Quaternions in Relativity*

PETER RASTALL

Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada
and
Department of Mathematics, University of Texas, Austin, Texas

CONTENTS

FOREWORD

Quaternion notation, far from being an outmoded Victorian fad, is shown to be a concise and perspicuous formalism in field theory. Clouds of indices are in great part evaporated, features which depend on the arbitrary choice of a coordinate system sink into the background, and physically significant relationships are emphasized.

This paper is written for those who know special relativity fairly well and Riemannian geometry fairly poorly. In Part I, after the historical introduction, there is an account of quaternion algebra which, although brief, should be sufficient preparation for anyone unacquainted with the subject. The notation will seem familiar to most physicists because quaternion basis elements are chosen to have the same multiplication law as Pauli spin matrices.

It is hoped that the work will prove useful to quantum field theorists, but it has a classical bias. We do not deal with the possibility of replacing complex numbers by quaternions in quantum mechanics,[1] nor with recent attempts to describe elementary particles by means of quaternions.[2]

I. INTRODUCTION

I.1 Historical Remarks

If the use of quaternions is advantageous, why have they been used so little? To answer this question we give a short, tendentious history.

It was recognized early that special relativity can be elegantly written in quaternion notation.[3,4] However, the formalism was never popular: perhaps because it is slightly harder to manipulate quaternions than tensors, but also because the Lorentz-transformation properties of a tensor are explicitly shown by its indices.

After the invention of the quantum mechanics of spinning particles there was a renewed attempt to introduce quaternions. This was natural because the Pauli spin matrices prove to be quaternion basis elements in a thin disguise. Again the attempt failed, and the matrix notations of Pauli and Dirac were almost universally adopted. There was here no question of Lorentz-transformation properties being shown more explicitly in one formalism than in the other, but only that it is less trouble to multiply a column vector by a matrix than to multiply a matrix by a

* This work was supported in part by a grant from the National Research Council of Canada, and by U.S. Air Force grant AF-AFOSR-454-63.

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936); T. Kaneno, Prog. Theoret. Phys. (Kyoto), **23**, 17 (1960); D. Finkelstein, J. M. Jauch, S. Schiminovitch, and D. Speiser, J. Math. Phys. **3**, 207 (1962), and **4**, 788 (1963); D. Finkelstein, J. M. Jauch, and D. Speiser, J. Math. Phys. **4**, 136 (1963).
[2] E. J. Schremp, Phys. Rev. **99**, 1603 (1959); G. R. Allcock, Nucl. Phys. **27**, 204 (1961).
[3] L. Silberstein, *The Theory of Relativity* (Macmillan and Company Ltd., London, 1924), 2nd ed., p. 148, footnote.
[4] F. Klein, Physik. Z. **12**, 17 (1911).

matrix—which is effectively what one has to do in writing, say, the Dirac equation in quaternion form.[5,6]

Another rival formalism was the spinor calculus.[7] This is completely explicit, in that transformation properties are shown by sets of indices. Four kinds of index are needed (co and contravariant, dotted and undotted), which most people find it tedious to juggle. Since one can readily indicate transformation properties without using indices, the spinor calculus has been little used in either flat or curved space–time.[8]

The Dirac equation in its γ-matrix version was soon generalized to the case of Riemannian space–time.[9] The γ matrices, which can be chosen to be constants in flat space–time, became functions of the space–time coordinates in the generalized equation. Further, there is the possibility of submitting the generalized equation to coordinate-dependent similarity transformations. This rather complicated state of affairs may be brought to order by introducing a reference tetrad of vectors at each event of space–time, and limiting oneself at each event to the set of similarity transformations which correspond to possible rotations of the tetrad. One can avoid talking about reference tetrads, but at the price of simplicity: compare the papers of Schrödinger,[9] and Bargmann,[10] and see Refs. 11, 12.

Coordinate-dependent γ matrices are fairly complex objects, but that is not their worst feature. More serious is that they have no obvious physical interpretation, so that by using them one loses much intuitive insight. It is here that the quaternion formalism has a decisive advantage. One can write field equations such as the Dirac equation in terms of quaternions σ^μ which correspond to generalizations of the Pauli spin matrices, and which at the same time have a clear physical meaning as the vectors of the reference tetrad. The situation should be compared with that in spinor calculus, where also one introduces tetrads $\sigma^\mu{}_{A\dot{B}}$, but where the physical significance is obscured by the abstractness of the spinor space to which the suffixes A, $\dot{B}$ refer. These differences between the three formalisms are characteristic; and also characteristic is the way in which the quaternion formalism combines the advantages of the other two. At least, so we try to demonstrate.

[5] C. Lanczos, Z. Physik **57**, 447, 474, 484 (1929).

[6] A. W. Conway, Proc. Roy. Soc. (London) **A 162**, 145 (1937).

[7] L. Infeld and B. L. van der Waerden, Sitzber. Preuss. Akad. Wiss., Physik.-Math. Kl., 380 (1933); W. L. Bade and H. Jehle, Revs. Modern Phys. **25**, 714 (1953); A. Peres, Nuovo Cimento Suppl. **24**, 389 (1962).

[8] P. G. Bergmann, Phys. Rev. **107**, 624 (1957), introduced a matrix version of spinor calculus which can also be regarded as a matrix representation of part of our quaternion theory.

[9] E. Schrödinger, Sitzber. Preuss. Akad. Wiss., Physik.-Math. Kl., 105 (1932), gives references to earlier work by Weyl, Fock, etc.

[10] V. Bargmann, Sitzber. Preuss. Akad. Wiss., Physik.-Math. Kl., 346 (1932).

[11] D. R. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).

[12] J. G. Fletcher, Nuovo Cimento **8**, 451 (1958).

I.2 Quaternion Algebra

In this section the necessary results of quaternion algebra are written in a notation familiar to physicists.[13,14]

Definitions

A quaternion $A=(A_1, A_2, A_3, A_4)$ is an ordered quadruple of complex numbers.[15] The A_α are the components of A (lower-case Greek indices from α through ι have the range 1, 2, 3, 4, and from κ through ω have the range 0, 1, 2, 3. Lower-case Latin indices have the range 1, 2, 3). Equality of two quaternions is equivalent to equality of their corresponding components: $A=B$ if and only if $A_\alpha=B_\alpha$. We shall sometimes put $\mathbf{A}=(A_1, A_2, A_3)$ and $A=(\mathbf{A}, A_4)$. It causes no confusion to write $A=\mathbf{A}$ when $A_4=0$; nor is it misleading to write the zero quaternion as $0=(0, 0, 0, 0)$.

The multiplication of a quaternion by a complex number λ, and the addition and subtraction of quaternions, are defined by

$$\begin{aligned}\lambda A&=(\lambda A_1, \lambda A_2, \lambda A_3, \lambda A_4),\\ A+B&=(A_1+B_1, A_2+B_2, A_3+B_3, A_4+B_4),\\ A-B&=A+(-1)B.\end{aligned} \tag{1}$$

Addition of quaternions is commutative and associative. Multiplication of quaternions by complex numbers is commutative, associative, and distributive. It follows that if one defines quaternion basis elements σ^α by $(\sigma^\alpha)_\beta=\delta_{\alpha\beta}$, so that

$$\begin{aligned}\sigma^1&=(1, 0, 0, 0), & \sigma^2&=(0, 1, 0, 0),\\ \sigma^3&=(0, 0, 1, 0), & \sigma^4&=(0, 0, 0, 1),\end{aligned} \tag{2}$$

then $A=A_\alpha\sigma^\alpha$ and $\mathbf{A}=A_m\sigma^m$ for any quaternion A (sum over repeated lower-case indices).

The product AB of quaternions A and B is itself a quaternion. The product is distributive, and for any complex number λ it satisfies

$$(\lambda A)B=A(\lambda B)=\lambda(AB). \tag{3}$$

The products of the basis elements σ^α are defined by

$$\sigma^p\sigma^q=-i\delta_{pq}\sigma^4+i\epsilon_{pqr}\sigma^r, \tag{4}$$

$$\sigma^4\sigma^\alpha=\sigma^\alpha\sigma^4=i\sigma^\alpha, \tag{5}$$

[13] L. Brand, *Vector and Tensor Analysis* (John Wiley & Sons, Inc., New York, 1947), gives a fuller, but still concise account of the subject in the traditional notation.

[14] F. Gürsey, Nuovo Cimento **3**, 988 (1956), uses a notation very similar to ours.

[15] Quaternions are often defined to be ordered quadruples of *real* numbers. Our quaternions are what Hamilton called biquaternions.

where ϵ_{pqr} is the permutation symbol. From Eq. (5) it follows that the quaternion σ^4 has the same properties as the complex number i. It is therefore permissible to write $\sigma^4=i$. With this change of notation, Eq. (4) becomes the well-known multiplication rule for the Pauli spin matrices. It was to achieve such a correspondence that we chose to use the σ^α rather than the traditional quaternion basis elements (which are $-i\sigma^\alpha$).

From these rules one finds that the product of two arbitrary quaternions is

$$AB=A_pB_p-A_4B_4+i(A_4B_r+A_rB_4+\epsilon_{pqr}A_pB_q)\sigma^r. \quad (6)$$

[Note that $(AB)_4=-i(A_pB_p-A_4B_4)$]. The product is not in general commutative, but is associative.

Algebraic Operations

The adjoint of a quaternion $A=(\mathbf{A}, A_4)$ is the quaternion $\bar{A}=(-\mathbf{A}, A_4)$. In particular, for the basis elements σ^α,

$$\bar{\sigma}^l=-\sigma^l, \qquad \bar{\sigma}^4=\sigma^4, \quad (7)$$

so that

$$\bar{A}=A_\alpha\bar{\sigma}^\alpha=\bar{A}_\alpha\sigma^\alpha. \quad (8)$$

From Eq. (6) we find that for any A, B

$$\bar{A}\bar{B}=\bar{B}\bar{A}, \quad (9)$$

which is to say that the operation of taking the adjoint is anti-automorphic.[16] From (4), (5), (7), one gets

$$2\bar{A}=\sigma^\alpha A\sigma^\alpha=\bar{\sigma}^\alpha A\bar{\sigma}^\alpha, \quad (10)$$

$$4A_\gamma=\sigma^\alpha A\bar{\sigma}^\gamma\bar{\sigma}^\alpha=\sigma^\alpha\bar{\sigma}^\gamma A\bar{\sigma}^\alpha$$
$$=\bar{\sigma}^\alpha A\bar{\sigma}^\gamma\sigma^\alpha=\bar{\sigma}^\alpha\bar{\sigma}^\gamma A\sigma^\alpha. \quad (11)$$

The norm of a quaternion A is defined to be $A\bar{A}$. The norm behaves as a complex number,

$$A\bar{A}=\bar{A}A=-A_\alpha A_\alpha. \quad (12)$$

It follows that $(AB)(\overline{AB})=(A\bar{A})(B\bar{B})$: the norm of a product is the product of the norms of the factors. If $A\bar{A}=0$, one says that A is null or singular. If $A\bar{A}=1$, then A is unimodular.

