THE USE OF ALGEBRAIC STRUCTURES IN PHYSICS

By C. W. KILMISTER and D. J. NEWMAN†

Communicated by H. Bondi

Received 20 April 1960, *and in revised form* 20 March 1961

ABSTRACT. The hypothesis of the existence of abstract structures in physics, which has been made in recent papers, is made precise. Difficulties in previous formulations are removed and insight into Maxwell electrodynamics is obtained. General relativity is related to other field theories which deal with flux from sources, and new field equations for general relativity are proposed.

1. *Introduction*

1·1. There are a number of experimental observations in physics which can be termed 'structural', and the purpose of investigating algebraic structures is to find a uniform theory of such observations. The present paper seeks to exploit the structural features of the concept of a *field* and in so doing to relate the two best understood field theories—Maxwell's electromagnetism and general relativity.

1·2. There have been earlier attempts at such a theory by each of the authors and this paper forms a union of two lines of work. Kilmister has considered, in conjunction with E. W. Bastin, certain algebraic structures in physics (1), (2), (3). Newman (8), (9) has formulated a theory which he calls 'entity-structure', and it appears that this is an algebraic structure of the type considered in Bastin and Kilmister (1). As a result, the connexion of Newman's theory with experiments is of an unusual kind and the nature of this connexion is not clear in his papers.

Newman observed that the theory-language structures of Bastin and Kilmister and of Eddington, being algebraic, still left very considerable latitude in interpretation. To reduce this latitude, so as to have more definite evidence to make our theory refutable, we must proceed to a more complex structure. Newman, in keeping with Eddington's occasional introduction of infinitesimals, introduced an algebraic-differential structure. The algebraic part of this structure was fully worked out, but only an inadequate treatment of the differential structure has so far been given (Newman (8)).

1·3. In the present paper we avoid lengthy philosophical discussions by simply postulating the existence of abstract structures which have the following properties:

(*a*) they have an explicit algebraic formulation which can be determined;

(*b*) a given structure defines a set of physical experiments which are represented by its algebra;

(*c*) each abstract structure provides, by means of its algebraic formulation, a mathematical setting for those laws of physics which explain the results of the associated set of experiments.

These statements will be collectively termed the 'structural postulate'.

† Now at Queen Mary College, London.

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

Only certain sets of experiments—those which physicists usually regard as being 'fundamental'—are included in this scheme. For example, an experiment to measure the specific heat of lead could only be included if it were possible to describe all the relevant properties of the lead algebraically; and it would be far too complicated to do this precisely.

The structural postulate provides a new outlook regarding the 'truth' of physical laws. Effects outside the structural specification of an experiment have to be eliminated in some way if its results are to be 'true to the structure' and so obey the corresponding laws. This is a criterion for a good experiment. We therefore have no *absolute* measure for the 'truth' of a law—no law can be expected to hold outside its structural context.

The relationship between the structural algebras and the experiments they represent is not discussed in this paper (see, for example, Newman (9)). It is well known, however, that any physical theory has a range of experimental applicability, but this range is usually regarded as being indeterminate. The structural postulate states that, for fundamental theories, this range is always determinate and suggests a technique for finding it.

1·4. In what follows we shall use the following purely algebraic results taken from Newman (9):

Measurements are to be described in terms of two sets of structural coefficients, E_μ and E'_μ ($\mu = 1, 2, 3$). The set E_μ represents real or virtual comparisons of spatial displacements, and satisfies

$$\mathrm{E}_\mu \mathrm{E}_\nu = E_\kappa, \quad \mathrm{E}_\mu^2 = -1 \quad (\mu, \nu, \kappa \text{ cyclic in } 1, 2, 3) \tag{1·4·1}$$

so that the E_μ are just the quaternion units. The set E'_μ represents real or virtual comparisons of time displacements, and satisfies

$$\mathrm{E}_2'^2 = \mathrm{E}_3'^2 = -\mathrm{E}_1'^2, \quad \mathrm{E}'_2\mathrm{E}'_3 = \mathrm{E}'_1, \quad \mathrm{E}'_1\mathrm{E}'_3 = \mathrm{E}'_2, \quad \mathrm{E}'_2\mathrm{E}'_1 = \mathrm{E}'_3. \tag{1·4·2}$$

Both these sets of symbols have been shown (Newman (9), § 8) to have a particular matrix representation in which those representing actual comparisons are diagonal, the remaining two symbols of each set being non-diagonal. It was also shown (end of § 6) that displacements and therefore structural coefficients, can carry only real numerical factors if the distinction between space and time is to be maintained.

It is now apparent that these symbols are isomorphic with the structural coefficients which are introduced in the earlier paper by Newman. This has already been noted in his other paper. In the earlier paper, however, no distinction was made between the E_μ which represent real spatial comparisons and those which represent virtual spatial comparisons. In fact, the multiplication table (1·4·1) does not serve to distinguish any one of the symbols E_μ in the spatial reference frame (compare Newman (8), § 2), since we can impose the condition that any expression

$$\mathrm{E}_R = l\mathrm{E}_1 + m\mathrm{E}_2 + n\mathrm{E}_3 \quad (l, m, n \text{ scalars})$$

can be given a diagonal matrix representation and therefore represent a real spatial comparison.

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

A similar interpretation can be given to the E′. In this case, however, only two of the coefficients can possibly represent real time comparisons (E'_2 or E'_3). The coefficient E'_1 is distinguished by the multiplication table (1·4·2) or, equivalently, by the isomorphism between the three E_μ and iE'_2, iE'_3, $-E'_1$. We cannot make E'_1 diagonal if its matrix representation is to be real, and Newman has shown that the physical interpretation of the E′ (as comparisons of time displacements) impose this condition.

