VECTOR BUNDLES AND CONNECTIONS IN PHYSICS AND MATHEMATICS: SOME HISTORICAL REMARKS

V. S. Varadarajan

To Seshadri, for his seventieth birthday

1. Introduction. I have known Seshadri for many years. I still remember the first time when I came across his name in a Séminaire Chevalley volume on Variétés de Picard. I was at the Indian Statistical Institute at that time and was browsing in the library one day when I noticed that a certain Conjeevaram S. Seshadri had given a series of talks in that seminar. I myself had been born in Conjeevaram (or Kancheepuram, to give the Sanskrit name for this ancient and historic South Indian city), and so the fact that someone, who had grown up in the same milieu as I, had reached a high level in the world of mathematics was not only a thrill but a big source of inspiration for me. Of course it was only many years later that I came to understand a little more of what he had done and the reason why he is regarded with a lot of respect and admiration.

Although a large part of Seshadri's work belongs to the theory of vector bundles on curves, his total work is much broader and has had a substantial impact in many areas. Concerning the vector bundle theory, his famous papers, with M. S. Narasimhan[1] and by himself[2], inaugurated the modern theory of holomorphic vector bundles on compact Riemann surfaces. He then extended the whole theory to the algebraic context in any characteristic. He was the first one to give substantial evidence to the Mumford conjecture on geometric reductivity of reductive algebraic groups by proving it for $GL(2)$ (it was fully proved by Haboush a little later). His work with Lakshmibai and others on flag manifolds of the classical groups extended the classical results on the grassmannians and pioneered the so–called standard monomial theory. His lectures and articles and influential scholarship have had a deep impact on his colleagues and students. After he left the Tata Institute of Fundamental Research where he had been a major force for many decades, he came back to Madras to organize a strong group of young and active mathematicians, working under the auspices of the SPIC, a private consortium of industrial companies. Unlike many of us who went abroad, he remained in India

and tried to exert himself in creating and maintaining a visible and active Indian school of mathematics. All of these achievements have been recognized by many awards and honors, both from India and from other countries. It is a privilege to be asked to contribute to this volume in celebration of his seventieth birthday and I am very happy that I have an opportunity to say something on this occasion.

I had much hesitation when Sridharan asked me to contribute an article to this volume because my own work does not have much overlap with Seshadri's. In the end I decided that it may be of some interest to explain, in a manner partly historical and partly pedagogical, how the concept of vector bundles and connections on them evolved from the point of view of both physics and mathematics. Although the matters I propose to discuss are perhaps only tangential to the main themes in Seshadri's work, they do touch upon them in a surprising manner as we shall see at the end.

Before I proceed it is important to have at least an informal understanding of the nature of the objects we will be dealing with. Roughly speaking, a *vector bundle* is a vector space depending on a parameter which varies on a manifold M; it is then said to be defined *over* M. It is assumed that we can, over small pieces of M, take bases that vary nicely with the parameter; here nicely refers to some smoothness, for instance C^∞ or holomorphic, depending on the context, and the vector bundle will be correspondingly called C^∞ or holomorphic.

C^∞ vector bundles arise naturally in modern physics, especially when the manifold M is spacetime or some extension of it. When M is spacetime, its points may be thought of as representing the locations of particles. These particles obey the laws of quantum physics and therefore have a much richer internal structure than their classical counterparts. For instance they have properties such as spin, isospin, color, charm, etc. These internal properties are encoded by a vector space equipped with an action by a unitary group. When the particle moves, it takes this internal vector space with itself. It is then clear that a mathematical treatment involving these particles has to be in the context of vector bundles over M. However the appearance of vector bundles dates back to a few years before the discovery of quantum theory, when such internal structures of particles were not known. So tracing the development of this concept has some interest.

Even in the informal approach that we have taken above, it must be admitted that the concept of a vector bundle is not very intuitive, although the physical example given above certainly does make it more acceptable. In differential geometry a manifold has vector bundles naturally defined over it–the bundle of tangent vectors, tensors, especially exterior differential forms, and so on. But there are also other situations where vector bundles arise that have nothing to do with these

geometrically derived bundles. One of the most interesting examples of a "non–geometric" vector bundle is obtained when we consider a system of linear first order partial differential equations which satisfy the Frobenius condition of integrability. The equations are of the form

$$\frac{\partial u}{\partial x^\mu} + A_\mu u = 0 \qquad (1 \le \mu \le n)$$

where the A_μ are smooth $r \times r$ matrix valued functions defined on an open set $M \subset \mathbf{R}^n$ satisfying the Frobenius condition

$$F_{\mu\nu} := A_{\nu,\mu} - A_{\mu,\nu} + [A_\mu, A_\nu] = 0 \qquad (1 \le \mu, \nu \le n)$$

on M, and u is the unknown function whose values are column vectors having r components. Here, as usual,

$$A_{\mu,\nu} = \frac{\partial A_\mu}{\partial x^\nu}, \quad [A_\mu, A_\nu] = A_\mu A_\nu - A_\nu A_\mu$$

The holomorphic version with M an open set in $\mathbf{C}^n$ and A_μ holomorphic is even more important for us later. The Frobenius condition is necessary and sufficient that for any point $m \in M$ and any r–vector $u_0 \in \mathbf{C}^n$ there is a solution u of this system in a neighborhood of m with $u(m) = u_0$. At each point of M we can then attach the r–dimensional vector space of initial values of the local solutions at that point and obtain a natural vector bundle on M. Clearly this vector bundle is quite different in nature from the geometric bundles of tangent vectors and tensors. Notice that when M has dimension 1, we have ordinary linear differential equations and there is no need of any integrability condition.

To use this notion of a vector bundle in a free and flexible manner it is necessary to have an apparatus of differential calculus on them. For instance one should be able to differentiate a vector that varies with the vector space in a smooth manner. It is not enough to differentiate the components of this vector in a basis that one is allowed to choose because these bases do not have any intrinsic significance. In fact, any intrinsic definition of derivative of a varying vector implies that one is able to compare the vectors attached to distinct but neighboring points of the manifold M. Thus what is needed to construct a coordinate invariant differential calculus is the existence of isomorphisms between the vector spaces at neighboring points, indeed, points that are infinitely near each other. Such a structure is called a *connection* on the vector bundle.

This article develops two themes. The first, taking up §2–§5, explains how the attempts to understand fundamental issues in physics played a truly important role

in the evolution of the idea of a C^∞ vector bundle on a manifold and the concept of connections on such bundles. These ideas eventually led to the modern concept of a *gauge field* which dominates much of the high energy physics of today. Indeed, it is now universally accepted that a consistent field theory of elementary particles and their interactions must be a gauge theory. But in the early days this was not so obvious, and some very remarkable ideas of Weyl, Dirac, Aharanov–Bohm, and Yang–Mills were responsible for the decisive emergence of this view point[3], as I shall explain in §2–§5. The second theme is taken up in §6 where I discuss the evolution of the concept of *holomorphic* vector bundles whose genesis is in function theory and differential equations on Riemann surfaces. On the surface these two themes, one coming from differential geometry and physics and the other from function theory, appear to have no common aspect. But remarkably they do have a deep interaction and I shall explain at the end of the article how the physics and mathematics come together in surprising ways.

I shall try to give a short account of these matters; however the reader should be warned that this account is far from being complete. Indeed, the topic can be treated adequately only in a course of several lectures. I should also put in a disclaimer to the effect that I am not a professional historian of science and so the reader should look upon this effort only as a pedagogical one aiming to trace briefly the development of the basic concepts of vector bundles and connections. To keep my discussion as self–contained and elementary as possible and to make the ideas accessible a wider audience, especially to students, I have gone over many things that are familiar to experts. I hope I will be forgiven for this.

2. The work of Weyl and his discovery that the electromagnetic vector potential is a connection on a suitable bundle on spacetime and the electromagnetic field is the curvature of this connection.

The creation of the theory of special and general relativity by Einstein in the years 1905–1916 had a profound impact on mathematics and mathematicians. The astonishing fact that gravitation is just a manifestation of the curvature of spacetime made a deep impression on mathematicians like Elie Cartan and Hermann Weyl. This was especially the case with Weyl. Weyl's initial work, as is appropriate for a student of Hilbert at that time, was concerned with analysis and spectral theory. By 1916 he had made profound contributions to the spectral theory of differential operators, the asymptotic distribution of their eigenvalues, and to questions of uniform distribution mod 1. But he had also pursued simultaneously a geometrically motivated line of thought, as can be inferred from the fact that his epoch–making book *Die Idee der Riemanschen Fläschen* was published in 1913. It is clear that his mind, which was universal in its scope and tremendously attracted to fundamental questions of mathematics and physics that had a deep philosophical component to

them, was fascinated by the role of differential geometry in the understanding of nature. He had thought deeply on the work of the Italian geometers, notably Ricci and Levi–Civita, on the calculus of tensors on Riemannian manifolds. But I am not sure if anyone except Weyl was thinking seriously at that time about combining, at a foundational level, the two streams of thought–that of Einstein on general relativity, and those of the Italian geometers, on analysis and algebra on Riemannian manifolds or pseudo Riemannian manifolds. The pseudo Riemannian case, whose theory does not differ in many essential aspects from the Riemannian case, became important after the work of Minkowski and Einstein because of the fact that spacetime is not Riemannian but pseudo Riemannian of signature (+,-,-,-), i.e., a manifold which has a smooth bilinear form on the tangent spaces having this signature at each point. In what follows the term Riemannian is used to include the pseudo Riemannian case also.

From the very beginning it was clear that the basic difficulty in doing physics or mathematics on manifolds that do not possess an intrinsically defined coordinate system was that of making sure that the results that one obtains by working in a single coordinate system remained valid in all of them. The traditional way to get around this issue is by insisting that the fundamental laws should be formulated in the language of tensors and their derivatives. However we have seen above, somewhat informally to be sure, that doing coordinate–invariant differential calculus of tensors requires the existence of a connection, or, isomorphisms between the tangent spaces at neighboring points. Such isomorphisms can be constructed on a Riemannian manifold by what is called *parallel transport,* as was first discovered by Levi–Civita. Thus, a connection may also be thought of as a way of moving tangent vectors along curves on the manifold in a canonical manner. If $t \mapsto \gamma(t)$ $(0 \leq t \leq 1)$ is a curve in M and Y is a tangent vector to M at the point $p = \gamma(0)$, we can find vectors $Y(t)$ tangent to M at $\gamma(t)$, uniquely determined by Y and depending linearly and isomorphically on Y; $Y(t)$ is the parallel transport of Y along the curve γ. These isomorphisms $Y \mapsto Y(t)$ between the tangent spaces at $p = \gamma(0)$ and $\gamma(t)$, allow one to define differentiation of vector and tensor fields along the curve γ. This is the idea of *covariant differentiation of vectors and tensor fields.* In fact, the three concepts

connection, parallel transport, covariant differentiation

are essentially equivalent.