The reciprocal of any A which is not null is defined by $A^{-1}=\bar{A}/(A\bar{A})$. The reciprocal has the properties

$$AA^{-1}=A^{-1}A=1, \qquad (AB)^{-1}=B^{-1}A^{-1}. \quad (13)$$

The Hermitian conjugate of a quaternion A is the quaternion $A^\dagger=(A_1^*, A_2^*, A_3^*, -A_4^*)$, where the asterisk denotes complex conjugation. The operation of Hermitian conjugation is anti-automorphic,

$$(AB)^\dagger=B^\dagger A^\dagger. \quad (14)$$

A quaternion A is Hermitian if $A^\dagger=A$.

The complex reflection of a quaternion A is the quaternion $A^\times=(-A_1^*, -A_2^*, -A_3^*, -A_4^*)$. Complex reflection is an automorphic operation,

$$(AB)^\times=A^\times B^\times. \quad (15)$$

The effect of applying any two of the operations $^-$, †, $^\times$, to a quaternion is the same as that of applying the third. For example,

$$(A^\times)^\dagger=(A^\dagger)^\times=\bar{A}. \quad (16)$$

Note too that, for any complex number a,

$$(\overline{aA})=a\bar{A}, \qquad (aA)^{-1}=a^{-1}A^{-1}, \qquad (aA)^\dagger=a^*A^\dagger,$$
$$(aA)^\times=a^*A^\times. \quad (17)$$

(In the second equation we must require $a\neq0$, $A\bar{A}\neq0$).

The scalar product $A\cdot B$ of quaternions A and B is defined by

$$A\cdot B=-\tfrac{1}{2}(A\bar{B}+B\bar{A})=-\tfrac{1}{2}(\bar{A}B+\bar{B}A)=A_\alpha B_\alpha. \quad (18)$$

In particular one has $\mathbf{A}\cdot\mathbf{B}=A_kB_k$, $A\cdot A=-A\bar{A}=-\bar{A}A$, and the important commutation relations

$$\sigma^\alpha\bar{\sigma}^\beta+\sigma^\beta\bar{\sigma}^\alpha=\bar{\sigma}^\alpha\sigma^\beta+\bar{\sigma}^\beta\sigma^\alpha=-2\sigma^\alpha\cdot\sigma^\beta=-2\delta_{\alpha\beta}. \quad (19)$$

It has already been pointed out that the multiplication rule for the σ^p is the same as that for the Pauli spin matrices. By making the correspondences

$$\sigma^1\leftrightarrow\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad \sigma^2\leftrightarrow\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},$$
$$\sigma^3\leftrightarrow\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \qquad \sigma^4\leftrightarrow\begin{pmatrix}i & 0\\ 0 & i\end{pmatrix}, \quad (20)$$

so that for a general quaternion A

$$A=A_\alpha\sigma^\alpha\leftrightarrow\begin{pmatrix}A_3+iA_4, & A_1-iA_2\\ A_1+iA_2, & -A_3+iA_4\end{pmatrix}=\mathfrak{A}, \quad (21)$$

an isomorphism is established between the algebra of quaternions and that of 2×2 matrices. Writing the determinant of $\mathfrak{A}$ as $\det\mathfrak{A}$, the adjoint as $\bar{\mathfrak{A}}$, the inverse as $\mathfrak{A}^{-1}$, and the Hermitian conjugate as $\mathfrak{A}^\dagger$, it is easy to show that $A\bar{A}=\det\mathfrak{A}$, $\bar{A}\leftrightarrow\bar{\mathfrak{A}}$, $A^{-1}\leftrightarrow\mathfrak{A}^{-1}$, $A^\dagger\leftrightarrow\mathfrak{A}^\dagger$, and also that $2iA_4=A+\bar{A}=\mathrm{Sp}\,\mathfrak{A}$.

II. QUATERNIONS IN SPECIAL RELATIVITY

II.1 Lorentz Vectors

Events in flat space–time may be specified by coordinates x_α, where the x_k are real Cartesian spatial

[16] It is sometimes useful to define an operator K_m which changes the sign of only the m component of a quaternion: $K_1A=1(-A_1, A_2, A_3, A_4)$, etc. This operator is anti-automorphic, while the product K_mK_n is automorphic, and $K_1K_2K_3A=\bar{A}$. In the matrix representation which is introduced at the end of this section, the matrices that correspond to K_2A and $K_2A^\dagger$ are the transpose and complex conjugate, respectively, of the matrix that corresponds to A.

coordinates in an inertial frame of reference and $x_4 = ict$ is the imaginary time coordinate. We define the Hermitian space–time quaternion x by $x = x_\alpha \sigma^\alpha$. A homogeneous Lorentz transformation of x corresponds to a homogeneous linear transformation of the x_α which leaves x Hermitian and does not change its norm:

$$x'_\alpha = \Lambda_{\alpha\beta} x_\beta, \qquad x^\dagger = x, \qquad x'^\dagger = x', \qquad x'\bar{x}' = x\bar{x}, \qquad (22)$$

where the $\Lambda_{\alpha\beta}$ are complex numbers independent of x. It follows from (22) that Λ_{44} and the Λ_{kl} are real, that Λ_{4k} and Λ_{k4} are imaginary, and that $\Lambda_{\alpha\gamma}\Lambda_{\beta\gamma} = \Lambda_{\gamma\alpha}\Lambda_{\gamma\beta} = \delta_{\alpha\beta}$. Any 4×4 matrix Λ whose elements $\Lambda_{\alpha\beta}$ satisfy these conditions will be called a *Lorentz matrix*.

Any Hermitian quaternion which transforms in the same way as a spacetime quaternion under restricted Lorentz transformations is called a *Lorentz vector* (*restricted* here means *homogeneous*, *proper*, *and orthochronous*). Apart from the trivial cases $y' = \pm y$, no Lorentz transformation of a Lorentz vector y can be written in the form $y' = Ry$; and any Lorentz transformation of the form $y' = RyS$ can be written as $y' = \pm QyQ^\dagger$, where Q is unimodular ($Q\bar{Q} = 1$). Further, by taking only the plus sign in the last equation one can obtain all restricted Lorentz transformations.[4,17] Space inversion corresponds to taking the adjoint or the complex reflection.

If we assume that a unimodular Q exists such that, for all y,

$$y' = y'_\alpha \sigma^\alpha = \Lambda_{\alpha\beta} y_\beta \sigma^\alpha = QyQ^\dagger, \qquad (23)$$

then by comparing coefficients of y_β one has $\Lambda_{\alpha\beta}\sigma^\alpha = Q\sigma^\beta Q^\dagger$, and multiplying this equation by $\bar{\sigma}^\gamma\bar{\sigma}^\beta$ and using Eq. (11) gives

$$Q = C\Lambda_{\alpha\beta}\sigma^\alpha\bar{\sigma}^\gamma\bar{\sigma}^\beta, \qquad (24)$$

where $C = 1/4(Q^\dagger)_\gamma$. For C to be defined, one must choose γ so that $(Q^\dagger)_\gamma \neq 0$. That this is always possible follows from the assumption that Q is unimodular. The value of C^2 is found by substituting (24) in $Q\bar{Q} = 1$.[18]

Define $\partial_\alpha = \partial/\partial x_\alpha$, $\partial'_\alpha = \partial/\partial x'_\alpha$, where x is a space–time quaternion. The differential operator $\partial = \sigma^\alpha\partial_\alpha$ is then of particular importance. When x undergoes the restricted Lorentz transformation $x \to x' = QxQ^\dagger$, the operator transforms like a Lorentz vector:

$$\partial' = \sigma^\alpha \partial'_\alpha = Q\sigma^\alpha Q^\dagger \partial_\alpha. \qquad (25)$$

To prove (25), one needs the equation

$$(\bar{Q}\sigma^\alpha\bar{Q}^\dagger)_\beta = (Q\sigma^\beta Q^\dagger)_\alpha, \qquad (26)$$

which follows from (10) and (11).

If y and z are Lorentz vectors, the quaternions $y\bar{z}$ and $\bar{y}z$ transform under restricted Lorentz transformations according to the rules

$$y'\bar{z}' = Qy\bar{z}\bar{Q}, \qquad \bar{y}'z' = \bar{Q}^\dagger\bar{y}zQ^\dagger. \qquad (27)$$

It follows that the complex number $y\bar{z} + z\bar{y} = \bar{y}z + \bar{z}y = -2y_\alpha z_\alpha$ is a Lorentz invariant. In particular, $z_\alpha z_\alpha$ and $\partial_\alpha z_\alpha$ are Lorentz invariants, and $\partial_\alpha\partial_\alpha$ is a Lorentz invariant operator. If we write $w = y\bar{z}$, then what we have shown is that $w_4 = iy_\alpha z_\alpha$ is a Lorentz invariant. However, $w\bar{w} = -w_\alpha w_\alpha$ also is a Lorentz invariant, and so therefore is $w_p w_p$.

Maxwell's Equations

As an illustration of the formalism, consider the Lorentz vector a, the potential of the electromagnetic field. From Eqs. (6), (7),

$$\bar{\partial}a = \bar{\sigma}^\alpha\partial_\alpha a = -\partial_\alpha a_\alpha + i[\partial_4 a_r - \partial_r a_4 - \epsilon_{pqr}\partial_p a_q]\sigma^r$$
$$= (-\mathbf{E} - i\mathbf{B},\ i\partial_\alpha a_\alpha), \qquad (28)$$

where $\mathbf{E}$ is the electric field and $\mathbf{B}$ the magnetic induction (cf. Ref. 3, pp. 46, 200, 217). We impose the invariant Lorentz condition $\partial\cdot a = \partial_\alpha a_\alpha = 0$. The four Maxwell equations are then equivalent to

$$\partial(\mathbf{E} + i\mathbf{B}) = \partial_\alpha\partial_\alpha a = -j, \qquad (29)$$

where j is the current-density Lorentz vector.[19] Finally, identifying $E_p + iB_p$ with the w_p of the last paragraph, we see that $(E_p + iB_p)(E_p + iB_p)$ is a Lorentz invariant.

II.2 Spinors

Spinors will be defined to be quarternions with specified Lorentz transformation properties; but they also will be required to satisfy algebraic conditions (just as Lorentz vectors are required to satisfy the condition of Hermiticity). The conditions are that the quaternions should belong to certain ideals of the quaternion ring.[20]

Ideals of Quaternions

A quaternion E is said to be idempotent if it satisfies

$$E^2 = E. \qquad (30)$$

It follows from (30) that $\bar{E}$ is idempotent if and only if E is, and that the only idempotent quaternions, apart

---

[17] M. A. Naimark, *Les Représentations Linéaires du Groupe de Lorentz* (Dunod, Paris, 1962).