1·5. It has been shown ((8), § 6) that the symbolic scheme outlined above is just sufficient to represent a general relativistic displacement in the form

$$d\mathbf{X} = E'_3 c\,dt + E'_2 E_\mu dx^\mu,$$

where μ is summed over 1, 2, 3. We shall find it convenient to use the alternative four-dimensional notation defined by $F_\mu = E'_2 E_\mu = E_\mu E'_2$, $F_4 = E'_3$ so that

$$d\mathbf{X} = F_i dx^i, \tag{1·5·1}$$

where i is summed over 1, 2, 3, 4 and $dx^4 = c\,dt$. Using (1·4·1) and (1·4·2) in conjunction with (1·5·1) we have

$$d\mathbf{X}^2 = c^2 dt^2 + dx^\mu dx_\mu = \eta_{ij} dx^i dx^j = ds^2 \quad \text{(say)},$$

where $\eta_{ij} = \text{diag}(-1, -1, -1, 1)$ is the metric of special relativity; and if we define $\mathbf{V} = d\mathbf{X}/ds$, then $\mathbf{V}^2 = 1$. Now we are used to seeing the nature of space in the simple motion of a body (cf. Bastin and Kilmister (1), p. 284) and we can use the idea of such a motion, as they did, to give us a connexion of our abstract structure with ordinary physical language. We interpret $\mathbf{V}$ as the 'momentum' of a particle, following Newman (8).

It can be written in the alternative forms

$$\mathbf{V} = \beta\left(E'_3 + E'_2 \frac{\mathbf{v}}{c}\right) = F_i \frac{dx^i}{ds}, \tag{1·5·2}$$

where $\beta = c\,dt/ds$ and $\mathbf{v} = d\mathbf{x}/dt$.

1·6. In four-dimensional notation we have

$$\eta_{ij} = \tfrac{1}{2}(F_i F_j + F_j F_i), \tag{1·6·1}$$

and we can define

$$F^i = \eta^{ij} F_j,$$

$$F_{ij} = \tfrac{1}{2}(F_i F_j - F_j F_i),$$

$$F_5 = F_1 F_2 F_3 F_4 = \frac{1}{4!}\epsilon^{ijkl} F_i F_j F_k F_l,$$

$$F^5 = F^1 F^2 F^3 F^4,$$

$$F^{5i} = \frac{1}{3!}\epsilon^{ijkl} F_j F_k F_l.$$

The following algebraic results can then easily be derived

$$(F_5)^2 = -1, \quad F_{5i} = -F_{i5} = F_5 F_i,$$

$$F_i F^{5i} = -4F^5, \quad F_5 = -F^5.$$

(We distinguish both between upper and lower suffixes 5 and between the suffixes $5i$ and $i5$.)

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

Similarly
$$\mathrm{F}_5\mathrm{F}_{ij} = \mathrm{F}_{ij}\mathrm{F}_5 = -\tfrac{1}{2}\epsilon_{ijkl}\mathrm{F}^{kl},$$
$$\mathrm{F}_{ij}\mathrm{F}_{kl} - \mathrm{F}_{kl}\mathrm{F}_{ij} = 2\{\mathrm{F}_{il}\eta_{jk} - \mathrm{F}_{ik}\eta_{jl} + \mathrm{F}_{jk}\eta_{il} - \mathrm{F}_{jl}\eta_{ik}\},$$
$$\mathrm{F}_{ij}\mathrm{F}_k - \mathrm{F}_k\mathrm{F}_{ij} = 2\{\mathrm{F}_i\eta_{jk} - \mathrm{F}_j\eta_{ik}\},$$
and
$$\mathrm{F}_{ij}\mathrm{F}_k = \eta_{jk}\mathrm{F}_i - \eta_{ik}\mathrm{F}_j - \epsilon_{ijkl}\mathrm{F}^{5l}.$$

1·7. To prevent confusion we may note here that the E′ form a counter-example to the theorem of Bastin and Kilmister ((2), § 5·3), which states that the only algebras over the real field exemplifying (in the sense of that paper) the quadratic group S are its Frobenius algebra, and real quaternions. The error in the proof of the theorem is connected with the ambiguous sign of the square roots in the case (*b*). In fact the result should be that the only algebras exemplifying S are those stated or those derived from them by replacing the basis $(1, e_\alpha)$ by $(1, e_1, ie_2, ie_3)$.

2. *Field equations*

2·1. We wish now to use our algebraic structures in an analysis of the idea of a field. This will then be applied to the formulation of field equations in the form of integral relations between the flux of field over a closed surface and the total source strength within the surface. Such formulations are familiar to us for the Newtonian gravitational field, and for the electrostatic field.

2·2. Classically a 'field' is a derived concept which we introduce to explain the primary experimental result that bits of the physical world affect each other in various ways. Not only would the analysis of a set of n particles, each producing accelerations in all the others, be very complex; it would also be very difficult to write in a Lorentz covariant way. So we analyse the system into particles, which act as sources for fields, which in turn act on other particles. The fields act on the other particles according to the equations of motion; and they are related to their sources by the field equations. Historically the oldest such system is Newtonian gravitation, in which the equation of motion is, in a familiar notation,
$$m\ddot{\mathbf{r}} = \mathbf{P},$$
and the field equations are
$$\operatorname{curl}\mathbf{P} = 0, \quad \operatorname{div}\mathbf{P} = -4\pi G\rho. \tag{2·2·1}$$
The first field equation expresses the conservation of energy, and the second is equivalent to the flux theorem:
$$\int \mathbf{P}.\mathbf{dS} = -4\pi G\int \rho\, d\tau. \tag{2·2·2}$$
This result enables us to think about the field in terms of tubes of force, etc.

