General Theory of Relativity

P.A.M. Dirac
Florida State University

1975

1 Special Relativity

For the space-time of physics we need four coordinates, the time t and three space coordinates x, y, z. We put

$$t = x^0, x = x^1, y = x^2, z = x^3,$$

so that the four coordinates may be written x^μ, where the suffix μ takes on the values 0, 1, 2, 3. The suffix is written in the upper position so that we may maintain a "balancing" of the suffixes in all the general equations of the theory. The precise meaning of "balancing" will become clear a little later.

Let us take a point close to the point that we originally considered and let its coordinates be $x^\mu + dx^\mu$. The four quantities dx^μ which form the displacement may be considered as the components of a vector. The laws of special relativity allow us to make linear nonhomogeneous transformations of the coordinates, resulting in linear homogeneous transformations of the dx^μ. These are such that, if we choose units of distance and of time such that the velocity of light is unity,

$$\left(dx^0\right)^2 - \left(dx^1\right)^2 - \left(dx^2\right)^2 - \left(dx^3\right)^2 = 0 \tag{1}$$

is invariant.

Any set of quantities A^μ that transform under a change of coordinates in the same way as the dx^μ form what is called a *contravariant vector*. The invariant quantity

$$\left(A^0\right)^2 - \left(A^1\right)^2 - \left(A^2\right)^2 - \left(A^3\right)^2 = (A, A) \tag{2}$$

may be called the squared length of the vector. With a second contravariant vector we have the scalar product invariant

$$A^0B^0 - A^1B^1 - A^2B^2 - A^3B^3 = (A, B) \tag{3}$$

In order to get a convenient way of writing such invariants we introduce the device of lowering suffixes. Define

$$A_0 = A^0,\ A_1 = -A^1,\ A_2 = -A^2,\ A_3 = -A^3. \tag{4}$$

Then the expression on the left side of (2) may be written $A_\mu A^\mu$, in which it is understood that a summation is to be taken over the four values of μ. With the same notation we can write (3) as $A_\mu B^\mu$ or else $A^\mu B_\mu$.

The four quantities $A\mu$ introduced by equation (4) ma also be considered as the compoenents of a vector. Their transformation laws under a change of coordinates are somewhat different from those of A^μ, because of the diferences in sign, and the vector is called a *covariant vector*.([1])

From two contravariant vectors A^μ and B^μ we may form the sixteen quantities $A^\mu B^\nu$. The suffix ν, like all the Greek suffixes appearing in this work, also takes on the four values 0, 1, 2, 3. Those sixteen quantities form the components of a tensor of the second rank. It is sometimes called the outer product of the vectors A^μ and B^ν, as distinct of the scalar product (3) which is called the inner product.

The tensor $A^\mu B^\nu$ is rather a special tensor because there are special relations between its components. But we can add together several tensors constructed in this way to get a general tensor of the second rank; say

$$T^{\mu\nu} = A^\mu B^\nu + A'^\mu B'^\nu + A''^\mu B''^\nu + \cdots \tag{5}$$

The important thing about the general tensor is that under a transformation of coordinates its components transform in the same way as the components of $A^\mu B^\nu$.

We may lower one of the suffixes in $T^{\mu\nu}$ by applying the lowering process to each of the terms on the right hand side of (5). Thus we may form $T_\mu{}^\nu$ or $T^\mu{}_\nu$.

In $T_\mu{}^\nu$ we may set $\nu = \mu$ and get $T_\mu{}^\mu$. This is to be summed over the four values of μ. A summation is always implied over a suffix that occurs twice in a term. Thus $T_\mu{}^\mu$ is scalar. It is equal to $T^\mu{}_\mu$.

We may continue this process and multiply more than two vectors together, taking care that their suffixes are all different. In this way we can construct tensors of higher rank. If the vectors are all contravariant, we get a vector with all its suffixes upstairs. We may then lower any of the suffixes and so get a general tensor with an number of suffixes uppstairs and any number downstairs.

We may set a downstsir suffix equal to an upstairs one. We then have to sum over all values of this suffix. The suffix becomes a dummy. We are left with a tensor having two fewer effective suffixes than the original one. This process is called *contraction*. Thus, if we start with the fourth rank tensor $T_{\mu\nu\rho}{}^\sigma$, one way of contracting it is to put $\sigma = \rho$, which gives the second rank tensor $T_{\mu\nu\rho}{}^\rho$, having only sixteen components, arising from the four values of μ and ν. We could contract again to get the scalar $T_{\mu\mu\rho}{}^\rho$, with just one component.

[1]In fact, if we set $A_\mu = g_{\mu\nu} A^\nu$ where

$$g_{\mu\nu} = g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

we can see that the coefficients of the transformation $A'_\mu = \Lambda_\mu{}^\nu A_\nu$ are linearly related to the coefficients of the transformation $A'^\mu = \Lambda^\mu{}_\nu A^\nu$.

At this stage one can appreciate the balancing of suffixes. Any effective suffix occurring in an equation appears once and only once in each term of the equation, and always upstairs or only downstairs. A suffix appearing twice in a term is a dummy, and it must occur once upstairs and once downstairs. It may be replaced by any other Greek letter not already mentioned in the term. Thus ${T_{\mu}}_{\nu\rho}{}^{\rho} = {T_{\mu}}_{\nu\alpha}{}^{\alpha}$. A suffix must never occur more than twice in a term.

2 Oblique axes

Before passing to the formalism of general relativity it is convenient to consider an intermediate formalism–special relativity referred to oblique rectilinear axes.

