

Physics 570

When is a Manifold Curved: Covariant Derivatives and Curvature

I. Action of “directional derivatives” on Vectors: an Affine Connection

As one moves on a manifold, along a curve with tangent vector \tilde{u} , we write the derivative, in that direction, of a scalar function, $f \in \mathcal{F}$, as $\tilde{u}(f) = u^\mu f_{,\mu}$. We now want to generalize this operator to ask how vectors, and other sorts of tensors, vary as we move along that same curve on the manifold. The problem is complicated because we have already chosen a basis field, $\{\tilde{e}_\mu\}|_P$ in each of the vector spaces \mathcal{T}_P above the points, P , on the manifold. However, we have no real requirements on “the way these basis vectors point,” in “nearby” vector spaces, except that they change in a continuous (smooth) way. Therefore the physics-related problem lies in ways to specify how they change, as one passes from a vector space, at the point $P \in \mathcal{M}$, say, to some nearby point. In principle, it should be some generalization of Taylor’s theorem, for the values of $f(x + \delta x)$, when one knows the values of the function at x as well as its various derivatives. Therefore, the first step in answering such a question is to determine the first derivative of the basis vectors at the point P . Of course the derivative may vary as one changes the direction in which one is going; i.e., we need a generalization of the directional derivative, for our basis vectors. Given the direction in question, i.e., a vector tangent to some curve, we can think of this as a mapping that should give us the (first) derivative of the basis vector in that direction. The ideas here are trying to show the independence of the general notion of affine connection and the metric tensor.

The ideas that accomplish this task are named either as “the covariant derivative” or “the affine connection.” Calling it the covariant derivative reminds us that we want some generalization of the derivative with the property that the rate of change of a tensor is again a tensor, i.e., it should vary covariantly. On the other hand, in general an affine relation is one that relates two geometric entities that can be smoothly “translated” one into the other, via a change in origin, which is surely what is going on when one moves from one vector space to another one. As there are in fact many approaches to these notions, let me also commend

to you the discussions in Carroll’s text, at the beginning of chapter 3, and in Padmanabhan’s text that begin with physically-motivated discussions in §4.5 and then keep recurring §4.6 with a nice figure there.

The following are reasonable requirements for an operator to be called the **covariant derivative of a vector in some direction, specified by another vector**, which reduces to our earlier notion of directional derivative when it acts only on functions,

$$\nabla : \mathcal{T} \otimes \mathcal{T} \longrightarrow \mathcal{T} \quad , \quad (1.1)$$

$\nabla_{\tilde{u}} \tilde{v} \Big|_P \longleftarrow$ the (covariant) derivative, at the point P , of \tilde{v} in the direction \tilde{u} .

- (a) linear for addition: $\nabla_{\tilde{u}}(\tilde{v} + \tilde{w}) = \nabla_{\tilde{u}}\tilde{v} + \nabla_{\tilde{u}}\tilde{w} \quad ,$
- (b) a derivation for products in the upper argument: $\nabla_{\tilde{u}}(f\tilde{v}) = f\nabla_{\tilde{u}}\tilde{v} + [\tilde{u}(f)]\tilde{v} \quad ,$
- (c) purely linear in the first (lower) argument: $\nabla_{f\tilde{u}+\tilde{w}}\tilde{v} = f\nabla_{\tilde{u}}\tilde{v} + \nabla_{\tilde{w}}\tilde{v} \quad .$

Since this object is linear in its first argument, we could consider the quantity $\nabla \tilde{v}$ as a $\left\{ \begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right\}$ -tensor since, it is awaiting one more vector—which is the role of a 1-form—in order to give back the vector desired. Therefore, we can say that $\nabla \tilde{v}$ is an element of $\Lambda \otimes \mathcal{T}$, and re-write the important part of Eqs. (1.1) above—the part dealing with its behavior as a derivation—in the following simple form:

$$\nabla(f\tilde{v}) = f \nabla \tilde{v} + df \otimes \tilde{v} \quad (1.2)$$

One should of course immediately ask just why one cannot use the ordinary partial derivative definition to determine the derivative of a vector at some point; i.e., why can we simply not use $\tilde{u}(\tilde{v}) \equiv \{u^i \partial v^j / \partial x^i\} \partial_{x^j}$ to describe the rate of change, since, after all, the requirements above make it clear that we do intend to take this approach for scalar functions. The answer that this does not work well at all is that it doesn’t transform well under a change of coordinates on the manifold. Recall that under a change of coordinates, where we choose some new set of coordinates $y^\alpha = y^\alpha(x^\mu)$, then the components of a vector transform as

$$\tilde{u} = u^\mu \partial_{x^\mu} = u'^\alpha \partial_{y^\alpha} \quad \iff \quad u^\mu = \frac{\partial x^\mu}{\partial y^\alpha} u'^\alpha \equiv X_\alpha^\mu u'^\alpha \quad . \quad (1.3a)$$

Therefore, the partial derivative of these components would transform as

$$\partial_{y^\beta} u'^\alpha = \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial}{\partial x^\nu} \left(\frac{\partial y^\alpha}{\partial x^\mu} u^\mu \right) = \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial y^\alpha}{\partial x^\mu} (\partial_{x^\nu} u^\mu) + \frac{\partial x^\nu}{\partial y^\beta} \frac{\partial^2 y^\alpha}{\partial x^\nu \partial x^\mu} u^\mu . \quad (1.3b)$$

This behavior is not at all how a tensor is supposed to transform. The first term is fine, is linear in the tensor components, and has the correct transformation matrices for the two different types of indices it has. However, the second term in the expression above is definitely not linear in the tensor in question. Presumably one of the qualities that the covariant derivative must have is to ensure that this extra term does not appear, i.e., that it is cancelled out by something else.

The requirements already given are insufficient to uniquely determine the covariant derivative; we need a way to further specify them. This is usually done by first specifying the action of the covariant derivative on the elements of a given basis set, $\{\tilde{e}_\mu\}$, where I have used the appropriate symbol to denote a completely arbitrarily chosen set of basis vectors for my tangent vectors. Since the transformation of the partial derivatives of the components of a tangent vector, calculated above, did not produce another tensorial quantity, we need to find some different way of thinking about how to create a tensorial quantity that has the character of a directional derivative. Since the partial derivative approach does indeed work for functions, i.e., the differential of a function is a 1-form, the problem must lie with the fact that when we considered a transformation for the components we did not also consider a transformation for the basis vectors. Yet a different way of thinking about that is that we need to know how to align the basis vectors in very nearby vector spaces; therefore, we begin to resolve this problem by giving a definition for the directional derivatives of the basis vectors themselves. The directional derivative of a basis vector is again a tangent vector; therefore, it must be a linear combination of the original basis vectors. Depending on one's point of view, and noting that the covariant derivative depends on the direction in question in a way that is linear with respect to both addition and scalar multiplication, i.e., tensorially, we may write down the defining equations for that derivative in several different, but totally equivalent, ways. We

begin by giving a name to the derivative of a fixed basis vector, say \tilde{e}_μ , in the direction of some other basis vector, say \tilde{e}_ν . Since the result must again be a vector it must be capable of being written out in terms of linear combinations of the original basis vectors. We give the names $\Gamma^\lambda_{\mu\nu}$ to those coefficients, a set of m^3 quantities:

$$\nabla_{\tilde{e}_\nu} \tilde{e}_\mu \equiv \nabla_\nu \tilde{e}_\mu \equiv \Gamma^\lambda_{\mu\nu} \tilde{e}_\lambda \in \mathcal{J} \quad , \quad (1.4a)$$

However, our general requirements for a covariant derivative tell us that it should be linear in the direction; therefore, we may rewrite this equation for a more general direction:

$$\begin{aligned} \nabla_{\tilde{u}} \tilde{e}_\mu &= u^\nu \nabla_{\tilde{e}_\nu} \tilde{e}_\mu = u^\nu \Gamma^\lambda_{\mu\nu} \tilde{e}_\lambda \equiv \{\mathfrak{L}^\lambda_\mu(\tilde{u})\} \tilde{e}_\lambda \in \mathcal{J} \quad , \\ \mathfrak{L}^\lambda_\nu(\tilde{e}_\mu) &= \Gamma^\lambda_{\nu\mu} \quad , \text{ or, equivalently } \mathfrak{L}^\lambda_\mu = \Gamma^\lambda_{\mu\nu} \omega^\nu \quad , \end{aligned} \quad (1.4b)$$

where it is reasonable to have created a set of m^2 1-forms, \mathfrak{L}^λ_μ , to act on the direction as given by the tangent vector \tilde{u} , since we know that this process is linear in that argument, and the $\{\omega^\nu\}$ constitute the basis for 1-forms reciprocal to the basis for tangent vectors, $\{\tilde{e}_\mu\}$, that we are using.