2·3. It is well known that these results can be more conveniently expressed in an algebraic form by using quaternions. Using now e_1, e_2, e_3 for the quaternion units, we define the surface element
$$dS = e_2 dy\, e_3 dz + e_1 dx\, e_2 dy + e_3 dz\, e_1 dx$$
$$= e_1 dy\, dz + e_2 dz\, dx + e_3 dx\, dy$$
and the volume element
$$d\tau = e_1 e_2 e_3 dx\, dy\, dz = -dx\, dy\, dz.$$
If now we write
$$\nabla = e_1\frac{\partial}{\partial x} + e_2\frac{\partial}{\partial y} + e_3\frac{\partial}{\partial z},$$

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

so that, writing P for the quaternion corresponding to $\mathbf{P}$,

$$\nabla P = -\operatorname{div}\mathbf{P} + \operatorname{curl}\mathbf{P},$$

the field equations become simply

$$\nabla P = 4\pi G\rho, \tag{2·3·1}$$

and the flux theorem is

$$\int dS P = -\int d\tau \nabla P$$
$$= -4\pi G \int d\tau \rho. \tag{2·3·2}$$

Although this simple example is not Lorentz invariant it contains two important lessons. The field equation (2·3·1) now contains both the original equations, one of which is usually included by the device of introducing a potential. Also the equation suggests a generalization in which ρ, P are replaced by general quaternions

$$\rho + \alpha_\mu e_\mu, \quad \phi + P.$$

The field equations (2·3·1) then state, in vector notation,

$$\operatorname{div}\mathbf{P} = -4\pi G\rho, \quad \operatorname{curl}\mathbf{P} + \operatorname{grad}\phi = -4\pi G\boldsymbol{\alpha}. \tag{2·3·3}$$

2·4. The field at a point is known, via the equations of motion, by the behaviour of test-particles (that is, particles not affecting the field to be measured). To specify such a particle is of course to specify a great deal about the abstract structure in which we are working, and so to delimit the extent of our observations. Therefore, two cases arise:

(i) the change of motion of the test particle is independent of the nature of the particle, so far as the given abstract structure is concerned, or

(ii) this independence does not hold.

In the first case we can by making measurements at just one point find nothing whatever about the field, since all the test-particles are acted on similarly. This is, of course, just the case of the principle of equivalence for the gravitational field. This field only has significance for two points; historically in Newtonian gravitation the two points chosen were a source-point and a field-point (as when one discusses the orbits of the planets and uses the sun as origin). An alternative approach, which has the advantage of preserving the analysis into fields and particles existing separately, is to consider two nearby points in the field (the 'equation of geodesic variation' method). In any case at least two points must be specified to make a meaningful statement about the field; we call such fields *two-fields*.

In case (ii) it is possible to determine something about the field by using different test-particles at one point. It may not be possible to determine the field completely in this way (e.g. a combined electromagnetic and gravitational field), but we distinguish a 'pure case' in which this is so, and call such a field a *one-field*. Although the two-field is in some ways conceptually simpler, our ideas of a classical field have mostly been developed with the electromagnetic field, which is a one-field. It is therefore easiest to begin with the analysis of Lorentz-invariant one-fields. (A typical non-relativistic one-field is simply that described in § 2·2, and there is no more to be said.)

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

2·5. We may notice that the distinction between one-fields and two-fields also divides abstract structures up into two classes, according to whether they are relevant to one or two particle observations. The appropriate structural algebra to describe the Lorentz-invariant test-particle has been given in § 1·5. The equation of motion of the test-particle described by $\mathbf{V}$ must be of at least the second order (to allow for starting conditions corresponding to arbitrary position and velocity). Hence we assume it to be of the form

$$\frac{d\mathbf{V}}{ds} = \mathbf{G} = G^a \mathbf{F}_a, \tag{2·5·1}$$

where G depends both on the field and on the particle. Since $\mathbf{V}^2 = 1$ we have at once

$$V_a G^a = 0$$

so that G depends on V. Two alternative forms of G which satisfy this condition are

$$G^a = H^{ap} V_p \quad \text{or} \quad G^a = (\delta^a_u - V^a V_u) H^u \quad (\text{where } H^{ap} = -H^{pa}).$$

These are completely equivalent, for one may write

$$H_u = H_{uv} V^v, \quad H_{uv} = H_u V_v - H_v V_u,$$

without restricting G if H^u, H^{uv} are allowed to depend on $\mathbf{V}$. The two simplest cases to consider are:

(i) fields of type I, where $G^a = H^{ap} V_p$ and H^{ap} is independent of $\mathbf{V}$;

(ii) fields of type II, where $G^a = (\delta^a_u - V^a V_u) H^u$ and H^u is independent of $\mathbf{V}$.

A field which falls into neither class may be called of type III, but we shall not consider these here.

2·6. The quantities H^{ap}, H^u are not the field quantities, since they depend on the nature of the test-particle, for example, its charge-mass ratio (although, in types I and II, not on its velocity). To determine their connexion with the field quantities we need a further assumption about the sources of the field. Since the test-particle is here specified by its velocity $\mathbf{V}$ and certain intrinsic parameters we shall suppose that the sources of the field are made up of numbers of such particles. The source term in the algebra must then be constructed in an invariant manner from a four-vector

$$J^a = \rho V^a,$$

where ρ is a scalar specifying the number of particles in the source. If we limit ourselves to 16-fold algebra the most general source-term is then

$$\mathbf{J} = \phi + \mathbf{F}_a J^a,$$

where ϕ is a scalar function of $J_a J^a$. (For the only invariant functions of J^a are its magnitude, and J^a itself. More complex covariant quantities, e.g. $J_a J_b$ are not expressible in the 16-fold algebra.)