If we make a transformation to oblique axes, each of the dx^{μ} mentioned in (1) becomes a linear function the new dx^{μ} and the quadratic form (1) becomes a general quadratic form in the new dx^{μ}. We may write it

$$g_{\mu\nu} dx^{\mu} dx^{\nu} \tag{6}$$

with summation understood over both μ and ν. The coefficients $g_{\mu\nu}$ appearing here depend on the system of oblique axes. Of course we take $g_{\mu\nu} = g_{\nu\mu}$ because any difference of $g_{\mu\nu}$ and $g_{\nu\mu}$ would not show up in the quadratic form (6). There are thus ten independent coefficients $g_{\mu\nu}$.[2]

A general contravariant vector has four components A^{μ} which transform like the dx^{μ} under any transformation of the oblique axes. Thus

$$g_{\mu\nu} A^{\mu} A^{\nu}$$

is invariant. It is the squared length of the vector A^{μ}.

Let B^{μ} be a second contravariant vector; then $A^{\mu} + \lambda B^{\mu}$ is still another, for any value of the number λ. It's squared length is

$$g_{\mu\nu} \left(A^{\mu} + \lambda B^{\mu}\right) \left(A^{\nu} + \lambda B^{\nu}\right) = g_{\mu\nu} A^{\mu} A^{\nu} + \lambda \left(g_{\mu\nu} A^{\mu} B^{\nu} + g_{\mu\nu} A^{\nu} B^{\mu}\right) + \lambda^2 g_{\mu\nu} B^{\mu} B^{\nu}.$$

This must be an invariant for all values of λ. It follows that the term independent of λ and the coefficients of λ and λ^2 must separately be invariants. The coefficient of λ is

$$g_{\mu\nu} A^{\mu} B^{\nu} + g_{\mu\nu} A^{\nu} B^{\mu} = 2 g_{\mu\nu} A^{\mu} B^{\nu}$$

[2] In general the number of independent components in a symmetric matrix of $n \times n$ dimensions is equal to the number of unordered pairs of n elements. From the total number of ordered pairs we subtract n pairs with identical elements, and we are left with $n^2 - n$. We divide the last number by two to obtain the number of unordered pairs with distinct elements. The total number of uordered pairs is then

$$T_n = \frac{1}{2} n \cdot (n+1)$$

which is the nth triangular number. Then we notice that T_{n-1} is the number of independent components of an antisymmetric matrix of $n \times n$ dimensions. In the case of four dimensions this number is $T_3 = 6$.

since in the second term in the left we can interchange μ and ν and then set $g_{\mu\nu} = g_{\nu\mu}$. Thus we find that $g_{\mu\nu}A^\mu B^\nu$ is an invariant. It is the scalar product of A^μ and B^μ.

Let g be the determinant of $g_{\mu\nu}$. It must not vanish; otherwise the four axes would not provide independent directions in space-time and would no be suitable as axes. For the orthogonal axes of the preceding section the diagonal elements of $g_{\mu\nu}$ are 1, -1, -1, -1 and the nondiagonal elements are zero. Thus $g = -1$. With oblique axes g must still be negative, because the oblique axes can be obtained from the orthogonal ones by a continuous process, resulting ig g varying continuosly, and g cannot pass through the value zero.

Define the covariant vector A_μ, with a downstairs suffix, by

$$A_\mu = g_{\mu\nu}A^\nu. \tag{7}$$

Since the determinant g does not vanish, these equations can be solved for A^ν in terms of the A_μ. Let the result be

$$A^\nu = g^{\mu\nu}A_\mu. \tag{8}$$

Each $g^{\mu\nu}$ equals the cofactor of the corresponding $g_{\mu\nu}$, divided by the determinant itself. It follows that $g^{\mu\nu} = g^{\nu\mu}$.

Let us substitute for the A^ν in (7) sus valores dados por (8). We must replace the dummy μ in (8) by some other Greek letter, say ρ, in order not to have three μ's in the same term. We get

$$A\mu = g_{\mu\nu}g^{\nu\rho}A_\rho.$$

Since this equation must hold for any four quantities A_μ, we can infer

$$g_{\mu\nu}g^{\nu\rho} = g^\rho_\mu, \tag{9}$$

where

$$\begin{aligned} g^\rho_\mu &= 1 \quad \text{for } \mu = \rho, \\ &= 0 \quad \text{for } \mu \neq \rho. \end{aligned} \tag{10}$$

The formula (7) may be used to lower any upper suffix occurring in a tensor. Similarly, (8) can be used to raise any downstairs suffix. If a suffix is lowered and raised agian, he result is the same as the original tensor, on account of (9) and 10. Note that g^ν_μ just produces a substitution of ρ for μ or of μ for ρ,

$$g^\rho_\mu A^\mu = A^\rho,$$

and of μ for ρ,

$$g^\rho_\mu A_\rho = A_\mu.$$

if we apply the rule to raising a suffix to the μ in $g_{\mu\rho}$ we get

$$g^\alpha{}_\nu = g^{\alpha\mu}g_{\mu\nu}.$$

This agrees with (9) if we take into accoun that in $g^{\alpha}{}_{\nu}$ we may write the suffixes one above the other because of the symmetry of $g_{\mu\nu}$. Further we may raise the suffix ν by the same rule and get

$$g^{\alpha\beta} = g^{\nu\beta} g^{\alpha}_{\nu},$$

a result which follows immediately form (10). The rules for raising and lowering suffixes apply to all the suffixes in $g_{\mu\nu}$, g^{μ}_{ν}, $g^{\mu\nu}$.