If we then go one step further, backward, we may remove the explicitly-presented tangent vector, \tilde{u} , and instead present it as a 1-form waiting to be given its desired tangent vector, so that we have the simpler form, now a $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ tensor:

$$\nabla \tilde{e}_\mu = \mathfrak{L}^\lambda_\mu \otimes \tilde{e}_\lambda \in \Lambda \otimes \mathcal{J} . \quad (1.5)$$

Collectively the m^2 1-forms labeled by \mathfrak{L}^λ_μ are referred to as “*the connection 1-forms*,” and were originally introduced by Cartan. (Be sure and note the order of the two lower indices on $\Gamma^\lambda_{\mu\nu}$, the components of the connection 1-form, in the 3rd and 4th lines!) Of course it must be so that the two remaining indices **do not** constitute indices for some tensorial object, since they must transform in such a way as to cancel out the non-tensorial transformation of the partial derivative that constitutes the remainder of the covariant derivative. It is straightforward, if

perhaps slightly lengthy, to show the following transformation laws for them, following the format of Eqs. (1.3):

$$\tilde{\Gamma}^\mu{}_\nu \equiv \Gamma^\mu{}_{\nu\lambda} \varpi^\lambda = (X^{-1})^\alpha{}_\nu X^\mu{}_\beta \tilde{\Gamma}^\beta{}_\alpha + X^\mu{}_\beta d\{(X^{-1})^\beta{}_\nu\}. \quad (1.6)$$

Given these coefficients we can use linearity and the fact that the operator in question is a derivation to give us a complete formula for an arbitrary vector field, $\tilde{v} \in \mathcal{T}$, and an arbitrary directional field, $\tilde{u} \in \mathcal{T}$, treating the components of \tilde{v} as scalars for which, according to the rules above in Eqs. (1.1), we simply use the ordinary exterior derivative:

$$\nabla_{\tilde{u}} \tilde{v} = \nabla_{\tilde{u}}(v^\lambda \tilde{e}_\lambda) = \{\tilde{u}(v^\lambda) + v^\nu \tilde{\Gamma}^\lambda{}_\nu(\tilde{u})\} \tilde{e}_\lambda = \tilde{e}_\lambda u^\nu \{v^\lambda{}_{;\nu} + \Gamma^\lambda{}_{\mu\nu} v^\mu\} \equiv \tilde{e}_\lambda u^\nu v^\lambda{}_{;\nu}, \quad (1.7a)$$

or, without yet giving it the direction vector \tilde{u}

$$\nabla \tilde{v} = \tilde{e}_\lambda \otimes \{dv^\lambda + \tilde{\Gamma}^\lambda{}_\mu v^\mu\} = \tilde{e}_\lambda \otimes \varpi^\nu \{v^\lambda{}_{;\nu} + \Gamma^\lambda{}_{\mu\nu} v^\mu\} \equiv \tilde{e}_\lambda \otimes \varpi^\nu \{v^\lambda{}_{;\nu}\}, \quad (1.7b)$$

and the two subscripted symbols $v^\lambda{}_{;\nu}$ and $v^\lambda{}_{;\nu}$ are common abbreviations:

$$v^\lambda{}_{;\nu} \equiv \tilde{e}_\nu(v^\lambda), \quad v^\lambda{}_{;\nu} \equiv v^\lambda{}_{,\nu} + \Gamma^\lambda{}_{\mu\nu} v^\mu. \quad (1.7c)$$

The first abbreviation, with the ‘‘comma,’’ is simply a generalization of the usual symbol for partial derivatives so that it denotes the action of **any** basis of tangent vectors on functions, even though the basis vectors may no longer be holonomic, i.e., it may not be just a partial derivative with respect to some coordinate, but, instead the linear combination of the actions of partial derivatives that constitutes the definition of the (possibly non-holonomic) basis vector. The second abbreviation, with the ‘‘semi-colon,’’ is referred to as ‘‘the components of the covariant derivative of the vector \tilde{v} in the direction specified by the ν -th basis vector, \tilde{e}_ν . When the v^λ are the components of a $\begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ tensor, then the $v^\lambda{}_{;\nu}$ are the components of a $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ tensor, as was originally desired.

I note that **the usual approach to specifying an affine connection** is to give rules by which one determines the values of these 1-forms, so as to satisfy various constraints that are

believed to be more fundamental. For instance, although one may have an affine connection whether or not there is a metric on the manifold, nonetheless one may ask that the covariant derivative leave the metric alone, i.e., require that the metric should be covariantly constant. One may also impose conditions on it that seem plausible for all vectors. We will discuss all this further when we introduce the metric nearer the end of this section of notes. For now we will suppose that the m^2 1-forms that determine the covariant derivative have been given, and we want to extend this to act on other tensorial quantities.

II. Action of the covariant derivative on Differential Forms and other Tensors

We may extend this definition to also act on 1-forms by **requiring the covariant derivative to commute with the operation of 1-forms on tangent vectors**. Since the action of a 1-form on a tangent vector, say $\mathcal{Q}(\tilde{v})$, is a function, and we already know how to calculate directional derivatives of functions, we may write

$$\tilde{u}\{\mathcal{Q}(\tilde{v})\} = \nabla_{\tilde{u}}\{\mathcal{Q}(\tilde{v})\} \equiv (\nabla_{\tilde{u}}\mathcal{Q})(\tilde{v}) + \mathcal{Q}(\nabla_{\tilde{u}}\tilde{v}) \quad . \quad (2.1)$$

Since every 1-form is a linear combination of the basis 1-forms, we now use Eq. (1.5) to determine the covariant derivative of the basis 1-forms by considering the special case when $\tilde{u} \rightarrow \tilde{e}_\nu$, $\mathcal{Q} \rightarrow \varpi^\lambda$ and $\tilde{v} \rightarrow \tilde{e}_\mu$:

$$\begin{aligned} 0 = \nabla_\nu \delta_\mu^\lambda &= \nabla_\nu \{\varpi^\lambda(\tilde{e}_\mu)\} = \{\nabla_\nu \varpi^\lambda\}(\tilde{e}_\mu) + \varpi^\lambda(\nabla_\nu \tilde{e}_\mu) = \{\nabla_\nu \varpi^\lambda\}(\tilde{e}_\mu) + \Gamma^\lambda_{\mu\nu} \quad , \\ &\implies \{\nabla_\nu \varpi^\lambda\}(\tilde{e}_\mu) = -\Gamma^\lambda_{\mu\nu} \quad , \\ &\implies \nabla_\nu \varpi^\lambda = -\Gamma^\lambda_{\mu\nu} \varpi^\mu \quad \Rightarrow \quad \nabla \varpi^\lambda = -\Gamma^\lambda_{\mu\nu} \otimes \varpi^\mu . \end{aligned} \quad (2.2)$$

Therefore the covariant derivative of a general 1-form, $\mathcal{Q} \equiv \alpha_\mu \varpi^\mu$, may be written

$$\begin{aligned} \nabla_{\tilde{u}}\mathcal{Q} &= \{\tilde{u}(\alpha_\mu) - \alpha_\lambda \Gamma^\lambda_{\mu\nu}(\tilde{u})\} \varpi^\mu = u^\nu \{\alpha_{\mu,\nu} - \alpha_\lambda \Gamma^\lambda_{\mu\nu}\} \varpi^\mu \equiv u^\nu \{\alpha_{\mu;\nu}\} \varpi^\mu \in \Lambda , \\ &\text{or, if still awaiting a tangent vector for the direction,} \end{aligned} \quad (2.3)$$

$$\nabla \alpha = \{d\alpha_\mu - \alpha_\lambda \Gamma^\lambda_{\mu\nu}\} \otimes \varpi^\mu \equiv \varpi^\nu \otimes \varpi^\mu \{\alpha_{\mu;\nu}\} \equiv \varpi^\nu \otimes \varpi^\mu \{\alpha_{\mu,\nu} - \alpha_\lambda \Gamma^\lambda_{\mu\nu}\} \in \Lambda \otimes \Lambda .$$

where we see that the “difference” between the action of the covariant derivative of 1-forms and tangent vectors amounts to a difference in sign, and summing on the upper index of Γ^λ_μ instead of the lower one.

We may now easily extend the action of the covariant derivative operator to spaces of arbitrary tensors by simply requiring that **it satisfy the product rule** when it is asked to act **on tensor products**. Therefore, for example, we have

$$\nabla(\tilde{u} \otimes \tilde{v}) \equiv \nabla\tilde{u} \otimes \tilde{v} + \tilde{u} \otimes \nabla\tilde{v} . \quad (2.4)$$

This causes the covariant derivative of a tensor product of two tangent vectors to have two positive terms with connection 1-forms, Γ^λ_μ , the covariant derivative of a tensor product of a single tangent vector and a single 1-form to have one positive and one negative “ Γ -term.” A reasonably generic tensor—chosen arbitrarily to be of rank $\{\frac{1}{3}\}$, as is the Riemann curvature tensor—would have the components of its covariant derivative as follows:

$$R^a{}_{bcd;e} = R^a{}_{bcd,e} + \Gamma^a{}_{ge} R^g{}_{bcd} - \Gamma^g{}_{be} R^a{}_{gcd} - \Gamma^g{}_{ce} R^a{}_{bgd} - \Gamma^g{}_{de} R^a{}_{bcg} . \quad (2.5)$$

III. Parallel Displacements of Functions and of Vectors, and the notion of Geodesics

Directional derivatives answer the question, “What is the rate of change of a function as it is moved to different places on the manifold, in some particular direction, as specified by moving it along some given curve, which has that direction as its tangent vector?” Covariant derivatives answer the same question for vectors, 1-forms, and more complicated tensors. The usual Taylor series expansion is of course invoked to do this. **The purpose of doing this is, eventually, to learn some properties of the underlying manifold itself.**

We first do such an expansion for ordinary (scalar) functions. For $P \in U \subseteq \mathcal{M}$, $f \in \mathcal{F}|_U$, and $\Gamma(\lambda)$ a curve such that $\Gamma(0) = P$, the Taylor expansion for f gives

$$f[\Gamma(\lambda)] = f[\Gamma(0)] + \lambda \left\{ \frac{d}{d\lambda} f[\Gamma(\lambda)] \right\} \Big|_{\lambda=0} + \frac{1}{2} \lambda^2 \left\{ \frac{d^2}{d\lambda^2} f[\Gamma(\lambda)] \right\} \Big|_{\lambda=0} + \dots \equiv e^{\lambda \tilde{u}} f|_P , \quad (3.1)$$

where we interpret the exponential function as simply its entire power series. You might recall the perhaps-easier but very similar expansion for functions of one real variable,

$$e^{a\partial_x} f(x) = f(x) + a \frac{df}{dx}(x) + \frac{1}{2} a^2 \frac{d^2 f}{dx^2}(x) + \dots = f(x + a) . \quad (3.1a)$$

The covariant derivative allows us to do the same sorts of things for vectors, 1-forms, etc.; i.e., we can actually talk about comparing the values of a vector field within the vector spaces at different points. However, in this section, we set up some apparatus for moving around on curves and first consider the case of functions. This will lead us to the notion of the *torsion* of a manifold; in the next section we will do the same thing for vectors, which will lead us to the notion of the *curvature* of a manifold. Firstly, we say that a vector field, \tilde{v} , at a point P , is **parallelly propagated** along a curve, $\Gamma(\lambda)$ with tangent vector $\tilde{u} \equiv \frac{d}{d\lambda} \Gamma(\lambda)$, if its covariant derivative in the direction \tilde{u} , evaluated at the point P , is exactly zero—i.e., if it’s not changing at that point, in that direction:

■ Definition for parallel propagation of a vector \tilde{v} along a direction \tilde{u} :

$$\nabla_{\tilde{u}} \tilde{v} = 0 . \quad (3.2)$$

In words this merely says that the relative angle of the tangent vector along the curve and the vector being propagated along that curve does not change!