[18] We have assumed the existence of a unimodular Q; but it is also possible to prove by direct calculation that, for any restricted Lorentz transformation Λ, there exists a unimodular Q of the form (24) which satisfies (23). See A. J. Macfarlane, J. Math. Phys. **3**, 1116 (1962), who also gives references. Note however that Macfarlane's argument is incomplete, since he implicity assumes that $Q_4 \neq 0$.

[19] Other forms of Maxwell's equations and references to earlier work are given by M. Sachs and S. L. Schwebel, J. Math. Phys. **3**, 843 (1962).

[20] J. Blaton, Z. Physik **95**, 337 (1935).

from 0 and 1, are those which satisfy the equations

$$E_kE_k=\tfrac{1}{4}, \qquad E_4=-\tfrac{1}{2}i. \tag{31}$$

From (31), the idempotents E and $\bar{E}$ satisfy

$$\bar{E}E=E\bar{E}=0, \qquad E+\bar{E}=1. \tag{32}$$

The set of quaternions φ which satisfy

$$\varphi E=\varphi \tag{33}$$

is called the *left ideal generated by the idempotent E*. (Since $E+\bar{E}=1$, Eq. (33) is equivalent to $\varphi\bar{E}=0$). Similarly, the set of quaternions χ which satisfy

$$E\chi=\chi \tag{34}$$

is called the *right ideal generated by the idempotent E*. Any quaternion A may be uniquely written as the sum of a quaternion belonging to the left ideal generated by E and a quaternion belonging to the left ideal generated by $\bar{E}$, for by (32),

$$A=A(E+\bar{E})=AE+A\bar{E}, \tag{35}$$

and $(AE)E=AE^2=AE$, etc. Similarly one can write $A=EA+\bar{E}A$. The proofs of the uniqueness of the decompositions are trivial.

Definition of Spinors

A spinor is defined to be a quaternion φ which satisfies Eq. (33) for some idempotent E ($E\neq 0$ and $E\neq 1$), and which transforms according to the law

$$\varphi\rightarrow\varphi'=Q\varphi, \tag{36}$$

when the space–time quaternion x undergoes the restricted Lorentz transformation $x\rightarrow x'=QxQ^\dagger$. Since φ' satisfies $\varphi'E=\varphi'$, the condition (33) is covariant with respect to the restricted Lorentz transformations (36).

The freedom one has in the choice of E may be useful for describing the internal degrees of freedom of elementary particles. However, in this paper we simply choose $E=E_\pm=\frac{1}{2}(1\pm\sigma^3)$, so that $\bar{E}_\pm=\frac{1}{2}(1\mp\sigma^3)=E_\mp$, and (33) reduces to

$$\varphi\sigma^3=\pm\varphi; \tag{37}$$

or equivalently, in terms of components, to

$$\varphi_4=\mp i\varphi_3, \qquad \varphi_1=\pm i\varphi_2. \tag{38}$$

If a and b are complex numbers, and if φ satisfies (37) when the upper sign is taken, then $\varphi(a\sigma^1+b\sigma^2)$ satisfies it when the lower sign is taken.

The reason for choosing $E_\pm=\frac{1}{2}(1\pm\sigma^3)$ is that the matrix representation (21) of a φ satisfying (37) is particularly simple:

$$\varphi\leftrightarrow\begin{pmatrix}\varphi_3(1\pm1), & i\varphi_2(-1\pm1)\\ i\varphi_2(1\pm1), & \varphi_3(-1\pm1)\end{pmatrix}. \tag{39}$$

The first column vanishes if we choose the lower signs, and the second vanishes if we choose the upper. The connection between our treatment of spinors as quaternions and the usual treatment, in which they are two-dimensional vectors, is therefore established.[21]

Take the complex reflection of Eqs. (36), (37), and write $\varphi'^\times=\varphi^{\times\prime}$:

$$\varphi^\times\rightarrow\varphi^{\times\prime}=Q^\times\varphi^\times, \tag*{(36)$^\times$}$$

$$\varphi^\times\sigma^3=\mp\varphi^\times. \tag*{(37)$^\times$}$$

We call $\varphi^\times$ a *conjugate spinor*. Similarly, taking the adjoint or the Hermitian conjugate of (36), (37) gives

$$\bar{\varphi}\rightarrow\bar{\varphi}'=\bar{\varphi}\bar{Q}, \tag*{(36)$^-$}$$

$$\sigma^3\bar{\varphi}=\mp\bar{\varphi}; \tag*{(37)$^-$}$$

$$\varphi^\dagger\rightarrow\varphi^{\dagger\prime}=\varphi^\dagger Q^\dagger, \tag*{(36)†}$$

$$\sigma^3\varphi^\dagger=\pm\varphi^\dagger. \tag*{(37)†}$$

We call $\bar{\varphi}$ an *adjoint spinor*, and $\varphi^\dagger$ an *adjoint conjugate spinor*.

In conventional spinor algebra, a spinor with an undotted superfix corresponds to our φ, and one with an undotted suffix to our $\bar{\varphi}$. Spinors with dotted superfixes and suffixes correspond to $K_2\varphi^\dagger$ and $K_2\varphi^\times$, respectively, where $K_2A=A-2A_2\sigma^2$ for any quaternion A. It is usually possible, as here, to indicate the transformation law of a given spinor without using indices.

Fermion Field Equations

The Weyl equation in quaternion form is

$$\partial\varphi^\times=0, \tag{40}$$

where covariance under restricted Lorentz transformations requires that $\varphi^\times$ be a conjugate spinor (and therefore that φ be a spinor). Taking the complex reflection of (40) and using (7) gives the neutrino equation

$$\bar{\partial}\varphi=0. \tag{41}$$

The Dirac equation is equivalent to

$$\begin{aligned}(\partial-i\epsilon a)\varphi^\times&=i\kappa\theta,\\ (\bar{\partial}-i\epsilon\bar{a})\theta&=i\kappa\varphi^\times,\end{aligned} \tag{42}$$

where φ and θ are spinors, a is the electromagnetic potential, $\kappa=mc/\hbar$, $\epsilon=e/\hbar c$, and e and m are the charge and rest mass of the particle.

III. QUATERNIONS IN RIEMANNIAN SPACE–TIME

To understand the rest of this paper, one needs a small amount of Riemannian geometry.[22] Section III.1

[21] It is tempting to speculate that the two sorts of spinor, corresponding to the two choices of sign, may refer to different particles: say electrons and μ mesons, or two kinds of neutrino.

[22] See, for example, J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1952), Chaps. 1–3.

is introductory in character, and may be skimmed by the expert for its notation.

III.1 Tetrad Formalism

It makes for simplicity later on if one adopts, not the most usual formulation of Riemannian geometry, but the so-called tetrad formalism, in which an orthonormal tetrad of vectors is introduced at each event of space–time.[23,24] All that we do here is write down some definitions, establish the notation, and show the connection between the tetrad formalism and the usual one.

Tetrads and Metric Tensors

Events in space–time are denoted by upper-case Latin letters P, Q, etc. The value of a function f at the event P is denoted by f_P or $(f)_P$. An event may be labeled by four *real* coordinates x^μ (recall that μ has the range 0, 1, 2, 3).

At each event of space–time one introduces a tetrad of contravariant vectors with components $\sigma^\mu{}_\alpha$. Here α (range 1, 2, 3, 4) labels the vector, and μ labels its components. The $\sigma^\mu{}_k$ (range of k is 1, 2, 3) are real, while the $\sigma^\mu{}_4$ are imaginary. The determinant of the matrix whose elements are $\sigma^\mu{}_\alpha$ is assumed not to vanish. One can then find unique quantities $\sigma_{\mu\alpha}$ such that

$$\sigma_{\mu\alpha}\sigma^\mu{}_\beta=\delta_{\alpha\beta}. \tag{43}$$

The $\sigma_{\mu k}$ are real and the $\sigma_{\mu 4}$ are imaginary.

Under the coordinate transformation $x^\mu\rightarrow x^{\mu'}$, the $\sigma^\mu{}_\alpha$ transform, by definition, as contravariant vectors[25]

$$\sigma^{\mu'}{}_\alpha=(\partial x^{\mu'}/\partial x^\nu)\sigma^\nu{}_\alpha, \tag{44}$$

and it follows from (43) and (44) that the $\sigma_{\mu\alpha}$ transform as covariant vectors,

$$\sigma_{\mu'\alpha}=(\partial x^\nu/\partial x^{\mu'})\sigma_{\nu\alpha}. \tag{45}$$

The covariant and contravariant metric tensors are defined by

$$g_{\mu\nu}=\sigma_{\mu\alpha}\sigma_{\nu\alpha},\qquad g^{\mu\nu}=\sigma^\mu{}_\alpha\sigma^\nu{}_\alpha. \tag{46}$$

They are symmetric, and because of (43) they satisfy

$$g_{\mu\lambda}g^{\nu\lambda}=\delta_\mu{}^\nu, \tag{47}$$

$$g_{\mu\nu}\sigma^\nu{}_\alpha=\sigma_{\mu\alpha},\qquad g^{\mu\nu}\sigma_{\nu\alpha}=\sigma^\mu{}_\alpha. \tag{48}$$

Generalizing (48), we define the covariant vector f_μ corresponding to a given contravariant f^μ by $f_\mu=g_{\mu\nu}f^\nu$; and then from (47), $f^\mu=g^{\mu\nu}f_\nu$. This is just as in the usual formalism.

The metric tensor is uniquely determined by (46) once the $\sigma^\mu{}_\alpha$ are chosen. However, the converse is not true: choosing the $g^{\mu\nu}$ at each event does not determine the $\sigma^\mu{}_\alpha$ uniquely. One easily shows that the tetrads $\sigma^\mu{}_\alpha$ and $\sigma'^\mu{}_\alpha$ give rise to the same $g^{\mu\nu}$ if and only if

$$\sigma'^\mu{}_\alpha=\Lambda_{\alpha\beta}\sigma^\mu{}_\beta, \tag{49}$$

where Λ is a Lorentz matrix (cf. Section II.1), which may now be a function of the x^μ.[26]

Inertial Coordinates

We now introduce a class of special coordinate systems in a neighborhood of each event. This may seem an unnecessarily complicated procedure, but it will help us when we try to generalize the results of Part II.

Start from the theorem[22] that in a neighborhood of any event P there exist real coordinates $\mathfrak{X}^{(P)\mu}$ in terms of which the metric tensor at P has diagonal elements $(-1, 1, 1, 1)$, zero nondiagonal elements, and zero first derivatives. Make the trivial coordinate transformation $\mathfrak{X}^{(P)}{}_k=\mathfrak{X}^{(P)k}$, $\mathfrak{X}^{(P)}{}_4=i\mathfrak{X}^{(P)0}$, so that the transformed metric tensor $\mathcal{G}^{(P)}{}_{\alpha\beta}$ satisfies at P

$$(\mathcal{G}^{(P)}{}_{\alpha\beta})_P=\delta_{\alpha\beta},\qquad (\partial\mathcal{G}^{(P)}{}_{\alpha\beta}/\partial\mathfrak{X}^{(P)}{}_\gamma)_P=0. \tag{50}$$

A coordinate system in which Eqs. (50) are satisfied is said to be *inertial* (*at* P). We use this word because such coordinate systems correspond physically to locally inertial frames of reference. The point of introducing an imaginary $\mathfrak{X}^{(P)}{}_4$ is to avoid having to raise and lower the indices $\alpha\cdots\iota$.

