising from the fact that multiplication by each of the linear coordinates $z_1, \dots, z_4$ maps $\tilde{E}(-m)$ to $\tilde{E}(-m+1)$. Formally we note that $z_1, \dots, z_4$ are a basis of the holomorphic sections over $P_3(C)$ of the bundle L^{-1}. Converting this statement into one on S^4 provides a link between the differential equations for various integers m. Knowledge of the equation for $m \neq 2$ can then be used to help identify the equation for $m = 2$. In more concrete terms our second-order operator has to annihilate elements of the form $z_\alpha \varphi$ where φ satisfies the appropriate Dirac equation (corresponding to $m = 3$). As shown in [29] this information is enough to identify our equation as

(3.3)
$$\left(\nabla^* \nabla + \frac{R}{6}\right)\varphi = 0$$

where ∇ is the covariant derivative of φ, ∇^* its adjoint relative to the standard metric of S^4, and φ is now considered as a pure scalar function (densities being identified with scalars by using the volume form of S^4).

An important consequence of (3.3) is that it has no global solutions on P_3 except $\varphi = 0$. This is because the operator $\nabla^*\nabla + R/6$ is positive just as in Section 2. We conclude that

$$(3.4) \qquad\qquad H^1(P_3(C), \tilde{E}(-2)) = 0 \ .$$

Unlike the case when $\tilde{E}$ is trivial (3.4) is not a general result of algebraic geometry. There are many vector bundles V which do not satisfy (3.4). All that general theory tells us is that

$$H^1(P_3, V(-m)) = 0 \quad \text{if } |m| > m_0(V)$$

where m_0 is a sufficiently large integer. The proof of (3.4) depends crucially on the fact that $\tilde{E}$ comes from a unitary bundle on S^4.

For $m > 2$ by composing the first-order operator with its adjoint one gets a second-order operator which can also be shown to be positive. This leads to the further result that:

$$(3.5) \qquad\qquad H^1(P_3(C), \tilde{E}(-m)) = 0 \quad m \geqslant 2 \ .$$

However, as we shall see in Chapter VII, (3.5) follows easily from (3.6). The reason is that multiplication by a monomial of degree $m - 2$ in $z_1, ..., z_4$ shifts us from $H^1(P_3(C), \tilde{E}(-m))$ to $H^1(P_3(C), \tilde{E}(-2))$.

So far we have explained that sheaf cohomology groups correspond to solutions of various linear equations in the background field of Yang-Mills instantons. For quantum calculations one needs, not only the solutions of the homogeneous linear equation, but the full Green's function or « propagator ». Using the explicit description of instantons given in Chapter II explicit and fairly simple formulae have been derived for some of these Green's functions [13] [15].

4. – The 't Hooft Ansatz.

This section is a digression in which we shall explain how solutions of the anti-self-duality equations can be constructed from solutions of linear equations. This point of view was explained in [4] and it leads to the 't Hooft Ansatz and natural generalizations of it.

We shall consider the group $SL(2, C)$, ignoring for the present reality constraints. Equivalently we work with 2-dimensional vector bundles with a skew form. According to the results of Chapter IV anti-self-dual $SL(2, C)$ connections on $U \subset S^4$ correspond to holomorphic vector bundles on $\tilde{U} \subset P_3(C)$ which are trivial along all fibres P_x for $x \in U$. One way to construct such vector bundles is to produce extensions of line-bundles, in other words to use only the triangular matrices in $SL(2, C)$ for gauge transformations. As explained in Section 1 such extensions are classified by elements of a sheaf cohomology group. Thus we can construct $SL(2, C)$ bundles by first picking a line-bundle N on $\tilde{U}$ and then choosing an element Φ of $H^1(\tilde{U}, N^2)$. This gives a vector bundle V with N as sub-bundle and N^* as quotient.

In particular we can take N to be the standard line-bundle L or one of its powers L^m. Then to every $\Phi \in H^1(\tilde{U}, L^{2m}) = H^1(\tilde{U}, \mathcal{O}(-2m))$ we get a holomorphic vector bundle over $\tilde{U}$. To descend to S^4 we need this vector bundle to be trivial on each P_x with $x \in \tilde{U}$. To investigate this question we consider a fixed projective line P_1 and ask which elements of $H^1(P_1, \mathcal{O}(-2m))$ give the trivial bundle. Note for example that the zero element gives the direct sum $L^m \oplus L^{-m}$ which is never trivial (for $m \neq 0$). As we shall see, a general element always gives the trivial bundle but special elements give rise to direct sums $L^r \oplus L^{-r}$ for $r = 1, 2, ..., m$.

For simplicity consider first the case $m = 1$, then $H^1(P_1, \mathcal{O}(-2))$ is one-dimensional and every non-zero Φ gives the trivial bundle [1]. Consider next the case $m = 2$, then $H^1(P_1, \mathcal{O}(-4))$ is a 3-dimensional space. In Section 1 we saw that a basis is given by the transition functions (or cocycles) z, z^2, z^3. A more invariant description of the space $H^1(P_1, \mathcal{O}(-4))$ is to say that it is the dual of the space of quadratic forms in the two homogeneous coordinates (z_1, z_2) of P_1. This arises because of the multiplication

$$H^0(P_1, \mathcal{O}(2)) \otimes H^1(P_1, \mathcal{O}(-4)) \to H^1(P_1, \mathcal{O}(-2)) \cong C .$$

Using our basis it is easy to verify that this multiplication gives the required duality. This is in fact a very special case of the much more general duality theorem of Serre [39]. Thus if $P_1 = P_1(V)$, i.e. V is the copy of C^2 whose lines are represented by points of P_1, we have a natural isomorphism

$$(4.1) \qquad H^1(P_1, \mathcal{O}(-4)) \cong \mathcal{S}^2(V)^* \cong \mathcal{S}^2(V^*) .$$

Hence elements $\Phi \in H^1(P_1, \mathcal{O}(-4))$ represent quadratic forms on V^*. Then Φ defines the trivial bundle over P_1 if and only if the corresponding quadratic form on V^* is non-singular (i.e. not a perfect square) [1]. If $\Phi \neq 0$ but corresponds to a perfect square the bundle is isomorphic to $L \oplus L^{-1}$.

The generalization of this to arbitrary $m > 0$ is now fairly straightforward. We have a natural isomorphism

(4.2) $$H^1\big(P_1,\, \mathcal{O}(-2m)\big) \cong \mathfrak{T}^{2m-2}(V^*)\,.$$

If $\Phi \in H^1(P_1,\, \mathcal{O}(-2m))$, let f be the corresponding homogeneous polynomial of degree $2m-2$ on V^*. Any such polynomial can be written as a linear combination of perfect powers:

$$f = \sum_{i=1}^{m-2} \lambda^i u_i^{2m-2}$$

where $\lambda_i \in C$ and u_i is a linear form on V^*. The general result proved in [1] is that Φ defines the trivial bundle unless f can be expressed as a combination of less than $m-2$ perfect powers (or is a limit of such). More generally if f is a combination of $m-2-r$ perfect powers (or a limit of such), with r being maximal, then Φ defines the bundle $L^r \oplus L^{-r}$.

