Vol. B10 (1979) ACTA PHYSICA POLONICA No 3

A SHORT INTRODUCTION TO THE MATHEMATICAL FORMULATION OF GAUGE THEORY

BY TSOU SHEUNG TSUN

Mathematical Institute, Oxford University*

(Presented at the XVIII Cracow School of Theoretical Physics, Zakopane, May 26 — June 8, 1978)

Some mathematical concepts used in the formulation of gauge theory are gathered together with emphasis on their physical and geometric interpretation.

For the last couple of years or so many physicists have been convinced that the language of fibre bundles is the correct language to use in gauge theory [1]. A very simple case where a non-trivial bundle occurs is that of the potential surrounding a magnetic monopole, as shown by the following theorem.

Theorem. Consider a magnetic monopole of strength $g \neq 0$ at the origin. Then there does not exist a vector potential A for the monopole magnetic field that is everywhere defined and singularity-free on a sphere of arbitrary radius surrounding the monopole.

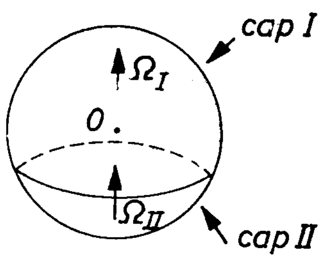

Fig. 1. A sphere surrounding a magnetic monopole at the origin

Proof. Suppose such a potential A exists. Then divide the sphere into two caps I and II by a parallel, Fig. 1. Then by Stokes' theorem, the total magnetic flux through cap I is

$$\Omega_{\mathrm{I}} = \oint A_\mu dx^\mu.$$

Similarly

$$\Omega_{\mathrm{II}} = \oint A_\mu dx^\mu,$$

* Address: Mathematical Institute, Oxford University, 24–29 St Giles's, Oxford OX1 3LB, England.

where the contour is taken around the same parallel in both cases. Thus the R.H.S. give zero, while by assumption the total flux out of the sphere $4\pi g$ is non-zero; hence a contradiction.

Since the sphere is arbitrary, we get Dirac's string of singularities. But these singularities are not physical. The trouble lies in trying to build a potential out of a *trivial bundle*. We shall come back to this example later.

Wu and Yang [2] gave a translation table between the language of gauge theory and that of fibre bundles, an adaptation of which is given in Table I. The essential point is to exhibit a one-to-one correspondence between the prominent ingredients in both theories.

TABLE I

Translation table between gauge field terminology and fibre bundle terminology, adapted from Ref. [2]

Gauge field terminology	Fibre bundle terminology
space of phase factors	bundle space
spacetime	base space
gauge group	structure group
gauge type	principal bundle
gauge potential	connection on a principal bundle
field strength	curvature of the connection
electromagnetism	connection on a U(1)-bundle
electromagnetism without monopole	connection on a trivial U(1)-bundle
electromagnetism with monopole	connection on a non-trivial U(1)-bundle
Dirac's monopole quantization	classification of U(1)-bundles according to first Chern class
Yang-Mills theory	connection on an SU(2)-bundle
instanton	real algebraic bundle on CP_3
instanton number	second Chern class

In gauge field theory we have spacetime over which a gauge field is defined via a matrix (i.e. Lie-algebra)-valued potential, enjoying a gauge freedom given by a gauge group. In fibre bundle theory we have a base space over which a bundle space is built, with a structure group acting on each fibre of the bundle space, and, in the case of a principal fibre, a connection which is a Lie-algebra-valued one-form that gives rise to the curvature two-form. From this parallelism one can easily see that a gauge theory with gauge group G can be described by a principal fibre bundle over (a model of) spacetime with the same group G as structure group.

For the convenience of readers who are not familiar with such notions, we have gathered below some of the fundamental definitions, adapted from various standard textbooks [3].

Definition 1. A *coordinate bundle* B is a collection of the following:

1) a topological space E, called the bundle (or total) space,

2) a topological space X, called the base space,
3) a map $\pi: E \to X$, called the projection,
4) a topological space Y called the typical fibre,
5) a topological transformation group G of Y, called the (structure) group of the bundle, such that $gy_1 = gy_2$, for all $y_1, y_2 \in Y$, implies that g is the identity element of G,
6) a family $\{V_j\}_{j\in J}$ of open sets covering X, the V_j's being called coordinate neighbourhoods (or patches), and
7) for each $j \in J$, a homeomorphism

$$\phi_j : V_j \times Y \to \pi^{-1}(V_j),$$

called the coordinate function. The set $\pi^{-1}(x)$ is called the fibre above x.