3 Curvilinear coordinates

We now pass on to a system of curvilinear coordinates. We shall deal with quantities which are located at a point in space. Such a quantity may have various components, which are then referred to the axes at that point. There may be a quantity of the same nature at all points in space. It then becomes a field quantity.

If we take such quantity Q (or one of its components if it has several), we can differentiate it with respect to any of the four coordinates. We write the result

$$\frac{\partial Q}{\partial x^{\mu}} = Q_{,\mu}$$

A downstairs suffix preceded by a comma will always denote a derivative in this way. We put the suffix μ downstairs in order to balance the upstairs μ in the denominator on the left. We can see that the suffixes balance by noting that the change in Q, when we pass from a point x^{μ} to the neighboring point $x^{\mu} + \delta x^{\mu}$ is

$$\delta Q = Q_{,\mu} \delta x^{\mu}. \tag{11}$$

We shall have vectors and tensors located at a point, with various components referring to the axes at that point. When we change our system of coordinates, the components will change according to the same laws as in the preceding section, depending on the change of axes at the point concerned. We shall have a $g_{\mu\nu}$ and a $g^{\mu\nu}$ to lower and raise suffixes, as before. But *they are no longer constants*. They vary from point to point. They are field quantities.

Let us see the effect of a particular change in the coordinate system. Take new curvilinear coordinates x'^{μ}, each a function of the four x^{μ}'s. They may be written more conveniently $x^{\mu'}$, with the prime attached to the suffix rather than the main symbol.

Making a small variation in the x^{μ}, we get the four quantities δx^{μ} forming a contravariant vector. Referred to the new axes, this vector has the components

$$\delta x^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^{\nu}} \delta x^{\nu} = x^{\mu'}_{,\nu} \delta x^{\nu}$$

with the notation of (11). This gives the law for he transformation of any contravariant vector A^{ν}; namely,

$$A^{\mu'} = x^{\mu'}_{,\nu} A^{\nu} \tag{12}$$

Interchanging the two systems of axes and changing the suffixes, we get

$$A^{\lambda} = x^{\lambda}_{,\mu'} A^{\mu'} \tag{13}$$

We know from the laws of partial differentiation that

$$\frac{\partial x^{\lambda}}{\partial^{\mu'}} \frac{\partial^{\mu'}}{\partial x^{\nu}} = g^{\lambda}_{\nu}. \tag{14}$$

To see how a covariant vector B_μ transforms, we use the condition that $A^\mu B_\mu$ is invariant. Thus with the help of (13)

$$A^{\mu'} B_{\mu'} = A^{\lambda} B_{\lambda} = x^{\lambda}_{,\mu'} A^{\mu'} B_{\lambda}$$

This result must hold for all values of the four $A^{\mu'}$; therefore we can equate the coefficients of $A^{\mu'}$ and get

$$B_{\mu'} = x^{\lambda}_{,\mu'} B_{\lambda} \tag{15}$$

We can now use the formulas (12) and (15) to transform any tensor with any upstairs and downstairs suffixes. We just have to use coefficients like $x^{\mu'}_{,\nu}$ for each upstairs suffix and $x^{\lambda}_{\nu'}$ for each downstairs suffix and make all the suffixes balance. For example

$$T^{\alpha'\beta'}{}_{\gamma'} = x^{\alpha'}_{,\lambda} x^{\beta'}_{,\mu} x^{\nu'}_{,\gamma} T^{\lambda\mu}{}_{\nu}. \tag{16}$$

Any quantity that transforms according to this law is a tensor. This may be taken as the definition of a tensor.

It should be noted that it has a meaning for a tensor to be symmetrical or antisymmetrical between two suffixes like λ and μ, because this property of symmetry is preserved with the change of coordinates.([3])

The formula (14) may be written

$$x^{\lambda}_{,\alpha'} x^{\beta'}_{,\nu} g^{\alpha'}_{\beta'} = g^{\lambda}_{\nu}$$

It just shows that g^{λ}_{ν} is a tensor. We have also, for any vectors A^μ, B^ν,

$$g_{\alpha'\beta'} A^{\alpha'} A^{\beta'} = g_{\mu\nu} A^{\mu} B^{\nu} = g_{\mu\nu} x^{\mu}_{,\alpha'} x^{\nu}_{,\beta'} A^{\alpha'} A^{\beta'}.$$

Since this holds for all values of $A^{\alpha'}$, $B^{\beta'}$, we can infer

$$g_{\alpha'\beta'} = x^{\mu}_{,\alpha'} x^{\nu}_{,\beta'} g_{\mu\nu}. \tag{17}$$

---

[3]In fact if $T^{\mu\nu}$ is tensor and ($T^{\mu\nu} = \pm T^{\nu\mu}$). By the transformation law

$$T^{\mu'\nu'} = x^{\mu'}_{,\alpha} x^{\nu'}_{,\beta} T^{\alpha\beta}.$$

Using the property of symmetry

$$T^{\mu'\nu'} = \pm x^{\mu'}_{,\alpha} x^{\nu'}_{,\beta} T^{\beta}\alpha = \pm x^{\nu'}_{,\beta} x^{\mu'}_{,\alpha} T^{\beta}\alpha$$

and reordering the dummy suffixes

$$T^{\mu'\nu'} = \pm x^{\nu'}_{,\alpha} x^{\mu'}_{,\beta} T^{\alpha}\beta = \pm T^{\nu'\mu'}.$$

This shows that $g_{\mu\nu}$ is a tensor. Similarly, $g^{\mu\nu}$ is a tensor. They are called the *fundamental tensors.*

If S is any scalar field quantity, it can be considered as a function of the four x^μ or the four $x^{\mu'}$. From the laws of partial differentiation

$$S_{,\mu'} = S_{,\lambda} x^\lambda_{,\mu'}.$$

Hence the $S_{,\lambda}$ transform like the B_λ of equations (15) and thus *the derivative of a scalar field is a covariant vector field.*

4 Nontensors

We can have a quantity $N^\mu_{\nu\rho\cdots}$ with various up and down suffixes, which is not a tensor. If it is a tensor, it must transform under a change of coordinate system according to the law exemplified by (16). With any other law it is a nontensor. A tensor has the property that if all the components vanish in a system of coordinates, they vanish in every system of coordinates. This may not hold for a nontensor.