We know therefore the precise conditions to be imposed on an element $\Phi \in H^1(\tilde{U},\, \mathcal{O}(-2m))$ in order to get a trivial bundle on all P_x with $x \in U$ and so to descend to a bundle on U with anti-self-dual $SL(2,\, C)$-connection. By the results described in Section 2, Φ corresponds to a section of the vector bundle $\mathfrak{T}^{2m-2}(S^+)$ satisfying the appropriate linear differential equation. Here S^+ denotes the half-spin bundle corresponding to instantons, whereas S^- corresponds to anti-instantons. We have also used the metric on S^4 or R^4 to identify S^+ with its dual $(S^+)^*$, although to exhibit full conformal invariance this should not be done.

To sum up we have given an implicit construction for anti-self-dual $SL(2,\, C)$ connections starting from a section φ of $\mathfrak{T}^{2m-2}(S^+)$ satisfying the relevant differential equation. To get a solution this section must be everywhere « general » i.e. not a combination of less than $(m-2)$ perfect powers.

For $m = 1$, φ is a scalar field satisfying the Laplace equation and nowhere zero. For $m = 2$, φ is a self-dual solution of Maxwell's equations which is nowhere the square of a spinor.

The case $m = 1$ is the Ansatz employed by 't Hooft. The explicit formulae for $m = 2$ are given in [4]. It is important to note that the solution of the non-linear Yang-Mills equations obtained in such a way may have a larger domain of regularity than the linear field φ from which we start. For example it is well-known, for the 't Hooft Ansatz, that a $1/r^2$ singularity for φ disappears when we construct the Yang-Mills field. Geometrically this corresponds to the fact that the reduction from $SL(2,\, C)$ to the triangular matrices may only be valid in a smaller open set. The situation is quite clear if we work on the whole of $P_3(C)$ as we shall now explain.

Let W be a 2-dimensional holomorphic vector bundle over $P_3(C)$. Although W may have no globally defined holomorphic section, other than zero, general theorems tell us that $W \otimes L^{-m} = W(m)$ has a non-zero section if m is sufficiently large. Let s be such a section, then $s = 0$ on some closed algebraic subariety Σ of $P_3(C)$ and $s \neq 0$ on the complementary open set. On this open set s generates a trivial sub-line-bundle of $W(m)$ and so exhibits $W(m)$ as an extension. Tensoring back by L^m then shows that, over this open set W has L^m as a sub-line-bundle. If W is an $SL(2, C)$ bundle then the quotient is necessarily L^{-m} and we are in the standard situation described above. Thus the subvariety Σ given by $s = 0$ will appear as a singularity of the extension element Φ, or the corresponding linear field φ, but is not a singularity of the vector bundle W.

If m is the least integer for which $W(m)$ has global sections then one can show that Σ necessarily has complex dimension 1, i.e. it is an algebraic curve. Conversely given any curve Σ one can describe (a) when it arises from a bundle W as above and (b) how to reconstruct W from Σ when the conditions of (a) are satisfied [23]. In particular when Σ is connected it uniquely determines W. If Σ has several connected components Σ_i then W depends on non-zero constants c_i (up to a common factor).

When $m = 1$, Σ has to be a collection of $k + 1$ disjoint lines (where $k = c_2(W)$ is the anti-instanton number). If we take these lines to be real, i.e. fibres of $P_3(C) \to S^4$, we recover the 't Hooft Ansatz depending on $(k + 1)$ points of S^4 and corresponding weights $c_1, ..., c_{k+1}$. When $m = 2$, Σ has to be a collection of disjoint elliptic curves, whereas for $m > 2$, Σ involves rather special curves of high genus.

For the bundles given by the Horrocks construction of Chapter V one can show that one need only take $m > \sqrt{3k+1} - 1$. Hartshorne [27] has shown that this bound is best possible in the sense that smaller values will not give all $- k$-instantons.

We conclude with a few further comments on the Ansatz for general m. In order to give an explicit formula for the potential one must choose a definite gauge and it is desirable to pick the gauge as simply as possible. Geometrically this means that we have to pick a basis of $H^0(P_x, W_x)$ depending smoothly on $x \in R^4$. Recall that in the appropriate open set W is given as an extension or exact sequence of bundles

$$0 \to L^m \to W \to L^{-m} \to 0 .$$

Restricting to P_x and taking cohomology we get an exact sequence of vector spaces:

$$0 \to H^0(P_x, W_x) \to H^0(P_x, \mathcal{O}(m)) \xrightarrow{\delta} H^1(P_x, \mathcal{O}(-m)) \to 0$$

Now $H^0(P_x, \mathcal{O}(m)) \cong \mathfrak{I}^m(S_x^+)$ and

$$H^1(P_x, \mathcal{O}(-m)) \cong \mathfrak{I}^{m-2}(S_x^+)^*$$

and the map δ between them is given by φ_x. Thus if $W = \tilde{E}$, the fibre $E_x = H^0(P_x, W_x)$ appears as the kernel of the homomorphism

$$\varphi_x \colon \mathfrak{I}^m(S_x^+) \to \mathfrak{I}^{m-2}(S_x^+)^* .$$

Now over R^4 the bundle S^+ has a natural trivialization and so therefore do $\mathfrak{I}^m(S^+)$ and $\mathfrak{I}^{m-2}(S^+)$. However the kernel of φ_x clearly varies with x and so the explicit description of the potential is best given by using the natural basis of the larger space $\mathfrak{I}^m(S^+)$. Note that this only applies for $m \geqslant 2$ because for $m = 1$ the second space $\mathfrak{I}^{m-2} = 0$. These remarks explain the apparent singularity in the formula for $m = 2$ given in [4]: this occurs because a basis for $\mathrm{Ker}\,\varphi_x$ has been taken by projecting a sub-set of basis vectors of S^+, and different subsets are needed for different positions of x in R^4.

5. – Relation with Radon transform.

In this section we make yet another digression to explain how the Penrose transform is related to other well-known transforms.

We recall first the classical Radon transform which associates to a function f on R^3 its integral over affine planes. Thus to each f we obtain a transformed function φ defined on the space of planes π by:

$$(5.1) \qquad\qquad \varphi(\pi) = \int_\pi f .$$

The space of all planes in R^3 is a 3-dimensional manifold which can be identified with $P_3(R)$ with a point removed. There is an inversion formula which defines f in terms of an integral of φ. The Radon transform is closely related to the Fourier transform and is therefore useful for solving constant coefficient linear differential equations [20].

An obvious generalization of the Radon transform is to associate to f its integral over affine lines l in R^3. Thus we define φ by

$$(5.2) \qquad\qquad \varphi(l) = \int_l f .$$

It is a function on the space of all lines in R^3. This space is a subspace of the space of all lines in $P_3(R)$, namely the real Klein quadric (cf. Chapter III). This quadric has signature $(3, 3)$ and contains two systems of real projective planes. Our subspace is obtained by omitting one plane of one system.

In (5.1) and (5.2) the measure on π or l, with respect to which the integration is performed, is the usual Euclidean area or length respectively. Consequently both transforms are compatible with the group of Euclidean motions, this group acting naturally on the space of planes or lines. It is possible however to extend both transforms to the whole projective 3-space $P_3(R)$, compatibly with the action of $SL(4, R)$, provided we replace functions by appropriate types of density. In general a k-th root of a density will be called a k^{-1}-density. Then in (5.1) φ, f must be taken as $\frac{1}{4}$ and $\frac{3}{4}$ densities respectively, while in (5.2) φ, f must be $\frac{1}{4}$ and $\frac{1}{2}$ densities. More generally if we integrate over $(p-1)$-planes in R^{p+q-1} the relevant densities are $1/(p+q)$ and $p/(p+q)$.