■ Definition of a **geodesic**:

One of the very useful things one can do in this regard is to begin with a complete basis of vectors, i.e., a tetrad used to make measurements, and then parallel transport them along someone’s worldline. The simplest sort of a worldline for a material object is motion with no forces acting on it. We will refer to the worldline of such an object as a *geodesic*. (Of course in Newtonian mechanics, or in special relativity, we would characterize such a motion as a “straight line.”

We therefore promulgate a definition of “a straight line,” as a (local piece of a) curve with the property that its direction is always the same; we will refer to such a curve, locally, as a **a**

geodesic. Since the direction cannot change, the covariant derivative of the tangent vector, parallelly propagated along itself, must always be proportional to itself: $\nabla_{\tilde{u}}\tilde{u} = \phi\tilde{u}$, where ϕ is some (scalar) function, of proportionality, defined in the neighborhood under scrutiny. This is surely in agreement with the general physical notion that if the “acceleration” of an object is in the same direction as its motion, it never changes direction.

However, given such an equation, it is simple to determine a new choice of parameter for the curve such that the new function ϕ is simply zero. We refer to such parameters as *affine parameters*. If some parametrization has a non-zero function $\phi(\lambda)$, we may find a “better choice,” $s = s(\lambda)$, by solving the equation $\frac{d^2s}{d\lambda^2} = \phi(\lambda)\frac{ds}{d\lambda}$. When everything is transformed to this new variable the function ϕ will have been transformed to zero. Considering the equation to determine s , we can see that one may still change the affine parameter to another one by finding a different solution of $d^2s/dt^2 = 0$. The solution of this is then straightforward, and says that all affine parameters are related one to another via an equation of the form $s' = as + b$, where a and b are constants, i.e., they simply determine a “choice of zero” and a “choice of constant scale length.” Therefore, modulo these two constant “choices,” the affine parameter for a geodesic is uniquely determined. In special relativity, we have already been using the proper length along a spacelike curve as an affine parameter, and it will continue to be good. Likewise, we have been using the proper time of an observer as an affine parameter along her worldline. This will continue to be the correct choice, although it is worthwhile to remember that timelike displacements along straight lines are curves of maximal time extent. Using this newer approach we will also be able to determine useful affine parameters along null curves.

As commented above, a timelike observer may well want to parallel transport her basic, spatial triad along with her, to continue to make measurements that are reliable. Let us first consider the case where this observer is moving in “free fall,” i.e., under the action of no forces, and therefore along a geodesic, which means that the tangent vector to her worldline satisfies the differential equations

$$0 = \nabla_{\tilde{u}}\tilde{u} = \tilde{e}_\mu \left\{ \frac{d}{d\tau}u^\mu + \Gamma^\mu{}_{\nu\lambda}u^\nu u^\lambda \right\} = \tilde{e}_\mu \left\{ \frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu{}_{\nu\lambda} \frac{dx^\nu}{d\tau} \frac{dx^\lambda}{d\tau} \right\}. \quad (3.3)$$

Our observer then wants to choose some initial tetrad of basis vectors, $\{\tilde{V}_{(\alpha)}\}_1^4$, for the purposes of her measurements, and will choose them to be an orthonormal set, i.e.,

$$\eta^{\alpha\beta} V_{(\alpha)}^\mu V_{(\beta)}^\nu \Big|_{\tau=0} = \eta^{\mu\nu} . \quad (3.4)$$

She will want to use her own timepiece to measure local time, which is the same as saying that she should take \tilde{u} as $\tilde{V}_{(4)}$. She then wants to have 3 other (spatial) basis vectors, satisfying Eqs. (3.4), that are parallel propagated along her worldline, i.e., they need to satisfy Eqs. (3.2) above, to be parallel propagated:

$$\nabla_{\tilde{u}} \tilde{V}_{(\alpha)} \equiv \tilde{e}_\mu \left\{ \frac{d}{d\tau} V_{(\alpha)}^\mu + V_{(\alpha)}^\lambda \Gamma^\mu_{\lambda\eta} u^\eta \right\} = 0 , \quad \alpha = 1, 2, 3, 4 . \quad (3.5)$$

where it is clear that the condition $\tilde{V}_{(4)} = \tilde{u}$ is consistent with these requirements.

When we will introduce, soon, a metric into our considerations, and also insist that the metric be parallelly propagated along any geodesic, a completely equivalent but quite different-sounding definition/property for a geodesic can be determined, namely that for any two **nearby** points, and every curve joining them, the geodesic is the one that *extremizes* the distance along that curve. For spacelike-separated points this extremal will be a minimum, while for timelike-separated points this extremal will be a maximum.

■ Physical Interpretation of the Connections:

It is appropriate at this point to interject some physics into the discussion. We know that a particle moving under no forces moves along a straight line. Using the equivalence principle this means that a particle moving **only under gravitational forces** will follow a geodesic. Writing the location of the particle, as a function of its proper time, as $x^\mu(\tau)$, it follows that the tangent vector to its motion is its world velocity, $u^\mu \equiv dx^\mu/d\lambda$. This vector must now be parallelly propagated along itself, i.e., we have $0 = (\nabla_{\tilde{u}} \tilde{u})^\mu$, which we now evaluate in the (instantaneous) rest frame of the particle, so that u^μ is just zero except for the 4th component, which is 1, and then solve the equation above for the instantaneously-measured acceleration:

$$\frac{d^2 x^\mu}{d\tau^2} = -\Gamma^\mu_{44} . \quad (3.6)$$

We know that this acceleration is caused, depending on one’s interpretation, either by the curvature of the spacetime or by some gravitational field, we see that the components of that field are simply given by $-\Gamma^\mu_{44}$, as measured in that instantaneous rest frame. It is worth remembering, as well, that we could always have used an orthonormal basis in this rest frame, so that we could use this to lower the index μ . We will show a bit later on that, in an orthonormal basis, the connection is always skew-symmetric on its first two indices—when all of them are “lower.” Therefore, we only really need to understand this when μ takes on only the values m , which vary from 1 to 3. This tells us that, indeed, the components of the connection are physical quantities; they are the 3 components of the gravitational field, $g^m = -\Gamma^m_{44}$.

■ Fermi-Walker Transport (along worldlines), Padmanabhan, p. 181-186:

If an observer is not moving along a geodesic, but is accelerated, for whatever reason, then the use of parallel transport of one’s initial tetrad is insufficient to allow the observer to continue to use his local time as the measure of time along a parallelly-transported tetrad. Therefore, Fermi was motivated to introduce a more general sort of transport of vector fields along worldlines—now referred to as Fermi-Walker transport—depends on the acceleration of the observer in question. The idea is that there is a 2-plane spanned by \tilde{u} and $\tilde{a} \equiv d\tilde{u}/d\tau = \nabla_{\tilde{u}}\tilde{u}$, and it is therefore reasonable that \tilde{u} and \tilde{a} may “rotate” in that 2-plane, but that vectors that began normal to that 2-plane should continue to do so as they are propagated along. (Something similar occurs, in ordinary Newtonian mechanics in 3-space, when one considers, for instance, an object rotating around some axis, ω , and considers the acceleration determined by the cross product with ω .)

At any event, we say that a vector, \tilde{V} , is Fermi-Walker transported along a worldline with 4-velocity \tilde{u} , having 4-acceleration \tilde{a} provided that

$$\nabla_{\tilde{u}}\tilde{V}^\mu = a^{[\lambda}u^{\mu]}V_\lambda . \tag{3.7a}$$

If we apply this to the case where \tilde{V} is just chosen as \tilde{u} itself, we see that it just gives the proper relationship between \tilde{u} and \tilde{a} , remembering that the two of those are always orthogonal. As well

we see that if some vector is orthogonal to both of \tilde{u} and \tilde{a} , i.e., does not have any component in the 2-plane they span, then this mode of transport will maintain that orthogonality. It is also useful to consider the case when the tetrad in question is determined via a set of 1-forms. Then we can write the equation above in the following form, where $\tilde{w} \equiv w^\mu \tilde{e}_\mu$ is the desired vector, and $\underline{w} \equiv g_{\mu\nu} w^\mu \underline{\omega}^\nu$ is the 1-form associated with that vector, and, similarly, we consider the 1-forms associated with the 4-velocity and the 4-acceleration:

$$\nabla_{\tilde{u}} \underline{w} = (\underline{u} \wedge \underline{a}) (\tilde{w}), \quad (3.7b)$$

where the wedge product plays the role that a cross product would have in 3 dimensions. We will have reason to utilize this approach when considering some interesting cases later on.