For defining this connection, known as the *Levi–Civita connection,* Levi–Civita originally assumed that the Riemannian manifold M was imbedded in a higher dimensional euclidean space. Let M be a Riemannian manifold whose metric comes

from the flat metric of an ambient space E which is Euclidean. Let us be given a vector field Y defined on M near p, where p be a point of M and let X a tangent vector to M at p. Then the covariant derivative at p of Y in the direction X can be defined as follows: first extend Y to a vector field Y' defined in a neighborhood of p in E, compute the directional derivative $(\partial(X)Y')_p$ (in E) at p in the direction of X, the directional derivative being applied to each component of Y'; the covariant derivative of Y at p in the direction X is then the orthogonal projection of $(\partial(X)Y')_p$ on the tangent space to M at p. Since only the direction X is involved, it is enough if Y is specified just on a curve γ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$. If $Y(t)$ are vectors tangent to M at $\gamma(t)$ (notation as above), the condition that they are parallel transports of $Y(0)$ is that the covariant derivative of $Y(t)$ in the direction of $\dot{\gamma}(t)$ is 0 for all t. It is not difficult to show that these conditions translate to linear first order ordinary differential equations for $Y(t) = (Y^\mu(t))$ of the form

$$\frac{dY^\mu}{dt} + \sum_{\lambda\nu} \Gamma^\mu_{\lambda\nu}(\gamma(t))\dot{\gamma}^\lambda(t)Y^\nu(t) = 0, \qquad Y^\mu(0) = Y^\mu \qquad (PT)$$

The functions $\Gamma^\mu_{\lambda\nu}$ which appear as coefficients of these differential equations, are the Christoffel symbols defined in terms of the first order derivatives of the coefficients of the metric tensor. In the geometrical picture described above, differentiation in E which is related to parallel movement of vectors is adjusted to produce a movement of tangent vectors on M, hence the name parallel transport.

It was Weyl who apparently made the observation[4] that the formulae for the parallel transport make sense on *abstract* (as opposed to *imbedded*) Riemannian manifolds. Perhaps one should not put too much emphasis on this point because most classical geometers never worried about global aspects of differential geometry; however, for Weyl, who was one of the pioneers in the global view of manifolds as illustrated by his work in the theory of Riemann surfaces, such a line of thinking was natural.

Another discovery of his, namely that it was not necessary to require that a manifold have a metric in order to define parallel transport on it, was much more fundamental. Weyl's observation was that the notion of parallel transport on a manifold could be taken as an *axiomatic starting point.* He thus introduced the concept of an *affinely connected manifold* on which an (affine) connection was defined axiomatically. He formulated this as the statement that parallel transport was to be defined on the manifold by the equations (PT). The functions $\Gamma^\mu_{\lambda\nu}$ were prescribed *a priori* together with their law of transformation under change of coordinates. The transformation law is in fact determined uniquely by the requirement that the parallel transport defined by (PT) is independent of the choice of

coordinates. If we write the equations (PT) heuristically as

$$dY^\mu = -\sum_{\lambda\nu} \Gamma^\mu_{\lambda\nu} Y^\nu dx^\lambda$$

it suggests itself that we introduce the *matrix* of 1–forms given by

$$A = \left(\sum_{1\le\lambda\le n} \Gamma^\mu_{\lambda\nu} dx^\lambda \right)_{1\le\mu,\nu\le n} \qquad (n = \dim(M))$$

We then say that the matrix A together with its transformation law *defines the connection* on M. Although the elements of A have been indexed as if they are tensors, a connection is *not* a tensor, that is, the $\Gamma^\mu_{\lambda\nu}$ are not the components of a tensor; but the *difference* of two connections, in particular, any *infinitesimal variation* of a connection, is a tensor. So the connections form an *affine space* rather than a linear space. The condition

$$\Gamma^\mu_{\lambda\nu} = \Gamma^\mu_{\nu\lambda}$$

is independent of the coordinate system used since it is manifestly equivalent to the vanishing of a difference of two connections and so is a tensor equation. In this case the connection is said to be *torsionless.* The Levi–Civita connection is rediscovered in two ways: by fiat, giving the formulae for $\Gamma^\mu_{\lambda\nu}$ in terms of the metric tensor, and intrinsically, through the important result that *it is the only connection which is torsionless and has the property that parallel transport preserves the length of the vector.* For the general affine connections the curvature tensor can be defined imitating what is done in the Riemannian case. In fact, in terms of the matrices A_μ the curvature can be viewed as a matrix F of 2–forms given by

$$F = (1/2) \sum_{\rho\tau} F_{\rho\tau} dx^\rho dx^\tau = \sum_{\rho<\tau} F_{\rho\tau} dx^\rho dx^\tau$$

where

$$F_{\rho\tau} = A_{\tau,\rho} - A_{\rho,\tau} + [A_\rho, A_\tau]$$

For the Levi–Civita connection of a Riemannian manifold the curvature is the usual Riemann–Christoffel curvature tensor and its vanishing is the necessary and sufficient condition that the metric is flat, namely that in a suitable coordinate system it has the form

$$(dx^1)^2 + \ldots + (dx^p)^2 - (dx^{p+1})^2 - \ldots - (dx^{p+q})^2$$

For a general manifold with an affine connection as described above, the vanishing of F is the necessary and sufficient condition that the manifold equipped with the connection is locally isomorphic to the affine space $\mathbf{R}^n$ where parallel transport is just the usual one; this is equivalent to saying that there is a coordinate system locally in which covariant differentiation is just ordinary differentiation.

This liberation of the concept of a connection from its metric origins, although extremely simple and natural from hindsight (which we know has always 20/20 vision!), turned out to be one of the most important things that Weyl did in differential geometry[5]. Indeed, once the idea of axiomatically introducing connections is accepted, it is easy to realize that one should not be limited to transporting just tangent vectors and tensors along curves, and that one should also think of *axiomatically* attaching vectors to points of the manifold and transporting them using connections *axiomatically defined* in the appropriate way. But I am running ahead of the story.

One cannot be certain at what point in Weyl's thinking his ideas about the foundations of differential geometry merged with his reflections on their role in the foundations of general relativity. It is quite possible that these two strands of thought coexisted in his mind from the beginning. Nevertheless the next element in the story was Weyl's observation that requiring parallel transport to preserve the lengths of vectors was *not natural* either from the mathematical or the physical point of view. Mathematically it was not natural because one is already assuming that parallel transport can change the direction of the vector and so one can ask why it should keep its length unchanged[6]. On the other hand, it was not natural from the physical point of view either. In fact, the numerical values of the lengths of vectors can be determined only after one chooses a unit of length. Therefore, if we suppose that the base manifold is space or spacetime and that observers are located at various points of it, the assumption of length invariance of vectors under parallel transport meant that *all observers agree on what a unit of length is, even though they may be widely separated in space and time*. Weyl insisted that this was too strong an assumption, and that the proper assumption should be that observers can choose the unit of length only at their location, and that, while they can communicate their choice along paths to other observers, this transference of the unit of length may not be path–independent. Indeed, this should not be surprising, especially since transference of the direction of vectors was already assumed to be path–dependent. It was by reasoning in this manner that Weyl arrived at his remarkable observation[7]: Riemannian geometry cannot be considered as a completely pure infinitesimal geometry because the (metric) enables us to compare, with respect to their length, not only two vectors at the same point, but also the vectors at any two points. *A truly infinitesimal geometry must recognize only the*

principle of transference of a length from one point to another point infinitely near to the first (italics as in the original).

So, from Weyl's point of view, the metric of spacetime was determined only up to a scale at each point and that the possible scales at a point of M is represented by a copy of $\mathbf{R}_{>0}$, the multiplicative group of positive real numbers. Weyl made the assumption that after choosing a unit of length at a point, an observer can transfer this choice along curves to other observers. Clearly this transport of *scales* was of a very different nature from parallel transport in Riemannian geometry where only tangent vectors and tensors were transported. It is here that his work on connections that are defined without any reference to a metric became a guiding principle for him. Using his experience with affinely connected manifolds Weyl realized that such a transport of scales could be introduced in the same way; all he had to do was to introduce the matrix of 1–forms analogous to the $(\sum_{\lambda\nu}\Gamma^{\mu}_{\lambda\nu}dx^{\lambda})$ of his affine geometry, and use them to define the differential equations that describe the transport process. Since the set of scales has dimension 1, one has to define a single 1–form A locally on M, given in local coordinates by

$$A = \sum_{\mu} A_{\mu} dx^{\mu}$$

for suitable functions $A_{\mu}(\mu = 0, 1, 2, 3)$. The transport of scales $s(t)$ along a curve $\gamma(t \mapsto \gamma(t))(0 \leq t \leq 1, \gamma(0) = p, \gamma(1) = q)$ connecting two points p and q of the manifold (assumed for simplicity to be in the same coordinate neighborhood) is then determined by the differential equation

$$\frac{ds}{dt} + \sum A_{\mu}(\gamma(t))\dot{\gamma}^{\mu}(t)s(t) = 0$$

which is solved by

$$s(t) = \exp\Big\{ -\oint_{\gamma_t} \sum_{\mu} A_{\mu} dx^{\mu} \Big\} = \exp\Big\{ -\oint_{\gamma_t} A \Big\}$$

where γ_t is the path γ from 0 to t and $\oint_{\gamma_t}$ is the line integral of A along γ_t. It is important to realize that A is a 1–form as far as coordinate transformations are concerned, but it would change when one considers *transformations that come from a change of scales.* Let the scales be changed locally by a function

$$g : x \longmapsto g(x) > 0$$

Then in the new units the function s changes over to s' where

$$s'(t) = s(t)g(\gamma(t))$$

If we require that the differential equation for s' has the same form as the equation for s with A replaced by A', then we must require that A should change to the 1–form A' where

$$A' = A - d(\log g)$$

Weyl described this process as a *change of gauge,* and was led to his *principle of gauge invariance* which asserted that *the fundamental physical laws should be invariant, not only under coordinate transformations, but also, in addition, under changes of gauge,* otherwise called *gauge transformations.*

Notice that the exterior derivative dA of A is independent of the gauge and is a *globally defined* 2–form. Indeed, as the change of gauge changes A to $A+d(\log g)$, the 2–form $F = dA$ is independent of change of gauge and so is globally defined on M. The components of F are

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} = A_{\nu,\mu} - A_{\mu,\nu}$$

Clearly F is the *curvature* of this connection.