In this extended form (5.2) therefore transforms the $\frac{1}{2}$-density f on $P_3(R)$ to the $\frac{1}{4}$-density φ on the real Klein quadric $Q_4(R)$. Since $\dim Q_4 = 4 > \dim P_3$ it is unreasonable to expect a universal inversion formula, as for (5.1). Instead one expects φ to satisfy some condition in order for it to be the transform of an f. In fact this condition is that φ be a solution of the conformally invariant second order equation $\Delta\varphi = 0$, where Δ is the indefinite analogue of the conformal Laplacian [21], [45]. Note that $Q_4(R)$ is the conformal compactification of the space R^4 with the $(2, 2)$ signature. Thus Δ is an « ultra-hyperbolic » operator, i.e. one which in flat space is given by

$$\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2} - \frac{\partial^2}{\partial x_4^2}.$$

The group $SL(4, R)$ acts on $Q_4(R)$ by conformal transformations and preserves the operator Δ.

In these transforms we have so far ignored regularity questions of f, φ. The precise correspondence holds either for C^∞ functions or for real analytic functions. Note that, unlike the positive definite case when Δ is elliptic, the equation $\Delta\varphi = 0$ does not imply analyticity of φ.

If we deal with analytic functions then the transform (5.2) can be looked at in the complexification. Thus, if we extend f as a complex analytic function to some neighbourhood of $P_3(R)$ in $P_3(C)$, then (5.2) becomes a Cauchy integral. This leads naturally to the sheaf cohomology occurring in the Penrose transform as we shall now explain.

To begin with let us recall that for the complex projective line $P_1(C)$,

we have

(5.3) $$H^1\big(P_1(C),\, \mathcal{O}(-2)\big) \cong C \;.$$

In section 1 this isomorphism was given in terms of explicit representative cocycles, and this can be made more intrinsic by recalling that $\mathcal{O}(-2) \cong \Omega^1$ the sheaf of holomorphic differentials on $P_1(C)$ (note that the tangent bundle of $P_1(C)$ is L^{-2}). If $\omega = f(z)\,dz$ is a holomorphic differential defined near the circle $|z| = 1$, we can take $\omega = \omega_{0\infty}$ to define a 1-cocycle relative to the standard covering of $P_1(C)$ by the open sets U_0, U_∞ given by $|z| < 1 + \varepsilon$, $|z| > 1 - \varepsilon$. The isomorphism (5.3) is then given by the contour integral

(5.4) $$\omega \to \frac{1}{2\pi i}\int \omega$$

taken around $|z| = 1$ (oriented in the conventional manner as the boundary of $|z| \leqslant 1$).

We now return to the transform (5.2) in which f, as explained earlier, should be taken as a $\frac{1}{2}$-density on $P_3(R)$. If we fix the orientation of $P_3(R)$ a density on $P_3(R)$ can be identified with an exterior differential 3-form and hence with a section of the real line-bundle L^4. Thus f can be viewed as a section of L^2 and complexified accordingly.

Next we fix a line l_0 in $P_3(R)$ and consider all nearby lines l, parametrized by a small open set $V \subset Q_4(R)$. The complexifications of the lines l fill out an open set U in $P_3(C)$ as one can verify. If we orient l_0, and extend this orientation by continuity to all nearby l, we can define open subsets U_0, U_∞ of U swept out by the (ε-enlarged) upper and lower hemispheres of each $P_1(C)$ in our family. If ε is small enough $U_0 \cap U_\infty$ will be close to $P_3(R)$ and hence within the domain of definition of f. Hence taking $f_{0\infty} = f$ we get an element

(5.5) $$(f) \in H^1\big(U,\, \mathcal{O}(-2)\big)\;.$$

By the Penrose transform, as explained in Section 2, (f) defines a $\frac{1}{4}$-density φ, on the appropriate open set of $Q_4(C)$, satisfying $\Delta\varphi = 0$. The value of φ at the point parametrizing a given $P_1(C)$ is given by restricting (f) to this $P_1(C)$ and using (5.3). Comparing (5.2) and (5.4) we see that they coincide, up to the factor $2\pi i$. Thus the Penrose description of solutions of $\Delta\varphi = 0$ via sheaf cohomology classes can be considered as the natural complex-analytic description of the integral transform (5.2).

Although the transform (5.2), taken globally on the whole of $P_3(R)$, is invertible (i.e. $\varphi \equiv 0 \Rightarrow f \equiv 0$), the same is not true locally. Thus if f, defined originally in U, extends holomorphically to U_0 or U_∞, or more

generally is the difference of two such extensions, then $(f) = 0$ and $\varphi \equiv 0$ in V. This clarifies the role of the sheaf cohomology group which precisely absorbs the local non-invertibility, in other words $\varphi \equiv 0$ in $V \Rightarrow (f) = 0$ in U.

It is perhaps worth emphasizing that the construction of the sheaf cohomology class (f) from the $\frac{1}{2}$-density f depended on orienting the lines l, so that U_0 and U_∞ could be unambiguously defined. Now for topological reasons we cannot coherently and continuously orient all real lines l in $P_3(R)$: the space $Q_4(R)$ is not simply-connected. Thus a global f does *not* define an element of the global cohomology group $H^1(P_3(C), \mathcal{O}(-2))$. This averts the contradiction that would otherwise arise from the vanishing of this cohomology group and the global invertibility of (5.3).

CHAPTER VII

Theorems on Algebraic Bundles

1. – Cohomology of the Horrocks construction.

In Chapter V we gave the construction of Horrocks for producing symplectic vector bundles on $P_3(C)$. In this chapter we shall prove a theorem due to Barth [8] which shows how the Horrocks bundles may be characterized in cohomological terms. Combined with the results of the preceding chapter on the vanishing of appropriate sheaf cohomology groups for instanton bundles this will finally prove that the Horrocks construction gives all instanton bundles.

We begin in this section by examining the Horrocks construction in greater detail and deducing its cohomological properties. At this stage, and also in the next section, we work purely over the complex numbers. Reality questions do not enter until later.

Throughout this chapter we shall be making extensive use of the machinery and standard results of sheaf cohomology which have only been lightly touched on in previous chapters. Inevitably therefore this chapter will be more technical, but we shall try to recall basic facts as and when they are needed. In any case, it is the arguments of this chapter which exhibit the full power of cohomology theory and the reader may find it instructive to try, where possible, to reinterpret the methods in real 4-space terms. We shall at the appropriate stage make some comments in this direction.

We recall that the Horrocks construction for $Sp(n, C)$-bundles, as explained in Chapter V, Section 2, starts from a linear map

$$(1.1) \qquad\qquad A(z)\colon W \to V$$

with appropriate properties. Here $\dim W = k$, $\dim V = 2k + 2n$, $z = (z_1, ..., z_4)$ and V has a non-degenerate skew-form. For each $z \neq 0$ the

image $U_z = A(z)W \subset V$ is assumed k-dimensional and isotropic, i.e. U_z is contained in its polar space U_z^0. We then put $E_z = U_z^0/U_z$ which gives a vector space of dimension $2n$ with an inherited non-degenerate skew-form. This is the fibre over $(z) \in P_3(C)$ of the required bundle E.