The coordinate functions are required to satisfy the following conditions:

8) $\pi\phi_j(x, y) = x$, for all $x \in V_j$, $y \in Y$,
9) if the map $\phi_{j,x}: Y \to \pi^{-1}(x)$ is defined by

$$\phi_{j,x}(y) = \phi_j(x, y),$$

then for each pair $i, j \in J$, and each $x \in V_i \cap V_j$, the homeomorphism

$$\phi_{j,x}^{-1}\phi_{i,x} : Y \to Y$$

coincides with the action of a (unique by (5)) element of G, and

10) for all pairs $i, j, \in J$, the map

$$g_{ji} : V_i \cap V_j \to G$$

defined by $g_{ij}(x) = \phi_{j,x}^{-1}\phi_{i,x}$, called a coordinate transformation or transition function, is continuous.

Remarks (*i*) In cases of physical interest, we are interested in the case in which X is spacetime, a differentiable manifold. In the above definition, we then replace "topological space" by "differentiable manifold", "topological group" by "Lie group", and "continuous map" by "differentiable map". Here for simplicity, "differentiable" means C^∞-differentiable.

(*ii*) The above definition depends on a given covering, or, in the case of manifolds, on local coordinates. In the next definition we get rid of this undesirable dependence.

Definition 2. Define an equivalence relation between bundles B and B', by saying that B is equivalent to B', in symbols $B \sim B'$, if and only if they have the same bundle space, base space, projection, typical fibre and group, and their coordinate functions $\{\phi_j\}$, $\{\phi'_k\}$ satisfy the conditions that

$$\bar{g}_{kj}(x) = \phi'^{-1}_{k,x}\phi_{j,x}, \quad x \in V_j \cap V'_k$$

coincides with the action of an element of G, and the map

$$\bar{g}_{kj} : V_j \cap V'_k \to G$$

so obtained is continuous (or differentiable as the case may be). Under such an equivalence, a *fibre bundle* is defined to be an equivalence class of coordinate bundles.

Remark: notwithstanding this abstract definition, when we speak about a particular fibre bundle, we shall often refer to one representative, i.e. one particular coordinate bundle in the equivalence class. The important fact is that the concept of a fibre bundle is independent of the particular local coordinate system one may choose to work with.

Definition 3. A *principal fibre bundle* is a fibre bundle in which the typical fibre Y coincides with the group G, and G acts on itself by left translation.

Definition 4. A *trivial bundle* is one in which $E = X \times Y$.

Definition 5. A (cross) section s of a fibre bundle is a continuous map

$$s : X \to E,$$

such that $\pi \circ s =$ identity on X.

Before we go on to give some examples illustrating the above definitions, it would perhaps be helpful to paraphrase these definitions in a more intuitive way. The (Cartesian) product of two spaces X and Y occurs naturally when one wants to represent the function $y = f(x)$ as a graph. In this sense a fibre bundle is just a generalization of ordinary products, and sections are the objects corresponding to graphs. Put another way, fibre bundle language gives us a means of coordinatizing locally certain spaces that are not naturally a product of two spaces. Looked at in a small enough neighbourhood of any arbitrary point, the total space is a product. As one varies from point to point, these products are "patched up" in a way compatible with the continuity requirements of the theory (i.e., continuous, differentiable, etc.). To picture a non-trivial bundle as distinct from a trivial one it is useful to have in mind the difference between a Möbius band and an ordinary ring.

Now for some examples.

1) Trivial or product bundle: here one has

$$E = X \times Y, \qquad \pi(x, y) = x.$$

One needs just one coordinate patch, and the transition function is the identity. The group G is reduced to the identity element. Sections are just graphs of maps of X into Y.

2) If G is a Lie group and Y is a closed Lie subgroup of G, then one can form the quotient $X = G/Y$, with the natural projection

$$\pi : G \to X.$$

This gives a principal fibre bundle.