As already observed, these ideas of Weyl represented a big departure from classical geometry where only vectors like tangent vectors, cotangent vectors, tensors and so on, which are geometrically associated to the underlying manifold, were transported. Here Weyl was breaking new ground by considering the possible scales at the various points as a *bundle* on spacetime, introducing connections on this bundle, and their curvatures as 2–forms defined on spacetime. *If I am not mistaken, this is the first occurrence in physics and differential geometry of a non–geometric vector bundle and connections on it;* the fact that scales form a multiplicative group rather than a vector space is not important since the map $x \mapsto \log x$ converts this multiplicative group into the additive group of real numbers. In modern terms the scale bundle may be thought of either as a principal bundle for the group of positive real numbers or a vector bundle of rank 1. Thus one can describe the situation as follows: there is a natural bundle, the *scale bundle,* on M with group $\mathbf{R}_{>0}$, and the communication of the choice of scales along a curve γ on M is simply a curve γ' in this bundle that lies above γ. Moreover this lifting of γ is nothing but parallel transport in this bundle, described by a connection on it. The curvature of this connection is then a well defined 2–form on M.

Before proceeding further let me mention as an aside that the generalization of these ideas to any vector bundle in place of the scale bundle is immediate although Weyl did not take this step. The connection is defined by a *matrix–valued* 1–form

$$A = \sum_{\mu} A_\mu dx^\mu \qquad A_\mu = \left(\Gamma^a_{\mu b}\right)_{1\leq a,b\leq r}$$

The vectors attached to a point of the manifold form a vector space of dimension r. Unlike the case of the tangent bundle these vectors may have nothing to do with the tangent space and so we index them by the symbols a, b. If the vector spaces attached to the points of M are equipped with an action of a compact Lie group G, one makes the further assumption that the A_μ take values in the Lie algebra of G, i.e., the matrices $(\Gamma^a_{\mu b})$ lie in the Lie algebra of G. Parallel transport is defined by

$$\frac{ds^a}{dt} + \sum_b \Gamma^a_{\mu b}(\gamma(t)) s^b(t) = 0 \qquad (PT)$$

and covariant differentiation is given by the formula

$$\left(\nabla_{\partial_\mu} s\right)^a = \frac{\partial s^a}{\partial x^\mu} + \sum_b \Gamma^a_{\mu b} s^b$$

where (s^a) is a local section of the bundle. This can also be written concisely in the form

$$\nabla_{\partial_\mu} = \partial_\mu + A_\mu \qquad A_\mu = (\Gamma^a_{\mu b})_{1\leq a,b\leq r}$$

where the operator on the right acts on sections as

$$s \longmapsto \partial_\mu s + A_\mu s$$

However, when $\dim(M) \geq 2$, the operators ∇_{∂_μ} and ∇_{∂_ν} do not in general commute for distinct μ, ν. More precisely, let

$$F_{\mu\nu b}{}^a = \left(\nabla_{\partial_\mu}\nabla_{\partial_\nu} - \nabla_{\partial_\nu}\nabla_{\partial_\mu}\right)^a_b$$

The matrix $F_{\mu\nu}$, with entries $F_{\mu\nu b}{}^a$ can be calculated to be

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

One should think of the $F_{\mu\nu}$ as the coefficients of a matrix–valued 2–form F, the *curvature (form) of the connection,* defined by

$$F = (1/2)\sum_{\mu\nu} F_{\mu\nu} dx^\mu dx^\nu = \sum_{\mu<\nu} F_{\mu\nu} dx^\mu dx^\nu$$

where $A_{\mu,\nu}$ is $\partial A_\mu/\partial x^\nu$ and $[C,D] = CD - DC$. Its vanishing is equivalent to the statement that parallel transport is *locally* path independent, or, equivalently, there is a coordinate system locally in which covariant differentiation is the same as ordinary differentiation. These local data go hand in hand with the analog of Weyl's gauge transformations, namely the maps

$$x \longmapsto g(x)$$

of the part of the manifold being described locally into the group G. Obviously these should be called *gauge transformations* also. Finally, under the gauge transformation g the matrix A changes to the matrix A' where

$$A' = gAg^{-1} - dg\ g^{-1}$$

The presence of the term $dg\ g^{-1}$ shows that this is *not* a linear transformation and so A is not a tensorial object, but the difference of two connections is tensorial. The formula connecting A and A' is of course obtained from the requirement that the equations (PT) describing parallel transport have intrinsic significance. Moreover, if $F' = dA' + A' \wedge A'$, we have,

$$F' = gFg^{-1}$$

Thus F is tensorial. In the case of the scale bundle we have

$$A' = A - d(\log g), \qquad F' = F$$

as we have seen earlier. In this case F is a globally defined 2–form. In the general case, to get globally defined forms we must take *traces* to get rid of the factors involving g. More precisely, let us take the *characteristic polynomial* of F,

$$\det\left(TI - \frac{1}{2\pi i}F\right) = T^r + c_1(F)T^{r-1} + \ldots + c_r(F)$$

We must remember that the exterior forms of *even* degree form a *commutative algebra* under exterior multiplication, and so F may be viewed as a matrix with entries from this commutative algebra; thus there is no difficulty in computing its characteristic polynomial. The coefficient $c_j(F)$ is an exterior form of degree $2j$. It is *globally defined* on M. It is a remarkable fact (the theorem of Chern–Weil) that the c_j are *closed, and that their cohomology classes do not depend on the connection.* Thus they are invariants of the vector bundle. These constructions make sense for real as well as complex vector bundles but have to be modified a

little when we work with a bundle with group G. For a complex vector bundle the cohomology classes of the c_j are called *Chern classes.* For a bundle with structure group as $U(r)$ the cohomology classes of the c_j are actually *integral classes,*. This is a far–reaching generalization of the Dirac monopole quantization where one is dealing with $c_1(F) = F$, as we shall see in §4 below. Because of the gauge theoretic origin of these notions the Chern classes are important in physics where they often arise as *topological charges.*

Let us now return to the original discussion. Weyl had therefore been led to a natural generalization of Riemannian geometry from physical considerations. Sometimes this geometry is called a *Weyl geometry,* and the manifolds with a Weyl geometry, *Weyl manifolds.* Weyl geometry is based on two things: first, a metric, which is meaningful only up to scale changes that could be spacetime dependent, i.e., a *conformal metric* in modern terminology[8]; second, a connection on the scale bundle of the manifold. In his calculations which were couched in local terms Weyl simply used the product bundle $M \times \mathbf{R}_{>0}$. But this is not a restriction, as all principal $\mathbf{R}_{>0}$–bundles on M are trivial, i.e., isomorphic to the product bundle. Indeed, the $\mathbf{R}_{>0}$–bundles are essentially the same as $\mathbf{R}$–bundles, and the latter are classified by $H^1(M, \mathcal{S})$ where $\mathcal{S}$ is the sheaf of real smooth functions on M; and this sheaf is a so–called fine sheaf and so all of its cohomologies in degrees ≥ 1 are 0.

Just as in the case of Riemannian geometry, in Weyl geometry there is a canonical connection on the base manifold, the so–called *Weyl connection,* which is the analogue of the Levi–Civita connection of a Riemannian manifold. The Weyl connection is determined by requiring that it is torsionless and that the lengths of vectors is preserved under parallel transport, *after adjusting for scale changes.* This means that

$$\frac{d}{dt}\left(s(t)^{-1}\sum_{\mu\nu} g_{\mu\nu}(\gamma(t))Y^{\mu}(t)Z^{\nu}(t)\right) = 0$$

where $(Y^{\mu}(t))$ and $(Z^{\nu}(t))$ are tangent vectors at $\gamma(t)$ that remain parallel as t varies, and $s(t)$ is the scale factor at $\gamma(t)$ defined by parallel transport of scales governed by the differential equations described earlier. From this it follows easily that the coefficients of the Weyl connection are given by the formulae

$$\Gamma^{\mu}_{\lambda\nu} = {}^0\Gamma^{\mu}_{\lambda\nu} + \frac{1}{2}\left(\delta^{\mu}_{\nu}A_{\lambda} + \delta^{\mu}_{\lambda}A_{\nu} - g_{\lambda\nu}A^{\mu}\right)$$

where ${}^0\Gamma^{\mu}_{\lambda\nu}$ are the components of the Levi–Civita connection of the metric g and $A^{\mu} = g^{\mu\tau}A_{\tau}$ as usual (see [4], 206 or [7], 125, equations (49) and (50)).

All this was quite interesting but what excited Weyl tremendously was that he saw in his new geometry *a way to unify electromagnetism with gravitation.* Let me now explain this in a little more detail. At the time we are talking about, there were just *two* field theories in nature; the *electromagnetic field,* which obeyed the Maxwell equations in flat Minkowski spacetime, and the *gravitational field,* which was described by the Einstein equations involving the curvature tensor on an underlying Riemannian manifold.