Since $A(z)$ is linear in z we can view (1.1) as a homomorphism of vector bundles over $P_3(C)$:

$$(1.2) \qquad\qquad A: W(-1) \to V$$

where $W(-1) = W \otimes L$. In other words the bundle $U \subset V$ is isomorphic to $W(-1)$. By duality V/U^0 is then isomorphic to $W^*(1)$. We can exhibit all bundles relevant to the construction in the following display of exact sequences:

$$(1.3)$$

$$
\begin{array}{ccccccc}
 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \\
0 \to & W(-1) \to & Q^* & \to & E & \to 0 \\
 & \| & \downarrow & & \downarrow & & \\
0 \to & W(-1) \xrightarrow{A} & V & \to & Q & \to 0 \\
 & & \downarrow{\scriptstyle A^*} & & \downarrow & & \\
 & & W^*(1) & = & W^*(1) & & \\
 & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & &
\end{array}
$$

where $Q = V/U$, $Q^* = U^0$. This diagram is self-dual, and the duality is skew in an appropriate sense, being induced by that of V.

In dealing with sheaf cohomology groups on $P_3(C)$ we shall, where no confusion can arise, omit the symbol $P_3(C)$ and simply write $H^q(S)$ for the cohomology of a sheaf S. The cohomology of the bundles E, Q in (1.3) is then given by

PROPOSITION (1.4).

$$H^0(Q) \cong V \qquad\qquad H^0(E(-n)) = 0 \quad \text{for } n \geqslant 1$$

$$H^0(Q(-n)) = 0 \quad \text{for } n \geqslant 1 \qquad H^1(E(-n)) = 0 \quad \text{for } n \geqslant 2$$

$$H^1(Q(n)) \quad = 0 \quad \text{for all } n \qquad\quad H^1(E(-1)) \cong W^* .$$

PROOF. The middle row of (1.3), together with the cohomology properties of $W(-1)$ and V give at once the results on Q [we use here the vanishing of $H^q(\mathcal{O}(n))$ for $q = 1, 2$ and all n, for $q = 0$ and $n < 0$]. Using this information on Q in the last column we then deduce the results on E.

As noted in Chapter V the second Chern class $c_2(E)$ can be read off from (1.3) and we get $c_2(E) = k$. More formally if x is the standard generator of $H^q(P_3, Z)$ the total Chern polynomials are given by:

$$c(W(-1)) = (1 - x)^k, \quad c(W^*(1)) = (1 + x)^k, \quad c(V) = 1$$

hence

$$c(Q) = (1 - x)^{-k}$$

and

$$c(E) = (1 - x)^{-k}(1 + x)^{-k} = (1 - x^2)^{-k} = 1 + kx^2$$

showing that $c_2(E) = k$.

Returning now to (1.3) let us use the last column to investigate $H^1(E(n))$ for $n \geqslant -1$. We see that

$$H^1(W^*(n + 1)) \twoheadrightarrow H^1(E(n))$$

is surjective. Using the identification of W^* given by (1.4) this can be replaced by the surjectivity of

$$H^1(E(-1)) \otimes H^0(\mathcal{O}(n + 1)) \twoheadrightarrow H^1(E(n)) .$$

Thus, if we introduce the graded module

$$M = \oplus M_n, \quad M_n = H^1(E(n))$$

over the ring of polynomials in $z_1, \ldots, z_4$, we see that M has the properties:

$$(1.5) \quad \begin{cases} M_n = 0 \text{ for } n < -1 \text{ and } n \text{ sufficiently large} \\ \dim M_{-1} = k \\ M_{-1} \text{ generates } M . \end{cases}$$

The properties (1.5) of the module M reflect the simple nature of the Horrocks construction. There are in fact vector bundles E given by more complicated constructions, in which A is not assumed linear in $z_1, \ldots, z_4$, for which the associated module M is more complicated. This is part of the general theory of Horrocks but fortunately for our purposes the simple case leading to (1.5) is sufficient.

2. – Theorem of Barth.

We shall now give Barth's theorem [8] which gives sufficient criteria on a symplectic bundle E in order that it should arise from the Horrocks construction. We consider first the following assumption on E:

(2.1)   *For some line l in P_3, the restriction of E to l is trivial.*

This assumption is related to the notion of semi-stability: for example when E is 2-dimensional (2.1) is equivalent to semi-stability and to the vanishing of $H^0(E(-1))$ [7]. Note that (2.1) implies automatically that E is trivial on the « general » line of P_3 (see Chapter IV). For bundles coming from S^4, via the basic construction of Chapter IV, (2.1) is certainly satisfied since E is trivial on all real lines, i.e. the fibres of $P_3(C) \to S^4$.

Our second assumption is that E satisfies the vanishing theorem:

$$(2.2) \qquad\qquad H^1(E(-2)) = 0$$

which is satisfied by bundles coming from the Horrocks construction (see (1.4)), and also for bundles corresponding to anti-self-dual $Sp(n)$-connections on S^4 by the results of Chapter VI.

Barth's theorem asserts that (2.1) and (2.2) are sufficient, namely

(2.3) THEOREM. *Let E be a symplectic vector bundle on $P_3(C)$ satisfying (2.1) and (2.2). Then E arises by the Horrocks construction from a linear map*

$$A(z)\colon W \to V$$

unique up to isomorphism.

The uniqueness in (2.3) means that, if (A, W, V) and (A', W', V') give isomorphic symplectic bundles, there are isomorphisms $W \to W'$, $V \to V'$ (preserving skew-forms) taking A into A'.

The idea of the proof of Theorem (2.3) is to show that the display (1.3) can be canonically reconstructed from E alone. As a preliminary we shall first show that the module $M = \bigoplus_n H^1(E(n))$ satisfies the conditions (1.5), a necessary consequence of (1.3).

Take any plane P_2 containing the line l, then E is trivial on the general line in this P_2 and so, for $n < 0$, $E(n)$ can have no non-zero section over this P_2. Now consider the standard exact sequence

$$0 \to E(-1) \to E \to E|P_2 \to 0 \ .$$

Tensoring with $\mathcal{O}(n)$ and taking cohomology gives the long exact sequence

$$\to H^0(P_2, E(n)) \to H^1(P_3, E(n-1)) \to H^1(P_3, E(n)) \to$$

Taking $n = -2, -3, \ldots$ in turn and using the vanishing of $H^0(P_2, E(n))$ we deduce inductively that $H^1(P_3, E(n)) = 0$ for all $n \leqslant -2$, which is the first part of (1.5). To prove the last part we take coordinates in P_3 so that l is given by $z_1 = z_2 = 0$. Then we have the exact sequence, resolving $\mathcal{O}_l$

$$(2.4) \qquad 0 \to \mathcal{O}(-2) \xrightarrow{\alpha} \mathcal{O}(-1) \oplus \mathcal{O}(-1) \xrightarrow{\beta} \mathcal{O} \to \mathcal{O}_l \to 0$$

where $\alpha = (z_2, -z_1)$ and $\beta = (z_1, z_2)$. The image of β is the ideal sheaf J of the line l. Tensoring (2.4) with $E(n)$, and taking cohomology we deduce the exact sequences

$$(2.5) \begin{aligned} &\to H^1(E(n-1)) \oplus H^1(E(n-1)) \to H^1(J(n)) \to H^2(E(n-2)) \to \\ &\qquad\qquad\qquad\qquad \| \\ &\to H^1(J(n)) \to H^1(E(n)) \to H^1(l, E(n)) \to 0 \ . \end{aligned}$$