The flux equation is now easily proved to be

$$\int d^3x \mathfrak{F} = \int d^4x D\mathfrak{F}, \tag{2·6·1}$$

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

where
$$d^3x = \mathbf{F}_2\mathbf{F}_3\mathbf{F}_4 dx^2 dx^3 dx^4 + \ldots$$
$$= -(\mathbf{F}_{15} dx^2 dx^3 dx^4 + \ldots),$$
$$d^4x = \mathbf{F}_1\mathbf{F}_2\mathbf{F}_3\mathbf{F}_4 dx^1 dx^2 dx^3 dx^4$$
$$= \mathbf{F}_5 dx^1 dx^2 dx^3 dx^4,$$

and
$$D = \mathbf{F}^u \frac{\partial}{\partial x^u}.$$

Hence the field equations are
$$D\mathfrak{F} = 4\pi\mathbf{J}. \tag{2·6·2}$$

Since
$$D^2\mathfrak{F} = \Box^2\mathfrak{F} = 4\pi D\mathbf{J},$$

we have a solution (subject to appropriate conditions at infinity; see, for example, Synge (11), p. 408)
$$\mathfrak{F} = A'\int_{N'} D\mathbf{J}'\, d\omega' + (1-A')\int_{N''} D\mathbf{J}''\, d\omega'', \tag{2·6·3}$$

where A' is arbitrary and the integrations are over the past and future null cones. Hence if we write
$$\mathfrak{F} = D\mathbf{A}, \quad \text{then} \quad \mathbf{A} = \psi + \mathbf{F}_a A^a,$$

and $\mathbf{A}$ has the same form as $\mathbf{J}$.

2·7. Now we notice that the field splits up, in a Lorentz invariant fashion, into a true scalar potential ψ, and vector-potential A^a. Similarly $\mathfrak{F}$ splits into

$$\left.\begin{aligned} \mathfrak{F}_u &= \partial_u \psi, \\ \mathfrak{F}_0 + \tfrac{1}{2}\mathbf{F}_{uv}\mathfrak{F}^{uv} &= -\partial_u A^u + \tfrac{1}{2}(\partial_u A_v - \partial_v A_u)\,\mathbf{F}^{uv}. \end{aligned}\right\} \tag{2·7·1}$$

We cannot, in an invariant fashion, derive a field of type I for an (isotropic) particle from the scalar field; neither can we derive a field of type II from the vector field. (The proof of these statements is immediate.)

Let us first consider fields of type I and suppose accordingly that they are to be derived from the vector-potential field alone. In the special Lorentz frame in which the particle is at rest the six-vectors H_{uv}, $\mathfrak{F}_{uv}$ form pairs of three-vectors $\mathbf{H}_1, \mathbf{H}_2, \mathfrak{F}_1, \mathfrak{F}_2$. We have a functional dependence
$$H_{uv} = L_{uv}(\mathfrak{F}_{ps}).$$

First suppose that $\mathfrak{F}_1$, $\mathfrak{F}_2$ are not parallel. Then we can certainly write
$$\mathbf{H}_1 = \alpha_{11}\mathfrak{F}_1 + \alpha_{12}\mathfrak{F}_2 + \alpha_{13}\mathfrak{F}_1 \wedge \mathfrak{F}_2,$$
$$\mathbf{H}_2 = \alpha_{21}\mathfrak{F}_1 + \alpha_{22}\mathfrak{F}_2 + \alpha_{23}\mathfrak{F}_1 \wedge \mathfrak{F}_2,$$

where the α_{ij} may depend on the magnitudes of $\mathfrak{F}_1$, $\mathfrak{F}_2$. But, paying attention to the Lorentz invariance, we must have $\alpha_{13} = \alpha_{23} = 0$, and
$$H_{mn} = \alpha\mathfrak{F}_{mn} + \beta\mathfrak{F}^*_{mn}, \tag{2·7·2}$$

where α, β depend only on the invariants of $\mathfrak{F}_{mn}$ and $\mathfrak{F}^*_{mn}$ is the dual of $\mathfrak{F}_{mn}$. In the case when the effects of two fields on a single particle are additive, α, β must be constants. If $\mathfrak{F}_1$, $\mathfrak{F}_2$ are parallel then $\mathfrak{F}_{ab} = i\mathfrak{F}^*_{ab}$ so that $\mathfrak{F}_{ab}$ cannot be real; in any case the same form results for H_{uv}.

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

We can give these equations several interpretations, according to values of α, β.

(i) If $\alpha = 1$, $\beta = 0$ always we have a model of a special relativistic theory of gravitation (since $\alpha = 1$ implies the principle of equivalence). Such a field is then unobservable at one point, but if we consider the source as well we can solve many problems. For example, it is this field which gives the frequently-quoted result of an advance of perihelion of Mercury of one-sixth of the observed value (see, for example, Stephenson and Kilmister (10), p. 77).