IV. Tensorial Tools to Measure the lack of Flatness of a Manifold

Two important tensors give reasonably specific information concerning the deviations of the manifold away from being just a flat \mathbb{R}^n . These are referred to as the *torsion tensor*, \mathbf{T} , and the *curvature tensor*, \mathbf{R} . These tensors depend on our choice of an affine connection, independent of a metric. For this reason, no metric has yet been introduced into the space (or spacetime) being considered. We are however “building up” to the point where we will be able to recognize the physical interaction of the metric and the connection, which may then relate the derivatives of the metric and the curvature.

The standard, classical version of Einstein’s general relativity spends time studying the curvature tensor, because it will be seen to be a local, covariant measure of those motions of test particles that we usually ascribe to “tidal gravitational fields”—those gravitational fields that vary from one point to another. On the other hand, this same point of view “assumes” that the torsion tensor must surely be just identically zero; i.e., we don’t attempt to measure it, but, instead, define it to be zero because of prior “philosophical or metaphysical” knowledge, concerning, perhaps, the way that functions should behave. In several other, more complicated theories of gravity, where plausible interactions of the gravitational field with local spinorial matter fields—classic examples

are due to Cartan and A. Trautman—the torsion tensor couples to whatever spinor fields that may exist in the matter of the system under study. Therefore we will at least spend a little bit of extra time describing both of these objects, and describing how to use them to look at the structure of the manifold and its various tensor bundles.

1. Preliminaries on a Geometrical Understanding of the Commutators of Vector Fields

We now define a **congruence of curves** as a **continuous family** of ordinary curves, say $\Gamma_s(\lambda)$. The parameter λ varies **along** any individual curve, while the parameter s labels just which curve it is that we are considering. A trivial example is the set of curves over \mathbb{R}^2 —with the usual (x, y) coordinates—given by $\Gamma_s(\lambda) = (\lambda, s) \in \mathbb{R}^2$. These are simply straight lines parallel to the \hat{x} -axis, in \mathbb{R}^2 , intercepting the \hat{y} -axis at the value s ; therefore, we may conceive of each one of these curves as having the same tangent vector but beginning at a different point on the manifold, the point with coordinates $(0, s)$. A common, physical example is a “bundle” of parallel light rays emanating, say, from a large searchlight, creating a “path” of light through a night sky.

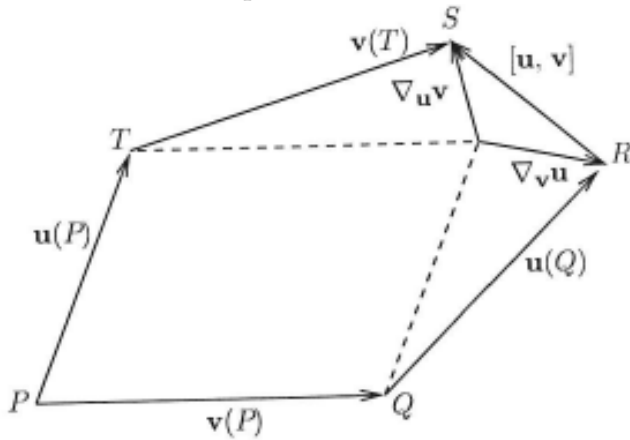
Given two *congruences* of curves, we will use them to create small closed paths in some neighborhood and then to parallel propagate functions, vectors, and other tensors around these paths. Beginning at a particular point, $P_0 \in \mathcal{M}$, we can specify a pair of curves through P_0 , and then another pair of curves, with the same functional form as the first two, i.e., with their tangent vectors simply the representatives at some other points of the two tangent vector fields that generated the first pair of curves. This process then gives us an area bounded by some four curves, the idea being to outline something like a small “rectangle.” Unfortunately, it may well be that these curves are not completely parallel—in general one would only expect this of curves of coordinate axes; therefore, as described below in a bit more detail, it may be necessary to select a “special” additional curve that “closes up” our rectangle into a closed area. It is claimed that this additional curve is precisely the curve whose tangent vector is

the *Lie commutator* of the tangent vectors of the two curves, namely the vector given by the following definition of its action on functions:

$$[\tilde{u}, \tilde{v}]f \equiv \tilde{u}\{\tilde{v}(f)\} - \tilde{v}\{\tilde{u}(f)\} . \quad (4.0)$$

We now want to give a proof of this fact, which will be needed for further discussions. Therefore, we begin by re-naming the members of a congruence of curves, labelling them by the point at which they begin. More precisely, instead of writing the curves as $\Gamma_s(\lambda)$, for some particular value of s , we will label the particular curve that “begins” at some point $P \in \mathcal{M}$ by the label $\Gamma(\lambda; P)$. We anticipate that curves which begin at two distinct points will determine two distinct curves, provided of course that the second point, say Q , does not lie on the curve that begins at the point P . We may then consider two congruences of curves. Two members of the first congruence are denoted by $\Gamma(\lambda; P)$ and $\Gamma(\lambda; Q)$, both of which have the tangent vector $\tilde{u} = d/d\lambda$, but of course at two different points, so that they are labeled $\tilde{u}(P)$ and $\tilde{u}(Q)$ on the diagram below.

Then two members of the second congruence will be denoted by $\Delta(\mu; P)$ and $\Delta(\mu; T)$, both of which have tangent vector $\tilde{v} = d/d\mu$. As well we have chosen these curves so that \tilde{v}_P is not parallel to \tilde{u}_P for any P in the neighborhood, U , where we are studying this problem. This structure is then sufficient to describe the picture given just below, where the dotted lines simply show what the original piece of curve would have looked like if it were simply just moved over to the second point.



Everything begins at the arbitrary point, $P \in U \subseteq \mathcal{M}$. We then take members of the two different congruences that begin at that point, $\Gamma(\lambda; P)$ and $\Delta(\mu; P)$, and move away from P along them. Following along the first curve, originally in the direction \tilde{u}_P , until the parameter has increased by some (small) amount, $\delta\lambda$, we come to some point $T \in \mathcal{M}$. Alternately, we follow along the direction \tilde{v}_P until the parameter increases by the small value $\delta\mu$, and we come to a different point $Q \in \mathcal{M}$. Then, back at T , we follow along the curve $\Delta(\mu; T)$, in the direction \tilde{v}_T until the parameter increases by the value $\delta\mu$, arriving at S . Likewise, beginning at Q , we may follow along the curve $\Gamma(\lambda; Q)$ initially in the direction \tilde{u}_Q until the parameter increases by the value $\delta\lambda$, arriving eventually at R . At this point, one wonders, aloud, as to whether or not S and R are the same point on the manifold! Were the manifold flat, **and** were the two curves straight, then this would surely be the case. On the other hand, in the generic case, it is surely not so.

To answer this question on the manifold, we select an arbitrary function, $f \in \mathfrak{F}|_U$, and use the sketch above to aid our visualization of the following scheme, where we use Taylor series to propagate it from P to S , and **also** from P to R :

$$\begin{aligned} f(T) &= f(P) + (\delta\lambda)(\tilde{u}(f))|_P + \frac{1}{2}(\delta\lambda)^2(\tilde{u}[\tilde{u}(f)])|_P + \dots, \\ f(Q) &= f(P) + (\delta\mu)(\tilde{v}(f))|_P + \frac{1}{2}(\delta\mu)^2(\tilde{v}[\tilde{v}(f)])|_P + \dots, \end{aligned} \tag{4.1}$$

This is sufficient algebra to allow us to determine the expression we actually want, namely the difference $f(S) - f(R)$:

$$\begin{aligned} f(S) - f(R) &= [f(S) - f(T)] + [f(T) - f(P)] - [f(R) - f(Q)] - [f(Q) - f(P)] \\ &= \left. \begin{array}{l} \{ \text{Taylor series at } T \} \\ \text{series from } P \end{array} \right\} + \text{Taylor series at } P \\ &\quad - \left. \begin{array}{l} \{ \text{Taylor series at } Q \} \\ \text{series from } P \end{array} \right\} - \text{Taylor series at } P \\ &= \dots = (\delta\lambda \delta\mu)(\tilde{u}[\tilde{v}(f)]|_P - (\tilde{v}[\tilde{u}(f)]|_P) + \dots \equiv (\delta\lambda \delta\mu)[\tilde{u}, \tilde{v}]_P(f) + \dots, \end{aligned} \tag{4.2}$$

where it was necessary to insert the two Taylor series from Eqs. (4.1) into those expansions in Eqs. (4.2), so that everything was then evaluated at the original point, P_0 . We see that in fact

the two points are not the same, and that it is exactly the Lie commutator of the two vector fields that gives us the difference of the values of the function at the two points—of course only to lowest order in the small quantities, $\delta\lambda$ and $\delta\hat{\mu}$. It is this commutator that measures the lack of “matching up” of the two sets of curves, thereby justifying Eq. (4.0). To rephrase, relative to the above picture, we describe the following two trajectories:

- 1) first follow the curve with tangent vector \tilde{u}_P for $\delta\lambda$, to T , and then \tilde{v}_T for $\delta\mu$, to S , this will be equivalent to the second trajectory, described as
- 2) first follow the curve with \tilde{v}_P for $\delta\mu$, to Q , and then \tilde{u}_Q for $\delta\lambda$, to R , **and then, further**, a curve with tangent vector $[\tilde{u}, \tilde{v}]_P$ for parameter distance $\delta\lambda \delta\mu$, which will finally bring us to S .

Since both these paths take us from P to point S , we may create **a closed path** beginning and ending at P , by first following, say, the second one above, and returning along the negative of the first one. Having such an explicit description of an arbitrary closed curve, we will begin “dragging” various geometric objects around these curves to find out what happens to them. Although, on the manifold itself, we actually get back to the beginning, it is **not** the case that functions, vectors, etc., also return to themselves.