Since E is trivial on l, $H^1(l, E(n)) = 0$ for $n \geqslant -1$, and by Serre duality $H^2(E(n-2))$ is dual to $H^1(E(-n-2))$ and so vanishes for $n \geqslant 0$ (as proved above): note that we have used $E \cong E^*$ here. Hence (2.5) asserts that the map

$$M_{n-1} \oplus M_{n-1} \xrightarrow{(z_1, z_2)} M_n$$

is surjective for all $n \geqslant 0$. This implies that M_{-1} generates M as a module (and is a somewhat stronger statement since only two of the variables are required). Finally to check the dimension of $M_{-1} = H^1(E(-1))$ we note that all the other cohomology groups $H^q(E(-1))$, $q \neq 1$, vanish. For $q = 2$ this follows from Serre duality and the vanishing of M_{-3}, for $q = 0$ it follows from the triviality of E on general lines and for $q = 3$ we apply Serre duality and reduce to the same argument. Thus $\dim M_{-1}$ can be computed from the Riemann-Roch theorem which evaluates

$$\sum_{q=0}^{3} (-1)^q \dim H^q(E(-1))$$

in terms of the Chern classes. Since $c_1(E) = 0$, $c_2(E) = k$ we get $\dim M_{-1} = a + bk$, where a, b are constants independent of E. These can either be obtained from the detailed Riemann-Roch formula or more simply we take

explicit examples arising from $-k$-instantons. Either way we find $\dim M_{-1} = k$ completing the verification of (1.5).

We turn now to the construction of the display (1.3) and we begin with the last column. Defining $W^* = H^1(E(-1))$ and using the interpretation of H^1 in terms of extensions (cf. Chapter VI) we construct the extension

$$(2.6) \qquad\qquad 0 \to E \to Q \to W^*(1) \to 0$$

corresponding to the identity element of

$$H^1\big(W \otimes E(-1)\big) = W \otimes W^* = \operatorname{End}(W) \,.$$

The effect of this is that, in the cohomology sequence of (2.6) tensored with $\mathcal{O}(-1)$

$$\to H^0(W^*) \overset{\delta}{\to} H^1\big(E(-1)\big) \to H^1\big(Q(-1)\big) \to H^1(W^*)$$

the coboundary δ is the identity. Since $H^1(W^*) = 0$ it follows that $H^1(Q(-1)) = 0$. If we tensor (2.6) in general with $\mathcal{O}(n)$ then for $n \leqslant -2$, we deduce

$$H^1\big(Q(-n)\big) \cong H^1\big(E(-n)\big) = 0$$

while for $n \geqslant 0$ we get

$$(2.7) \qquad W^* \otimes H^0\big(\mathcal{O}(n+1)\big) \overset{\delta}{\to} H^1\big(E(n)\big) \to H^1\big(Q(n)\big) \to 0 \,.$$

In (2.7) δ can be identified with the module multiplication from M_{-1} to M_n, and the surjectivity just proved (i.e. (1.5)) tells us that $H^1(Q(n)) = 0$. Thus we have proved

$$(2.8) \qquad\qquad H^1\big(Q(n)\big) = 0 \text{ for all } n \,.$$

We have now constructed the last column of (1.3) and established the key property (2.8) of the bundle Q. Dualizing we get the first row of (1.3). The cohomology sequence of this first row tells us that

$$H^1\big(Q^*(n)\big) \cong H^1\big(E(n)\big) \quad \text{ for all } n \,.$$

Interpreted in terms of extensions (with $n = -1$) this shows that there is an extension of Q^* by $W^*(1)$ compatible with each such extension of E. In particular this gives us the middle column of (1.3), for a suitable vector bundle V.

From the middle row of (1.3), using (2.8), we deduce

$$(2.9) \qquad\qquad H^1\big(V(n)\big) = 0 \quad \text{for all } n \,.$$

By Serre duality (2.8) implies that $H^2(Q^*(n)) = 0$ for all n, and hence from the middle column of (1.3) we get

$$(2.10) \qquad\qquad H^2\big(V(n)\big) = 0 \quad \text{for all } n \,.$$

From (2.9) and (2.10), we can now deduce that V is a trivial bundle. We require the following special case of the general theory of Horrocks [28].

PROPOSITION (2.11). *Let V be a vector bundle over P_3 such that $H^1\big(V(n)\big) = = H^2\big(V(n)\big) = 0$ for all n. Then V is isomorphic to a direct sum of line-bundles.*

PROOF. Fix $P_1 \subset P_2 \subset P_3$ and let

$$(2.12) \qquad\qquad V|_{P_1} \cong K|_{P_1}$$

where $K \cong L^{n_1} \oplus \ldots \oplus L^{n_m}$ is a sum of line-bundles. It will be enough to shows that the isomorphism φ of (2.12) extends to P_3 as a homomorphism, because the points where φ is not an isomorphism form an algebraic surface (local equation $\det \varphi = 0$) not meeting P_1 and so necessarily vacuous. We show that φ extends first to P_2 then to P_3 by using the exact sequences

$$0 \to V(n-1)|_{P_2} \to V(n)|_{P_2} \to V(n)|_{P_1} \to 0$$
$$0 \to V(n-1) \to V(n) \to V(n)|_{P_2} \to 0 \,.$$

The second sequence and the vanishing of $H^1\big(V(n)\big)$, $H^2\big(V(n)\big)$ shows that $H^1(P_2, V(n)) = 0$ and that every section of $V(n)$ over P_2 extends to P_3. The vanishing of $H^1(P_2, V(n)) = 0$ applied to the first sequence shows similarly that every section of $V(n)$ over P_1 extends to P_2. Since φ (or rather φ^{-1}) is a direct sum of sections $V(n_i)$, the proposition follows.

We now apply (2.11) to the bundle V constructed above. To prove finally that the integers n_i are all zero, i.e. that V is trivial, it is enough to restrict V to one line l. But when (1.3) is restricted to l, since $E|l$ is trivial, all the extensions split and it follows that $V|l$ is trivial.

We have now reconstructed the display (1.3) from E. It remains only to show that it is skew-symmetric and in particular that V has a canonical skew-form. But dualizing (1.3) and using the skew duality $E \cong E^*$ we get an isomorphic diagram, showing that $V \cong V^*$. With a little more care

(cf. [8]) one checks that this isomorphism is skew, completing the proof of Barth's theorem. The uniqueness of the display (1.3) follows from the canonical nature of our reconstruction, in which no arbitrary choices were involved.

It is clear from our proof that exactly the same argument works for an orthogonal bundle, i.e. a vector bundle E with a non-degenerate quadratic form. The duality of the corresponding diagram to (1.3) is now symmetric.