(ii) If $\beta = 0$ but α varies for different particles and is independent of $\mathfrak{F}_{uv}$, we have a theory almost identical with the usual Maxwell electrodynamics, and α is the charge-mass ratio. The field equations become

$$\left.\begin{aligned} \partial_m \mathfrak{F}^{*mn} &= 0, \\ \partial_m \mathfrak{F}^{mn} + \partial^n \mathfrak{F}_0 &= 4\pi J^n, \end{aligned}\right\} \tag{2·7·3}$$

where
$$\mathfrak{F}_{mn} = \partial_m A_n - \partial_n A_m, \tag{2·7·4}$$

and
$$\mathfrak{F}_0 = -\partial_m A^m.$$

The usual Maxwell theory imposes conservation of sources in the form

$$\partial_u J^u = 0.$$

This results in the present theory if the Lorentz gauge condition $\mathfrak{F}_0 = 0$ is imposed on A^u, or more generally if $\Box^2 \mathfrak{F}_0 = 0$, but not otherwise. In general, we have instead

$$4\pi \partial_u J^u = -\Box^2(\partial_u A^u),$$

so that
$$J^u + \frac{1}{4\pi} \Box^2 A^u \tag{2·7·5}$$

is the conserved vector. These equations represent a modification of Maxwell's which allow for creation of charge. Another such modification was suggested recently by Lyttleton and Bondi (7), but the present one differs from theirs in still allowing the description of the field in terms of flux from sources.

(iii) If $\beta \neq 0$ we have a Maxwell-type theory in which also free poles are allowed.

2·8. For fields of type II we must have

$$H_u = \alpha \mathfrak{F}_u = \alpha \partial_u \psi,$$

where α may depend on the magnitude of $\mathfrak{F}_u$ but in the simplest cases does not. Hence

$$\frac{dV^m}{ds} = \alpha[\mathfrak{F}^m - V^m(V_n \mathfrak{F}^n)]. \tag{2·8·1}$$

This is most easily interpreted by writing it in the form

$$\frac{d}{ds}(mV^u) = m\alpha \mathfrak{F}^u, \tag{2·8·2}$$

where $m = m_0 e^{\alpha\psi}$. The ψ is then visualized as potential energy, which increases the mass m accordingly. The field equations are

$$\partial_m \mathfrak{F}^m = 4\pi\phi, \quad \partial_m \mathfrak{F}_n - \partial_n \mathfrak{F}_m = 0.$$

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

If $\alpha = 1$ always we have another candidate for a special relativistic theory of gravitation; this one gives about one-third of the observed perihelion advance (Stephenson and Kilmister (10), p. 71).

3. *Interpretation of fields of type* I

3·1. The fields of type I gave rise in § 2·7 to well-known theories, which have equations of motion with a ready interpretation. We have seen in Newman ((8), § 7) that all the automorphisms of the F-algebra are inner automorphisms of the form

$$\bar{\mathbf{F}}_i = \mathbf{Q}\mathbf{F}_i\mathbf{Q}^{-1}, \tag{3·1·1}$$

where $\mathbf{Q}$ is some non-singular symbol. (Inner automorphisms are termed 'rotations' by Eddington.) Now, given a certain restricted set of $\mathbf{Q}$'s in (3·1·1) we can write

$$d\mathbf{X} = \mathbf{F}_i dx^i = \bar{\mathbf{F}}_i d\bar{x}^i. \tag{3·1·2}$$

This restricted set of $\mathbf{Q}$'s was investigated by Newman ((8), § 7) and found to correspond to the proper Lorentz group. The relationship between the infinitesimals dx^i and $d\bar{x}^i$ is determined by the actual $\mathbf{Q}$ used in (3·1·1).

The point of interest is this. In the transformation defined by (3·1·1) and (3·1·2) the quantity $d\mathbf{X}$ is left unaltered. We can therefore interpret $d\mathbf{X}$ as a *physical* quantity —left invariant for arbitrary changes in the reference frame $\mathbf{F}_i$. An alteration in the *physical* quantity $d\mathbf{X}$ will be defined by

$$d\bar{\mathbf{X}} = \mathbf{Q}d\mathbf{X}\mathbf{Q}^{-1} = \bar{\mathbf{F}}_i dx^i = \mathbf{F}_i d\bar{x}^i, \tag{3·1·3}$$

and we note that $$d\bar{\mathbf{X}}d\bar{\mathbf{Y}} = \mathbf{Q}d\mathbf{X}d\mathbf{Y}\mathbf{Q}^{-1},$$

and in particular $$d\bar{\mathbf{X}}^2 = d\mathbf{X}^2 = ds^2.$$

Thus a physical change as defined by (3·1·3) does not alter the algebraic structure of the systems, nor the value of an interval ds. We shall now show that this determines the equation of motion.

3·2. Suppose a physical change of the velocity vector is of the form

$$\mathbf{V} \to \bar{\mathbf{V}} = \mathbf{V} + d\mathbf{V} = \mathbf{Q}\mathbf{V}\mathbf{Q}^{-1},$$

where $\mathbf{Q}$ is an infinitesimal rotation, say

$$\mathbf{Q} = 1 + \tfrac{1}{2}\mathfrak{F}ds.$$

Then $$\frac{d\mathbf{V}}{ds} = \tfrac{1}{2}(\mathfrak{F}\mathbf{V} - \mathbf{V}\mathfrak{F}). \tag{3·2·1}$$

If $\mathfrak{F}$ is independent of $\mathbf{V}$ and depends only on the position of the particle we may identify it with a field. Since by (3·1·2) $\mathbf{V}$ has the special form

$$\mathbf{V} = \mathbf{F}_i \frac{dx^i}{ds},$$

$\mathfrak{F}$ must be restricted to the form

$$\mathfrak{F} = \tfrac{1}{2}\mathbf{F}_{ij}l^{ij} + l \tag{3·2·2}$$

(the scalar term l contributing nothing to $d\mathbf{V}/ds$). The equation (3·2·1) then has the form of a type I equation of motion.