In the electromagnetic theory, spacetime is $\mathbf{R}^{1,3}$ with the Minkowski flat metric. The electromagnetic field, which is usually described by a skew–symmetric matrix containing the electric and magnetic fields, is an exterior 2–form F

$$F = (1/2)\sum_{\mu\nu} F_{\mu\nu} dx^\mu dx^\nu = \sum_{\mu<\nu} F_{\mu\nu} dx^\mu dx^\nu$$

The Maxwell equations are then written concisely as

$$dF = 0, \qquad d * F = 0$$

where d is the exterior derivative and $*$ is the Hodge $*$–operator. In $\mathbf{R}^{1,3}$ we can write $F = dA$ for a 1–form

$$A = \sum_{\mu} A_\mu dx^\mu$$

which is uniquely determined up to an additive term of the form df for a scalar function f; A is called the *vector potential*. This change from F to A gets rid of the equation $dF = 0$. It is then well known that the second equation is the Euler equation for the action

$$\mathcal{A} = \int F \wedge *F dm = (1/4)\int \sum_{\mu\nu} F_{\mu\nu} F^{\mu\nu}$$

where the electromagnetic field F is the 2–form

$$F = dA = (1/2)\sum_{\mu\nu} F_{\mu\nu} dx^\mu dx^\nu, \qquad F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}$$

Since the field equations involve only F the vector potential has no physical significance.

In the Einstein theory spacetime is a pseudo Riemannian manifold M of dimension 4 and its metric itself is a dynamical object satisfying the equations in empty space which are concisely written as

$$R_{\mu\nu} = 0$$

where $R_{\mu\nu}$ is the Ricci tensor. They are the Euler equations for the Einstein–Hilbert action

$$\mathcal{A} = \int_M \text{Rsc } dm = \int_M \text{Rsc } \sqrt{-\det g}\; d^4x$$

where Rsc is the Ricci scalar. From a purely aesthetic point of view it was natural to ask whether one can unify these two theories; this was the starting point of the long but still unfinished quest for a *unified field theory* begun by Einstein and carried on by Weyl and many others.

Let me return now to the scale bundle. The fact that the parallel transport of scales can be defined by a connection expressed locally by a 1–form suggested to Weyl the possibility of identifying this connection with the electromagnetic potential; more precisely, if

$$A = \sum_\mu A_\mu dx^\mu$$

is the 1–form in local coordinates, then the A_μ could be identified with the components of the vector potential. This suggestion is reinforced by the observation, mentioned earlier, that the 2–form dA is globally defined and is the local description of the curvature of the connection. Clearly the identification of the connection with the vector potential goes hand in hand with the identification of its curvature with the electromagnetic field. This was what Weyl did.

The dynamical state of the Weyl universe is described by the conformal metric together with the connection on the scale bundle, the *connection* being identified with the electromagnetic vector potential and the *curvature* of this connection with the electromagnetic field. The equations of motion of the Weyl universe were required to be invariant not only under coordinate transformations as in Einstein's theory, but under gauge transformations of scales also. How can one write such equations of motion? The work of Einstein and Hilbert on gravitation gave the clue to Weyl and he wrote down a Lagrangian that was based on the curvature of the Weyl connection. The Euler equations derived from the Weyl Lagrangian gave coupled gravitational and electromagnetic equations, and the gauge invariance of the equations followed from the invariance of the Lagrangian under the gauge transformations determined by smooth positive functions g that represent the scale changes at the various points of M. The presence of the gauge transformations meant that along with tensors one should also consider *tensor densities,* and organize them according to their *weights,* where a tensor density is said to be of *weight* e if it is a tensor that gets multiplied by g^e under the gauge transformation g (e is an integer but can be negative). For instance, the metric tensor is of weight 1, and the curvature tensor $R_{\lambda\mu\nu\rho}$ of the Weyl connection is also of weight 1, while

the mixed curvature tensor $R^{\lambda}_{\ \mu\nu\rho}$ is of weight 0, i. e., gauge invariant. Weyl took the Lagrangian W given by

$$W = R^{\mu}_{\ \nu\rho\sigma} R_{\mu}^{\ \nu\rho\sigma}$$

and defined the corresponding action by

$$\mathcal{A} = \int_M W dm = \int W \sqrt{(-1)^q g} dx \qquad (g = \det(g_{\mu\nu}))$$

where q is the number of negative terms in the diagonal form of the metric. Since $\sqrt{(-1)^q g}$ has weight $n/2$ where n is the dimension of M, and W has weight -2, it follows that this Lagrangian is gauge invariant only if $-2 + n/2 = 0$, i.e., when $n = 4$! Using this Lagrangian Weyl wrote down the corresponding (Euler) equations of motion. It turns out that electromagnetism is still described in Weyl's theory by the Maxwell equations for the curvature F of the connection on the scale bundle, namely,

$$dF = 0, \qquad d * F = 0$$

However the fact that the curvature of the Weyl connection combined both the metric and the electromagnetic connection led to gravitational equations that were more complicated than Einstein's. Weyl took the point of view that only in the absence of the electromagnetic field it is true that gravitation is described by the Einstein equations, and that when both electromagnetism and gravitation are present, the Einstein equations are only an approximation to the more complicated but exact gravitational equations because the gravitational constant is enormously smaller than the electron radius.

The theory of Weyl described above was the first important example of a unified field theory. Weyl developed his ideas in three papers published during the years 1918–1920[9]. From his comments in his various expositions of these ideas it is obvious that Weyl was extremely happy with his ideas that suggested a *differential geometric* origin of electromagnetism, and the possibility of using this to solve the problem of a unified field theory. However there were serious objections to the Weyl theory, especially from Einstein. Einstein pointed out that if the scale transference is assumed to be path–dependent (which is necessary in order to have electromagnetic phenomena), then an object changes its size when transported around a closed curve. In particular, if the object is a clock, it will keep time differently after a transport along a closed curve. In other words, Einstein argued that the period of a clock will depend on its past history, making objective measurements impossible even for a single observer. Thus Einstein's objections cast doubts on scale transference as the source of electromagnetism[10]. Weyl answered these objections to some extent but he eventually gave up the idea that his geometry was the answer to the

problem of unifying electromagnetism and gravitation. Nevertheless, because of the beauty and simplicity of his theory, Weyl remained committed to the concept of the gauge principle on bundles over spacetime and the related concept that electromagnetism should be viewed as a connection on such bundles[11]. I wish to note here that it may be of interest to explore Einstein's objections further by installing the clocks on the *loop space* of M.

3. The rise of quantum theory. Electromagnetism as a connection on the phase bundle. The birth of quantum mechanics in 1925 introduced new themes into physics at the foundational level and Weyl had a major role in the identification and elucidation of some of these. Of particular interest for the present discussion is the modification of the nature of electromagnetism that he proposed as a consequence of quantum mechanics. He realized that quantum theory, with its introduction of the Planck's constant $\hbar$, *made it possible to choose a universal unit of length.* So far as he was concerned, this was the decisive argument in abandoning scale transfer as the source of electromagnetism. But in quantum theory the electron wave functions are undetermined only up to a phase. This phase however was the same at all spacetime points. Weyl had the idea to *make the phase depend on the spacetime points.* Thus he was led to the idea that the scale bundle can be replaced by the *phase bundle* and that one should do *phase transfer* along closed curves. The phases at a point form a group isomorphic to $U(1)$, the multiplicative group of complex numbers of absolute value 1. His proposal was to view the electromagnetic vector potential as a *connection on the phase bundle.* The phase bundle is just $\mathbf{R}^{1,3} \times U(1)$ viewed as a principal bundle for the group $U(1)$. The connection is defined locally by the 1–form

$$-i \sum_{\mu} A_\mu dx^\mu$$

(here $i = \sqrt{-1}$, and we use natural units with $e = \hbar = c = 1$). The appearance of i in the above formula indicates clearly that this is a *quantum* theory. The parallel transport of the phases corresponding to this connection is now defined by the differential equation

$$\frac{ds}{dt} - i \sum_{\mu} A_\mu(\gamma(t))\dot{\gamma}^\mu(t)s(t) = 0$$

which is solved by

$$s(t) = \exp\left\{ i \oint_{\gamma_t} \sum_{\mu} A_\mu dx^\mu \right\}$$

This is the same as the earlier formula on the scale bundle except for the factor i in front of the line integral. The gauge transformations are still defined by functions on M but now with values in $U(1)$:

$$g : x \longmapsto g(x) \in U(1)$$

Locally on M (and globally too, if M is simply connected) we can write

$$g = e^{i\beta}$$

for smooth real functions β which is how Weyl denoted them. The connection form A changes to A' under the gauge transformation g by

$$A' = A - d(\log g)$$

(since $\log g = i\beta$ is defined locally up to an additive constant, $d(\log g)$ is defined without any ambiguity). He discussed all of these matters thoroughly in a remarkable paper[12] in 1929. Apart from what we have talked about, his treatment of electromagnetism allowed Weyl to give a gauge–theoretic explanation of the conservation of electric charge. However phases are quantum mechanical quantities and so Weyl's description of electromagnetism is essentially quantum mechanical. In particular, classical gravitation does not come into the picture at all. So the problem of unification became open again. I do not have the time to go into the various attempts to construct unified field theories since then (see however the end of §5).

Interestingly enough, a $U(1)$–bundle is topologically more subtle than the scale bundle which, as we have noted already, is always trivial. Indeed, the same sheaf–theoretic argument given earlier can be adapted to show that the principal $U(1)$–bundles on a manifold M are classified by $H^2(M, \mathbf{Z})$. In modern terminology the element $d(P)$ of $H^2(M, \mathbf{Z})$ corresponding to a $U(1)$–bundle P is its *degree* or its *Chern class;* its image in $H^2(M, \mathbf{R})$ is determined as the cohomology class of $(2\pi i)^{-1}\Omega$ where Ω is the curvature of an arbitrary connection on P, by the Chern–Weil theorem. If M is a vector space or more generally contractible, this is again 0. But there are M such that $H^2(M, \mathbf{Z})$ is not 0 (see below); new electromagnetic phenomena then become possible because of the *topology* of M. Weyl himself did not consider these *topological effects*. They were first considered by Dirac in his theory of magnetic monopoles in 1931 and by Aharanov–Bohm in 1959 when they proposed their famous experiment to test whether the electromagnetic potentials have physical significance (see the following sections). In both cases the topology of M created the circumstances for new electromagnetic phenomena. The interaction

between quantum physics and topology started by Dirac and Aharanov–Bohm has become much more thorough nowadays. For instance the Chern classes give rise to *topological charges* in nonabelian gauge theory which are direct generalizations of the Dirac charge of the magnetic monopole.