In conclusion we should point out that a more general theorem than (2.3) is given in [17] (and has been further generalized in [9]) which gives necessary and sufficient conditions for a bundle to arise from the Horrocks construction. Condition (2.2) is still kept but (2.1) is relaxed. In terms of the module M, conditions (1.5) are still fulfilled but it is not assumed that 2 of the 4 variables $(z_1, ..., z_4)$ are enough to generate M from M_{-1}. For instanton bundles condition (2.1) is satisfied and therefore the extra generality of [17] is not needed.

3. – Reality constraints.

We come now to the question of imposing the necessary reality constraints on the Horrocks construction to produce instanton bundles. In Chapter V, § 2 we imposed a Galois action σ on the triple (A, V, W), with $\sigma^2 = -1$ on V and $\sigma^2 = 1$ on W, which enabled us to obtain an $Sp(n)$-bundle over S^4 with anti-self-dual connection. We want now to show the converse, i.e. that every such connection arises in this way.

According to Chapter IV, § 2 every such connection corresponds to a holomorphic vector bundle E over $P_3(C)$ with

(i) a holomorphic symplectic structure

(ii) an anti-linear σ, compatible with (i), with σ on $P_3(C)$, satisfying $\sigma^2 = -1$, and inducing a positive hermitian form on sections on $\tilde{E}$ along all real lines of $P_3(C)$.

Since $\tilde{E}$ is trivial on all real lines of $P_3(C)$ and satisfies the vanishing condition $H^1(\tilde{E}(-2)) = 0$ (Chapter VI, § 3) we can apply Barth's theorem of the previous section and deduce that $\tilde{E}$ is constructed canonically from a triple (A, V, W). The anti-linear map σ then induces an antilinear map on V. This follows from the uniqueness part of Barth's theorem: we compare $\tilde{E}$ and the complex conjugate of $\sigma^*(\tilde{E})$. The uniqueness also shows that, on V, $\sigma^2 = -1$ and is compatible with the symplectic form.

To derive all the conditions imposed in Chapter V, § 2 we need finally to show that the hermitian form induced by σ on V is positive. Now recall

that, for any $(z) \in P_3(C)$ we have an orthogonal decomposition:

$$V = U_z \oplus E_z \oplus U_{\sigma z}$$

where $U_{\sigma z}^{\perp} = U_z^{\circ} = U_z \oplus E_z$. The hermitian form restricted to U_z is therefore definite. Applying σ shows that the sign of this definite form is the same in $U_{\sigma z}$ as in U_z. The form is positive on E_z because of its original definition. Hence the form is either positive on all V or else it has signature $(2n, 2k)$. But in the latter case the fibres E_z would lie in the positive cone and the bundle would be topologically trivial (deformable to a fixed positive subspace), contradicting the fact that $c_2(\tilde{E}) = k \neq 0$.

This completes the proof that the construction of Chapter V, made explicit in Chapter II, gives all anti-instantons for $Sp(n)$. Exactly analogous arguments work for the orthogonal group. Finally for the unitary group $U(n)$ we can embed this in $SO(2n)$ and consider an additional anti-linear J. Again using the uniqueness part of Barth's theorem we deduce that we get an operator J on V with the appropriate properties.

4. – The Drinfeld-Manin description.

In the proof of Barth's theorem in §2 we reconstructed the display (1.3) from the symplectic vector bundle E over P_3. The vector space W^* was identified with the cohomology group $H^1(E(-1))$. The vector space V can also be identified as a cohomology group in such a way that the linear map A (or its dual A^*) acquires a simple cohomological interpretation. This enables one to reformulate Barth's theorem rather more elegantly by directly exhibiting the triple (A, V, W) in terms of the cohomology of E. This reformulation has the advantage of making the additional reality conditions more transparent.

We shall explain how to interpret V cohomologically. For fuller details and a more systematic account see [18].

From the middle column of (1.3) we get the cohomology sequence:

(4.1) $0 \rightarrow H^0(Q^*) \rightarrow H^0(V) \rightarrow H^0(W^*(1)) \rightarrow H^1(Q^*) \rightarrow H^1(V)$.

Since V is a trivial bundle $H^0(V) \cong V$, $H^1(V) = 0$.

From the top row we have

$$H^q(Q^*) \cong H^q(E), \quad q = 0, 1$$

and finally we recall that

$$W^* \cong H^1\big(E(-1)\big) .$$

Substituting these into (4.1) we get

$$(4.2) \qquad 0 \to H^0(E) \to V \to H^1(E(-1)) \otimes H^0(\mathcal{O}(1)) \xrightarrow{\mu} H^1(E) \to 0$$

where μ can be identified with the natural module multiplication of the module M.

On the other hand we have a natural exact sequence on P_3:

$$0 \to \mathcal{O}(-1) \to C^4 \to T(-1) \to 0$$

where T is the tangent bundle. Dualizing and tensoring with $E(-1)$ gives

$$0 \to E \otimes \Omega^1 \to (C^4)^* \otimes E(-1) \to E \to 0 .$$

The cohomology sequence of this is

$$(4.3) \qquad 0 \to H^0(E) \to H^1(E \otimes \Omega^1) \to H^1\big(E(-1)\big) \otimes (C^4)^* \xrightarrow{\mu} H^1(E)$$

where μ is the natural multiplication $\big($recall that $H^0(\mathcal{O}(1))$ is the space of linear forms $(C^4)^*\big)$.

Comparing (4.2) with (4.3) suggests that there should be a natural isomorphism

$$(4.4) \qquad\qquad\qquad V \cong H^*(E \otimes \Omega^1) .$$

This is clear if $H^0(E) = 0$ which is the case for instanton bundles not containing a trivial summand (e.g. for $Sp(1)$-bundles and $k \neq 0$). In the general case V splits as a direct sum

$$(4.5) \qquad\qquad\qquad V \cong H^0(E) \oplus V'$$

using the skew-form on V. Now $H^1(E \otimes \Omega^1)$ has a natural skew-form given by the cup-product (and the skew-form on E):

$$(4.6) \qquad\qquad H^1(E \otimes \Omega^1) \otimes H^1(E \otimes \Omega^1) \to H^2(\Omega^2) \cong C$$

and so it decomposes analogously to (4.5). This gives us the isomorphism (4.4) and one must then check directly that this isomorphism is compatible with the skew-forms.

Thus the linear map A^* associated to E can be identified with the map

$$H^1(E \otimes \Omega^1) \to H^1(E(-1)) \otimes H^0(\mathcal{O}(1))$$

occurring in (4.3).

Any additional structure on E such as σ or J then passes naturally to these cohomology groups. Note that in the orthogonal case the multiplication (4.6) is symmetric.