Using (50) one finds

$$(g^{\mu\nu})_P=\left(\frac{\partial x^\mu}{\partial\mathfrak{X}^{(P)}{}_\alpha}\frac{\partial x^\nu}{\partial\mathfrak{X}^{(P)}{}_\beta}\mathcal{G}^{(P)}{}_{\alpha\beta}\right)_P=\left(\frac{\partial x^\mu}{\partial\mathfrak{X}^{(P)}{}_\alpha}\frac{\partial x^\nu}{\partial\mathfrak{X}^{(P)}{}_\alpha}\right)_P. \tag{51}$$

It follows from (49) that one may obtain the $(\partial x^\mu/\partial\mathfrak{X}^{(P)}{}_\alpha)_P$ from the $(\sigma^\mu{}_\alpha)_P$ by a Lorentz transformation. Thus, by performing a suitable inhomogeneous Lorentz transformation of the $\mathfrak{X}^{(P)}{}_\alpha$ we can find coordinates $x^{(P)}{}_\alpha$ with the following properties. (i) The $x^{(P)}{}_\alpha$ have their origin at P, and are inertial at P [Eqs. (50) hold]. (ii) The $x^{(P)}{}_k$ are real and $x^{(P)}{}_4$ is imaginary. (iii) The transformation coefficients $\partial x^\mu/\partial x^{(P)}{}_\alpha$ are the $\sigma^\mu{}_\alpha$ at P

$$(\partial x^\mu/\partial x^{(P)}{}_\alpha)_P=(\sigma^\mu{}_\alpha)_P. \tag{52}$$

Such $x^{(P)}{}_\alpha$ are said to *correspond to* the tetrad $\sigma^\mu{}_\alpha$. The $x^{(P)}{}_\alpha$ are not uniquely determined by these conditions; but if $x^{(P)}{}_\alpha$ and $\hat{x}^{(P)}{}_\alpha$ are two sets of coordinates inertial at P which correspond to the same $\sigma^\mu{}_\alpha$, then $\hat{x}^{(P)}{}_\alpha=x^{(P)}{}_\alpha+O(x^{(P)}{}_\beta x^{(P)}{}_\beta{}^*)$ as $x^{(P)}{}_\alpha x^{(P)}{}_\alpha{}^*\rightarrow 0$. The meaning of Eq. (52) is that the vector with components $\sigma^\mu{}_\alpha$ (α fixed) is tangential at P to the $x^{(P)}{}_\alpha$ coordinate curve.

---

[23] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1926), Chap. 3; F.A.E. Pirani, Acta Phys. Polon. **15**, 389 (1956); C. Møller, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Skr. **1**, No. 10 (1961). Also see A. E. Levashev and O. S. Ivanitskaya, Acta Phys. Polon. **23**, 647 (1963).

[24] For some purposes it is better to use tetrads of null vectors: see R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962); E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[25] We use here the notation of J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954).

[26] Be clear on the notation: the $\sigma^{\mu'}{}_\alpha$ are got from the $\sigma^\mu{}_\alpha$ by a coordinate transformation $x^\mu\rightarrow x^{\mu'}$ (they are two descriptions of the same tetrad), while the $\sigma'^\mu{}_\alpha$ are got from the $\sigma^\mu{}_\alpha$ by a Lorentz transformation (they are two different tetrads described in the same x^μ coordinate system). We do not write $\sigma^\mu{}_{\alpha'}$ for the Lorentz-transformed tetrad, because later the suffixes α are suppressed.

We have implicitly assumed, and always assume, that the Jacobian of the transformation from the x^μ to the $x^{(P)}{}_\alpha$ does not vanish in some neighborhood of P.[27] This, together with (43) and (52), implies

$$(\partial x^{(P)}{}_\alpha/\partial x^\mu)_P=(\sigma_{\mu\alpha})_P. \tag{53}$$

Local Lorentz Transformations

If $(\sigma^\mu{}_\alpha)_P$ and $(\sigma'^\mu{}_\alpha)_P$ are two tetrads at the event P which give rise to the same $(g^{\mu\nu})_P$, then they are related by a Lorentz transformation like (49)

$$(\sigma'^\mu{}_\alpha)_P=(\Lambda_{\alpha\beta})_P(\sigma^\mu{}_\beta)_P, \qquad (49)_P$$

where $(\Lambda_{\alpha\gamma})_P(\Lambda_{\beta\gamma})_P=\delta_{\alpha\beta}$. Suppose that $x^{(P)}{}_\alpha$ and $x'^{(P)}{}_\alpha$ are coordinate systems inertial at P which correspond, respectively, to $\sigma^\mu{}_\alpha$ and $\sigma'^\mu{}_\alpha$. Then from (52) and the assumption that the determinants of the $\sigma^\mu{}_\alpha$ and the $\sigma'^\mu{}_\alpha$ are nonzero, one shows that

$$x'^{(P)}{}_\alpha=(\Lambda_{\alpha\beta})_P x^{(P)}{}_\beta+O(x^{(P)}{}_\gamma x^{(P)}{}_\gamma{}^*)$$

as

$$x^{(P)}{}_\gamma x^{(P)}{}_\gamma{}^*\to 0. \tag{54}$$

We call transformations of the type (54) *local Lorentz transformations* (*in the neighborhood of* P). They are important in Sec. III.3, where we demand that physical laws should be covariant with respect to rotations (i.e. Lorentz transformations) of the reference tetrads.

Christoffel Symbols

We adopt the notation $\partial f/\partial x^\kappa=f_{,\kappa}$, and similarly for any suffix from κ through ω; and we put $\partial f/\partial x^{(P)}{}_\alpha=f_{,\alpha}$, and similarly for any suffix from α through ι. However, one must not write $f_{,\alpha}$ if it is unclear which set of inertial coordinates $x^{(P)}{}_\alpha$ one is dealing with. In derivatives of the form $\partial f^{(P)}/\partial x^\mu$ or $\partial f^{(P)}/\partial x^{(P)}{}_\alpha$ of quantities $f^{(P)}$ which depend on the choice of inertial coordinate system, we make the convention that the event P and the inertial coordinate system associated with it stay fixed throughout the differentiation.

The transformation law for the metric tensor can now be written

$$g_{\mu\nu}=x^{(P)}{}_{\alpha,\mu}x^{(P)}{}_{\beta,\nu}g^{(P)}{}_{\alpha\beta}. \tag{55}$$

Differentiate (55) with respect to x^λ at P, and use $(g^{(P)}{}_{\alpha\beta,\lambda})_P=(g^{(P)}{}_{\alpha\beta,\gamma}x^{(P)}{}_{\gamma,\lambda})_P=0$, which holds because the $x^{(P)}{}_\alpha$ at inertial at P:

$$(g_{\mu\nu,\lambda})_P=x^{(P)}{}_{\alpha,\mu\lambda}x^{(P)}{}_{\alpha,\nu}+x^{(P)}{}_{\alpha,\mu}x^{(P)}{}_{\alpha,\nu\lambda}. \tag{56}$$

The Christoffel symbols of the first kind at P are therefore

$$[\mu\nu,\lambda]_P=\tfrac{1}{2}(g_{\mu\lambda,\nu}+g_{\nu\lambda,\mu}-g_{\mu\nu,\lambda})=(x^{(P)}{}_{\alpha,\mu\nu}x^{(P)}{}_{\alpha,\lambda})_P. \tag{57}$$

Multiply by $g^{\pi\lambda}$ and use (43), (46), (52), (53) to find the Christoffel symbols of the second kind:

$$\begin{Bmatrix}\pi\\ \mu\nu\end{Bmatrix}_P=(g^{\pi\lambda}[\mu\nu,\lambda])_P=(x^{(P)}{}_{\alpha,\mu\nu}x^\pi{}_{,\alpha})_P. \tag{58}$$

Since $x^{(P)}{}_{\alpha,\mu}x^\pi{}_{,\alpha}=\delta_\mu{}^\pi$, one has $x^{(P)}{}_{\alpha,\mu\nu}x^\pi{}_{,\alpha}+x^{(P)}{}_{\alpha,\mu}x^\pi{}_{,\alpha\nu}=0$, and (58) becomes

$$\begin{Bmatrix}\pi\\ \mu\nu\end{Bmatrix}_P=-(x^{(P)}{}_{\alpha,\mu}x^\pi{}_{,\alpha\nu})_P. \qquad (58)'$$

Absolute and Covariant Derivatives

Let $f_{\mu\cdots}{}^{\lambda\cdots}$ be a tensor with respect to transformations of the x^μ. Its components in the $x^{(P)}{}_\alpha$ coordinate system are

$$f^{(P)}{}_{\alpha\cdots\gamma\cdots}=x^\mu{}_{,\alpha}\cdots x^{(P)}{}_{\gamma,\lambda}\cdots f_{\mu\cdots}{}^{\lambda\cdots}, \tag{59}$$

where $x^\mu{}_{,\alpha}=\partial x^\mu/\partial x^{(P)}{}_\alpha$. If $f_{\mu\cdots}{}^{\lambda\cdots}$ is defined along a curve $\mathfrak{C}$ and is differentiable with respect to an invariant parameter v on $\mathfrak{C}$, then the absolute derivative of $f_{\mu\cdots}{}^{\lambda\cdots}$ with respect to v at P is defined to be the tensor $g_{\mu\cdots}{}^{\lambda\cdots}$, of the same type as $f_{\mu\cdots}{}^{\lambda\cdots}$, whose components at the event P in the $x^{(P)}{}_\alpha$ coordinate system are

$$(g^{(P)}{}_{\alpha\cdots\gamma\cdots})_P=[(d/dv)f^{(P)}{}_{\alpha\cdots\gamma\cdots}]_P. \tag{60}$$

Transforming (60) to the x^μ coordinate system and introducing the notation $g_{\mu\cdots}{}^{\lambda\cdots}=(\delta/\delta v)f_{\mu\cdots}{}^{\lambda\cdots}$, one finds by (52), (53), that

$$[(\delta/\delta v)f_{\pi\cdots}{}^{\sigma\cdots}]_P$$
$$=(\sigma_{\pi\alpha}\cdots\sigma^\sigma{}_\gamma\cdots(d/dv)\,(x^\mu{}_{,\alpha}\cdots x^{(P)}{}_{\gamma,\lambda}\cdots f_{\mu\cdots}{}^{\lambda\cdots}))_P. \tag{61}$$

(In such equations it often does no harm to omit the suffix P). If one expands the derivative in (61), writes $d(x^\mu{}_{,\alpha})/dv=x^\mu{}_{,\alpha\tau}dx^\tau/dv$, etc., and uses (43) and (58), one gets the usual expression for the absolute derivative in terms of the Christoffel symbols.