Weyl's vision, that *the electromagnetic potential is a connection on a* $U(1)$*–bundle on spacetime whose curvature is the electromagnetic field, and that electromagnetism is a gauge field theory with structure group* $U(1)$, has proved to be prophetic. Weyl's ideas have survived to this day, through the many generalizations and modifications of the foundations of physics demanded by the quantum revolution.

4. Dirac's theory of monopoles. Weyl's idea that the wave function of the electron allowed spacetime-dependent changes in its phase attracted Dirac and he explored this theme. He replaced the principal $U(1)$–bundle by its associated *complex* line bundle and noticed that it acquired a hermitian structure in a natural manner. Let us call this line bundle L. For Dirac the wave functions were now *sections* of L; of course he did not express his ideas in this language but what he was doing amounted exactly to this. Since the fibers have a hermitian structure, the absolute values of the sections at the points of spacetime are well defined independently of the gauge transformations. This led Dirac to introduce what we would now call the Hilbert space $\mathcal{H}(L)$ of *square integrable sections* of L and identify its unit vectors with the states of the system as usual.

Dirac used this more subtle framework to discuss *magnetic monopoles* in one of the most famous papers he ever wrote[13] in 1931. His starting point was the observation that the conventional description of electromagnetism was not completely symmetric between electricity and magnetism. Thus, while an electric charge localized at a point is a perfectly viable concept, both theoretically and experimentally, the same is not true for a magnetic pole localized at a point. All magnets in nature are dipoles; moreover, any treatment of a magnetic monopole using Maxwell equations in the conventional way immediately runs into a contradiction. Indeed, if $\mathbf{B}$ is the magnetic field, and we assume that we have a vector potential that gives rise to the field, the Maxwell equations imply that $\mathbf{B} = d'A$ where d' is exterior differentiation in space so that the *magnetic flux* through any closed 2–dimensional surface is 0. *Thus there are no magnetic charges.* However there is a 2–form on $\mathbf{R}^3 \setminus \{0\}$ which is a candidate for the field of a stationary magnetic monopole located at the origin, namely,

$$F_g = \mathbf{B}_g = \frac{g}{4\pi r^3}(xdydz + ydzdx + zdxdy) \qquad (*)$$

where g is a nonzero constant. It is easy to check that F is the unique *closed* 2–form up to the scalar factor g, which is invariant under rotations, and that it satisfies

the equation $d * F_g = 0$, so that it is a solution of the Maxwell equations

$$dF_g = 0, \qquad d * F_g = 0$$

One can now verify that the flux of the field through any closed surface is 0 if the surface does not enclose 0 but is equal to g if the surface encloses 0, so that F can be taken as the magnetic monopole field of charge equal to g located at the origin. The earlier discussion shows that this field cannot come from a vector potential defined in the conventional way. But Dirac discovered that things are different if we consider the Maxwell equations on a *nontrivial* $U(1)$–bundle on $M = \mathbf{R}^3 \setminus \{0\}$, the natural space for treating a monopole located at the origin (we consider the stationary situation so that time can be ignored), and ask whether there is a connection on such a line bundle whose curvature is the field $(*)$. First of all nontrivial principal $U(1)$–bundles exist on M; indeed M is contractible to S^2 and so $H^2(\mathbf{R}^3 \setminus \{0\}, \mathbf{Z}) \simeq H^2(S^2, \mathbf{Z})$ and the latter is isomorphic to $\mathbf{Z}$. Thus the principal $U(1)$–bundles (or line bundles with group $U(1)$) on M are parametrized by *integers.* Let L_n be the bundle corresponding to the nonzero integer n; we can think of it as a principal bundle or a line bundle. Dirac showed that there is a connection on L_n whose curvature Ω_n satisfies

$$\frac{1}{2\pi i}\Omega_n = F_n = \mathbf{B}_n \qquad (MMPF)$$

and so can appropriately called the magnetic monopole of charge n. Moreover, as we have remarked earlier, for *any* connection on L_n its curvature Ω has the property that its cohomology class is the same as that of $\mathbf{B}_n$. In other words, only the magnetic monopole fields of *integral* charge n can be obtained from L_n. This is the modern version of the arguments that allowed Dirac to conclude that *the magnetic monopole charge is quantized*. The Dirac quantization of magnetic charge is a consequence of the classification of principal $U(1)$–bundles on M by $\mathbf{Z}$ and the Chern–Weil theorem which permits the determination of the Chern class as the class of $(2\pi i)^{-1}\Omega$ where Ω is the curvature of any connection on the bundle. Dirac's work marks the first emergence of Chern classes as topological charges. More remarkably, Dirac showed further that this magnetic charge quantization implied that *the electric charge is also quantized.* No other explanation of the quantization of the electric charge has been given yet.

Dirac proceeded to study the quantum mechanical motion of an electron in the field of a magnetic monopole defined above. The Hilbert space is the space $\mathcal{H}(L)$ defined above where L is the nontrivial hermitian line bundle on M with the monopole connection defined on it. To get a gauge invariant Schrödinger equation

Dirac replaced the usual partial derivatives ∂_μ by the *covariant partial derivatives* ∇_{∂_μ}:

$$\partial_\mu \longrightarrow \nabla_{\partial_\mu} = \partial_\mu - iA_\mu$$

The problem that Dirac considered was to determine the energy states of the energy operator in which the covariant Laplacian replaced the ordinary Laplacian. There is no space here to go in any more detail into this question[14].

Although magnetic monopoles have yet to be discovered, Dirac's monopole idea has proved to be surprisingly persistent. It is believed (as already pointed out by Dirac in his 1931 paper) that the difficulty of finding monopoles is due to the enormous energies needed to separate the dipoles, and that at an early time in the evolution of the universe they must have been around. Indeed, monopoles have to be admitted in many GUT's (Grand Unification Theories), and have recently played a fundamental role in the celebrated Seiberg–Witten theory.

5. Yang–Mills and the gauge principle. The Aharanov–Bohm idea. The next major step in the development of the gauge concept was the 1954 paper of Yang–Mills[15]. Yang and Mills were trying to understand the conservation of isotopic spin in interactions of nuclei and wanted to see if this can be done in a manner analogous to charge conservation in electrodynamics. Their starting point was the observation that the proton and neutron are just two states of a single entity (called the nucleon, an idea already proposed by Heisenberg in 1932). There are many reasons why this is a good idea–for instance, their masses are nearly identical, they have the same spin, and so on. To formulate this idea mathematically one introduces an internal space for the nucleon, called the isotopic spin space, which is a 2–dimensional complex Hilbert space with the standard action of $SU(2)$. When electromagnetic effects are neglected the observer cannot distinguish between the proton and neutron and so is free to choose any state in the isotopic spin space as defining the charged state. This freedom of choice for the observer led Yang and Mills to the principle that physical laws should be invariant under all the rotations of the isotopic spin space. But what made their work a major breakthrough was their insistence, that *to be compatible with the field concept in physics, these rotations of the isotopic spin space should be allowed to be different for observers situated at different points of spacetime.* They were thus led to define an *isotopic gauge transformation* as an arbitrary smooth $SU(2)$–valued function g defined on Minkowski spacetime ($= \mathbf{R}^{1,3}$), and to formulate the principle that all physical laws (when the electromagnetic field is absent) should be invariant under all isotopic gauge transformations g and the corresponding transformations of the nucleon wave functions

$$\psi \longmapsto \psi' = g\psi.$$

From this assumption they proceeded to obtain a theory which, for all intents and purposes, was completely analogous to the Weyl theory except that $SU(2)$ had replaced $U(1)$, i.e., one had now a *nonabelian gauge theory.*

Expressed in modern language, Yang and Mills worked with the $SU(2)$–bundles $\mathbf{R}^{1,3} \times SU(2)$ on Minkowski space $\mathbf{R}^{1,3}$ and the associated hermitian vector bundles corresponding to unitary representations of $SU(2)$. (Since the base manifold is a vector space and hence contractible, all the principal bundles are product bundles anyway.) The gauge concept then led to the introduction of connections on these bundles described by 1–forms on $\mathbf{R}^{1,3}$ with values in the Lie algebra of $SU(2)$,

$$A = -i \sum_{\mu} A_\mu dx^\mu \qquad (A_\mu = A_\mu^\dagger)$$

where $\dagger$ denotes the adjoint and the A_μ are 2×2 complex hermitian matrix functions. The corresponding field F was defined by them as the curvature of this connection,

$$iF = iF_A = dA + A \wedge A$$

The additional term $A \wedge A$ by which F differs from the formula for the field in electromagnetism is dictated for gauge invariance. In local coordinates

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

The dynamical equations are now written in the gauge context starting from a gauge invariant Lagrangian. The bundle under consideration has an adjoint operator analogous to the Hodge $*$–operator, also denoted by $*$, and one takes the action

$$\mathcal{A} = \int_M F \wedge *F dm = \int_M \sum_{\mu\nu} F_{\mu\nu} F^{\mu\nu} dm$$

It is easy to see that this is gauge invariant and so the corresponding Euler equations are also gauge invariant. These are the celebrated Yang–Mills equations (in the absence of matter). Globally they take the following form:

$$D_A * F_A = 0 \qquad (iF = dA + A \wedge A) \qquad (YM)$$

where D_A is *the covariant exterior derivative defined by the connection A itself.* One has automatiucally

$$D_A F_A = 0$$

which is the *Bianchi identity;* it is the analog of the equation $dF = 0 (F = dA)$ in the Maxwell theory. Thus the Yang–Mills equations generalize the Maxwell equations in empty space which are

$$dF = 0, \qquad d * F = 0, \qquad (F = dA) \tag{M}$$

where d is the usual exterior derivative.

It is indeed remarkable how similar this whole theory is to that of Weyl and Dirac. However there are major differences because the Yang–Mills theory is a nonabelian gauge theory. For electromagnetism on a principal $U(1)$–bundle, we have $D_A = d$ because the 1–forms take values in the Lie algebra of $U(1)$ which is $\simeq \mathbf{R}$. It is only when we go over to nonabelian gauge groups that the covariant exterior derivative depends on the connection so that the connection itself explicitly enters the field equations. The Yang–Mills field equations for the potentials A_μ are:

$$\sum_\nu (F^{\mu\nu}_{,\nu} + [A_\nu, F^{\mu\nu}]) = 0 \tag{YM}$$

In electromagnetism the second term is 0 because the commutators are 0, as we observed just a little earlier. But they do not vanish in nonabelian gauge theory. In other words, *the field equations* (YM) *contain the potentials in an essential manner; it is no longer true that the fields contain all the physical information.* This is in striking contrast with *classical electromagnetism* where the Maxwell equations (M) involve only the electromagnetic field. It is also noteworthy that the equations are *nonlinear* in the potentials.