For a nontensor we can raise and lower suffixes by the same rules as for a tensor. Thus, for example,

$$g^{\alpha\nu} N^\mu{}_{\nu\rho} = N^{\mu\alpha}{}_\rho.$$

The consistency of these rules is quite independent of the transformation laws to a different system of coordinates. Similarly, we can contract a nontensor by putting an upper and lower suffix equal.

We may have tensors and nontensors apparing together in the same equation. The rules for balancing suffixes apply equally to tensors and nontensors.

4.1 The Quotient Theorem

Suppose that $P_{\lambda\mu\nu}$ is such that $A^\lambda P_{\lambda\mu\nu}$ is a tensor *for any vector* A^λ. Then $P_{\lambda\mu\nu}$ is a tensor.

To prove it, write $A^\lambda P_{\lambda\mu\nu} = Q_{\mu\nu}$. We are given that this is a tensor; therefore

$$Q_{\beta\gamma} = Q_{\mu'\nu'} x^{\mu'}_{,\beta} x^{\nu'}_{,\gamma}.$$

Thus

$$A^\alpha P_{\alpha\beta\gamma} = A^{\lambda'} P_{\lambda'\mu'\nu'} x^{\mu'}_{,\beta} x^{\nu'}_{,\gamma}.$$

Since A^λ is a vector, we have from (12),

$$A^{\lambda'} = A^\alpha x^{\lambda'}_{,\alpha'}.$$

So

$$A^\alpha P_{\alpha\beta\gamma} = A^\alpha x^{\lambda'}_{,\alpha} P_{\lambda'\mu'\nu'} x^{\mu'}_{,\beta} x^{\nu'}_{,\gamma}.$$

This equation must hold for all values of A^α, so

$$P_{\alpha\beta\gamma} = P_{\lambda'\mu'\nu'} x^{\lambda'}_{,\alpha} x^{\mu'}_{,\beta} x^{\nu'}_{,\gamma}.$$

showing that $P_{\alpha\beta\gamma}$ is a tensor.

The theorem also holds is $P_{\alpha\beta\gamma}$ is replaced by a quantity with any number of suffixes, and if some of the suffixes are upstairs.

5 Curved Space

One can easily imagine a curved two-dimensional space as a surface immersed in Euclidean three-dimensional space. In the same way, one can have a curved four-dimensional space immersed in a flat space of a larger number of dimensions. Such a curved space is called a Riemann space. A small region of it is approximately flat.

Einstein assumed that physical space is of this nature and thereby laid the foundation for his theory of gravitation.

For dealing with curved space one cannot introduce a rectilinear system of axes. One has to use curvilinear coordinates, such as those dealt with in Section 3. The whole formalism of that section can be applied to curved space, because all the equations are local ones which are not disturbed by the curvature.

The invariant distance ds between a point x^μ and a neighboring point $x^\mu + dx^\mu$ is given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

like (6). ds is real for a timelike interval and imaginary for a spacelike interval.

With a network of curvilinear coordinates the $g_{\mu\nu}$, given as functions of the coordinates, fix all the elements of distance; so they fix the metric. They determine both the coordinate system and the curvature of the space.

6 Parallel Displacement

Suppose we have a vector A^μ located at a point P. **If the space is curved, we cannot give a meaning to a parallel vector at a different point** Q, as one can easily see if one thinks of the example of a curved two-dimensional space in a three-dimensional Euclidean space. However, if we take a point P' close to P, there is a parallel vector at P', with an uncertainty of the second order, counting the distance from P to P' as the first order. Thus we can give a meaning to displacing the vector A^μ from P to P' keeping it parallel to itself and keeping the length constant.

We can transfer the vector continuously along a path by this process of parallel displacement. Taking a path from P to Q, we end up with a vector at Q which is parallel to the original vector at P, **with respect to this path**. But a different path will give a different result. There is no absolute meaning to a parallel vector at Q. If we transport the vector at P by parallel displacement around a closed loop, we shall end up with a vector at P which is usually in a different direction.

We can get equations for the parallel displacement of a vector by supposing our four-dimensional physical space immersed in a flat space of a higher number of dimensions; say N. In this N-dimensional space we introduce rectilinear coordinates z^n $(n = 1, \cdots, N)$. Those coordinates do not need to be orthogonal, only rectilinear. Between two neighboring points there is an invariant distance ds given by

$$ds^2 = h_{mn} dz^m dz^n, \tag{18}$$

summed for n, $m = 1, 2, \cdots, N$. The h_{nm} are constants, unlike the $g_{\mu\nu}$. We may use them to lower suffixes in the N-dimensional space; thus

$$dz_n = h_{mn} dz^m.$$

Physical space forms a four-dimensional "surface" in the flat N-dimensional space. Each point x^μ in the surface determines a definite point y^n in the N-dimensional space. Each coordinate y^n is a function of the four x's; say $y^n(x)$. The equations of the surface would be given by eliminating the four x's from the N $y^n(x)$'s. There are $N - 4$ such equations.