If $f_{\mu\cdots}{}^{\lambda\cdots}$ is differentiable throughout a region which contains P, then the covariant derivative of $f_{\mu\cdots}{}^{\lambda\cdots}$ at P is defined to be the tensor $h_{\mu\cdots}{}^{\lambda\cdots}{}_\rho$ whose contravariant degree is the same as that of $f_{\mu\cdots}{}^{\lambda\cdots}$, whose covariant degree is one higher than that of $f_{\mu\cdots}{}^{\lambda\cdots}$, and whose components at the event P in the $x^{(P)}{}_\alpha$ coordinate system are

$$(h^{(P)}{}_{\alpha\cdots\gamma\cdots\delta})_P=[(\partial/\partial x^{(P)}{}_\delta)f^{(P)}{}_{\alpha\cdots\gamma\cdots}]_P. \tag{62}$$

Transforming (62) to the x^μ coordinate system, and introducing the notation $h_{\mu\cdots}{}^{\lambda\cdots}{}_\rho=f_{\mu\cdots}{}^{\lambda\cdots}{}_{|\rho}$, one finds

$$f_{\pi\cdots}{}^{\sigma\cdots}{}_{|\tau}=\sigma_{\pi\alpha}\cdots\sigma^\sigma{}_\gamma\cdots(\partial/\partial x^\tau)\,(x^\mu{}_{,\alpha}\cdots x^{(P)}{}_{\gamma,\lambda}\cdots f_{\mu\cdots}{}^{\lambda\cdots}). \tag{63}$$

Again the usual expression in terms of Christoffel symbols is got by expanding the derivative, etc.

The preceding examples of the tetrad formalism should be sufficient to establish the notation and to familiarize strangers with its use. The only other remark we make is that the Ricci identity

$$f_{\pi|\mu\nu}-f_{\pi|\nu\mu}=R^\lambda{}_{\pi\mu\nu}f_\lambda, \tag{64}$$

[27] Let it be understood that throughout this paper we exclude singular cases. All necessary derivatives are assumed to exist, and even "arbitrary" transformations are as smooth as we please.

which holds for any differentiable vector field f_π, and which may be used to define the mixed curvature tensor $R^\lambda{}_{\pi\mu\nu}$, gives when $f_\pi = \sigma_{\pi\alpha}$ the equation

$$R^\lambda{}_{\pi\mu\sigma} = \sigma^\lambda{}_\alpha (\sigma_{\pi\alpha|\mu\sigma} - \sigma_{\pi\alpha|\sigma\mu}). \tag{65}$$

It follows, using $R_{\lambda\pi\mu\sigma} = -R_{\pi\lambda\mu\sigma}$, that the Einstein field equations for empty space, $R_{\pi\mu} = R^\lambda{}_{\pi\mu\lambda} = 0$, are equivalent to

$$\sigma^\lambda{}_{\alpha|[\nu\lambda]} = 0, \tag{66}$$

where $f_{[\alpha\beta]} = \frac{1}{2}(f_{\alpha\beta} - f_{\beta\alpha})$.

III.2 The Quaternions σ^μ and σ_μ

The quaternions σ^μ and σ_μ are defined by

$$\sigma^\mu = \sigma^\mu{}_\alpha \sigma^\alpha, \qquad \sigma_\mu = \sigma_{\mu\alpha}\sigma^\alpha. \tag{67}$$

From (48) and (67)

$$\sigma_\mu = g_{\mu\nu}\sigma^\nu, \qquad \sigma^\mu = g^{\mu\nu}\sigma_\nu. \tag{68}$$

The $\sigma^\mu{}_\alpha$ and $\sigma_{\mu\alpha}$ are the α components of σ^μ and σ_μ, respectively, which agrees with the notation of Section I.2. Our previous device of writing σ^α with a superfix, so that one is not misled into thinking that the α denotes a quaternion component, is not needed here: only suffixes from α through ι (range 1, 2, 3, 4) are used for quaternion components, so σ_μ cannot be the μ quaternion-component of some quaternion σ. If ambiguity does threaten, one can always put brackets round a suffix to indicate that it does not denote a quaternion component: for example,

$$\sigma^{(P)}{}_{(\alpha)} = (\partial x^{(P)}{}_\alpha / \partial x^\mu)\sigma^\mu$$

is the reference tetrad expressed in the $x^{(P)}{}_\alpha$ coordinate system.

The reality conditions imposed on the $\sigma^\mu{}_\alpha$ and $\sigma_{\mu\alpha}$ are $\sigma^\mu{}_k{}^* = \sigma^\mu{}_k$, $\sigma^\mu{}_4{}^* = -\sigma^\mu{}_4$, $\sigma_{\mu k}{}^* = \sigma_{\mu k}$, $\sigma_{\mu 4}{}^* = -\sigma_{\mu 4}$. It follows from the definitions of Section I.2 that

$$\sigma^{\mu\dagger} = \sigma^\mu, \qquad \sigma_\mu{}^\dagger = \sigma_\mu,$$
$$\bar\sigma^\mu = \sigma^{\mu\times}, \qquad \bar\sigma_\mu = \sigma_\mu{}^\times. \tag{69}$$

The operations of raising and lowering indices commute with the operations $^-$, †, $^\times$: for example, $g_{\mu\nu}\bar\sigma^\nu = \bar\sigma_\mu$.

The scalar product of σ^μ and σ^ν is, by (18) and (46),

$$\sigma^\mu\cdot\sigma^\nu = -\tfrac{1}{2}(\sigma^\mu\bar\sigma^\nu + \sigma^\nu\bar\sigma^\mu) = -\tfrac{1}{2}(\bar\sigma^\mu\sigma^\nu + \bar\sigma^\nu\sigma^\mu)$$
$$= \sigma^\mu{}_\alpha\sigma^\nu{}_\alpha = g^{\mu\nu}. \tag{70}$$

Lowering one superfix in (70) gives $\sigma^\mu\cdot\sigma_\nu = g^\mu{}_\nu = \delta^\mu{}_\nu$; lowering both gives $\sigma_\mu\cdot\sigma_\nu = g_{\mu\nu}$.

The norm of σ^μ is

$$\sigma^\mu\bar\sigma^\mu = -\sigma^\mu{}_\alpha\sigma^\mu{}_\alpha = -g^{\mu\mu} \quad \text{(no summation on } \mu\text{)}; \tag{71}$$

and its reciprocal is

$$(\sigma^\mu)^{-1} = -\frac{\sigma^\mu}{g^{\mu\mu}} \quad \text{for} \quad g^{\mu\mu}M0 \quad \text{(no summation on } \mu\text{)}. \tag{72}$$

Similar results hold for σ_μ.

From (10), (11), (43), one shows that, for any quaternion A,

$$\sigma_\mu A\sigma^\mu = \bar\sigma_\mu A\bar\sigma^\mu = 2\bar A, \tag{73}$$

$$\sigma_\mu A\bar\sigma^\gamma\bar\sigma^\mu = \sigma_\mu\bar\sigma^\gamma A\bar\sigma^\mu = \bar\sigma_\mu A\bar\sigma^\gamma\sigma^\mu = \bar\sigma_\mu\bar\sigma^\gamma A\sigma^\mu = 4A_\gamma. \tag{74}$$

Eqs. (73), (74) remain valid if one everywhere raises the suffixes μ and lowers the superfixes μ.

Coordinate Transformations and the Operators ∂, $\bar\partial$

The σ^μ and $\bar\sigma^\mu$ are contravariant vectors.[28] This follows from (44) and from the fact that the σ^α and $\bar\sigma^\alpha$ are constants, unaffected by transformations of the x^μ. Similarly, the σ_μ and $\bar\sigma_\mu$ are covariant vectors. Transforming the σ^μ and σ_μ to an inertial coordinate system $x^{(P)}{}_\alpha$, one has

$$(\sigma^{(P)}{}_{(\alpha)})_P = (\sigma_{\mu\alpha}\sigma^\mu)_P = (\sigma^\mu{}_\alpha\sigma_\mu)_P = \sigma^\alpha. \tag{75}$$

It follows that if a quaternion $f_P = (f_\alpha)_P\sigma^\alpha$ is given at the origin P of an inertial coordinate system $x^{(P)}{}_\alpha$, then by defining a vector $(f_\mu)_P = (\sigma_{\mu\alpha}f_\alpha)_P$, one can write f_P in the invariant form $f_P = (f_\mu\sigma^\mu)_P$.

The operators $\partial = \sigma^\mu\partial_\mu$ and $\bar\partial = \bar\sigma^\mu\partial_\mu$ are invariants: $\sigma^\mu\partial_\mu = \sigma^{\mu'}\partial_{\mu'}$ and $\bar\sigma^\mu\partial_\mu = \bar\sigma^{\mu'}\partial_{\mu'}$ for any coordinate systems x^μ and $x^{\mu'}$. If a is an invariant quaternion, then ∂a and $\bar\partial a$ are also invariant quaternions, and $\partial\cdot a$ and $\bar\partial\cdot a$ are invariant complex numbers. If the coordinates x_α are inertial at the event P, then it follows from (75) that

$$\partial_P = (\sigma^\mu\partial_\mu)_P = \sigma^\alpha(\partial_\alpha)_P, \tag{76}$$

where $\partial_\alpha = \partial/\partial x_\alpha$. In a flat space–time one can find coordinates x_α which are inertial at every event. In such a coordinate system Eq. (76) is valid everywhere, so that $\partial = \sigma^\alpha\partial_\alpha$ in agreement with the definition of Section II.1.

Lorentz Transformation of σ_μ

It was shown in Sec. III.1 that a Lorentz transformation of the reference tetrad $(\sigma_{\mu\alpha})_P$ entails a local Lorentz transformation (54) of the coordinates inertial at P which correspond to $\sigma_{\mu\alpha}$. Omitting the suffixes P, we write the local Lorentz transformation corresponding to $\sigma'_{\mu\alpha} = \Lambda_{\alpha\beta}\sigma_{\mu\beta}$ as $x'^{(P)}{}_\alpha = \Lambda_{\alpha\beta}x^{(P)}{}_\beta + O_2$, where O_2 denotes terms of the second order in the $x^{(P)}{}_\alpha$; and provided that Λ describes a restricted Lorentz transformation, these equations are equivalent [cf. Eq. (23)] to

$$\sigma'_\mu = Q\sigma_\mu Q^\dagger, \tag{77}$$

$$x'^{(P)} = Qx^{(P)}Q^\dagger + \mathcal{O}_2, \tag{78}$$

for some unimodular Q. We have here written $\sigma'_\mu = \sigma'_{\mu\alpha}\sigma^\alpha$, $x^{(P)} = x^{(P)}{}_\alpha\sigma^\alpha$, $x'^{(P)} = x'^{(P)}{}_\alpha\sigma^\alpha$, and $\mathcal{O}_2$ is a quater-

[28] We always use the words *invariant, vector, tensor*, to refer to quantities which behave in the appropriate manner with respect to general transformations of the x^μ. When we are concerned with Lorentz transformations of the reference tetrads, we use the phrases *Lorentz invariant, Lorentz vector, Lorentz tensor*, as in Part II.

nion whose components are of second order in the $x^{(P)}{}_\alpha$. Note that both Λ and Q are invariants. One can multiply (77) by $g^{\mu\nu}$ and use (48) to get

$$\sigma'^\mu = Q\sigma^\mu Q^\dagger, \tag{79}$$

where $\sigma'^\mu = \sigma'^\mu{}_\alpha \sigma^\alpha$.