2. A pair of Preliminary Analytic Results concerning Commutators

To give good proofs of these important results we need some useful mathematical expressions concerning the the commutator of two vector fields, and the action of the exterior derivative of a 1-form on such a commutator. As the derivations are somewhat tedious, I will present them here in smaller type—so that you may omit the reading of them if you so desire—**but** I will highlight the important resulting formulae by giving their equation numbers on the left-hand side.

To discover a simple analytic formula for the commutator, we first note that the function $\tilde{u}[\tilde{v}(f)]$ is of course not the result of any tangent vector acting on f , since it involves second derivatives. However, it is straightforward to show that the “Lie commutator” of two tangent vectors is just again some other tangent

vector, which we may describe in the following way, relative to some, perhaps non-holonomic basis set, $\{\tilde{e}_\mu\}$:

$$(4.3) \quad \begin{aligned} [\tilde{u}, \tilde{v}] &= \tilde{u}(v^\mu) \tilde{e}_\mu - \tilde{v}(u^\mu) \tilde{e}_\mu + u^\mu v^\nu [\tilde{e}_\mu, \tilde{e}_\nu], \text{ or} \\ [\tilde{u}, \tilde{v}] &= \{\tilde{u}(v^\mu) - \tilde{v}(u^\mu) + u^\rho v^\sigma C_{\rho\sigma}{}^\mu\} \tilde{e}_\mu = \{u^\nu v^\mu_{,\nu} - v^\nu u^\mu_{,\nu} + u^\rho v^\sigma C_{\rho\sigma}{}^\mu\} \tilde{e}_\mu, \end{aligned}$$

where we have used the definition of the commutator of two (in general anholonomic) basis vectors, and the prescription for subscripts with commas to denote partial derivatives created by basis vectors. See also p. 67 of Csrrroll's text.

Our next analytic step is to uncover a link between the commutator of two vectors and the action of the exterior derivative, $d\mathcal{Q} \in \Lambda^2$ on the same pair of vectors. We want to show that for any $\alpha \in \Lambda$, there is an action, on $\tilde{u}, \tilde{v} \in \mathcal{J}$ given by

$$(4.4) \quad d\mathcal{Q}(\tilde{u}, \tilde{v}) = \tilde{u}[\mathcal{Q}(\tilde{v})] - \tilde{v}[\mathcal{Q}(\tilde{u})] - \mathcal{Q}([\tilde{u}, \tilde{v}]) \quad .$$

Before proving this, notice that in the simplest case, where we take \tilde{u} and \tilde{v} to be just two members of a commuting basis set, say $\{\partial_i, \partial_j\}$, then their commutator, $[\partial_i, \partial_j]$ is simply zero, so that the equation above says that $d\mathcal{Q}(\partial_i, \partial_j) = \partial_i[\mathcal{Q}(\partial_j)] - \partial_j[\mathcal{Q}(\partial_i)] = \partial_i\alpha_j - \partial_j\alpha_i = \partial_{[i}\alpha_{j]}$, just exactly as one would expect, i.e., these are the components of something like the curl of a vector with the components of \mathcal{Q} !

To proceed for a proof, we begin by expanding the right-hand side of the desired result, Eq. (4.4). We also do this for the special case that the basis set chosen is actually holonomic, i.e., the elements of the basis set all commute with each other; since the result is a completely tensorial equation, its validity will not depend upon any properties of a particular choice of basis, although the length of the proof will indeed be somewhat shorter as a result:

$$\begin{aligned} \tilde{u}[\mathcal{Q}(\tilde{v})] - \tilde{v}[\mathcal{Q}(\tilde{u})] - \mathcal{Q}([\tilde{u}, \tilde{v}]) &= u^\mu (\alpha_\nu v^\nu)_{,\mu} - v^\nu (\alpha_\mu u^\mu)_{,\nu} - \alpha_\nu u^\mu v^\nu_{,\mu} + \alpha_\mu v^\nu u^\mu_{,\nu} \\ &= u^\mu v^\nu \alpha_{\nu,\mu} - v^\nu u^\mu \alpha_{\mu,\nu} = u^\mu v^\nu \alpha_{[\nu,\mu]} \end{aligned}$$

However, we may now consider re-writing the left-hand side of our Eq. (4.3) as follows, using the basis of 1-forms, $\{\varpi^\nu\}$, which, again, we are taking to be holonomic, since we must be consistent on the two sides of the equation:

$$\begin{aligned} d\mathcal{Q}(\tilde{u}, \tilde{v}) &= [d\{\alpha_\nu \varpi^\nu\}](\tilde{u}, \tilde{v}) = [(\alpha_{\nu,\mu} \varpi^\mu \wedge \varpi^\nu)](\tilde{u}, \tilde{v}) = \alpha_{[\nu,\mu]} [\varpi^\mu \otimes \varpi^\nu](\tilde{u}, \tilde{v}) \\ &= \alpha_{[\nu,\mu]} \varpi^\mu(\tilde{u}) \varpi^\nu(\tilde{v}) = \alpha_{[\nu,\mu]} u^\mu v^\nu \quad . \end{aligned}$$

Both sides have now been brought to the same form; very important parts of the proof are the cancellation of two terms in the first set of equalities, and the skew symmetry imposed by the exterior derivative in the second set. We will be able to use this information to create simple and usable forms of the equations for both the torsion and the curvature tensors. A particularly useful application of this formula—as is often the case—is its application when all of the objects involved are (appropriate) basis vectors. Therefore, in particular, let us now re-write Eq. (4.4) for the case when we choose $\mathcal{Q} \rightarrow \varpi^\lambda$, $\tilde{u} \rightarrow \tilde{e}_\mu$, $\tilde{v} \rightarrow \tilde{e}_\nu$:

$$(4.5) \quad \begin{aligned} d\varpi^\lambda(\tilde{e}_\mu, \tilde{e}_\nu) &= -\varpi^\lambda([\tilde{e}_\mu, \tilde{e}_\nu]) = -C_{\mu\nu}{}^\lambda \\ \implies d\varpi^\lambda &= -\frac{1}{2} C_{\mu\nu}{}^\lambda \varpi^\mu \wedge \varpi^\nu, \end{aligned}$$

where the last line gives us an explicitly-useful **expression for the exterior derivative of an arbitrary basis 1-form**, which was mentioned but not proved at Eqs. (6.4) in the (previous) handout, on manifolds, tangent vectors, and differential forms.

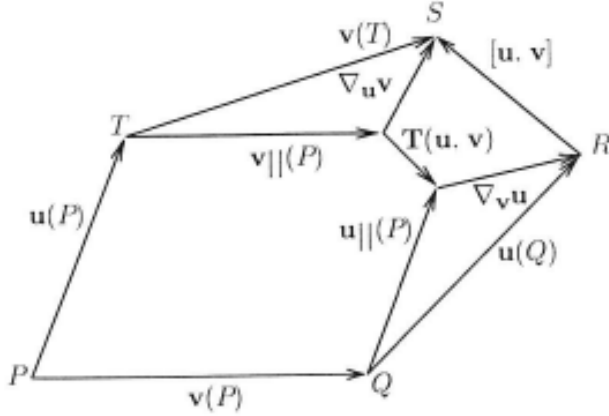
3. The Torsion Tensor, \mathbf{T}

Following the analogy of the differences of our curves discussed above, we now define a related quantity that is equipped to deal with both functions and vectors, by virtue of involving the covariant derivative in its formulation. The **torsion tensor** is defined as the following operator, on two tangent vectors, giving a result which is a tangent vector, i.e., it is an element of $\mathcal{T} \otimes \Lambda^2$, or, if you prefer, a rank $\left\{ \begin{smallmatrix} 1 \\ 2 \end{smallmatrix} \right\}$ tensor:

$$\mathbf{T}(\tilde{u}, \tilde{v}) \equiv (\nabla_{\tilde{u}}\tilde{v} - \nabla_{\tilde{v}}\tilde{u}) - [\tilde{u}, \tilde{v}] \quad . \quad (4.6)$$

It is worth referring back to the figure between Eqs. (4.0) and (4.1), that is drawn under the assumption that the torsion is zero, as is presumably appropriate at that level of the discussion. One can see the geometry allowing you to follow a path all the way around to where it began, by following the right-hand side of Eq. (4.6), with the left-hand side, therefore, as zero. To get a better view of the geometric role of the torsion when it is not zero, consider the re-drawing of that figure, with a non-zero value of the torsion, as presented just below, remembering that the \parallel subscript on a vector is not, necessarily, its true value at the point drawn, but rather the

value it would have were it to have been parallelly-propagated to that point. One can see that whereas the commutator $[\tilde{u}, \tilde{v}]$ closes up the figure when one has truly drawn \tilde{u} and \tilde{v} at the points propagated along the curves correctly, instead the torsion—more precisely, $\mathbf{T}(\tilde{u}, \tilde{v})$ —closes up the figure when one draws only those portions of \tilde{v} and \tilde{u} that have been parallelly propagated to those new points!



To show that the torsion is actually a tensor of rank $\left\{ \begin{smallmatrix} 1 \\ 2 \end{smallmatrix} \right\}$, one needs to show that it just lets (scalar) functions pass through. Since this is not true of covariant derivatives, it is not obvious that it would be true for the torsion; however, the proof follows from our determination of what happens to functions when followed around a curve closed up by the commutator of its tangent vectors. More precisely, that calculation allows one to show the following:

$$\begin{aligned} \mathbf{T}(\tilde{u}, f\tilde{v}) &= f\mathbf{T}(\tilde{u}, \tilde{v}) = \mathbf{T}(f\tilde{u}, \tilde{v}), \\ \text{or } \Delta f|_{P_0} &= \delta\lambda \delta\mu \mathbf{T}(\tilde{u}, \tilde{v})f|_P + \text{terms cubic in } \delta\lambda \text{ and/or } \delta\mu, \end{aligned} \tag{4.6a}$$

where Δf is defined as the difference between the two Taylor-series expansions of $f(P_3)$ determined by those two different routes. Therefore one might say that the torsion tensor measures the lack of “uniqueness” of Taylor series expansions. Because we do in fact believe that the Taylor series expansion for (smooth) functions gives unique answers, this gives us strong motivation for setting the Torsion tensor to zero, as is traditionally done in general relativity. Nonetheless, for right now, let us study the tensor a bit more.