By differentiating the $y^n(x)$ with respect to the parameters x^μ, we get

$$\frac{\partial y^n(x)}{\partial x^\mu} = y^n_{,\mu}.$$

For two neighboring points in the surface differing by δx^μ, we have

$$\delta y^n = y^n_{,x^\mu} \delta x^\mu \tag{19}$$

The squared distance between them is, from (18)

$$\delta s^2 = h_{mn} y^m_{,\mu} y^n_{,\nu} \delta x^\mu \delta x^\nu.$$

We may write it

$$\delta s^2 = y^n_{,\mu} y_{n,\nu} \delta x^\mu \delta x^\nu.$$

Hence

$$g_{\mu\nu} = y^n_{,\mu} y_{n,\nu}. \tag{20}$$

Take a contravariant vector A^μ in physical space, located at the point x. Its components A^μ are like the δx^μ of (19). Thus

$$A^n = y^n_{,\mu} A^\mu. \tag{21}$$

Now, shift the vector A^n, keeping it parallel to itself (which means, of course, keeping its components constant), to a neighboring point $x + dx$ in the surface. It will no longer lie in the surface at the new point, on account of the curvature of the surface. But we can project it on to the surface, to get a definite vector lying on the surface.

The projection process consists in splitting the vector into two parts, a tangential part and a normal part, and discarding the normal part. Thus

$$A^n = A^n_{\text{tan}} + A^n_{\text{nor}}. \tag{22}$$

Now, if K^μ denotes the components of A^n_{tan} referred to the x coordinate system in the surface, we have, corresponding to (21),

$$A^n_{\text{tan}} = K^\mu y^n_{,\mu}(x + dx), \tag{23}$$

with the coefficients $y^n_{,\mu}$ taken at the new point $x + dx$.

A^n_{nor} is defined to be orthogonal to every tangential vector at the point $x+dx$, and thus to every vector like the right-hand side of (23), no matter what the K^μ are. Thus([4])

$$A^n_{\text{nor}} y_{n,\mu}(x + dx) = 0.$$

If we now multiply (6.5) by $y_{n,\nu}(x + dx)$ the A^n_{nor} term drops out and we are left with

$$\begin{array}{rcl} A^n y_{n,\nu}(x + dx) & = & K^\mu y^n_{,\mu}(x + dx)_{n,\nu}(x + dx) \\ & = & K^\mu g_{\mu\nu}(x + dx) \end{array}$$

from (20). Thus, to the first order in dx

$$\begin{array}{rcl} K_\nu(x + dx) & = & A^n \left[y_{n,\nu}(x) + y_{n,\nu,\sigma} dx^\sigma\right] \\ & = & A^\mu y^n_{,\mu} \left[y_{n,\nu} + y_{n,\nu,\sigma} dx^\sigma\right] \\ & = & A_\nu + A^\mu y^n_{,\mu} y_{n,\nu,\sigma} dx^\sigma \end{array}$$

This K_ν is the result of parallel displacement of A_ν to the point $x + dx$. We may put

$$K_\nu - A_\nu = dA_\nu,$$

so dA_ν denotes the change in A_ν under parallel displacement. Then we have

$$dA_\nu = A^\mu y^n_{,\mu} y_{n,\nu,\sigma} dx^\sigma \tag{24}$$

7 Christoffel Symbols

By differentiating (20) we get (**omitting the second comma with two differentiations**)

$$\begin{array}{rcl} g_{\mu\nu,\sigma} & = & y^n_{,\mu\sigma} y_{n,\nu} + y^n_{,\mu} y_{n,\nu\sigma} \\ & = & y_{n,\mu\sigma} y^n_{,\nu} + y_{n,\nu\sigma} y^n_{,\mu} \end{array} \tag{25}$$

since we can move the suffix n freely up and down, on account of the constancy of the h_{mn}. Interchanging μ and σ in (25) we get

$$g_{\sigma\nu,\mu} = y_{n,\mu\sigma} y^n_{,\nu} + y_{n,\nu\mu} y^n_{,\sigma} \tag{26}$$

[4]Here, Dirac assumes that the metric is positive (or negative) which is not the case for the space-time of relativity. Therefore, this mathematical argument supports, at most, an analogy between Einstein's space-time and a Riemann space.

Interchanging ν and σ in (25)

$$g_{\mu\sigma,\nu} = y_{n,\mu\nu} y^n_{,\sigma} + y_{n,\sigma\nu} y^n_{,\mu} \tag{27}$$

Now take (25)+(27)−(26) and divide by 2. The result is

$$\frac{1}{2}\left(g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\nu\sigma,\mu}\right) = y_{n,\nu\sigma} y^n_{,\mu}. \tag{28}$$

Put

$$\Gamma_{\mu\nu\sigma} = \frac{1}{2}\left(g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\nu\sigma,\mu}\right) \tag{29}$$

It is called a Christoffel symbol of the first kind. It is symmetrical with respect to the last two suffixes. It is a nontensor. A simple consequence of (29) is

$$\Gamma_{\mu\nu\sigma} + \Gamma_{\nu\mu\sigma} = g_{\mu\nu,\sigma}. \tag{30}$$

We see now that (24) can be written as

$$dA_\nu = A^\mu \Gamma_{\mu\nu\sigma} dx^\sigma. \tag{31}$$

Al reference to the N-dimensional space has now disappeared, as the Christoffel symbol involves only the metric $g_{\mu\nu}$ of physical space.