The transformation of the operator $\partial = \sigma^\mu \partial_\mu$ which corresponds to the Lorentz transformation (79) is

$$\partial' = \sigma'^\mu \partial_\mu = Q\sigma^\mu Q^\dagger \partial_\mu. \tag{80}$$

If one uses (76) to write $\partial' = \sigma^\alpha(\partial'_\alpha)_P$, and (52) to write $Q\sigma^\mu Q^\dagger \partial_\mu = Q\sigma^\alpha Q^\dagger(\partial_\alpha)_P$, then (80) becomes identical, apart from the suffixes P, with Eq. (25).

When Q differs little from 1, we may write $Q = 1 + \delta q$ and neglect terms of the order of $(\delta q_\alpha)^2$. The unimodularity condition $Q\bar{Q} = 1$ is then equivalent to $\delta q + \overline{\delta q} = 0$ or to $\delta q_4 = 0$, and the transformation law (77) reduces to

$$\delta\sigma_\mu = \sigma'_\mu - \sigma_\mu = \delta q \sigma_\mu + \sigma_\mu (\delta q)^\dagger + O(\delta q_\alpha \delta q_\alpha^*)$$

as

$$\delta q_\alpha \delta q_\alpha^* \to 0. \tag{81}$$

We note that none of the Eqs. (77) through (81) is altered if, instead of requiring that Q be unimodular, we impose the weaker condition[29]

$$Q\bar{Q} = \bar{Q}Q = e^{2i\beta}, \tag{82}$$

where β is a real invariant. If β is of the same order as the δq_α, then to first order (82) is equivalent to $\delta q + \overline{\delta q} = 2i\beta$, or to $\delta q_4 = \beta$.

Skew-symmetric Tensors

The representation of real skew-symmetric tensors by quaternions which is given in this subsection will be needed when we calculate derivatives of the σ_μ.

Let $f_{\mu\nu}$ be a real skew-symmetric tensor. Define the invariant quaternion F by

$$F = f_{\mu\nu}\sigma^\mu \bar{\sigma}^\nu. \tag{83}$$

Because $f_{\mu\nu} = -f_{\nu\mu}$, we have $\bar{F} = -F$, or $F_4 = 0$. From Eqs. (69), (70) it follows that

$$4f_{\mu\nu}\sigma^\nu = F\sigma_\mu + \sigma_\mu F^\dagger, \tag{84}$$

$$8f_{\mu\nu} = \sigma_\nu(\bar{\sigma}_\mu F + F^\dagger \bar{\sigma}_\mu) - (F\sigma_\mu + \sigma_\mu F^\dagger)\bar{\sigma}_\nu, \tag{85}$$

and hence $f_{\mu\nu}$ is uniquely determined by F [that (85) is skew-symmetric can be seen by writing $\sigma_\nu \bar{\sigma}_\mu = \frac{1}{2}(\sigma_\nu\bar{\sigma}_\mu - \sigma_\mu\bar{\sigma}_\nu - 2g_{\mu\nu})$, etc.].

Another possibility is to define an invariant quaternion H by

$$H = -F^\dagger = f_{\mu\nu}\bar{\sigma}^\mu\sigma^\nu, \tag{86}$$

and put $F = -H^\dagger$ in (84) and (85).

For any given $f_{\mu\nu}$ we have found a quaternion F which satisfies (84). The most general F which satisfies (84) can always be written

$$F = f_{\mu\nu}\sigma^\mu\bar{\sigma}^\nu + i\alpha \tag{87}$$

for some quaternion α, and by substituting (87) in (84) we show that F is a solution if and only if α is a real number. When $\alpha \neq 0$, one naturally has $F \neq -\bar{F}$.

Derivatives of σ_μ

The definitions of absolute and covariant derivatives, which were given in the last section, apply unchanged to tensors which are also quaternions. Each quaternion-component of such a tensor is, of course, itself a tensor; and the requirements that a tensor should be differentiable along a curve or throughout a region must now be interpreted as meaning that these quaternion-components are so differentiable. Applying the definitions, we see that the covariant derivatives of σ^μ and σ_μ are

$$\sigma^\mu{}_{|\nu} = \sigma^\mu{}_{\alpha|\nu}\sigma^\alpha, \qquad \sigma_{\mu|\nu} = \sigma_{\mu\alpha|\nu}\sigma^\alpha, \tag{88}$$

where

$$\sigma^\mu{}_{\alpha|\nu} = \sigma^\mu{}_{\alpha,\nu} + \begin{Bmatrix} \mu \\ \nu\lambda \end{Bmatrix} \sigma^\lambda{}_\alpha, \qquad \sigma_{\mu\alpha|\nu} = \sigma_{\mu\alpha,\nu} - \begin{Bmatrix} \lambda \\ \mu\nu \end{Bmatrix} \sigma_{\lambda\alpha}.$$

It makes for ease of interpretation if, instead of using Christoffel symbols, one expresses the covariant derivatives of σ_μ in terms of rotation coefficients (see Eisenhart[23] p. 97, but notice that his rotation coefficients are all real). These are the invariants $\gamma_{\alpha\beta\gamma}$ defined by

$$\gamma_{\alpha\beta\gamma} = \sigma_{\mu\alpha|\nu}\sigma^\mu{}_\beta\sigma^\nu{}_\gamma. \tag{89}$$

Because of (43) they satisfy $\gamma_{\alpha\beta\gamma} = -\gamma_{\beta\alpha\gamma}$. Define the tensors $\gamma_{\pi\mu\nu}$ by

$$\gamma_{\pi\mu\nu} = \gamma_{\alpha\beta\gamma}\sigma_{\pi\alpha}\sigma_{\mu\beta}\sigma_{\nu\gamma}. \tag{90}$$

They are real, satisfy $\gamma_{\pi\mu\nu} = -\gamma_{\mu\pi\nu}$, and at P in a coordinate system inertial at P one has $(\gamma^{(P)}{}_{\alpha\beta\gamma})_P = \gamma_{\alpha\beta\gamma}$. Multiply (89) by $\sigma_{\pi\beta}\sigma_{\rho\gamma}$ and use (43), (88), (90):

$$\sigma_{\mu|\pi} = \gamma_{\lambda\mu\pi}\sigma^\lambda. \tag{91}$$

Define a quaternion Γ_π by

$$\Gamma_\pi = \tfrac{1}{4}\sigma_{\nu|\pi}\bar{\sigma}^\nu + i\alpha_\pi, \tag{92}$$

where the α_π are real and are chosen to transform as the components of a covariant vector (so that Γ_π is a covariant vector). From (91) and (92),

$$\Gamma_\pi = \tfrac{1}{4}\gamma_{\mu\nu\pi}\sigma^\mu\bar{\sigma}^\nu + i\alpha_\pi. \tag{93}$$

Since the $\gamma_{\mu\nu\pi}$ are real and skew-symmetric in μ and ν, the Γ_π satisfy all the conditions required of the F of Eq. (87), and therefore by (84),

$$\gamma_{\mu\nu\pi}\sigma^\nu = \Gamma_\pi\sigma_\mu + \sigma_\mu\Gamma_\pi{}^\dagger. \tag{94}$$

Further, Γ_π is the most general covariant vector solution of (94). Finally, substitute (94) in (91):

$$\sigma_{\mu|\pi} = -\Gamma_\pi\sigma_\mu - \sigma_\mu\Gamma_\pi{}^\dagger. \tag{95}$$

[29] H. Weyl, Z. Physik **56**, 330 (1929).

The absolute derivative of σ_μ with respect to a parameter v on a curve $\mathfrak{C}$ is, from (95),

$$\delta\sigma_\mu/\delta v=\sigma_{\mu|\pi}(dx^\pi/dv)=-\Gamma_v\sigma_\mu-\sigma_\mu\Gamma_v{}^\dagger, \qquad (96)$$

where $\Gamma_v=\Gamma_\pi dx^\pi/dv$ is an invariant. Multiply (96) by dv, and put $d(\sigma_\mu)=(\delta\sigma_\mu/\delta v)\,dv$, $dq=-\Gamma_v\,dv$:

$$d(\sigma_\mu)=(dq)\sigma_\mu+\sigma_\mu(dq)^\dagger. \qquad (97)$$

Equation (97) has been written to look like the differential version of (81), but it has a different interpretation. In a coordinate system which is inertial at the event $P=P(v)$ of $\mathfrak{C}$, the vector $d(\sigma_\mu)$ at P is $(\delta\sigma^{(P)}{}_{(\alpha)}/\delta v)\,dv=(d\sigma^{(P)}{}_{(\alpha)}/dv)\,dv=d\sigma^{(P)}{}_{(\alpha)}=\sigma^{(P)}{}_{(\alpha)}(v+dv)-\sigma^{(P)}{}_{(\alpha)}(v)$. In this way the vector $d(\sigma_\mu)$ determines the change in σ_μ corresponding to the displacement dv along $\mathfrak{C}$ from P. On the other hand, the vector $\delta\sigma_\mu$ in Eq. (81) is the change in σ_μ at P corresponding to the Lorentz transformation $Q=1+\delta q$. We conclude that *in a coordinate system inertial at P* the change in the reference tetrad produced by the displacement dv along $\mathfrak{C}$ from P is the same as that produced by the Lorentz transformation $1-\Gamma_v\,dv$ at P.