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

The unrestricted form of $\mathfrak{F}$ can be written as

$$\mathfrak{F} = \tfrac{1}{2}\mathrm{F}_{ij}l^{ij} + l + \mathrm{F}_5 l^5 + \mathrm{F}_i l^i + \mathrm{F}_{5i}l^{5i} \tag{3·2·3}$$

in four-dimensional notation. The action of such a force on V will produce a generalized velocity vector

$$\mathbf{W} = \tfrac{1}{2}\mathrm{F}_{ij}w^{ij} + w + \mathrm{F}_5 w^5 + \mathrm{F}_i w^i + \mathrm{F}_{5i}w^{5i} \tag{3·2·4}$$

restricted by the condition $\mathbf{W}^2 = 1$. We may note that if $w = 0$ then W can be transformed to the form V by one of the *unrestricted* transformations (3·1·1).

Allowing for the generalization of V to W, we can now write the most general form of the equation of motion (3·2·1) as

$$\frac{d\mathbf{W}}{ds} = \tfrac{1}{2}(\mathfrak{F}\mathbf{W} - \mathbf{W}\mathfrak{F}). \tag{3·2·5}$$

Newman assumed ((8), § 9) that the general law of motion had the form

$$\frac{d\mathbf{W}}{ds} = \mathfrak{F}\mathbf{W}.$$

This equation has two faults, as became evident there. The order of the terms on the right-hand side is arbitrary, and the resultant expression for $d\mathbf{W}/ds$ contains a large number of terms which have no apparent interpretation. We therefore replace this by (3·2·5), which has no arbitrariness, and we shall shortly find that all the terms in $d\mathbf{W}/ds$ have simple interpretations.

3·3. It is illuminating to determine the explicit restrictions imposed on the components of W by the relation $\mathbf{W}^2 = 1$.

From (3·2·4) we have

$$\begin{aligned}1 = \mathbf{W}^2 = {} & w^2 + w^i w_i + w^{5i}w^5{}_i + \tfrac{1}{2}w^{ij}w_{ij} - (w^5)^2 + \tfrac{1}{4}\mathrm{F}_5\epsilon_{ijkl}w^{ij}w^{kl} + 2\mathrm{F}_5 w^5 w \\ & + \mathrm{F}_5\mathrm{F}_{ij}(w^5 w^{ij} + 2w^{5i}w^j + w w^{ij}) + \mathrm{F}_5\mathrm{F}_{ijk}w^{ij}w^{5k} + 2\mathrm{F}_{5i}w^{5i}w + \mathrm{F}_{ijk}w^{ij}w^k + 2\mathrm{F}_i w^i w.\end{aligned}$$

Comparing coefficients:

$$\left.\begin{aligned}
&(a) \quad 1 = w^2 - (w^5)^2 + w^i w_i + w^{5i}w^5{}_i + \tfrac{1}{2}w^{ij}w_{ij}, \\
&(b) \quad 0 = -\epsilon_{ijkl}w^{ij}w^{5k} + 2w_l w, \\
&(c) \quad 0 = -\epsilon_{ijkl}w^{ij}w^k + 2w^5{}_l w, \\
&(d) \quad 0 = \tfrac{1}{4}\epsilon_{ijkl}w^{ij}w^{kl} + 2w^5 w, \\
&(e) \quad 0 = w^5 w^{ij} + w^{5i}w^j - w^{5j}w^i + w^{ij}w.
\end{aligned}\right\} \tag{3·3·1}$$

Equations (3·3·1 (*b*)–(*e*)) are satisfied identically when $\mathbf{W} = \mathbf{V}$, and (3·3·1 (*a*)) is then equivalent to $\mathbf{V}^2 = 1$.

3·4. As in § 2·3 the algebraic formulation suggests generalizations. It is easy to compute the components of the general equations

$$D\mathfrak{F} = \mathbf{J},$$

$$\frac{d\mathbf{W}}{ds} = \tfrac{1}{2}(\mathfrak{F}\mathbf{W} - \mathbf{W}\mathfrak{F}),$$

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

where $\mathbf{J} = \rho\mathbf{W}$, but it is not clear how to interpret the complete set. However, we know there are subsets (e.g. the electromagnetic case considered in 3·1 and 3·2) which can consistently be taken by themselves.

The two largest such subsets are

$$(a)\quad \mathbf{W} = W_5\mathbf{F}^5 + W_u\mathbf{F}^u, \qquad \mathfrak{F} = \tfrac{1}{2}\mathfrak{F}_{uv}\mathbf{F}^{uv} + \mathfrak{F}_{u5}\mathbf{F}^{u5};$$
$$(b)\quad \mathbf{W} = W_u\mathbf{F}^u + W_{u5}\mathbf{F}^{u5}, \qquad \mathfrak{F} = \tfrac{1}{2}\mathfrak{F}_{uv}\mathbf{F}^{uv} + \mathfrak{F}_5\mathbf{F}^5.$$

The corresponding field equations and equations of motion are

$$(a)\qquad \partial_m\mathfrak{F}^{m5} = J^5, \qquad \frac{dW^m}{ds} = \mathfrak{F}^{mn}W_n + W^5\mathfrak{F}_{n5};$$
$$\partial_m\mathfrak{F}_{n5} - \partial_n\mathfrak{F}_{m5} = 0,$$
$$\partial_m\mathfrak{F}^{mn} + \partial^n\mathfrak{F}_0 = J^n, \qquad \frac{dW^5}{ds} = \mathfrak{F}^{m5}W_m.$$
$$\partial_{(m}\mathfrak{F}_{ns)} = 0.$$
$$(b)\qquad \partial_m\mathfrak{F}^{mn} + \partial^n\mathfrak{F}_0 = J^n, \qquad \frac{dW_m}{ds} = \mathfrak{F}^5W_{m5} + \mathfrak{F}_{mn}W^n;$$
$$\partial_m\mathfrak{F}_5 + \epsilon_{mnsr}\partial^n\mathfrak{F}^{sr} = J_5, \qquad \frac{dW_{m5}}{ds} = \mathfrak{F}_5W_m + \mathfrak{F}_{mn}W^5.$$

Any smaller subsets are in fact the obvious subsets of these two sets.