Since $H^2(P_3, \Omega^2)$ maps isomorphically onto $H^2(P_2, \Omega^2)$ it follows that (4.6) factors through the restriction to P_2 where it coincides with Serre duality. The non-degeneracy of (4.6) therefore implies the injectivity of

$$H^1(P_3, E \otimes \Omega^1(P_3)) \to H^1(P_2, E \otimes \Omega^1(P_2)) \ .$$

Concerning restriction from P_3 to P_2 we draw attention to the following. Let P_2 and $\sigma(P_2)$ be considered as a degenerate quadric with real structure. In general for any real quadric $Q \subset P_3$ two real bundles E, F on P_3, corresponding to anti-instanton $Sp(n)$-bundles on S^4, are isomorphic if and only if their restrictions to Q are isomorphic. To see this we have to show that any isomorphism $\varphi: E \to F$ over Q extends as a homomorphism over P_3 (it is then necessarily an isomorphism by an argument used earlier). Now $G = \mathrm{Hom}\,(E, F)$ again corresponds to an anti-instanton bundle and so satisfies the vanishing condition $H^1(P_3, G(-2)) = 0$. Applied to the exact sequence

$$0 \to G(-2) \to G \to G|_Q \to 0$$

this shows precisely that every section of G over Q extends to a section over P_3. Taking $Q = P_2 \cup \sigma(P_2)$ we deduce that $E \cong F$ over P_3 (as real bundles i.e. commuting with σ) if and only if $E \cong F$ over P_2 and the isomorphism respects σ over the real line $P_1 = P_2 \cap \sigma(P_2)$. Since E, F are trivial over real lines the real structure over P_1 corresponds to a real structure or hermitian metric on C^{2n}. Thus we deduce that the map from the space of moduli of real $Sp(n)$ anti-instantons to the space of moduli of complex symplectic bundles over $P_2(C)$ has fibre the symmetric space $Sp(n, C)/Sp(n)$. For example when $n = 1$ this fibre is the hyperbolic 3-space. The $(8k - 3)$-manifold of $SU(2) - k$-instantons therefore fibres over an $8k - 6 = 2(4k - 3)$-manifold of bundles on $P_2(C)$. But a dimension count (cf. [3]) shows that $(4k - 3)$ is the complex dimension of the relevant space of (stable) bundles on $P_2(C)$. Hence, after factoring out by the hyperbolic 3-space, we get a complex structure on the real moduli space.

This result can be interpreted roughly as follows. Starting in R^4 we pick coordinates to identify it with C^2. Then an anti-self-dual connection gives a holomorphic bundle on C^2. If we require the bundle to extend to S^4 our holomorphic bundle extends to $P_2(C)$ and acquires a real structure at ∞. The surprising result is that this data uniquely determines the original solution: in other words it is enough to fix one set of complex coordinates if we work globally, whereas the main local result of Chapter IV was that we needed all complex coordinates to interpret the anti-self-duality equations. It would be very interesting to have a direct differential-geometric proof of this result. For this it would be necessary to prove the existence of a unique hermitian metric for the holomorphic bundle which satisfied appropriate conditions.

CHAPTER VIII

Further Problems

1. – Euclidean approach to instantons.

As we have seen, the proof that the Horrocks construction yields all instantons uses the full machinery of algebraic geometry. On the other hand the construction itself has a simple description in the Euclidean 4-space without resorting to the twistor picture. It is natural therefore to ask if we can produce an alternative proof of the completeness working only in the Euclidean 4-space. As a first step it seems necessary to describe the canonical data used in the Horrocks construction in Euclidean terms. In particular the vector spaces V and W should have such a description. We recall that, in sheaf cohomology terms, we have

$$W^* = H^1\big(E(-1)\big)$$

$$V \;= H^1(E \otimes \Omega^1)\,.$$

According to the results of Chapter VI we already know that $H^1\big(E(-1)\big)$ can be identified with the space of $E \otimes S^-$ satisfying the coupled Dirac equation (recall that with our sign conventions we produce anti-instantons): here S^- is the negative spin bundle over S^4. We shall now show how the other cohomology group can be similarly interpreted.

Using the standard metric of S^4 the basic anti-instanton bundle is precisely the bundle S^- and, when lifted to $P_3(C)$, it gives the holomorphic bundle orthogonal to the fibres as explained in Chapter V, § 1. Thus on $P_3(C)$ we have the exact sequence

$$0 \to \tilde{S}^- \otimes L^* \to T \to L^{-2} \to 0$$

or dualizing,

$$0 \to L^2 \to \Omega^1 \to \tilde{S}^- \otimes L \to 0\,.$$

Tensoring with E gives

$$0 \to E(-2) \to E \otimes \Omega^1 \to E \otimes \tilde{S}^-(-1) \to 0 \ .$$

Taking cohomology and using the vanishing of $H^q(E(-2))$ for $q = 1, 2$ (these are duals of each other) we see that

$$H^1(E \otimes \Omega^1) \cong H^1(E \otimes \tilde{S}^-(-1)) \ .$$

Thus $H^1(E \otimes \Omega^1)$ can be identified with the space of solutions of the Dirac equation on $(E \otimes S^-) \otimes S^-$ (we have for simplicity used the same symbol E for the holomorphic bundle on $P_3(C)$ and the corresponding anti-instanton bundle on S^4). Now sections of $S^- \otimes S^-$ can be naturally identified with pairs (f, ω) where f is a scalar function and ω is an anti-self-dual 2-form: this is a matter of examining the representations of Spin (4). Moreover the Dirac operator coupled to S^- can be identified with the Hodge operator $d + d^*$ on differential forms. The corresponding results hold if we now further couple to E provided d is replaced by its covariant analogue D. Thus we see that $H^1(E \otimes \Omega^1)$ can be identified with the space of pairs (f, ω) satisfying the equation

(1.1) $$Df = -D^*\omega$$

where f is a section of E and ω a section of $E \otimes \Omega^2_-$.

When E is irreducible it has no covariant constant sections except zero and so ω, in (1.1), uniquely determines f.

The geometrical description of the bundle E, arising from the Horrocks construction, gives it as a sub-bundle of the trivial bundle $S^4 \times V$. By orthogonal projection V then becomes identified with a certain space of sections of E. Conversely to find the embedding of E into $S^4 \times V$, it is enough to give this space of sections. It seems reasonable to conjecture that this space of sections is related to the pairs (f, ω) by interior and exterior products with the curvature F.

From this point of view one would have to show the appropriate non-degeneracy of this space of sections (so as to get an embedding of E into $S^4 \times V$). In addition one would have to verify that the original connection on E coincided with the connection induced from the embedding. This last part appears to be the hardest.

A direct interpretation of the self-dual Yang-Mills equations in terms of analytic function theory of a quaternionic variable has been discovered by F. Gürsey. It would be very interesting to investigate in detail the rela-

tion between this point of view and our complex methods. This might help to produce a proof of the completeness of the instanton construction in the Euclidean framework.

2. – General solutions of the Yang-Mills equations.

In these lectures I have concentrated on the instanton problems which correspond to absolute minima of the Yang-Mills functional on S^4. I shall now review what is known about general solutions of the Yang-Mills equations, corresponding to critical points which are not absolute minima.

First of all there is a recent result of Bourguignon, Lawson and Simons [11] which shows that there are no other local minima. It is at present unknown if other critical points exist, but [11] asserts that any such points must be unstable. The proof consists in showing that the second variation is indefinite.

The twistor interpretation of instantons does not immediately apply to other Yang-Mills solutions, but Witten [44] and Green et al. [24] have shown how to generalize Ward's ideas to the general case. However, the twistor interpretation is now more complicated and its potential has not yet been exploited.

Topological aspects of Yang-Mills theory related to ideas of Morse theory have been studied in [5]. In this and other respects there are close analogies with the non-linear σ model in 2 dimensions and these analogies suggest that no other critical points exist.

Finally, K. Uhlenbeck has recently shown [41] that square integrability on R^4 for Yang-Mills solutions automatically ensures the extension to S^4. More precisely there is a purely local result asserting that a smooth connection defined in a neighbourhood of 0 (but not at 0), which is locally square integrable and satisfies the Yang-Mills equations, automatically extends to a smooth connection defined at 0. In other words, point singularities are « removable ». Applied to the point at ∞, after a conformal transformation, this yields the extension from R^4 to S^4.