3) Let X be a manifold, E the set of all tangent vectors at all points of X, and π assign to every tangent vector its initial point in X. Then each fibre $\pi^{-1}(x)$ is the tangent space at x, denoted by $T_x(X)$. The resulting fibre bundle is called the *tangent bundle*, denoted $T(X)$. Each fibre is a vector space (of the same dimension), but since there is no unique way of mapping one vector space to another of the same dimension, G in this case is the full linear group. A section is just a vector field over X, see Fig. 2. The tangent bundle always admits a section. The zero section is an example.

The last example is one of vector bundles. Bundles are usually classified according to the typical fibre Y, and a *vector bundle* is one in which the fibres are vector spaces and the group is the corresponding general linear group. An important special case is the line bundle, where each fibre is just the complex line C. Vector bundles will be important when we come to the self-dual Yang-Mills theory.

So far we have gone through the minimum of concepts in topology. We still need some geometry. A "connection" is needed to transport from one point of a manifold to another objects such as vectors, tensors and values of wavefunctions. In bundle language,

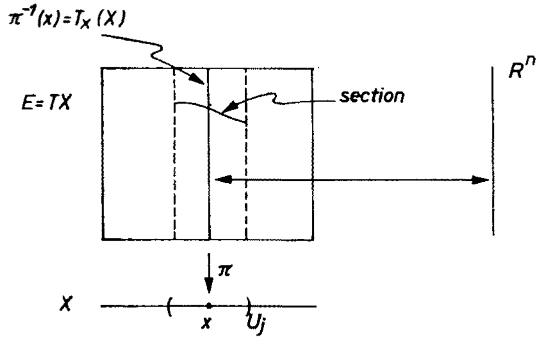

Fig. 2. Illustration of the tangent bundle over spacetime

we want to be able to go from fibre to fibre. (Along a fibre we know how, by virtue of the group action.) A connection can be defined in several ways. Using local coordinates one can define the coefficients of a *connection* Γ^i_{jk} on a manifold in terms of the covariant differentiation ∇u, where u is a vector field on X, satisfying the usual linearity and Leibnitz rules. If $\{e_i\}$ is a field of frames on X, then Γ^i_{jk} is given by

$$\nabla e_i e_j = \Gamma^k_{ij} e_k.$$

In particular for a Riemannian manifold one can take Γ^k_{ij} to be the Christoffel symbols.

Let us go back to the magnetic monopole. As we said, our trouble comes from trying to build a field out of a trivial bundle. By a theorem in topology we know that the U(1)-bundle is trivial if X is contractible. One way out is to assume that spacetime is not contractible, in particular, not homeomorphic to R^4 globally. One model [4] is to take $X \simeq R^2 \times S^2$ and to construct non-trivial U(1) bundles over S^2. These bundles are known and are characterized by integers. One simple example is the Hopf bundle:

$$U(2) \to U(2)/U(1) = S^3 \to S^2 \simeq CP_1.$$

The connection in terms of the Euler angles is

$$\omega = \frac{i}{2}(d\chi + \cos\theta d\varphi),$$

and the corresponding electromagnetic field

$$F = \tfrac{1}{2}\sin\theta d\varphi \wedge d\theta$$

describes a magnetic monopole of strength $g = \frac{1}{2}$.

Having gone through the basic concepts let us look at gauge theory, and instantons in particular. Instantons are the minimum action solutions of the SU(2) Yang-Mills fields in euclidean 4-space R^4 [5]. By imposing certain convergence conditions at infinity the problem is equivalent to working on the 4-sphere, which is the conformal compactification of R^4. If G is a simple Lie group, then the principal G-bundles over S^4 are classified by the third homotopy group $\pi_3(G)$, which is the integers Z. Hence we can label the principal bundles by an integer k:

$$\begin{array}{l} P_k \\ \downarrow G. \\ S^4 \end{array}$$

By choosing the orientation one can always arrange for $k > 0$. The case $k = 0$ is trivial. Thus *topology gives* us a number k, the so-called *instanton number*.

Before we can proceed further we have to feed in some differential geometry. Bundles in gauge theory are bundles with a connection A (and the corresponding curvature F). Having a connection enables us to parallelly propagate vectors around in spacetime.