Instead of applying the torsion to a function, if we simply write out the definition of $\mathbf{T}(\tilde{u}, \tilde{v})$, one may get a general formula for it in terms of the connection coefficients, Γ^j_{mn} and the commutation coefficients, C_{mn}^j :

$$\begin{aligned}\mathbf{T}(\tilde{u}, \tilde{v}) &= \tilde{e}_\mu \left\{ \tilde{u}(v^\mu) + v^\nu \mathfrak{L}^\mu_\nu(\tilde{u}) - \tilde{v}(u^\mu) - u^\nu \mathfrak{L}^\mu_\nu(\tilde{v}) - \tilde{u}(v^\mu) + \tilde{v}(u^\mu) - u^\lambda v^\nu C_{\lambda\nu}^\mu \right\} \\ &= \tilde{e}_\mu \left\{ v^\nu \mathfrak{L}^\mu_\nu(\tilde{u}) - u^\nu \mathfrak{L}^\mu_\nu(\tilde{v}) - u^\lambda v^\nu C_{\lambda\nu}^\mu \right\} \\ &= -u^\mu v^\nu \tilde{e}_\lambda \left\{ C_{\mu\nu}^\lambda - \Gamma^\lambda_{\nu\mu} + \Gamma^\lambda_{\mu\nu} \right\}.\end{aligned}\quad (4.6b)$$

However, the action of the tensor \mathbf{T} on the two arbitrary vectors could also have just been written out explicitly, in terms of the components of the [1,2]-tensor; therefore we may conclude from the above that

$$\mathbf{T}(\tilde{u}, \tilde{v}) \equiv u^\mu v^\nu \tilde{e}_\lambda T^\lambda_{\mu\nu} \implies T^\lambda_{\mu\nu} + C_{\mu\nu}^\lambda + \Gamma^\lambda_{[\mu\nu]} = 0, \quad (4.7)$$

providing a **simple relation between the torsion, the commutation coefficients, and the (skew part of the) components of the connection**. Or, differently phrased, we now have a way to determine the skew-symmetric part of the components of the affine connection, in terms of more physical quantities. One can also see that this also gives us a proof that \mathbf{T} is indeed a tensor, although one surely could have given a much more direct proof.

There is still one more (**very useful**) variant of the action of the torsion tensor which is often referred to as Cartan's First Structure Equations. To see how it comes about, we return to Eq. (4.6) for \mathbf{T} and expand that definition as follows:

$$\mathbf{T}(\tilde{u}, \tilde{v}) = \tilde{e}_\nu \left\{ \tilde{u}(v^\nu) + v^\mu \mathfrak{L}^\nu_\mu(\tilde{u}) - \tilde{v}(u^\mu) - u^\mu \mathfrak{L}^\nu_\mu(\tilde{v}) - \varpi^\nu([\tilde{u}, \tilde{v}]) \right\}.$$

As we can always write $v^\nu = \varpi^\nu(\tilde{v})$, we can re-write the quantity in the large brace above, i.e., the coefficient of \tilde{e}_ν , as

$$\tilde{u}[\varpi^\nu(\tilde{v})] - \tilde{v}[\varpi^\nu(\tilde{u})] - \varpi^\nu([\tilde{u}, \tilde{v}]) + \mathfrak{L}^\nu_\mu(\tilde{u}) \varpi^\nu(\tilde{v}) - \mathfrak{L}^\nu_\mu(\tilde{v}) \varpi^\nu(\tilde{u}).$$

However, in the first three terms of the above, we recognize the right-hand side of the identity concerning exterior derivatives of 1-forms, given in Eq. (4.4), for the case that the 1-form ϱ , there, is chosen as a basis 1-form, ϖ^ν , so that we can replace those three terms by simply $d\varpi^\nu(\tilde{u}, \tilde{v})$. As well, we see that the last two terms of our current equation are simply the same pair of 1-forms acting first on \tilde{u}, \tilde{v} , and then acting on \tilde{v}, \tilde{u} , and with a minus sign between them; i.e., they are skew-symmetric in their action on this pair of tangent vectors, which is the hallmark of the action of a 2-form on a pair of vectors. Therefore, we may now put all this together as the following equations

$$\mathbf{T}(\tilde{u}, \tilde{v}) = \tilde{e}_\nu \left\{ d\varpi^\nu(\tilde{u}, \tilde{v}) + (\mathbb{F}^\nu{}_\mu \wedge \varpi^\mu)(\tilde{u}, \tilde{v}) \right\} = \tilde{e}_\nu \left\{ d\varpi^\nu + \mathbb{F}^\nu{}_\mu \wedge \varpi^\mu \right\}(\tilde{u}, \tilde{v}) .$$

Since this is true for arbitrary tangent vectors, \tilde{u}, \tilde{v} , we conclude that it must be true as an identity between tensors, of rank $\left\{ \frac{1}{2} \right\}$, still awaiting those vectors to act upon.

It is this relationship which is originally due to Cartan:

Cartan's First Structure Equations

$$\mathbf{T} = \tilde{e}_\lambda \otimes \left\{ d\varpi^\lambda - \varpi^\mu \wedge \mathbb{F}^\lambda{}_\mu \right\} \tag{4.8a}$$

$$\text{or } \mathbf{T} = \tilde{e}_\lambda \otimes \left\{ d\varpi^\lambda - \Gamma^\lambda{}_{\mu\nu} \varpi^\mu \wedge \varpi^\nu \right\} = \tilde{e}_\lambda \otimes \left\{ d\varpi^\lambda - \frac{1}{2} \Gamma^\lambda{}_{[\mu\nu]} \varpi^\mu \wedge \varpi^\nu \right\} .$$

Inserting the form for $d\varpi^\lambda$ from Eqs. (4.5), we may re-write the torsion tensor in yet one more way, also somewhat useful:

$$\mathbf{T} = -\frac{1}{2} \tilde{e}_\lambda \otimes \left\{ C_{\mu\nu}{}^\lambda + \Gamma^\lambda{}_{[\mu\nu]} \right\} \varpi^\mu \wedge \varpi^\nu . \tag{4.8b}$$

Noting that there are theories of gravity other than general relativity that maintain non-zero the torsion tensor, we will nonetheless from now on continue with our study of general relativity itself, and set this tensor identically to zero, based on our beliefs concerning how functions should transform from point to point around closed paths on an arbitrary smooth manifold. This allows us to use Cartan's First Structure Equations as a direct statement of

the relationship between the skew portion of the affine connection and the (already skew) quantities given by the lack of commutativity of a non-holonomic basis set:

from now on the torsion is set to zero, which implies the following:

$$\begin{aligned} d\varpi^\lambda &= \varpi^\mu \wedge \Gamma^\lambda{}_\mu = -\Gamma^\lambda{}_\mu \wedge \varpi^\mu, & \Gamma^\lambda{}_{[\mu\nu]} &= -C_{\mu\nu}{}^\lambda, & (4.9) \\ d\alpha &= d(\alpha_\mu \varpi^\mu) = (d\alpha_\mu - \alpha_\lambda \Gamma^\lambda{}_\mu) \wedge \varpi^\mu = \alpha_{\mu;\nu} \varpi^\nu \wedge \varpi^\mu. \end{aligned}$$

The first line of these implications will in fact give us the “standard” method which we will use to calculate the affine connections, while the second line tells us that the exterior derivative of a p-form already includes the (skew portion of) covariant derivative, showing another of the useful facts about the use of the exterior derivative.

4. The Curvature of an Affine Connection

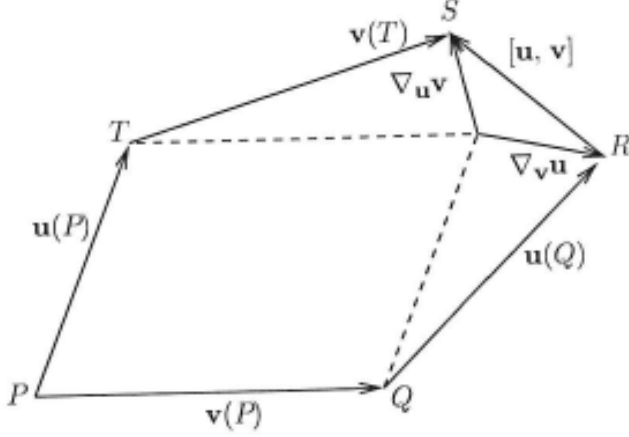
The commutator of two tangent vectors gives us enough geometrical information to “close-up” an area bounded by two members of two congruences of curves, therefore creating a “closed loop” as the *boundary of that area*; the torsion measures what happens to the values of a function taken around a closed loop. However, *the curvature* tells us what happens to the values of a vector as it is taken around a closed loop.