We can infer that the length of a vector is unchanged by parallel displacement. We have

$$\begin{aligned} d\left(g^{\mu\nu} A_\mu A_\nu\right) &= g^{\mu\nu} A_\nu dA_\mu + g^{\mu\nu} A_\mu dA_\nu + A_\mu A_\nu g^{\mu\nu}{}_{,\sigma} dx^\sigma \\ &= A^\mu dA_\mu + A^\nu dA_\nu + A_\mu A_\nu g^{\mu\nu}{}_{,\sigma} dx^\sigma \\ &= A^\mu A^\nu \Gamma_{\nu\mu\sigma} dx^\sigma + A^\nu A^\mu \Gamma_{\mu\nu\sigma} dx^\sigma + A_\mu A_\nu g^{\mu\nu}{}_{,\sigma} dx^\sigma \\ &= A^\mu A^\nu g_{\mu\nu,\sigma} dx^\sigma + A_\mu A_\nu g^{\mu\nu}{}_{,\sigma} dx^\sigma \end{aligned} \tag{32}$$

Now $g^{\alpha\mu}{}_{,\sigma} g_{\mu\nu} + g^{\alpha\mu} g_{\mu\nu,\sigma} = \left(g^{\alpha\mu} g_{\mu\nu}\right)_{,\sigma} = g^\alpha_{\nu,\sigma} = 0$. Multiplying by $g^{\beta\nu}$, we get

$$g^{\alpha\beta}{}_{,\sigma} = -g^{\alpha\mu} g^{\beta\nu} g_{\mu\nu,\sigma}. \tag{33}$$

This is a useful formula giving the derivative of $g^{\alpha\beta}$ in terms of the derivative of $g_{\mu\nu}$. It allows us to infer

$$A_\alpha A_\beta g^{\alpha\beta}{}_{,\sigma} = -A^\mu A^\nu g_{\mu\nu,\sigma}$$

and so the expression (32) vanishes. Thus the length of a vector is constant. In particular, a null vector (i.e. a vector of zero length) remains a null vector under parallel displacement.

The constancy of the length of the vector follows also from geometrical arguments. When we split up the vector A^μ into tangential and normal parts according to (22), the normal part is infinitesimal and is orthogonal to the tangential part. It follows that, to the first order, the length of the whole vector equals that of its tangential part.

The constancy of the length of any vector requires the constancy of the scalar product $g^{\mu\nu}A_\mu B_\nu$ of any two vectors A and B. This can be inferred from the constancy of the length of $A+\lambda B$ for any value of the parameter λ.

It is frequently useful to raise the first suffix of the Christoffel symbol so as to form

$$\Gamma^\mu_{\nu\sigma} = g^{\mu\lambda}\Gamma_{\lambda\mu\sigma}.$$

It is then called a Christoffel symbol of the second kind. It is symmetrical between its two lower suffixes. As explained in Section 4, this raising is quite permissible, even for a nontensor.

The formula (31) may be written

$$dA_\nu = \Gamma^\mu_{\nu\sigma} A_\mu dx^\sigma \tag{34}$$

It is the standard formula referring to covariant components. For a second vector B^ν we have

$$\begin{aligned} d(A_\nu B^\nu) &= 0 \\ A_\nu dB^\nu &= -B^\nu dA_\nu = -B^\nu \Gamma^\mu_{\nu\sigma} A_\mu dx^\sigma \\ &= -B^\mu \Gamma^\nu_{\mu\sigma} A_\nu dx^\sigma \end{aligned}$$

This must hold for any A_ν, so we get

$$dB^\nu = -\Gamma^\nu_{\mu\sigma} B^\mu dx^\sigma. \tag{35}$$

This is the standard formula for parallel displacement referring to contravariant components.

8 Geodesics

Take a point with coordinates z^μ and suppose it moves along a track; we then have a function of some parameter τ. Put $dz^\mu/d\tau = u^\mu$.

There is a vector u^μ at each point of the track. Suppose that as we go along the track the vector u^μ gets shifted by parallel displacement. Then the whole track is determined if we are given the initial point and the initial value of the vector u^μ. We just have to shift the initial point from z^μ to $z^\mu + u^\mu d\tau$, then shift the vector u^μ to this new point by parallel displacement, then shift the point again in the direction fixed by the new u^μ, and so on. Not only is the track determined, but also the parameter τ along it. A track produced this way is called a geodesic.

If the vector u^μ is a null vector, it always remains a null vector and the track is called a null geodesic. If the vector u^μ is initially timelike (i.e., $u^\mu u_\mu > 0$), it is always timelike and we have a timelike geodesic. Similarly if u^μ is initially spacelike ($u^\mu u_\mu < 0$), it is always spacelike and we have a spacelike geodesic.

We get the equations of a geodesic by applying (35) with $B^\nu = u^\nu$ and $dx^\sigma = dz^\sigma$. Thus

$$\frac{du^\nu}{d\tau} + \Gamma^\nu_{\mu\sigma} u^\mu \frac{dz^\sigma}{d\tau} = 0 \tag{36}$$

or

$$\frac{d^2z}{d\tau^2} = +\Gamma^{\nu}_{\mu\sigma}\frac{dz^\mu}{d\tau}\frac{dz^\sigma}{d\tau} = 0 \tag{37}$$

For a timelike geodesic we may multiply the initial u^ν by a factor so as to make its length unity. This merely requires a change in the scale of τ. The vector u^μ now always has the length unity. It is just the velocity vector $v^\mu = dz^\mu/ds$, and the parameter τ has become the proper time s.