It is not at all apparent what physical interpretation (if any) is to be given to set (*b*), but in set (*a*) the identification of $\mathfrak{F}_{uv}$, $\mathfrak{F}_{u5}$ as electromagnetic and gravitational field strengths, and W_u as momentum are immediate, and W_5 is seen to be the gravitational potential energy.

4. *The theory of two-fields*

4·1. Our definition of a two-field implies that at each point there exists a coordinate-system relative to which any particles at the point are unaccelerated. Locally the geometry in such a coordinate-system is Minkowskian, and so we can define the same algebraic structure as in § 2. But unless the two-field is zero such a coordinate-system is only possible locally; virtually the only possible geometry for the whole domain is the Riemannian one, and the $\mathbf{F}_u$ at each point can be represented as the local orthogonal components of certain functions of position $\mathscr{E}_u$, say, satisfying

$$\left.\begin{aligned}\mathscr{E}_u\mathscr{E}_v + \mathscr{E}_v\mathscr{E}_u &= 2g_{uv},\\ \mathscr{E}_u\mathscr{E}_v - \mathscr{E}_v\mathscr{E}_u &= 2\mathscr{E}_{uv}.\end{aligned}\right\} \tag{4·1·1}$$

We can then define the contravariant $\mathscr{E}^u$ and the affine connexion by

$$\mathscr{E}^u = g^{uv}\mathscr{E}_v, \quad \mathscr{E}_{u,v} = \Gamma^s_{uv}\mathscr{E}_s \tag{4·1·2}$$

with $\Gamma^s_{uv} = \Gamma^s_{vu}$. It follows that Γ^s_{uv} is the Christoffel bracket connexion. If any vector is written

$$\mathbf{A} = A^u\mathscr{E}_u,$$

then

$$\mathbf{A}_{,s} = \mathscr{E}_uA^u{}_{;s}, \tag{4·1·3}$$

where a comma denotes ordinary differentiation, and a semi-colon covariant differentiation.

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

4·2. Now consider two nearby orbits of test-particles from a field of such orbits defining everywhere in a domain the unit tangent $\mathbf{t} = t^a \mathscr{E}_a$. If $\mathbf{y} = y^a \mathscr{E}_a$ is the infinitesimal vector orthogonal to one orbit and intersecting the other, we have as usual

$$d\mathbf{y}/ds = y^a \mathbf{\Phi}_a,$$

where

$$\mathbf{\Phi}_a = \mathbf{t}_{,a} - \frac{dt_a}{ds}\mathbf{t}. \tag{4·2·1}$$

We expect an equation of motion of the second order, since we can choose an arbitrary starting point and value of $d\mathbf{y}/ds$; and we easily calculate

$$\frac{d^2\mathbf{y}}{ds^2} = \mathscr{E}^u R_{uvsp} t^v y^s t^p + y^s \mathbf{\Lambda}_s, \tag{4·2·2}$$

where

$$\mathbf{\Lambda}_s = (v_{r:s} - v_{s;r}) t^r \mathbf{t} + \mathbf{v}_{,s} - 2v_s \mathbf{v}, \tag{4·2·3}$$

and

$$\mathbf{v} = \frac{d\mathbf{t}}{ds}.$$

(Of course, if $v_s = 0$ we get the usual equation of geodesic variation.)

4·3. We should expect to get more terms than in the usual equation of geodesic variation because if any one-field is present, it is again represented (as its derivative) as a two-field. We can, however, recognize the part of $d^2\mathbf{y}/ds^2$ which represents the reducible two-fields produced in this way, since it will depend on the curvature of the original path, and not merely on its direction. This is the part represented by $\mathbf{\Lambda}_s$; now we can prove that a necessary and sufficient condition for $\mathbf{\Lambda}_s = \mathbf{0}$ is that $\mathbf{v} = \mathbf{0}$, which implies that the original orbits are geodesics. Such a condition is obviously sufficient. To prove the necessity take a scalar product with $\mathbf{t}$:

$$0 = t^p(v_{p;s} - v_{s;p}) + t^a v_{a;s}$$

(since $t^a v_a = 0$). Hence, in particular

$$t^a \frac{dv_a}{ds} = 0,$$

so that

$$v_s \frac{dt^s}{ds} = v_s v^s = 0.$$

But v_s cannot be null, since it is orthogonal to a time-like vector, which proves the result $v_s = 0$.

Thus if we can ignore the effect of one-fields we have the usual equation of geodesic variation, and it follows that the appropriate quantities to describe the two-field are the Riemann symbols R_{uvsp}.

4·4. In terms of the 256-fold algebra, which is the direct product of two sixteen-fold algebras, such a field can be written

$$\mathbf{R} = R_{uvsp} \mathscr{E}^{uv} \times \mathscr{E}^{sp}. \tag{4·4·1}$$

It is then possible to write the equation of motion (4·2·2) algebraically, but as it is rather complex there is no advantage in doing so. Since $R_{uvsp} = R_{spuv}$, it does not matter which set of $\mathscr{E}_u$ is used to describe the coordinates, volumes, etc. We accordingly

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

postulate the field equations by equating the flux of **R** across a closed surface to the total source within it, i.e.

$$\int d^3x\mathbf{R} = \int d^4x\mathbf{J}$$

$$= \int d^4xD\mathbf{R}, \tag{4·4·2}$$

where $$d^3x = (\mathscr{E}_2\mathscr{E}_3\mathscr{E}_4 dx^2dx^3dx^4 + \ldots) \times 1,$$

$$d^4x = \mathscr{E}_5 dx^1dx^2dx^3dx^4 \times \mathbf{1}.$$

The field equations are then $$D\mathbf{R} = \mathbf{J} \tag{4·4·3}$$

which reduce to $$R^{uvsp}{}_{;p} = J^{uvs}, \quad R_{uv(sp;t)} = 0, \tag{4·4·4}$$

where the brackets denote a sum of cyclic permutations. The second equation is of course an identity, and simply implies that the source **J** is not in fact so complex as it might be (as also in the electromagnetic case, where there are no free poles).