Now the geometry of spacetime is well studied, especially by relativists. There is one theory, the twistor theory [6], which looks at spacetime from a complex geometric point of view and is particularly suitable for dealing with massless spacetime fields. In fact, it was through Penrose's construction of the non-linear (self-dual) graviton [7] that Ward discovered his construction of the self-dual Yang-Mills solutions [8]. Let us have a closer look at twistors.

The basic philosophy of twistor theory is that complex numbers are more basic than real numbers and that massless particles more fundamental than massive ones. The former is borne out by the essential use of complex numbers in quantum mechanics and the latter fits in well with the sort of gauge theories considered here. Massless particles, moreover, are tied up with the conformal properties of spacetime, because, in a sense, scales do not matter.

Let us take first flat Minkowski space, i.e. "ordinary" spacetime with signature $(- - - +)$. Next we compactify it in a way that respects its conformal structure. The resulting manifold M has the topology of a torus, $S^3 \times S^1$. The conformal group C(1, 3) of M preserves null geodesics. Consider its identity-connected component, and denote it by $C_+^\dagger(1,3)$. We have the following local isomorphism:

$$SU(2,2) \xrightarrow{2-1} SO(2,4) \xrightarrow{2-1} C_+^\dagger(1,3). \tag{1}$$

The group SU(2, 2) acts on C^4. *Twistors* are representations of SU(2,2), and hence twistors of the lowest valence are given by four complex numbers Z^α, $\alpha = 0, 1, 2, 3$. In other words, twistors arise from the local isomorphism between the group SU(2,2) and the conformal group, and they are tailored to Minkowski space in the sense that the local isomorphisms (1) do not generalize to higher dimensions nor other signatures. This is a unique feature of twistors.

A more geometric way of looking at twistors is to complexify M into a complex manifold CM (still compact), which can be embedded as a quadric Q in CP_5, 5-dimensional complex projective space. Furthermore, this quadric Q can be identified, via the Klein representation, with lines in CP_3. On the other hand, consider the space of null geodesics of M, which, as a real manifold, is $S^2 \times R^3$, but which can be embedded as a hyperplane PN in the complex manifold CP_3. Now recall that twistors are given by four complex numbers. If we divide out by the (complex) scale, i. e., take "homogeneous coordinates", then we get the space of projective twistors PT, which can be shown to be precisely CP_3. Hence we have the following correspondence:

$$\{\text{point in } CM\} \Leftrightarrow \{\text{lines in } PT\}$$

$$\{\text{null geodesics in } M\} \leftrightarrow \{\text{points in } PN\}$$

This is pictorially shown in Fig. 3.

The significance of the preceding construction, as far as gauge theories are concerned, is to enable us to study geometric objects over projective space instead of fields over

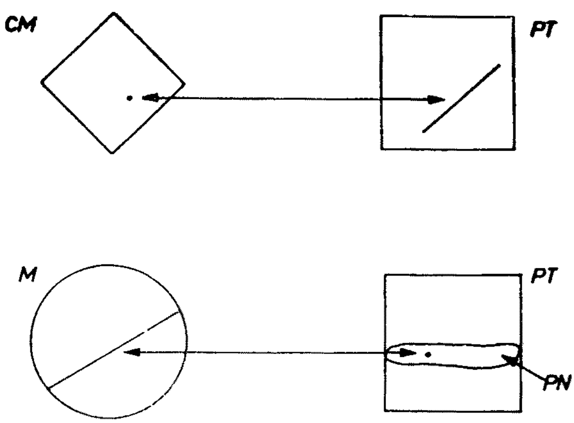

Fig. 3. Correspondence between geometric objects in spacetime and those in PT

Minkowski space. Mathematically this is a great advantage. Thus (anti-)self-dual Yang-Mills fields are in one-to-one correspondence, via the Penrose twistor transform, with certain complex two-dimensional holomorphic vector bundles over regions of CP_3, which we shall specify presently.