This is the place to make some remarks on what is nowadays called the *gauge principle.* This principle says that once the group of gauge transformations is given, everything in the theory is automatically determined. If we view the gauge group as the symmetry group of the system, this is very similar to the principle in classical geometry, first enunciated by Felix Klein in his Erlangen Program, that any geometry is completely determined by the group of congruent transformations. To make this more precise, let us consider the two basic features of the gauge theory–the wave functions ψ, and the gauge transformations $\psi \mapsto \psi' = g\psi$. Since

$$\frac{\partial(g\psi)}{\partial x^\mu} = g\left(\frac{\partial\psi}{\partial x^\mu} + g^{-1}g_{,\,\mu}\psi\right) \qquad (g_{,\,\mu} = \partial g/\partial x^\mu)$$

we can say that under the gauge transformation g the derivative ∂_μ changes to

$$\partial_\mu + g^{-1}g_{,\,\mu}$$

So, in order to obtain gauge covariant quantities, it is natural to *counteract* the effects of the gauge transformation by introducing fields A_μ and work with $\partial_\mu + A_\mu$ instead of ∂_μ. Such a choice will be covariant if

$$g(\partial_\mu + A_\mu) = (\partial_\mu + A'_\mu)g$$

where A'_μ is the counteracting term in the new gauge. If $A = \sum_\mu A_\mu dx^\mu$ this is just the condition

$$A' = gAg^{-1} - dg\ g^{-1}$$

which says that A transforms as a connection. *Thus the specification of the gauge group automatically forces one to introduce gauge potentials(connections) in the associated bundle and use covariant derivatives with respect to this connection instead of the usual derivatives.* Now the wave equations satisfied by the particle (Klein–Gordon, Dirac, Weyl, etc) usually come from a Lagrangian. Moreover, the same reasoning shows that one must make the change

$$\partial_\mu \longrightarrow \nabla_{\partial_\mu} = \partial_\mu - iA_\mu$$

in the Lagrangian of the electron wave function ψ mentioned earlier to have a gauge invariant Lagrangian. The new expression for the Lagrangian with this change is then gauge invariant; but what is more remarkable, it contains the terms that tell us how the field A interacts with the wave function ψ. In other words, the specification of the gauge group and the requirement of gauge invariance already force one to introduce the gauge potentials or connections as well as the form of the interaction of the gauge field with the matter field. As an example, let us consider the free Dirac field which is a spinor field ψ(a 4- component column vector) whose equation of motion are

$$(i\sum_\mu \gamma^\mu \partial_\mu + m)\psi = 0 \quad \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\delta_{\mu\nu}$$

arising out of the Lagrangian

$$\mathcal{L}_{\text{Dirac}}(\psi) = -(\psi^\dagger(i\sum_\mu \gamma^\mu \partial_\mu + m)\psi)$$

where the γ^μ are Dirac's γ-matrices and $\dagger$ is the complex conjugate. Now,

$$\mathcal{L}_{\text{Dirac}}(e^{i\beta}\psi) = \mathcal{L}_{\text{Dirac}}(\psi) + \left(\sum_\mu \psi^\dagger \gamma^\mu \beta_{,\mu}\psi\right)$$

In other words, $\mathcal{L}_{\text{Dirac}}$ by itself is not gauge invariant; but if we introduce compensating fields A_μ which change to $A_\mu - \beta_{,\mu}$ when ψ changes to $e^{i\beta}\psi$, and simultaneously replace ∂_μ by $\partial_\mu - iA_\mu$, then the new Lagrangian is gauge invariant. A simple calculation then shows that the new Lagrangian is

$$-\psi^\dagger \left(\sum_\mu i\gamma^\mu(\partial_\mu - iA_\mu) + m \right) \psi = \mathcal{L}_{\text{Dirac}}(\psi) - \psi^\dagger \sum_\mu \gamma^\mu A_\mu \psi$$

The second term in this expression gives the term that corresponds to the interaction of the gauge field A with the matter field ψ. The final expression for the Lagrangian with matter fields present is

$$\mathcal{L} = -(1/2)F \wedge *F - \psi^\dagger (i \sum_\mu \gamma^\mu \partial_\mu + m)\psi + \psi^\dagger \sum \gamma^\mu A_\mu \psi$$

In physics this entire set of ideas is called the *gauge principle.* It was quite explicit in the works of Weyl and Dirac on electromagnetism but was elevated to a universal principle after the work of Yang and Mills was fully understood.

In 1959, Aharanov and Bohm wrote a remarkable paper[16] in which they discussed the question *whether the potentials have physical significance.* They suggested that this is true even in electromagnetism. More precisely they predicted that *when M is not simply connected, electromagnetic effects could be present even if the electromagnetic field is* 0. To test this hypothesis they proposed an experiment, the famous *Aharanov–Bohm experiment.* In this experiment, $M = \mathbf{R}^3 \setminus L$ where L is a straight line in space which represents a solenoid which is ultra thin so that its magnetic field does not leak outside of L. Electrons are sent around the solenoid and then allowed to interact, producing interference patterns in a screen. Although in the region in which the electrons travel, namely M, the electromagnetic field is 0, nevertheless, they predicted that *the interference pattern on the screens would change when the flux inside the solenoid is varied.* This experiment was performed and this prediction verified. I cannot go into this in more detail here; see my book with Sundararaman[3] for a more detailed discussion[17]. The fact that Yang–Mills equations involve the potentials in an essential manner should be understood from this perspective.

There is one interesting point worth mentioning. In both the theories of Weyl and Yang–Mills the physical description is through the connections on a principal bundle on spacetime M with group G which is a compact Lie group ($G = U(1), SU(2)$). However, because of the possibility of making gauge transformations, two connections which are gauge equivalent represent the same physical

situation. Indeed, as is explained in [17], this is exactly the case in the Aharanov–Bohm experiment. In addition the connections must satisfy the appropriate equations, the Maxwell equations in the Weyl case and the Yang–Mills equation in the theory treated by Yang–Mills. So if $\mathcal{E}$ is the set of connections satisfying the appropriate equations and $\mathcal{G}$ is the group of gauge transformations, the states of the system at the classical level are represented by the points of the *quotient space*

$$\mathcal{M} = \mathcal{E}/\mathcal{G}$$

Of course one has to quantize this situation before an adequate gauge theory can be developed; but even this classical description is very interesting because it associates with any G–bundle a geometric object $\mathcal{M}$ which is important physically. $\mathcal{M}$ is a *moduli space* of great mathematical interest also. Indeed, it is precisely by studying $\mathcal{M}$ that new discoveries in the topology and geometry of 4–manifolds were made, a striking confirmation of the deep nature of the interaction between mathematics and physics in recent decades[18].

The Yang–Mills theory is a beautiful and natural generalization of the theories of Maxwell, Weyl and Dirac. It must be pointed out however that at the time of their discovery its relationship to the earlier themes was not very clear. It was only subsequently, in the 1970's, that the relationship of gauge theories to the differential geometry of principal and vector bundles, and their natural evolution starting from the themes explored by Weyl and Dirac, were finally understood[19].

The revolution started by Weyl, Dirac, and Yang–Mills, eventually led, in conjunction with a huge development of quantum field theory, to a unification of three of the basic forces in nature, electromagnetic, weak, and strong. However gravitation had remained outside this process. Nevertheless, for the past several years an intensive effort is being made to obtain a quantum theory of gravitation that will then include all forces and will be the unified field theory that has been the holy grail of theoretical physics since the discovery of general relativity by Einstein.

One starting point of a possible unified view of things is the same as Weyl's, namely a principal bundle on spacetime. The group of the bundle however is much bigger than $U(1)$ because *we have to admit many additional types of particles.* Moreover one has to consider not only gauge fields with this structure group but also matter fields. Now in Minkowski space, the typical matter field satisfies the Dirac equation in the massive case and the Weyl equation in the massless case, and so they are *spinor fields.* So to define them in the case when M is curved due to gravitation, we have to introduce spinor fields on curved manifolds. This imposes a topological restriction on M (it should be what is called a spin manifold); the matter fields then satisfy the natural generalization of the Dirac or the Weyl equation to curved spin

manifolds. There are additional restrictions on M also because in the classical limit of the vacuum state the theory should describe spacetime with no matter and so M should be Ricci flat (recall that the Einstein equations for matterless spacetime are that the Ricci tensor vanishes). The phenomenology of particle spectra also imposes additional restrictions and difficulties which have been overcome by assuming that M is coupled to a *very small* compact manifold (*compactification* in the terminology of the physicists) of 6 dimensions which is the underlying real manifold of a complex Calabi–Yau manifold of complex dimension 3. This is not the place even to begin talking about these matters but the reader should refer to the address of Witten to the International Congress of Mathematicians at Berkeley in 1986 for an inspiring discussion[20].

6. Holomorphic vector bundles and the meeting of the physics and the mathematics. I mentioned at the beginning that I was going to talk about matters that are tangential to the interests of Seshadri. Tangential means at least some contact and so far there is no hint that the ideas that I have discussed so far have anything to do with *holomorphic* vector bundles on Riemann surfaces! I would like to address this point, at least briefly, now. In keeping with the historical and pedagogical stance of this article let me start by giving a very brief discussion of the *holomorphic* vector bundles.

Although the concept of a holomorphic vector bundle is quite a modern one, the word modern should be taken to refer only to the view point from which the subject is looked at and the language in which its main results are formulated. This use of the modern view point and language is of course natural and appropriate because it is impossible, given what we know about topology and geometry in modern times, to continue to work in the confining paradigm of the ancients. But it is good to know, at least at the starting point, what the subject is from a classical perspective.