Equation (36) then becomes

$$\frac{dv^\mu}{ds} + \Gamma^{\mu}_{\nu\sigma}u^\nu u^\sigma = 0. \tag{38}$$

Equation (37) becomes

$$\frac{d^2z^\mu}{ds^2} + \Gamma^{\mu}_{\nu\sigma}\frac{dz^\nu}{ds}\frac{dz^\sigma}{ds} = 0. \tag{39}$$

We make the physical assumption that the world line of a particle not acted on by any forces, except gravitational, is a timelike geodesic. This replaces Newton's first law of motion. Equation (39) fixes the acceleration and provides the equations of motion.

We also make the assumption that the path of a ray of light is a null geodesic. It is fixed by equation (37) referring to some parameter along the path. The proper time s cannot now be used because ds vanishes.

9 The Stationary Property of Geodesics

A gepdesic that is not a null geodesic has the property that $\int ds$, taken along a sectio of the track with the end points P and Q, is stationary if one makes a small variation of the tack keeping the end points fixed.

Let us suppose that each point of the track, with the coordinates z^μ is shifted so that its coordinates become $z^\mu + \delta z^\mu$. If dx^μ denotes an element along the track,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu$$

Thus

$$\begin{aligned} 2ds\delta ds &= dx^\mu dx^\nu \delta g_{\mu\nu} + g_{\mu\nu}dx^\mu \delta dx^\nu + g_{\mu\nu}dx^\nu \delta dx^\mu \\ &= dx^\mu dx^\nu \delta g_{\mu\nu} + 2g_{\mu\lambda}dx^\mu \delta dx^\lambda. \end{aligned}$$

Now

$$\delta dx^\lambda = d(\delta x^\lambda).$$

Thus, with the help of $dx^\mu = v^\mu ds$,

$$\delta(ds) = \left(\frac{1}{2}g_{\mu\nu,\lambda}v^\mu v^\nu \delta x^\lambda + g_{\mu\lambda}v^\mu \frac{d\delta x^\lambda}{ds}\right) ds$$

Hence

$$\delta\int ds = \int \delta(ds) = \int \left[\frac{1}{2}g_{\mu\nu,\lambda}v^\mu v^\nu \delta x^\lambda + g_{\mu\lambda}v^\mu \frac{d\delta x^\lambda}{ds}\right] ds$$

By partial integration (of the second term), using the condition that $\delta x^\lambda = 0$ at the end points P and Q, we get

$$\delta \int ds = \int \left[\frac{1}{2} g_{\mu\nu,\lambda} v^\mu v^\nu - \frac{d}{ds} \left(g_{\mu\lambda} v^\mu \right) \right] \delta x^\lambda ds \tag{40}$$

The condition for this to vanish with arbitrary δx^λ is

$$\frac{d}{ds} \left(g_{\mu\lambda} v^\mu \right) - \frac{1}{2} g_{\mu\nu,\lambda} = 0 \tag{41}$$

Now

$$\begin{aligned} \tfrac{d}{ds} \left(g_{\mu\lambda} v^\mu \right) &= g_{\mu\lambda} \tfrac{dv^\mu}{ds} + g_{\mu\lambda,\nu} v^\mu v^\nu \\ &= g_{\mu\lambda} \tfrac{dv^\mu}{ds} + \tfrac{1}{2} \left(g_{\lambda\mu,\nu} + g_{\lambda\nu,\mu} \right) v^\mu v^\nu . \end{aligned}$$

Thus the condition (41) becomes

$$g_{\mu\lambda} \frac{dv^\mu}{ds} + \Gamma_{\lambda\mu\nu} v^\mu v^\nu = 0.$$

Multiplying this by $g^{\lambda\sigma}$, it becomes

$$\frac{dv^\sigma}{ds} + \Gamma^\sigma_{\mu\nu} v^\mu v^\nu = 0,$$

which is just the condition (38) for the geodesic.

This work shows that for a geodesic, (40) vanishes and $\int ds$ is stationary. Conversely, if we assume that $\int ds$ is stationary, we can infer that the track is a geodesic. Thus we may use the stationary condition as the definition of a geodesic, except in the case of a null geodesic.

10 Covariant Differentiation

Let S be a scalar field. Its derivative S, y is a covariant vector, as we saw in Section 3. Now let A_μ be a vector field. Is its derivative $A_{\mu,\nu}$ a tensor?

We must examine how $A_{\nu,\nu}$ transforms under a change of coordinate system. With the notation in Section 3, A_μ transforms to

$$A_{\mu'} = A_\rho x^\rho_{,\mu'}$$

like equation (15), and hence

$$\begin{aligned} A_{\mu',\nu'} &= \left(A_\rho x^\rho_{,\mu'} \right)_{,\nu'} \\ &= A_{\rho,\sigma} x^\sigma_{,\nu'} x^\rho_{,\mu'} + A_\rho x^\rho_{,\mu'\nu'} . \end{aligned}$$

The last term should not be here if we were to have the correct transformation law for a tensor. Thus $A_{\mu,\nu}$ is not a tensor.

We can, however, modify the process of differentiation so as to get a tensor. Let us take the vector A_μ at the point x and shift it to $x + dx$ by parallel displacement. It is still a vector. We may subtract it from the vector A_μ at $x+dx$ and the difference will be a vector. It is, to the first order

$$A_\mu(x+dx) - \left[A_\mu(x) + \Gamma^\alpha_{\mu\nu} A_\alpha dx^\nu\right] = \left(A_{\mu,\nu} - \Gamma^\alpha_{\mu\nu} A_\alpha\right) dx^\nu.$$

This quantity is a vector, for any vector dx^ν; hence, by the quotient theorem of Section 4, the coefficient

$$A_{\mu\nu} - \Gamma^\alpha_{\mu\nu} A_\alpha$$

is a tensor. One can easily verify directly that it transforms correctly under a change of coordinate system.