Lorentz Transformation of Γ_μ

Apart from the usual assumption that all necessary derivatives exist, the choice of reference tetrads in the last subsection was quite arbitrary. Suppose that σ_μ and σ'_μ are two sets of tetrads defined on the curve $\mathfrak{C}$ and related by the v-dependent Lorentz transformation $\sigma'_\mu=Q\sigma_\mu Q^\dagger$. In a coordinate system inertial at $P=P(v)$, the tetrads are $\sigma^{(P)}{}_{(\alpha)}=(\partial x^\mu/\partial x^{(P)}{}_\alpha)\sigma_\mu$ and $\sigma'^{(P)}{}_{(\alpha)}=(\partial x^\mu/\partial x^{(P)}{}_\alpha)\sigma'_\mu$. The displacement dv along $\mathfrak{C}$ from P, which corresponds to a Lorentz transformation $1-\Gamma_v\,dv$ of the $\sigma^{(P)}{}_{(\alpha)}$, corresponds similarly to a Lorentz transformation $1-\Gamma'_v\,dv$ of the $\sigma'^{(P)}{}_{(\alpha)}$. This latter transformation can be accomplished in three steps: from $\sigma'^{(P)}{}_{(\alpha)}(v)$ to $\sigma^{(P)}{}_{(\alpha)}(v)$, from $\sigma^{(P)}{}_{(\alpha)}(v)$ to $\sigma^{(P)}{}_{(\alpha)}(v+dv)$, and from $\sigma^{(P)}{}_{(\alpha)}(v+dv)$ to $\sigma'^{(P)}{}_{(\alpha)}(v+dv)$. The corresponding Lorentz transformations are $Q^{-1}(v)$, $1-\Gamma_v\,dv$, $Q(v+dv)$, so that $1-\Gamma'_v\,dv=Q(v+dv)(1-\Gamma_v\,dv)Q^{-1}(v)$, and hence

$$\Gamma'_v=Q\Gamma_vQ^{-1}-(dQ/dv)Q^{-1}. \qquad (98)$$

In (98) everything is evaluated at $P=P(v)$. One can, of course, add to (98) a term $i\gamma(v)$, where $\gamma(v)$ is real [cf. (93)].

By taking the curve $\mathfrak{C}$ to be arbitrary, we can deduce from (98) the relation between Γ_π and Γ'_π. Alternatively, use the defining equation analogous to (92):

$$\Gamma'_\pi=\tfrac{1}{4}\sigma'_{\nu|\pi}\bar{\sigma}'^\nu+i\alpha'_\pi. \qquad (99)$$

It follows from (62) that for any quaternions A and B one has $(AB)_{|\pi}=A_{|\pi}B+AB_{|\pi}$, and that the covariant derivative of an invariant quaternion is the same as the partial derivative. Since Q is an invariant,

$$\sigma'_{\nu|\pi}=(Q\sigma_\nu Q^\dagger)_{|\pi}=Q\sigma_{\nu|\pi}Q^\dagger+Q_{,\pi}\sigma_\nu Q^\dagger+Q\sigma_\nu Q^\dagger{}_{,\pi}. \qquad (100)$$

We impose on Q the condition (82), which is equivalent to $Q^{-1}=\bar{Q}e^{-2i\beta}$ and which implies $(Q^\dagger{}_{,\pi}\bar{Q}^\dagger)_4=-\beta_{,\pi}e^{-2i\beta}$. A short calculation using (74), (92), (99), and (100) then gives

$$\Gamma'_\pi=Q\Gamma_\pi Q^{-1}-Q_{,\pi}Q^{-1}+i(\alpha'_\pi-\alpha_\pi+\beta_{,\pi}). \qquad (101)$$

We define α'_π so that the last term in (101) vanishes:

$$\alpha'_\pi=\alpha_\pi-\beta_{,\pi}. \qquad (102)$$

Because α_π is by assumption a real covariant vector and β is a real invariant, we see that α'_π is a real covariant vector. If we write $\epsilon=e/\hbar c$ and identify $a_\pi=-\alpha_\pi/\epsilon$ as the electromagnetic potential, then (102) represents a gauge transformation.[29]

At any given event P we can choose $Q_P=1$, $(Q_{,\pi})_P=(\Gamma_\pi)_P$. Because Q is an invariant, these equations are covariant with respect to transformations of the x^μ. Then from (101) and (102) we have $(\Gamma'_\pi)_P=0$, and from (95), $(\sigma'_{\mu|\pi})_P=0$. At P in a coordinate system inertial at P the covariant derivative is the partial derivative, and the last equation becomes $(\sigma'^{(P)}{}_{(\alpha),\beta})_P=0$. The physical meaning of the transformation is therefore that it makes the vectors of the tetrad at an event near P parallel to the corresponding vectors of the tetrad at P. We say that after the transformation the tetrads are *locally aligned* with respect to the tetrad at P. This is of importance in the next section when we define aligned absolute and aligned covariant derivatives.

III.3 Physical Laws in Riemannian Space–Time

There are two kinds of freedom within the formalism that we have been developing:

(i) The coordinates x^μ may be transformed in an arbitrary manner.

(ii) The reference tetrads at every event may be subjected to independent Lorentz transformations.[30]

We impose the traditional requirement that any equation which correctly expresses a physical law must be covariant with respect to both sorts of transformation. This is quite restrictive, and consequently a sensible method for guessing the laws of curved Riemannian space–time is to try to write known laws of flat space–time in a form covariant with respect to the transformations (i) and (ii), and then to assume that these covariant laws are valid in general. We shall now examine this procedure more closely.

In flat space–time one can choose coordinates x_α which are inertial at every event [that is, by (50), the metric tensor is everywhere $g_{\alpha\beta}=\delta_{\alpha\beta}$]. When working in such a coordinate system one does not normally bother with reference tetrads; but at each event we shall introduce a tetrad whose vectors point in the x_α coordinate directions. From the last subsec-

[30] That is, in accordance with out general policy, independent but smoothly varying from event to event.

tion we see that these tetrads are everywhere locally aligned with respect to one another.

We suppose that a physical law of flat space–time is expressed in the x_α coordinate system by the equation $f=0$. To write this equation in a form which is covariant with respect to transformations of type (i) is commonly not hard: one can use the mechanism of tensor calculus. What we must think about is how to ensure its covariance with respect to the general tetrad transformations of type (ii).

The original equation $f=0$ must be covariant with respect to Lorentz transformations of the x_α,—otherwise it would not be acceptable as the expression of a physical law. Since we have identified the x_α coordinate directions with the directions of the vectors of the reference tetrads, we can say that $f=0$ is covariant with respect to those special transformations of type (ii) in which the reference tetrads undergo the same rotation at every event. We cannot however expect it to be covariant with respect to *general* transformations of type (ii) unless the Lorentz transformation of the quantity f_P is determined solely by the rotation of the tetrad at P, and not at all by the rotations of the neighboring tetrads. When this is the case, we say that the Lorentz transformation of f_P is *Q_P-determined* (recall that Q_P is the Lorentz-transformation quaternion at P, and that it determines the rotation of the tetrad at P).

How does one find out whether the Lorentz transformation of a quantity at P is Q_P-determined? We shall not attempt a profound answer, but simply *assume* that the Lorentz transformations of certain of the variables appearing in f_P are Q_P-determined. Further, when we come to reinterpret the covariant form of the equation $f_P=0$ as being valid in curved space–time, we shall assume that for such a variable the transformation which corresponds to a given Lorentz transformation of the tetrad at P is the same in the old, flat space–time, and the new, curved space–time interpretations.[31]

In the examples that we shall deal with, the Lorentz transformations of the quantities f_P which appear in the flat space–time equations are not Q_P-determined; but they are functions of Q_P-determined variables and of the partial derivatives of Q_P-determined variables with respect to the x_α. If, therefore, we can define a quantity which has the same Lorentz-transformation properties as the Q_P-determined variable Φ_P, and which reduces to $(\partial\Phi/\partial x_\alpha)_P$ at the event P in the case when the tetrads are locally aligned with respect to the tetrad at P, then the problem of writing the equation $f=0$ in a form which is covariant with respect to general transformations of type (ii) should be readily solved. The

[31] In flat space–time, where the tetrad vectors were chosen to point in the x_α coordinate directions, the Lorentz transformation of the tetrad at P is the same as the Lorentz transformation of the x_α. Similarly, in curved space–time, Eq. (77) and (78) express the relation between the Lorentz transformation of the tetrad at P and the local Lorentz transformation of the corresponding coordinates inertial at P.

quantity that we need is the aligned covariant derivative, defined in the next subsection.

Aligned Absolute and Aligned Covariant Derivatives

In Sec. III.1 we defined the absolute derivative of a tensor A, which is a tensor of the same type as A, and the covariant derivative, which is a tensor of the same contravariant degree as A and of covariant degree one higher. The definitions do not involve any particular choice of reference tetrads, but transformations of the tetrads need not correspond to any simple transformation of the derivatives. Now, for any Riemannian space–time, we define new quantities, the *aligned absolute derivative* and the *aligned covariant derivative* of the tensor A. They have, respectively, the same tensor transformation properties as do the absolute and covariant derivatives of A, but they transform in the same way as A when the tetrads undergo Lorentz transformations.

It was shown at the end of Sec. III.2 that, given an event P and a tetrad $(\sigma_\mu)_P$, one can always transform the neighboring tetrads so that $(\sigma'_{\mu|\nu})_P=0$. After the transformation the tetrads are said to be locally aligned with respect to the tetrad $(\sigma_\mu)_P$. Suppose that, as a result of the aligning transformation, the tensor A becomes A^Γ, where A^Γ is defined in some neighborhood of P. Then we define the *aligned absolute derivative of A at P* to be the absolute derivative of A^Γ at P, and the *aligned covariant derivative of A at P* to be the covariant derivative of A^Γ at P. We introduce the notations $\Delta A/\Delta v$ for the aligned absolute derivative of A with respect to a parameter v, and $A_{;\mu}$ for the aligned covariant derivative of A with respect to x^μ.

It follows from (60) that if one locally aligns the tetrads with respect to the tetrad at P, and chooses coordinates $x^{(P)}{}_\alpha$ inertial at P, then at P the aligned absolute derivative is the ordinary derivative. Similarly, from (62), the aligned covariant derivative with respect to $x^{(P)}{}_\alpha$ at P is the partial derivative with respect to $x^{(P)}{}_\alpha$. From this we get an intuitive idea of the meaning of the aligned derivatives.

The aligned absolute (covariant) derivative of a tensor A is the absolute (covariant) derivative of a tensor A^Γ of the same type as A, and is therefore a tensor of the same type as the absolute (covariant) derivative of A.

The aligned absolute and aligned covariant derivatives of a product of tensors are found by rules like that for ordinary derivatives. For any tensors A and B one has $(AB)_{;\mu}=A_{;\mu}B+AB_{;\mu}$, etc. The proofs are trivial if one first transforms to an inertial coordinate system.

It follows from the definitions that, if a tensor is unchanged by Lorentz transformations of the reference tetrads, then its aligned absolute (covariant) derivative is the same as its absolute (covariant) derivative. In particular, $g_{\mu\nu;\lambda}=g_{\mu\nu|\lambda}=0$ and $g^{\mu\nu}{}_{;\lambda}=g^{\mu\nu}{}_{|\lambda}=0$; and using the product rule of the last paragraph, we see that the operations of raising and lowering tensor indices commute with the operations of taking the aligned absolute or aligned covariant derivative.

To investigate the Lorentz transformation properties of the aligned absolute derivative, we make an arbitrary, but smooth, transformation of the reference tetrads on the curve $\mathcal{C}$. Take v to be a parameter on $\mathcal{C}$, write the events of $\mathcal{C}$ as $P=P(v)$, and let $Q=Q(v)$ be the Lorentz transformation quaternion. The Lorentz transformation of the tensor A, defined on $\mathcal{C}$, is written $A\rightarrow A'=L[Q]A$, where $L[Q]$ is an operator with the group property $L[Q^{(1)}]L[Q^{(2)}]=L[Q^{(1)}Q^{(2)}]$ for any Lorentz-transformation quaternions $Q^{(1)}$ and $Q^{(2)}$.