Since we are working in complexified Minkowski spacetime CM, we have to complexify the group SU(2) as well. This gives SL(2, C). The gauge potential A_a gives a connection ∇_a on an SL(2, C)-bundle P over CM. The covariant derivative operator is given by

$$D_a = \nabla_a + A_a,$$

and the curvature, which is the field strength, is

$$F_{ab} = 2\nabla_{[a}A_{b]} + [A_a, A_b]. \tag{2}$$

In this notation, the Yang-Mills equations can be written as

$$\nabla^a F_{ab} + [A^a, F_{ab}] = 0. \tag{3}$$

If we define $*F_{ab}$ the dual of F_{ab} by

$$*F_{ab} = \frac{1}{2}\varepsilon_{abcd}F^{cd},$$

where ε_{abcd} is the totally skew symmetric tensor, then the Bianchi identities are given by

$$\nabla^{a*}F_{ab}+[A^a, *F_{ab}] = 0. \tag{4}$$

Hence if the field is anti-self-dual, i. e.

$$*F_{ab} = -iF_{ab}, \tag{5}$$

the Yang-Mills equations are automatic consequences of the Bianchi identities.

We note that this treatment covers both the real Minkowski and the Euclidean cases, since both are real slices of CM. If we now specialize to the Euclidean case, i. e. to the so-called instanton solutions, then we consider vector bundles defined over the whole of CP_3[1], and put some reality conditions on them to recover the group SU(2) and not just SL(2, C). Such bundles have been studied by Horrocks [10], and the whole problem can now be reduced to linear algebra [9]. In particular, a quick parameter count shows that the dimension of the space of solutions is $8k$-3, where k is the instanton number.

Mathematically the linear algebra construction referred to is satisfying indeed. It has been generalized to the symplectic and orthogonal groups. The outstanding question for mathematicians is, it seems, the topology of the so-called space of moduli, the $8k$-3 dimensional manifold which represents the solution space of the self-dual Yang–Mills field. It appears that its topology is not at all well understood. In fact, for $k > 1$, it is not known whether the space is connected or not [11].

For physicists, however, the outstanding problem is to find general, i. e. non-self--dual, solutions of the Yang-Mills field. A step towards this direction has been made by considering bundles over PT, considered as the diagonal in $PT \times PT$, that can be extended to bundles over a certain type of neighbourhood of the diagonal [12]. However, beyond a purely formal construction nothing much is known yet.

I have greatly benefited from conversations and correspondence with Professors R. Penrose, A. Trautman and Dr. R. S. Ward, to whom I wish to express my gratitude. I would like also to thank Professor A. Białas and the organizers of the XVIIIth Kraków School for their invitation and kind hospitality. Finally, I wish to thank Professor J. Prentki for hospitality at the Theory Division at CERN, where most of this article was written.

REFERENCES

[1] As an example, see C. N. Yang *Annals of the New York Academy of Sciences* **294**, 86 (1977) also Ref. [2].

[2] T. T. Wu, C. N. Yang, *Phys. Rev.* **D12**, 3845 (1975).

[3] N. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press 1951; F. Hirzebruch, *Topological Methods in Algebraic Geometry*, Springer Verlag, Berlin 1966; A. Lichnerowicz, *Théorie globale des connexions et des groupes d'holonomie*, edizione Cremonese, Rome 1955.

---

[1] Actually some extra conditions are required to get rid of pathological cases, see Ref. [9].

[4] A. Trautman, *Int. J. Theor. Phys.* **16**, 561 (1977).
[5] There are many references. For a mathematical treatment, see M. F. Atiyah, R. S. Ward, *Commun. Math. Phys.* **55**, 117 (1977).
[6] R. Penrose, M. A. H. MacCallum, *Phys. Reports* **6C**, 241 (1972). For more up-to-date results, see recent issues of Twistor Newsletter, Mathematical Institute, Oxford University.
[7] R. Penrose, *Gen. Relativ. Gravitation* **7**, 31 (1976).
[8] R. S. Ward, *Phys. Lett.* **61A**, 81 (1977).
[9] M. F. Atiyah, N. Nitchin, V. G. Drinfeld, Yu. I. Manin, *Phys. Lett.* **65A**, 185 (1978).
[10] G. Horrocks, to be published. See also R. Hartshorne, *Commun. Math. Phys.* **59**, 1 (1978).
[11] M. F. Atiyah, private communication.
[12] E. Witten, Harvard preprint HUTP–78/A009, 1978; P. Yasskin, J. Isenberg, P. Green, University of Maryland preprint 1978.