Given a compact Riemann surface, one wants to study not only the (meromorphic) functions and differential forms on the surface but also those which are multiple–valued on the given surface but become single–valued when we go over to a covering surface. Thus, together with a given compact Riemann surface, one wants to have the entire tower of its covering surfaces in view. Weyl, one of the founders of the modern theory of Riemann surfaces, emphasized the deep analogy of this situation with the tower of extensions of an algebraic number field. The extensions of a number field are governed by the Galois groups and their representations, and classical number theorists had already obtained a profound understanding of the *abelian* part of the tower, namely the *subtower* of extensions with an *abelian* Galois group, otherwise called the *classfields*. In function theory the analog of the Galois

group (at least when the extension is unramified) is the fundamental group. But the classical theory of Riemann surfaces, with its emphasis on the *line integrals* on such surfaces, had access only to the covering surface whose covering group is the *homology* of the surface, which is the largest abelian quotient of the fundamental group. Weyl called the corresponding covering surface the *class surface,* reinforcing the arithmetic analogy with classfields. The multiple–valued functions that become single–valued on the class surface and transform according to *characters* of the homology group were an object of study in the classical theory. These characters constitute a complex torus, the so–called classical Jacobian variety, and the function theory on this torus is the theory of *abelian functions.*

Already Riemann had begun thinking of some *nonabelian* generalizations of this theme, at least on $\mathbf{P}^1$, in his work on the *monodromy* of regular singular differential equations. He considered linear ordinary differential equations of arbitrary order n on the extended complex plane $X = \mathbf{P}^1$ with rational coefficients. Let F be the set of singular points of the coefficients of the equation; it is convenient to assume that $\infty \in F$. At a point $x_0 \notin F$, the local (germs of)solutions to the equation around x_0 form a vector space V of dimension n. By the linearity of the equation one can continue the elements of V along any path not meeting F and the continuation depends only on the homotopy class of the path. Taking closed paths at x_0 it is now clear that we have an action of the fundamental group $\pi_1(X \setminus F, x_0)$ on V; the image of the fundamental group inside $GL(V)$ is the *monodromy group* of the equation in classical terminology. Riemann studied the hypergeometric equation from this point of view. Here the differential equation is of the second order, $F = \{0, 1, \infty\}$; the fundamental group of $X \setminus F = \mathbf{C} \setminus \{0, 1\}$ (with respect to some base point) is then the free group on two generators. In unpublished work that did not become available for some years after he died, Riemann considered also the generalizations to equations of higher order. However, in the general case, one has to impose some restrictions on the orders of the coefficients at their poles so that the solutions of the equations are linear combination of functions of the form $z^\lambda (\log z)^r$. This is equivalent to the requirement that the solutions, although possibly multiple–valued around the singularities, are nevertheless of *moderate growth* (in a suitable sense) at these points. One is thus led to the class of equations with rational coefficients having only *regular singularities;* these are also called *Fuchsian equations* in honor of Fuchs who treated them formally for the first time, Riemann's work coming to light only posthumously.

But the hypergeometric equation, and more generally, the Fuchsian equations, touch the theory of compact Riemann surfaces at a deeper level also. For instance, the periods of elliptic functions satisfy the hypergeometric equation, a fact that is seen from the Legendre formula for the periods of elliptic integrals in terms of the

hypergeometric series:

$$K := \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}} = \frac{\pi}{2} F(1/2, 1/2, 1; k^2)$$

Fuchs extended this result to hyperelliptic integrals and showed that if one considers a family of hyperelliptic integrals, their periods satisfy a Fuchsian equation. This was in fact a major discovery of Fuchs. The generalizations of these equations are called *Picard–Fuchs equations* and they have become important in physics recently.

Two questions about Fuchsian equations come up naturally. The first is whether the Fuchsian assumption imposes any restriction on the monodromy representation: given any representation ρ of $\pi_1(X\backslash, x_0)$ on $\mathbf{C}^n$, is there a Fuchsian equation with ρ as its monodromy representation? The second question is more delicate. If $y \in F$ is a singular point, the *local monodromy group* is the fundamental group of a punctured disk at y and so is isomorphic to $\mathbf{Z}$. Thus we have an action of $\mathbf{Z}$ on V which is determined by an automorphism L of V. Generically this action is determined by the eigenvalues

$$\alpha_1, \alpha_2, \ldots, \alpha_n$$

of L which are traditionally called the *local exponents* of the equation. The second question is then the following: suppose one is given all the local exponents at the singularities (these must satisfy certain natural relations), what is the structure of the space of all Fuchsian equations (up to isomorphism) with these singularities and local exponents? For equations of second order with $F = \{0, 1, \infty\}$ Riemann knew that the hypergeometric equation is the only one for fixed local exponents, and he developed the theory of hypergeometric functions from this point of view (the Riemann P–functions). The parameters that describe the equations with given local data are called *ancillary parameters* in classical language.

Hilbert formulated the first question as his twentyfirst problem in his famous 1900 address; this is nowadays called the Riemann–Hilbert(R–H) problem, not only in Hilbert's original formulation but also in its myriad modifications and variations. There is a little vagueness in the Hilbert formulation whether the differential equation is to be sought as a Fuchsian ordinary differential equation or as a first order system

$$\frac{du}{dz} = Au, \qquad A = \sum_k \frac{A_k}{z - a_k}$$

where the A_k are *constant matrices.* As for the second question, Riemann himself had calculated the number of ancillary parameters in his unpublished work, namely,

the moduli in his language, for the equations with fixed singularities and local exponents.

The functions envisioned by Riemann and Hilbert are not algebraic as in the classical theory but transcendental; they were called the *Riemann transcendentals.* The local solutions of the system generate a *differential field* and one can keep the arithmetic analogy going by thinking about these fields as forming a tower of *differential field extensions* with the Galois groups replaced by the differential Galois groups; however this point of view did not come in till quite a bit later. It is interesting to note that neither Riemann nor Hilbert made any effort to formulate these questions when the base manifold is taken to be any compact Riemann surface; in terms of the function fields, this is the relative point of view in which the field of rational functions is replaced by the field of meromorphic functions on a compact Riemann surface. That Hilbert, who was one of the first to recognize the importance of relative extensions in number theory (relative means over an arbitrary number field, as opposed to just the rational numbers), and who was also a champion of the analogy between number fields and function fields, did not formulate the R–H problem on an arbitrary compact Riemann surface, is striking.

In its classical formulation over $\mathbf{P}^1$, the R–H problem was studied and solved by several people under varied conditions–Schlessinger, Lappo–Danilevsky, Plemelj, Birkhoff, and so on. But further progress of the R–H problem on other Riemann surfaces had to await the modern development of differential geometry and topology. The formulation and the solution of the R–H problem for arbitrary compact Riemann surfaces was given by Röhrl[21] in 1957. Then Deligne took up this theme in 1970, and in a very influential monograph[22], gave the formulation of the R–H problem in the language of vector bundles and connections. But he went considerably beyond the original context by formulating the R–H problem in *all* dimensions and solving it completely, as well as treating the other classical themes such as the differential equations satisfied by the period matrices, in the more general context.

I shall briefly explain Deligne's formulation of the R–H problem in dimension 1 on an arbitrary compact Riemann surface X. One introduces the sheaf of local solutions of the differential equations which gives rise to a holomorphic vector bundle V on $X \setminus F$ where F is the set of singularities of the equations. There is a natural *holomorphic* connection ∇ on V whose horizontal sections are the local solutions. One can extend the vector bundle to the whole of X. One should add a condition about the moderate growth of the sections of the bundle at the singular points to make sure that the system is regular singular. Deligne's solution is that the assignment that takes a pair (V, ∇) with regular singularities at F to the monodromy action of $\pi_1(X \setminus F, x_0)$ is a functor that gives an equivalence of categories.

However some care has to be exercised before deducing the solution to the R–H problem in the classical setting of Hilbert from this result[23].

It must be noted that in dimension 1 all connections are *flat* while this is not so in higher dimensions and one has to restrict oneself to flat connections to formulate the question; in classical language, the partial differential equations must satisfy the Frobenius integrability conditions. These equations in higher dimensions are therefore very special. Such differential equations go back to the papers of Appel, Picard, and others where hypergeometric functions in several variables were first introduced[23].

However one cannot restrict oneself to the regular singular context all the time. Already in dimension 1, many physical problems led to meromorphic differential equations with *irregular singularities*–Bessel, Airy, confluent hypergeometric, and so on. Their systematic treatment was begun in the late nineteenth century by Fabry and Poincaré. Now a characteristic feature of Fuchsian equations is the fact that the local solutions can be computed formally–this is the essence of the well known Frobenius method of generalized power series and indicial exponents. This approach fails completely in the presence of irregular singularities. So the local theory in the irregular case is already considerably richer than in the regular case. Ponicaré began the process of bridging the gap between the formal and analytic theories with his theory of *asymptotic expansions* of solutions at an irregular singularity. For instance, in the theory of the Bessel equation, the singularity at infinity is irregular while the one at the origin is regular, and the central fact of the theory is the asymptotic structure at infinity of the solutions with given local exponents at the origin. For more general equations the asymptotic aspects are completely captured in the theory of the *Stokes phenomenon.* The modern point of view of these equations began with the work of Malgrange, followed by that of Deligne (expressed in a letter of Deligne to Malgrange), and then continued by Sibuya's work on the Stokes phenomenon. The work of Deligne, Malgrange, and Sibuya revealed that the Stokes phenomenon is really cohomological and is governed by the H^1 of a certain sheaf of noncommutative groups of flat solutions of the equations, called the *Stokes sheaf.* The analysis of the Stokes phenomenon in the classical setting by Balser, Jurkat, and Lutz[24] led to a construction of the *local* moduli space, namely, the moduli space of meromorphic connections with an irregular singularity at $z = 0$. The construction of the moduli *scheme* starting from the Stokes sheaf was accomplished by Babbitt and Varadarajan[25]. Many questions in the global theory of moduli are still open, although a substantial amount of work has been done[25].

Let me now return to the original theme of *algebraic functions* and their non-abelian aspects. If I am not mistaken, it was Weil's paper[26] in 1938 which brought into the foreground the algebraic theory together with the arithmetic analogy and

began the search for a better understanding of the *nonabelian* phenomena that govern the systems of functions that transform according to representations of the fundamental group. This is very similar to the R–H context except that there are no singularities. Here not only the classification of the systems corresponding to the various representations of the fundamental group was an objective, but also function theory on the space that parametrizes such systems, the "moduli space".