Using the second equation in the first gives at once that

$$R_{su;v} - R_{sv;u} = J_{uvs}, \tag{4·4·5}$$

where $$R_{ij} = R^k_{ijk},$$

which makes it clear that, in the absence of sources ($J_{uvs} = 0$), any solution of the field equations of general relativity is also a solution of the present equations.

On the other hand, suppose that

$$R_{sm;n} - R_{sn;m} = 0 \tag{4·4·6}$$

everywhere in a domain D containing a surface S on which $R_{sm} = 0$. Choose a coordinate system so that S is $x^0 = 0$, and let Greek suffixes take values 1, 2, 3. From (4·4·6) it follows at once that $R_{i\alpha;j} = 0$ at every point of S; using this result in the contracted Bianchi identity then gives $R_{00;0} = 0$. Hence $R_{ij;0} = 0$ at each point of S and so if we proceed a small distance along the normal to S, R_{ij} is still zero. Proceeding in this way we can prove that $R_{ij} = 0$ everywhere in D; this result is due to Lichnerowicz. Thus the equations (4·4·6) are not so much weaker than Einstein's as at first sight appears.

It might be thought that we could appeal to a limiting form of Lichnerowicz's result in which S is chosen as the sphere at infinity at all times, and then, if $R_{abcd} = 0$ there, $R_{ab} = 0$ everywhere. However, the proof of Lichnerowicz's result clearly does not apply, and it has been shown by A. H. Thompson that the limiting form of the theorem is false, since the equations (4·4·6) admit a static axially symmetric solution for two particles with no singularity joining them. In the linearized theory one finds easily an inverse square law field, as always, and also a direct distance law which may be repulsive. (This accounts for the axially symmetric solution but also makes experimental differentiation between the present equations and Einstein's very difficult, since the most distant matter produces the overwhelmingly large part of the direct distance law, and this matter is not subject to manipulation or even observation.)

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press

4·5. The source term J^{uvs} has 20 independent components, since $\epsilon_{mnsr}J^{nsr} \equiv 0$, and these are further restricted by 6 'conservation' identities

$$J^{mns}{}_{;s} = 0. \tag{4·5·1}$$

These identities are at first sight severe ones, whereas the corresponding identity in electromagnetic theory was simply the equation of continuity for charge. Thus if we imagined a medium made up of a large number of point-sources, of density ρ, with 'charge-current' vector J^s, we would have

$$(\rho J_m)_{;n} - (\rho J_n)_{;m} = 0.$$

But in fact such sources are not the appropriate ones to consider; and if we write down (e.g.) a Schwarzschild interior solution, the resultant source term must certainly satisfy (4·5·1) identically.

4·6. The field equations (4·4·4) thus form a somewhat weaker alternative to Einstein's equations $R_{uv} = 0$, but the difference does not affect any present experimental results. The weaker equations have been used by Lichnerowicz (6) in his quantization of the gravitational field; but he seems to have regarded them as a convenient mathematical alternative which would make quantization easier. In the same way the equations occur implicitly in Bel's investigation of the 'super-energy tensor' T_{ijkl} of the gravitational field when sources are present (Bel (4)), but they are then regarded as identities, J_{ijk} being defined by (4·4·5).

In the present paper we suggest that the general equations are (4·4·4), of which Einstein's equations are a special solution. And the reason for this suggestion is that the flux theorem, which leads to the weaker equations, is an immediate generalization to covariant form of the inverse square law in the form of Gauss's theorem. This bridges the hiatus in the usual derivations of the field equations noticed by Kilmister and Stephenson (5).

The work described in this paper was supported in part by Wright ADC, ARDC, USAF through its European Office.

REFERENCES

(1) BASTIN, E. W. and KILMISTER, C. W. *Proc. Cambridge Philos. Soc.* 50 (1954), 278.
(2) BASTIN, E. W. and KILMISTER, C. W. *Proc. Cambridge Philos. Soc.* 50 (1954), 439.
(3) BASTIN, E. W. and KILMISTER, C. W. *Proc. Cambridge Philos. Soc.* 51 (1955), 454.
(4) BEL, L. *C.R. Acad. Sci. Paris*, 248 (1959), 1297.
(5) KILMISTER, C. W. and STEPHENSON, G. *Nuovo Cimento*, 12 (1955), 361.
(6) LICHNEROWICZ, A. *C.R. Acad. Sci. Paris*, 247 (1958), 433.
(7) LYTTLETON, R. A. and BONDI, H. *Proc. Roy. Soc. London, Ser.* A, 252 (1959), 313.
(8) NEWMAN, D. J. *Proc. Roy. Irish Acad. Sect.* A, 59 (1958), 29.
(9) NEWMAN, D. J. *Mathematika*, 5 (1958), 146.
(10) STEPHENSON, G. and KILMISTER, C. W. *Special relativity for physicists* (London, 1958).
(11) SYNGE, J. L. *Relativity: the special theory* (Amsterdam, 1956).

KING'S COLLEGE
LONDON

https://doi.org/10.1017/S0305004100036008 Published online by Cambridge University Press