Without performing the transformation Q of the reference tetrads, one may locally align them with respect to the tetrad at the event $O=P(v_O)$ by means of a Lorentz transformation q. The aligned tensor is

$$A^{\Gamma}=L[q]A, \tag{103}$$

where, using the results and notation of Sec. III.2,

$$q=q(v)=1+(\Gamma_v)_O(v-v_O)+O(v-v_O)^2 \quad \text{as} \quad v\rightarrow v_O. \tag{104}$$

Alternatively, *after* performing the transformation Q one may locally align the tetrads with respect to the transformed tetrad at O by means of a Lorentz transformation q':

$$A'^{\Gamma}=L[q']A'=L[q']L[Q]A, \tag{105}$$

$$q'=q'(v)=1+(\Gamma'_v)_O(v-v_O)+O(v-v_O)^2 \quad \text{as} \quad v\rightarrow v_O. \tag{106}$$

The aligned absolute derivatives of A with respect to v at O before the transformation Q is

$$\left(\frac{\Delta A}{\Delta v}\right)_O=\left(\frac{\delta A^{\Gamma}}{\delta v}\right)_O=\left(\frac{\delta}{\delta v}(L[q]A)\right)_O, \tag{107}$$

and after the transformation Q the aligned absolute derivative of A' is

$$\left(\frac{\Delta A'}{\Delta v}\right)_O=\left(\frac{\delta A'^{\Gamma}}{\delta v}\right)_O=\left(\frac{\delta}{\delta v}(L[q']A')\right)_O. \tag{108}$$

From the group property of L, one has $L[q']A'=L[q']L[Q]A=L[q'Q]A$, and since by (98),

$$q'=Q_OqQ_O^{-1}-(dQ/dv)_OQ_O^{-1}(v-v_O)+O(v-v_O)^2, \tag{109}$$

we get, writing $Q(v)=Q_O+(dQ/dv)_O(v-v_O)+O(v-v_O)^2$ and using $q=1+O(v-v_O)$,

$$q'Q=Q_Oq+O(v-v_O)^2. \tag{110}$$

Thus

$$L[q'Q]A=L[Q_O]L[q]A+O(v-v_O)^2$$
$$=L[Q_O]A^{\Gamma}+O(v-v_O)^2,$$

and from (107), (108),

$$\left(\frac{\Delta A'}{\Delta v}\right)_O=\left(\frac{\delta}{\delta v}(L[Q_O]A^{\Gamma})\right)_O=L[Q_O]\left(\frac{\delta A^{\Gamma}}{\delta v}\right)_O$$
$$=L[Q_O]\left(\frac{\Delta A}{\Delta v}\right)_O. \tag{111}$$

The commutation of the operators $L[Q_O]$ and $\delta/\delta v$ is valid for all reasonably behaved $L[Q]$, and certainly for all the $L[Q]$ that we shall deal with. This is easiest seen by transforming to a coordinate system inertial at O (where $\delta/\delta v$ becomes d/dv).

Eq. (111) says that the Lorentz transformation law of the aligned absolute derivative $\Delta A/\Delta v$ of a tensor A is the same as that of A. Provided that A is differentiable throughout a region containing O, so that the curve $\mathcal{C}$ can be chosen arbitrarily, one proves at once that the Lorentz transformation law of the aligned covariant derivative $A_{;\mu}$ is also the same as that of A.

Relation of Aligned to Nonaligned Derivatives

For brevity, let *aligned derivative(s)* mean *aligned absolute or (and) aligned covariant derivative(s)*, and *nonaligned derivative(s)* mean *absolute or (and) covariant derivative(s)*. We now express aligned derivatives in terms of nonaligned. The principal application that we shall make of the results will be in evaluating the aligned derivatives of invariants. (A nonaligned derivative of an invariant is just an ordinary or partial derivative).

The aligned form A^{Γ} of the tensor A is given by Eqs. (103), (104). Expand the Lorentz transformation operator $L[q]$ about the point O (for which $v=v_O$) on the curve $\mathcal{C}$:

$$L[q]=1+K_O(v-v_O)+O(v-v_O)^2. \tag{112}$$

Substituting in (107), one finds the aligned absolute derivative at O in terms of the operator K_O

$$(\Delta A/\Delta v)_O=(\delta A/\delta v)_O+K_OA_O. \tag{113}$$

A simple special case of (113) is when A is a spinor φ. Instead of (36), we assume the slightly more general Lorentz transformation

$$\varphi\rightarrow\varphi'=\exp[i(N-1)\beta]Q\varphi, \tag{114}$$

where β is the real function such that $Q\bar{Q}=e^{2i\beta}$, and N is a positive or negative integer or zero. (In quantum theory, the charge of the particles described by φ is N times the electronic charge). From (93) and (104), using the definitions $\Gamma_v=\Gamma_\pi\, dx^\pi/dv$ and $\alpha_v=\alpha_\pi\, dx^\pi/dv$, we have

$$q\bar{q}=1+2i(\alpha_v)_O(v-v_O)+O(v-v_O)^2, \tag{115}$$

and substituting in $L[q]\varphi=(q\bar{q})^{\frac{1}{2}(N-1)}q\varphi$, one finds the K_O of Eq. (112). If we drop the suffixes O, Eq. (113) then gives

$$(\Delta\varphi/\Delta v)=(\delta\varphi/\delta v)+(i(N-1)\alpha_v+\Gamma_v)\varphi. \tag{116}$$

For $N=0$, the $i(N-1)\alpha_v$ term in (116) is cancelled by the "electromagnetic" $i\alpha_v$ term in Γ_v, which is what one would expect for neutral particles.

Starting from the transformation law

$$\varphi^{\times}\rightarrow\varphi^{\times\prime}=\exp[-i(N-1)\beta]Q^{\times}\varphi^{\times}, \qquad (117)$$

one shows similarly that $\Delta\varphi^{\times}/\Delta v$ is given by the complex reflection of (116), and it is also easy to prove that $\Delta\bar{\varphi}/\Delta v$ and $\Delta\varphi^{\dagger}/\Delta v$ are given by its adjoint and Hermitian conjugate, respectively.

If y is a Lorentz vector, then $L[q]y=qyq^{\dagger}$, and $K_O y_O=(\Gamma_v)_O y_O+y_O(\Gamma_v^{\dagger})_O$. Again dropping the suffixes O, Eq. (113) gives

$$(\Delta y/\Delta v)=(\delta y/\delta v)+\Gamma_v y+y\Gamma_v^{\dagger}. \qquad (118)$$

In particular, taking $y=\sigma_\mu$ and using (96), we have $\Delta\sigma_\mu/\Delta v=0$, which is also an immediate consequence of the definition of the aligned derivative.

The aligned covariant derivatives of φ and y with respect to x^μ are found by replacing the aligned and nonaligned absolute derivatives in (116) and (118) by the corresponding aligned and nonaligned covariant derivatives, and replacing α_v, Γ_v by α_μ, Γ_μ:

$$\varphi_{;\mu}=\varphi_{|\mu}+(i(N-1)\alpha_\mu+\Gamma_\mu)\varphi, \qquad (119)$$

$$y_{;\mu}=y_{|\mu}+\Gamma_\mu y+y\Gamma_\mu^{\dagger}. \qquad (120)$$

Covariant Field Equations

With the mathematical apparatus that we have assembled, it is straightforward to write the field equations of Part II in forms which are covariant with respect to transformations of the x^μ and Lorentz transformations of the reference tetrads. We define operators D and $\bar{D}$ by

$$DA=\sigma^\mu A_{;\mu}, \qquad \bar{D}A=\bar{\sigma}^\mu A_{;\mu}, \qquad (121)$$

for any tensor A. We write the variables that appear in the field equations—the electromagnetic potential a, the spinors φ, θ, etc.—as invariants. For example, we define a real vector a_μ such that $a=a_\mu\sigma^\mu$ [see Eq. (75) and the remarks that follow it]. When A is an invariant and the tetrads are locally aligned, one has $DA=\partial A$ and $\bar{D}A=\bar{\partial}A$, where $\partial=\sigma^\mu\partial_\mu$ and $\bar{\partial}=\bar{\sigma}^\mu\partial_\mu$. Thus we put the field equations of Part II in a generally covariant form by replacing ∂ by D and $\bar{\partial}$ by $\bar{D}$.

As an illustration, the Lorentz condition on the electromagnetic potential was written as $\partial\cdot a=O$. This is equivalent to $(\partial\bar{a})_4=O$; and according to our prescription it has the generally covariant form $(D\bar{a})_4=O$.

The generalized form of the Dirac equation (42) is

$$D\varphi^{\times}=i\kappa\theta, \qquad \bar{D}\theta=i\kappa\varphi^{\times}. \qquad (122)$$

The terms involving the electromagnetic potential are implicit in (122) provided that we allow Lorentz transformations in which Q satisfies only (82), and provided that the Lorentz transformation of φ is given by (114) and the Lorentz transformation of θ by

$$\theta\rightarrow\theta'=\exp[i(N'-1)\beta]Q\theta, \qquad (123)$$

for suitable N, N'. To show that (122) can be reduced to (42), substitute for $\bar{D}$ and D from (121), then use (119) and its complex reflection. Recall that in (42) the tetrads may be taken to be locally aligned, so that $\sigma_{\nu|\pi}=O$ and, by (92), $\Gamma_\pi=i\alpha_\pi$. We find

$$(\partial-iN\alpha)\varphi^{\times}=i\kappa\theta, \qquad (\bar{\partial}+iN'\bar{\alpha})\theta=i\kappa\varphi^{\times}, \qquad (124)$$

where $\alpha=\sigma^\mu\alpha_\mu$. Eq. (124) is the same as Eq. (42) provided that $N\alpha=-N'\alpha=\epsilon a$. If as before $\epsilon a=-\alpha$, then $N'=-N=1$.

The covariance of these field equations does not depend upon space–time being flat. They are therefore possible expressions of physical laws in curved space–time.

IV. CONCLUSION

The movement towards abstract algebraic and coordinate-independent formulations of physical theories, and away from particular matrix representations and special coordinate systems, is an increasingly popular one, and our work is in accord with it. Less popular, and seemingly opposed to this rarefied mathematical spirit, is our desire to make abstract concepts more concrete and imaginable. To pure mathematical minds the aim is unsympathetic. They are happy in their complex spaces, and would prefer to postulate an affine connection rather than to align tetrad vectors. It is a matter of taste. Those, however, who are prepared to exploit the accident of having been born in space–time may find this paper useful.

ACKNOWLEDGMENTS

I should like to thank Dr. P. G. Bergmann, Dr. F. A. Kaempffer, and Dr. W. Opechowski for their helpful comments on an earlier version of this work.