As an operator, the (affine) curvature takes the following form, requiring three distinct vectors given to it, and returning one vector back:

$$\mathbf{R} : \mathcal{T} \otimes \mathcal{T} \otimes \mathcal{T} \longrightarrow \mathcal{T}, \quad \mathbf{R}(\tilde{u}, \tilde{v}) \equiv \nabla_{\tilde{u}} \nabla_{\tilde{v}} - \nabla_{\tilde{v}} \nabla_{\tilde{u}} - \nabla_{[\tilde{u}, \tilde{v}]} \quad , \quad (4.10)$$

where the actual formulation above shows us that it is some sort of commutator of second covariant derivatives, waiting to be given its third vector, so that it can determine how that has changed when acted on in this way. We can see that it is an obvious extension, to act on vectors, of the torsion operator. On the other hand, also like the torsion tensor, as it turns out it is a tensor, multilinear in all of its arguments, where, again, this is not particularly obvious at first glance. From that point of view, let us work through the definition above in some detail, and determine how it one should view this action of the propagation of a vector

around a closed curve. We use the following figure and parallel propagate a vector, \tilde{w} around these two different paths to end at the same place, and of course eventually envision the entire process infinitesimally, since we are using Taylor series to propagate everything:



We begin by writing the Taylor series to evaluate \tilde{w} propagated from P to the point T , along the curve with tangent vector \tilde{u} , and then along to S via \tilde{v} :

$$\begin{aligned}\tilde{w}|_T &= \tilde{w}|_P + \delta\lambda (\nabla_{\tilde{u}}\tilde{w})|_P + \dots, \\ \tilde{w}|_S &= \tilde{w}|_T + \delta\eta (\nabla_{\tilde{v}}\tilde{w})|_T + \dots \\ &= \tilde{w}|_P + \delta\lambda (\nabla_{\tilde{u}}\tilde{w})|_P + \delta\eta (\nabla_{\tilde{v}}\tilde{w})|_T + \dots\end{aligned}\tag{4.11a}$$

Next, we want to propagate the vector along a different route to S , going from P to Q , to R , and eventually to S :

$$\begin{aligned}\tilde{w}|_Q &= \tilde{w}|_P + \delta\eta (\nabla_{\tilde{v}}\tilde{w})|_P + \dots, \\ \tilde{w}|_R &= \tilde{w}|_Q + \delta\lambda (\nabla_{\tilde{u}}\tilde{w})|_Q + \dots \\ &= \tilde{w}|_P + \delta\eta (\nabla_{\tilde{v}}\tilde{w})|_P + \delta\lambda (\nabla_{\tilde{u}}\tilde{w})|_Q + \dots, \\ \tilde{w}|_S &= \tilde{w}|_R + \delta\lambda\delta\eta (\nabla_{[\tilde{u}, \tilde{v}]} \tilde{w})|_R + \dots.\end{aligned}\tag{4.11b}$$

We may now determine the difference between propagating to S , via first \tilde{u} and then \tilde{v} , and propagating to S , via first \tilde{v} and then \tilde{u} , and lastly $[\tilde{u}, \tilde{v}]$:

$$\tilde{w}|_S \text{ via } T - \tilde{w}|_S \text{ via } Q \text{ and } R = -\delta\lambda\{(\nabla_{\tilde{u}}\tilde{w})|_Q - (\nabla_{\tilde{u}}\tilde{w})|_P\} + \delta\eta\{(\nabla_{\tilde{v}}\tilde{w})|_T - (\nabla_{\tilde{v}}\tilde{w})|_P\} - \delta\lambda\delta\eta (\nabla_{[\tilde{u}, \tilde{v}]} \tilde{w})|_R.\tag{4.11c}$$

However, using the first line of Eq. (4.11b) as a guide—since it “moves” objects from P to Q , as needed here—we can re-write the first expression above as

$$-\delta\lambda\{(\nabla_{\tilde{u}}\tilde{w})|_Q - (\nabla_{\tilde{u}}\tilde{w})|_P\} = -\delta\lambda\delta\eta(\nabla_{\tilde{v}}(\nabla_{\tilde{u}}\tilde{w})|_P)|_P . \quad (4.11d)$$

Next we can re-write the second expression, in Eq. (4.11c), where we need to move an object from P to T , using the first line of Eq. (4.11a) as a guide:

$$\delta\eta\{(\nabla_{\tilde{v}}\tilde{w})|_T - (\nabla_{\tilde{v}}\tilde{w})|_P\} = \delta\eta\delta\lambda(\nabla_{\tilde{u}}(\nabla_{\tilde{v}}\tilde{w})|_P)|_P . \quad (4.11e)$$

At this point, perhaps it is best to think about what we’re doing relative to the final result desired. As we can see all the terms so far involve $\delta\eta\delta\lambda$ plus higher-order terms, none of which have yet been written explicitly. That is reasonable because this is all supposed to be happening in some small neighborhood; therefore, the idea is to divide the final result by $\delta\lambda$ and also by $\delta\eta$ —which are the measures of the size of the area of this pseudo-parallelogram we are considering—and then to take the limit as these measures go to zero. Taking this limit will remove all terms of higher order, and take the area to zero as we desire. Therefore, when we now consider the last term in Eq. (4.11c), which is already of that order, even though it is evaluated at R , we can suppose a Taylor series for it, starting back at P , and then adding additional, higher-order terms to move it along, from Q to R . All those higher-order terms will vanish in this limit, so that it is just as if it were evaluated all the way back at P , which we now suppose. The final result is then, to this order, the following:

$$\Delta\tilde{w} \equiv \tilde{w}|_{S \text{ via } T} - \tilde{w}|_{S \text{ via } Q \text{ and } R} = \delta\lambda\delta\eta\{\nabla_{\tilde{u}}\nabla_{\tilde{v}} - \nabla_{\tilde{v}}\nabla_{\tilde{u}} - \nabla_{[\tilde{u},\tilde{v}]}\}\tilde{w}|_P = \delta\lambda\delta\eta R(\tilde{u},\tilde{v})\tilde{w} , \quad (4.11f)$$

where we have not bothered to write that there are still higher-order terms. However, as already discussed, that is not really a problem since the limit of the left-hand side divided by $\delta\lambda$ and by $\delta\eta$ is just the commutator of the two covariant derivatives acting on \tilde{w} , which gives me the exact result:

$$\{[\nabla_{\tilde{u}}, \nabla_{\tilde{v}}] - \nabla_{[\tilde{u},\tilde{v}]}\}\tilde{w} = R(\tilde{u},\tilde{v})\tilde{w} , \quad (4.12)$$

which is only a re-statement of Eq. (4.10), acting on a vector field; however, the process has shown us its meaning. It is also relevant to think about this result in a slightly different way. We determined the difference in value of \tilde{w} when taken from an initial point to a final point by two different routes. However, one could just as easily considered this as a propagation of \tilde{w} around a closed curve, which amounts to following the second curve backwards after the first one forwards, i.e., by starting at P , moving to T , then onward to S , and next to R , and next to Q , and finally back to P . One would determine exactly the same difference because, after all, we subtracted the second route's value from the first, which is the same as adding that route's value to the first when followed in the negative direction.

The next thing that should be shown is that even though this quantity involves several derivatives, i.e., derivation operators, it is actually a tensor. One could suppose that this is true by remembering our earlier discussions about dragging functions around a closed curve like this. Nonetheless, in principle one should use this same approach for a calculation of the effect on, say $f\tilde{w}$, where f is a locally-defined function. We recall that

$$\nabla_{\tilde{v}}f\tilde{w} = f\nabla_{\tilde{v}} + \{\tilde{v}(f)\}\tilde{w} .$$

Therefore we can write out the following 3 lines for the 3 distinct parts of the curvature operator acting on $f\tilde{w}$:

$$\begin{aligned} \nabla_{\tilde{u}}\nabla_{\tilde{v}}(f\tilde{w}) &= f\nabla_{\tilde{u}}\nabla_{\tilde{v}}\tilde{w} + \{\tilde{u}(f)\}\nabla_{\tilde{v}}\tilde{w} + \{\tilde{v}(f)\}\nabla_{\tilde{u}}\tilde{w} + \{\tilde{u}(\tilde{v}(f))\}\tilde{w} , \\ \nabla_{\tilde{v}}\nabla_{\tilde{u}}(f\tilde{w}) &= f\nabla_{\tilde{v}}\nabla_{\tilde{u}}\tilde{w} + \{\tilde{v}(f)\}\nabla_{\tilde{u}}\tilde{w} + \{\tilde{u}(f)\}\nabla_{\tilde{v}}\tilde{w} + \{\tilde{v}(\tilde{u}(f))\}\tilde{w} , \\ \nabla_{[\tilde{u},\tilde{v}]}(f\tilde{w}) &= f\nabla_{[\tilde{u},\tilde{v}]} \tilde{w} + \{[\tilde{u},\tilde{v}]f\}\tilde{w} . \end{aligned} \tag{4.13}$$

Subtracting the last two lines from the first one we see that the 2nd term of the first line cancels the third term of the second line, as well as the third term of the first line cancels the second term of the second line. In addition the second term of the last line cancels the last terms of the first and second lines. All this cancellation leaves only the first terms, on the right-hand side, of each line which are neatly summarized by the desired result:

$$R(\tilde{u},\tilde{v})(f\tilde{w}) = f R(\tilde{u},\tilde{v})\tilde{w} . \tag{4.14}$$

This means that the object so defined is actually a tensor, of $\left\{\frac{1}{3}\right\}$, since it takes 3 vectors and gives a result which is still a vector. Of course, this means that it will transform like one, and that we may write it out in terms of components relative to the appropriate choices of basis vectors:

$$\begin{aligned}\mathbf{R} &= \frac{1}{2}R^\lambda{}_{\mu\rho\sigma}\omega^\mu \otimes (\omega^\rho \wedge \omega^\sigma) \otimes \tilde{e}_\lambda ; \\ \mathbf{R}(\tilde{u}, \tilde{v})\tilde{w} &= \frac{1}{2}R^\lambda{}_{\mu\rho\sigma}u^{[\rho}v^{\sigma]}w^\mu\tilde{e}_\lambda = R^\lambda{}_{\mu\rho\sigma}u^\rho v^\sigma w^\mu\tilde{e}_\lambda ; \\ [\nabla_\sigma, \nabla_\rho]w^\lambda &= w^\lambda{}_{;[\rho\sigma]} = R^\lambda{}_{\mu\rho\sigma}w^\mu \quad ,\end{aligned}\tag{4.14}$$

where the last line is a re-statement of Eq. (4.12) using our realization that this object which appears to be a differential operator actually only determines the components of a tensor.