It is called the covariant derivative of A_μ and written

$$A_{\mu:\nu} = A_{\mu,\nu} - \Gamma^\alpha_{\mu\nu} A_\alpha. \tag{42}$$

The sign : before a lower suffix will always denote a covariant derivative, just as the comma denotes an ordinary derivative.

Let B_ν be a second vector. We define the outer product $A_\mu B_\nu$ to have the covariant derivative([5])

$$\left(A_\mu B_\nu\right)_\sigma = A_{\mu:\sigma} B_\nu + A_\mu B_{\nu:\sigma} \tag{43}$$

Evidently it is a tensor with three suffixes. It has the value

$$\left(A_\mu B_\nu\right)_\sigma = \left(A_{\mu,\sigma} - \Gamma^\alpha_{\mu\sigma} A_\alpha\right) B_\nu + A_\mu \left(B_{\nu,\sigma} - \Gamma^\alpha_{\nu\sigma} B_\alpha\right)$$

Let $T_{\mu\nu}$ be a tensor with two suffixes. It is expressible as a sum of terms like $A_\mu B_\nu$, so, its covariant derivative is

$$T_{\mu\nu:\sigma} = T_{\mu\nu,\sigma} - \Gamma^\alpha_{\mu\sigma} T_{\alpha\nu} - \Gamma^\alpha_{\nu\sigma} T_{\mu\alpha}. \tag{44}$$

The rule can be extended to the covariant derivative of a tensor $Y_{\mu\nu\cdots}$ with any number of suffixes downstairs

$$Y_{\mu\nu\cdots:\sigma} = Y_{\mu\nu\cdots,\sigma} - \text{ a } \Gamma \text{ term for each suffix.} \tag{45}$$

In each of these Γ terms we must make the suffixes balance, which is sufficient to fix how the suffixes go.

The case of a scalar is included in the general formula (45) with the number of suffixes in Y zero.

$$Y_{:\sigma} = Y_{,\sigma}. \tag{46}$$

Let's apply (44) to the fundamental tensor $g_{\mu\nu}$. It gives

$$\begin{aligned} g_{\mu\nu:\sigma} &= g_{\mu\nu,\sigma} - \Gamma^\alpha_{\mu\sigma} g_{\alpha,\nu} - \Gamma^\alpha_{\nu\sigma} g_{\mu\alpha} \\ &= g_{\mu\nu,\sigma} - \Gamma_{\nu\mu\sigma} - \Gamma_{\nu\mu\sigma} = 0 \end{aligned}$$

from (30). Thus the $g_{\mu\nu}$, count as constants under covariant differentiation.

[5]Notice that the definition of the covariant derivative of a second rank tensor is not motivated because the notion of *parallel displacement of tensor* doesn't make sense.

Formula (43) is the usual rule that one uses for differentiating a product. We assume this usual rule holds also for the covariant derivative of the scalar product of two vectors. Thus

$$(A^\mu B_\mu) : \sigma = A^\mu_{:\sigma} B_\mu + A^\mu B_{\mu:\sigma}.$$

We get, according to (46) and (42),

$$(A^\mu B_\mu) : \sigma = A^\mu_{:\sigma} B_\mu + A^\mu \left(B_{\mu,\sigma} - \Gamma^\alpha_{\mu\sigma} B_\alpha \right).$$

and hence

$$A^\mu_{,\sigma} B_\mu = A^\mu_{:\sigma} B_\mu - A^\alpha \Gamma^\mu_{\alpha\sigma} B_\mu.$$

Since this holds for any B_μ, we get

$$A^\mu_{:\sigma} B_\mu = A^\mu_{,\sigma} B_\mu + A^\alpha \Gamma^\mu_{\alpha\sigma} B_\mu, \tag{47}$$

which is the basic formula for the covariant derivative of a contravariant vector. The same Christoffel symbol occurs as in the basic formula (42) for a covariant vector, but now there is a + sign. The arrangement of suffixes is completely determined by the balancing requirement.

We can extend the formalism so as to include the covariant derivative of any tensor with any number of upstairs or downstairs suffixes. A Γ terms appears for each suffix, with a + sign if the suffix is upstairs an a − sign if it is downstairs. If we contract two suffixes in the tensor, the corresponding Γ terms cancel.

The formula for the covariant derivative of a product,

$$(XY)_{:\sigma} = X_{:\sigma} Y + X Y_{:\sigma}, \tag{48}$$

holds quite generally, with X and Y any kind of tensor quantities. On account of the $g_{\mu\nu}$ counting as constants, we can shift suffixes up or down before covariant differentiation and the result is the same as if we shifted them afterwards.

The covariant derivative of a nontensor has no meaning.

The laws of physics must be valid in all systems of coordinates. Thus must thus be expressible as tensor equations. Whenever they involve the derivative of a field quantity, it must be a covariant derivative. The field equations of physics must all be rewritten with the ordinary derivatives replaced by covariant derivatives. For example, the d'Alembert equation $\Box V = 0$ for a scalar V becomes, in covariant form

$$g^{\mu\nu} V_{:\mu:\nu} = 0.$$

This gives, from (42) and (46),

$$g^{\mu\nu} \left(V_{,\mu\nu} - \Gamma^\alpha_{\mu\nu} V_\alpha = 0. \right) \tag{49}$$

Even if one is working with flat space (which means neglecting the gravitational field) and one is using curvilinear coordinates, one must write one's equations in terms of covariant derivatives if one wants them to hold in all systems of coordinates.