Even though Weil had obtained deep insights in his studies, the time was not ripe for making decisive progress. It was only after the vector bundle aspect was placed at the center of the investigations that progress could be made. The major breakthrough came in the papers[1,2] of Narasimhan and Seshadri where the concept of *stability* of an algebraic vector bundle, introduced by Mumford in his geometric invariant theory, was clarified in terms of the *unitarity* of the corresponding representation of the fundamental group of the Riemann surface. The importance of stable vector bundles arises from the fact that they are generic, and to get a "good"moduli space one has to restrict oneself to the set of stable bundles (up to isomorphism of course) and compactify this set when necessary. In the new point of view the moduli space of vector bundles on a compact Riemann surface, properly defined, became the central object of study, emerging as a far–reaching generalization of the classical Jacobians of curves.

The moduli space of the bundles (in the refined sense I have tried to explain above) is a difficult object to apprehend. In particular understanding its topology already requires substantial effort. The Indian school of geometers who followed Seshadri and Narasimhan, as well as several others, made major contributions to this study. However, in a paper published in 1982, Atiyah and Bott[27] *used the theory of Yang–Mills equations and Morse theory to study the topology of these moduli spaces.* This coming together of physics and mathematics, here as well as in the gauge theoretic approach to questions of the geometry of 4–manifolds that I have mentioned earlier where the moduli space of Yang–Mills connections plays a fundamental role, is quite remarkable. The Atiyah–Bott approach to the cohomology of the moduli space of vector bundles, with its origins in differential geometry and the physics of gauge theories, has also proved very fruitful in higher dimensional generalizations of the vector bundle theory.

Acknowledgement. I wish to express my deep gratitude to my friends Robert J. Finkelstein and Donald G. Babbitt for conversations spread over many years on all the topics discussed in this article. These conversations were tremondously inspiring and contributed greatly to my understanding.

NOTES AND REFERENCES

[1] Ann. Math., 82(1965), 540.

[2] Ann. Math., 85(1967), 303.

[3] See the beautiful account of Yang in *Hermann Weyl, 1885–1985,* K. Chandrasekharan (Ed), Springer Verlag, 1986. I have profited a great deal by reading Yang's paper and have made use of it freely in writing this essay. A more detailed discussion is given in the forthcoming book *An Introduction to Gauge Theories,* by D. Sundararaman and V. S. Varadarajan. See also the articles in the Hermann Weyl International Congress volume *Exact Sciences and their Philosophical Foundations,* Wolfgang Deppert/Kurt Hübner, (Eds), Verlag Peter Lang, 1985.

[4] See Weyl's book *Space, Time, and Matter,* Dover, Note 10 to Chapter II.

[5] See §§14–15 as well as the Notes 9 and 10 to Chapter II of Weyl's book *Space, Time, and Matter,* Dover. These ideas were first developed in his paper *Reine Infinitesimalgeometrie,*, Math. Zeit. 2(1918), 384. A convenient place to look up Weyl's papers is the *Hermann Weyl Gesammelte Abhandlungen,* K. Chandrasekharan (Ed), Springer Verlag 1968), hereafter abbreviated as GA, I–IV. The paper referred to here is GA, II, 1.

[6] In an article reviewing 50 years of relativity, GA, IV, 421, Weyl says the following: *Beim Herumfahren eines Vektors längs einer geschlossenen Kurve durch fortgesetzte infinitesimale Parallelverscheibungkehrt dieser im allgemeinen in einer andern Lage zurück; seineRichtung hat sich geändert. Warum nicht auch seine Länge?*

[7] See page 203 of the paper *Gravitation and Electricity* in *The Principles of Relativity,* Dover, 201. The latter is a collection of reprints of the fundamental papers in the theory of relativity, containing the paper of Weyl, the papers of Einstein on special and general relativity, and the famous address of Minkowski in which spacetime occurs for the first time as a mathematical and physical construct.

[8] It is not surprising therefore that Weyl had a deep interest in *conformal geometry.* He constructed the analogue of Riemann's curvature tensor in conformal geometry which is nowadays called the *Weyl tensor.* Its vanishing is the necessary and sufficient condition that the manifold is *conformally flat, i.e., its metric can be transformed into a flat metric by a coordinate change followed by a rescaling* (theorem of Weyl–Schouten).

[9] The basic papers are GA II, 1; 29; 55. The second of these papers is the most important one from the point of view of physics. Its translation into English is the paper which is a part of the reprint volume referred to in [7]. In addition Weyl came back to this theme in several subsequent papers of which the most important are GA, III, 217; 229; 245.

[10] See Yang's discussion in reference [3].

[11] See the discussion in Dirac's paper, Proc. Roy. Soc. Lond., A333(1973), 403.

[12] This is the paper GA, III, 245; the papers GA, III, 217, and 229 are variations of this.

[13] Proc. Roy. Soc. Lond., A133(1931), 60.

[14] The problem here is to obtain the spectral decomposition of the covariant Laplacian acting on the Hilbert space of square integrable sections of the monopole line bundle. See[15], 493. For a treatment of this as well as the spectral analysis of the covariant Dirac operator on this line bundle using the point of view of induced representations see R. P. Langlands in Pac. Jour. Math., 126(1987), 145.

[15] See Phys. Rev., 96(1954), 191; also in *Selected Papers, 1945–1980, with Commentary, Chen Ning Yang,* W. H. Freeman and Company, 19; 172.

[16] See Aharanov–Bohm, Phys. Rev., 115(1959), 485. The experiment was performed by Chambers in 1960, see Phys. Rev. Lett., 5(1960), 3.

[17] Here the manifold is contractible to the circle and so its second integer cohomology is trivial. So only product bundles and globally defined vector potentials arise. As there is no electromagnetic field the connection has vanishing curvature, i.e., it is *flat.* So there is no phase change coming from parallel transport of phases along *small* closed paths but there may be phase changes if the loop is *big.* This is the familiar concept of *holonomy* of a flat connection in differential geometry; the holonomy around a closed path depends only on the homotopy class of the path and leads to a homomorphism of the fundamental group (integers) of the space into the group (circle group) of the bundle. So these homomorphisms are parametrized by a real number mod 1 and correspond to the various interference patterns obtained by the experimenter. Furthermore the connection is *determined up to gauge equivalence* by this homomorphism. So the *physical states of the system are really not the connections per se, but gauge equivalence classes of the connections* which are the points of the moduli space of connections satisfying Maxwell equations[15], 460.

[18] The gauge theoretic approach to the geometry of 4–manifolds is the theme of the book of Donaldson and Kronheimer, Oxford, 1990.

[19] See C. N. Yang[15], 457; 460; 493; 509; 519 and Yang's article in the *Chern Symposium,* Springer Verlag, 1979, 247.

[20] Proceedings of the Int.Cong.Math., Berkeley, 1986, I, 267. I remember that Witten began his lecture at the Congress with the remark that Weyl's attempted unification of electricity and gravitation failed because the forms of matter included in it were not diverse enough. It is of interest to note that Weyl himself had come to the same conclusion. In his preface to the first American printing of his book *Space, Time, and Matter,* he says: *Since then, a unitary field theory, so it seems to me, should encompass at least three fields: electromagnetic, gravitational and electronic. Ultimately the wave fields of other elementary particles will have to be included too–unless quantum physics succeeds in interpreting them all as different quantum states of one particle.*

[21] Math. Ann. 133(1957), 1.

[22] See Deligne's Springer Lecture Notes 163(1970).

[23] For expositions of some of these questions see my articles in Exp. Math., 9(1991), 97 and Bull. Amer. Math. Soc., 33(1996), 1, both of which are reviews. For recent work on the R–H problem see A. A. Bolibruch, Proc. Int. Cong. Math., 2(1994), 1159, Birkhäuser, Basel, 1995.

[24] See W. Balser, Jour. für reine und ang. Math., 318(1980), 51; W. Balser, W. Jurkat, and D. Lutz, Funkc. Ekvac., 22(1979), 257; W. Jurkat, Springer Lecture Notes in Math., 637(1978).

[25] See Malgrange, Springer Lecture Notes in Mathematics 712(1979), and Sibuya, Bull. Amer. Math. Soc., 83(1977). For the moduli problem see D. G. Babbitt and V. S. Varadarajan, Pac. Jour. Math., 108(1983), 1; Mem. Amer. Math. Soc., 55(1985), no. 325; Asterisqué, 169–170 (1989) as well as my review papers[23]. Roughly speaking, the moduli space is the first cohomology of the Stokes sheaf which is a sheaf of *complex unipotent nonabelian groups* and it is not immediately clear why it should have a nice structure as a variety nor why such a structure, if it exists, should be independent of the (Cěch) covering used to compute it. The basic idea in the Babbitt–Varadarajan approach to the moduli at an irregular singularity (which was based on a suggestion of Deligne) is to view the Stokes sheaf as a sheaf of complex unipotent group *schemes,* defined on the unit circle in the complex plane, so that its first cohomology can be viewed as a functor from commutative complex algebras with units into the pointed sets of the first cohomology of sheaves of groups on the unit circle; the existence of a scheme structure on the moduli space then becomes the question whether this functor is *representable* by a scheme. It turns out, and this is compatible with the Balser–Jurkat–Lutz theory, that this functor is actually represented by *affine space.* For global moduli the literature is too vast and I give only some scattered references. For the Fuchsian case, see K. Iwasaki, Pac. Jour. Math., 155(1992), 319, and the references therein. The subject of special functions from a modern viewpoint is the theme of the book *From Gauss to Painlevé* by K. Iwasaki, S. Shimomura, and M. Yoshida. For some other aspects of Fuchsian connections and the Stokes sheaf, see the book *Déformations isomonodromiques et variétés de Frobenius,* Savoirs Actuels, EDP Sciences/CNRS Éditions, 2002, by C. Sabbah.

[26] See *Andre Weil Oeuvres Scientifiques,* Springer Verlag, 1979, I, 185.

[25] See Atiyah's *Collected Works,* Oxford, 1988, 5, 265.

V. S. Varadarajan
Department of Mathematics
University of California
Los Angeles, CA 90095–1555
vsv@math.ucla.edu