REFERENCES

[1] M. F. ATIYAH, *Complex fibre bundles and ruled surfaces*, Proc. London Math. Soc., **5** (1955), pp. 407-434.

[2] M. F. ATIYAH - V. G. DRINFELD - N. J. HITCHIN - YU. I. MANIN, *Construction of instantons*, Phys. Lett., **65** A (1978), pp. 185-187.

[3] M. F. ATIYAH - N. J. HITCHIN - I. M. SINGER, *Self-duality in four dimensional Riemannian geometry*, Proc. Roy. Soc. Lond., A **362** (1978), pp. 425-461.

[4] M. F. ATIYAH - R. S. WARD, *Instantons and algebraic geometry*, Commun. Math. Phys., **55** (1977), pp. 117-124.

[5] M. F. ATIYAH - J. D. S. JONES, *Topological aspects of Yang-Mills theory*, Commun. Math. Phys., **61** (1978), pp. 97-118.

[6] W. BARTH, *Moduli of vector bundles on the projective plane*, Inventiones Math., **42** (1977), pp. 63-91.

[7] W. BARTH, *Some properties of stable rank-2 vector bundles on P_n*, Math. Ann. **226** (1977), pp. 125-150.

[8] W. BARTH - K. HULEK, *Monads and moduli of vector bundles*, Manuscripta Math., **25** (1978), pp. 323-347.

[9] A. BEILINSON, *The derived category of coherent sheaves on P^N*, Funk. Analiz., **12** (1978).

[10] A. BELAVIN - A. POLYAKOV - A. SCHWARTZ - Y. TYUPKIN, *Pseudoparticle solutions of the Yang-Mills equations*, Phys. Lett., **59** B (1975), pp. 85-87.

[11] J.-P. BOURGUIGNON - H. B. LAWSON - J. SIMONS, *Stability and gap phenomena for Yang-Mills fields*, Proc. Nat. Acad. Sci. U.S.A., **76** (1979), pp. 1550-1553.

[12] S. COLEMAN, *The uses of instantons*, Lecture Notes, Erice Summer School 1977 (to be published by Cambridge University Press).

[13] N. CHRIST - N. STANTON - E. J. WEINBERG, *General self-dual Yang-Mills solutions*, Columbia University preprint, 1978.

[14] E. CORRIGAN - D. B. FAIRLIE, *Scalar field theory and exact solutions to a classical $SU(2)$-gauge theory*, Phys. Lett., **67** B (1977), pp. 69-71.

[15] E. CORRIGAN - D. B. FAIRLIE - P. GODDARD - S. TEMPLETON, *A Green's function for the general self-dual gauge field*, preprint (E.N.S., Paris).

[16] V. G. DRINFELD - YU. I. MANIN, *Self-dual Yang-Mills fields on the sphere* Funk. Analiz., **12** (1978), pp. 78-79.

[17] V. G. DRINFELD - YU. I. MANIN, *Locally free sheaves on CP^3 associated to Yang-Mills fields*, Uspekhi Mat. Nauk., **33** (1978), pp. 241-242.

[18] V. G. DRINFELD - YU. I. MANIN, *Instantons and sheaves on CP^3*, Funk. Analiz., **13** (1979), pp. 59-74.

[19] V. G. DRINFELD - YU. I. MANIN, *A description of instantons* (preprint, 1978).

[20] I. M. GEL'FAND - G. E. SHILOV, *Generalized Functions*, vol. I, Academic Press.

[21] I. M. GEL'FAND - M. I. GRAEV - Y. SHAPIRO, *Integral geometry on k-dimensional planes*, Funk. Analiz., **1** (1967), pp. 15-31.

[22] P. GODDARD - J. NUYTS - D. OLIVE, *Gauge theories and magnetic charge*, Nuclear Physics, B **125** (1977), pp. 1-28.

[23] H. GRAUERT - G. MÜLICH, *Vektorbündel vom Rang 2 über dem n-dimensionalen komplex-projektiven Raum*, Manuscripta Math., **16** (1975), pp. 75-100.

[24] P. S. Green - J. Isenberg - P. Yasskin, *Non-self-dual gauge fields*, Phys. Lett., **78** B (1978), pp. 462-464.

[25] P. A. Griffiths, *The extension problem in complex analysis.* - II: *Embeddings with positive normal bundle*, Amer. J. Math., **88** (1966), pp. 366-446.

[26] A. Grothendieck, *Sur la classification de fibré holomorphe sur la sphère de Riemann*, Amer. J. Math., **79** (1957), pp. 121-138.

[27] R. Hartshorne (to appear).

[28] G. Horrocks, *Vector bundles on the punctured spectrum of a local ring*, Proc, London Math. Soc., **14** (1964), pp. 684-713.

[29] N. J. Hitchin, *Linear field equations on self-dual spaces* (to appear in Proc. Roy. Soc.).

[30] R. Jackiw - C. Nohl - C. Rebbi, *Classical and semi-classical solutions to Yang-Mills theory*, Proceedings 1977 Banff School, Plenum Press.

[31] R. Jackiw - C. Nohl - C. Rebbi, *Conformal properties of pseudo-particle configurations*, Phys. Rev., **150** (1977), pp. 1642-1646.

[32] R. L. Mills - C. N. Yang, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev., **96** (1954), p. 191.

[33] A. Newlander - L. Nirenberg, *Complex analytic coordinates in almost complex manifolds*, Ann. of Math., **65** (1957), pp. 391-404.

[34] R. Penrose, *Zero rest-mass fields including gravitation: asymptotic behaviour*, Proc. Roy. Soc., A **284** (1965), pp. 159-203.

[35] R. Penrose, *The Twistor programme*, Rep. Mathematical Phys., **12** (1977), pp. 65-76.

[36] J. Rawnsley, *On the Atiyah-Hitchin vanishing theorem for certain cohomology groups of instantons bundles*, Math. Ann., **241** (1979), pp. 43-56.

[37] A. S. Schwartz, *On regular solutions of Euclidean Yang-Mills equations*, Phys. Lett., **67** B (1977), pp. 172-174.

[38] J.-P. Serre, *Géométrie algébrique et géométrie analytique*, Ann. Inst. Fourier, **6** (1956), pp. 1-42.

[39] J.-P. Serre, *Un théorème de dualité*, Comm. Math. Helv., **29** (1955), pp. 2-26.

[40] G. 't Hooft, *On the phase transition towards permanent quark confinement.*

[41] K. Uhlenbeck, *Removable singularities in Yang-Mills fields*, Bull. Amer. Math. Soc., **1** (1979), pp. 579-581.

[42] R. S. Ward, *On self-dual gauge fields*, Phys. Lett., **61** A (1977), pp. 81-82.

[43] H. Weyl, *Gravitation und Elektrizität*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, **26** (1918), pp. 465-480.

[44] E. Witten, *An interpretation of classical Yang-Mills theory*, Phys. Lett., **77** B (1978), pp. 394-398.

[45] F. John, *The ultrahyperbolic differential equation with four independent variables*, Duke Math. J., **4** (1938), pp. 300-322.