It would now be useful to calculate explicitly the components of this curvature tensor; after all, we do know how to calculate the covariant derivative of a tangent vector field using the components of the affine connection. For the curvature tensor we must do this twice; therefore, it is to be expected that the (first) derivatives of those connection components will be involved. The calculation is quite lengthy, with the following results:

$$R^\lambda{}_{\mu\rho\sigma} = \Gamma^\lambda{}_{\mu[\sigma, \rho]} + \Gamma^\lambda{}_{\tau[\rho}\Gamma^\tau{}_{\mu\sigma]} - \Gamma^\lambda{}_{\mu\nu}C_{\rho\sigma}{}^\nu \quad .\tag{4.15}$$

Because we see results that are skew on pairs of indices, it strongly suggests that this should be re-written in terms of the connection 1-forms, rather than just their components. Cartan of course noticed this quite some time back, and so he formulated

Cartan's Second Structural Equations

which defines a collection of several independent 2-forms defined along the lines that Eq. (4.15) suggests. We define the collection of 2-forms, $\Omega^\lambda{}_\mu$, which mirror the way that Cartan set up the connection 1-forms:

$$\Omega^\lambda{}_\mu = d\tilde{\Gamma}^\lambda{}_\mu + \tilde{\Gamma}^\lambda{}_\rho \wedge \tilde{\Gamma}^\rho{}_\mu \quad .\tag{4.16}$$

These are then a way of looking at the last two indices of the curvature tensor as letting them define a 2-form. More explicitly we have the following relationships between the components

of our curvature tensor and Cartan's 2-forms, which makes it seem much more likely that this does define a tensor:

$$\begin{aligned} \mathbf{R} &\equiv \frac{1}{2} R^\lambda{}_{\mu\nu\eta} \tilde{e}_\lambda \otimes \varpi^\mu \otimes \varpi^\nu \wedge \varpi^\eta = \tilde{e}_\lambda \otimes \varpi^\mu \otimes \mathfrak{Q}^\lambda{}_\mu, \\ &\text{or } \mathfrak{Q}^\mu{}_\nu = \frac{1}{2} R^\mu{}_{\nu\lambda\eta} \varpi^\lambda \wedge \varpi^\eta. \end{aligned} \quad (4.17)$$

It is quite important to understand that there are quite a number of useful identities satisfied by the components of the curvature tensor. It is obvious that $R^a{}_{bcd}$ is skew-symmetric in its last two indices, which reduces the number of components that need to be determined. **When working with the assumption of zero torsion, and constant metric coefficients**, as is the usual case for us, one can see from Eqs. (5.9), in the next section, that the connection is skew in its first two indices—or, if you prefer, the connection 1-forms are skew-symmetric in their indices. Therefore, the equations just above make it clear that it would then be true that the curvature 2-forms are skew-symmetric in their indices, which is the same as saying that the entire curvature tensor is skew-symmetric in its first two indices as well as the second two. There are other symmetries available when there is a metric, and when the torsion is zero. We demonstrate some of those due to the vanishing of the torsion below, namely the (first) Bianchi identities, which comes from applying the exterior derivative to the First Structure Equations, as we will show below: To begin we recall that the action of the 1-form basis elements on a vector gives the components of that vector, i.e., if $\tilde{v} = v^\alpha \tilde{e}_\alpha$ then $\omega^\alpha(\tilde{v}) = v^\alpha$. Therefore, since the torsion tensor \mathbf{T} is an element of $\mathcal{T} \otimes \Lambda^2$, using Eq. (4.8a), we can write the vectorial components of the torsion as follows. Since the result is a 2-form, it is then appropriate to take its exterior derivative:

$$\begin{aligned} \omega^\alpha(\mathbf{T}) &= d\varpi^\alpha - \varpi^\beta \wedge \mathfrak{L}^\alpha{}_\beta = d\varpi^\alpha + \mathfrak{L}^\alpha{}_\beta \wedge \varpi^\beta, \\ \implies d\{\omega^\alpha(\mathbf{T})\} &= 0 + d\mathfrak{L}^\alpha{}_\beta \wedge \varpi^\beta - \mathfrak{L}^\alpha{}_\beta \wedge d\varpi^\beta = d\mathfrak{L}^\alpha{}_\beta \wedge \varpi^\beta + \mathfrak{L}^\alpha{}_\beta \wedge \mathfrak{L}^\beta{}_\rho \wedge \varpi^\rho \\ &= \{d\mathfrak{L}^\alpha{}_\beta + \mathfrak{L}^\alpha{}_\rho \wedge \mathfrak{L}^\rho{}_\beta\} \wedge \varpi^\beta = \mathfrak{Q}^\alpha{}_\beta \wedge \varpi^\beta, \end{aligned} \quad (4.18a)$$

where the second line comes by substituting for $d\varpi^\alpha$ in terms of the torsion once again, and, this time, remembering that we are assuming the torsion is zero, while the last line recalls the

definition of the two-forms $\Omega^\alpha{}_\beta$ from Eq. (4.16). However, Eqs. (4.17) gives us the relation between these 2-forms and the components of the Riemann tensor. When this is inserted here we have finally shown that the vanishing of the torsion is sufficient to guarantee

$$R^\mu{}_{\nu\lambda\eta}\varpi^\nu \wedge \varpi^\lambda \wedge \varpi^\eta = 0 = \frac{1}{3}\{R^\mu{}_{\nu\lambda\eta} + R^\mu{}_{\lambda\eta\nu} + R^\mu{}_{\eta\nu\lambda}\}\varpi^\nu \wedge \varpi^\lambda \wedge \varpi^\eta, \quad (4.18b)$$

where the last equality is written out to arrange the coefficients of our 3-form to have the same symmetry as the basis set, so that we can then set the coefficients themselves to zero. That result is then referred to as the

First Bianchi Identities :

$$R^\mu{}_{\nu\lambda\eta} + R^\mu{}_{\lambda\eta\nu} + R^\mu{}_{\eta\nu\lambda} = T^\mu{}_{\lambda\eta,\nu} + T^\mu{}_{\eta\nu,\lambda} + T^\mu{}_{\nu\lambda,\eta} \xrightarrow{\text{torsion-free}} 0, \quad (4.18c)$$

where in this “final” form I have resurrected the partial derivatives of the torsion that we have set equal to zero, simply so that you could see what they were—that were dropped in the middle of the calculation above since we were going to set them all to zero.

The component form of the first Bianchi identities is a linear relationship between different sets of components of the curvature tensor, minimizing the number of independent elements that need to be computed, or, if you prefer, minimizing the number of independent degrees of freedom that the curvature may have.

There is of course also a second set of such identities, which arise from the exterior derivative of the defining equation for the $\Omega^\lambda{}_\eta$ 2-forms, given in Eq. (4.16). We first differentiate that equation, and then back substitute for the exterior derivatives of the connection 1-forms via the 2-forms themselves:

$$\begin{aligned} d\Omega^\alpha{}_\beta &= dd\Gamma^\alpha{}_\beta + d\Gamma^\alpha{}_\rho \wedge \Gamma^\rho{}_\beta - \Gamma^\alpha{}_\rho \wedge d\Gamma^\rho{}_\beta \\ &= (\Omega^\alpha{}_\rho - \Gamma^\alpha{}_\sigma \wedge \Gamma^\sigma{}_\rho) \wedge \Gamma^\rho{}_\beta - (\Omega^\rho{}_\beta - \Gamma^\rho{}_\sigma \wedge \Gamma^\sigma{}_\beta) \wedge \Gamma^\alpha{}_\rho \\ &= \Omega^\alpha{}_\rho \wedge \Gamma^\rho{}_\beta - \Omega^\rho{}_\beta \wedge \Gamma^\alpha{}_\rho. \end{aligned} \quad (4.19a)$$

Moving those terms on the right over to the left, so that we have an expression involving the various 2-forms $\Omega^\alpha{}_\beta$ that vanishes, we see a rather unusual, and unexpected combination of

quantities. It seems as if it is a covariant derivative, treating the upper and lower indices as if they were tensor indices, even though they are only labels for the set of n^2 2-forms needed to describe the Riemann tensor. Cartan was “intrigued” by this and created what he called a “generalization” of the exterior derivative, d , denoted by D , which did indeed treat the labels of a set of p-forms as if they should be involved with the affine connections in this way, as you will see more explicitly in the equation just below for the second Bianchi identities:

Second Bianchi Identities :

$$D\mathcal{Q}^a{}_b \equiv d\mathcal{Q}^a{}_b + \mathcal{L}^a{}_c \wedge \mathcal{Q}^c{}_b - \mathcal{L}^e{}_b \wedge \mathcal{Q}^a{}_e = 0 , \quad (4.19b)$$

or in components $R^a{}_{bcd;e} + R^a{}_{bde;c} + R^a{}_{bec;d} = 0 . \quad (4.19c)$

This set of identities may also be looked at as a set of integrability conditions on the curvature tensor, which, when satisfied, allow integration backwards to determine the affine connection from which a given set of curvature components came.

This is also a good place to introduce the tensor with a great deal of physical significance— as we will learn later—namely *the Ricci tensor*, which is a (partial) trace of the Riemann tensor, i.e., a sum over all the allowed values of the upper index with one of the lower indices:

$$\mathcal{R}_{\nu\eta} \equiv R^\mu{}_{\nu\mu\eta} . \quad (4.20)$$

V. (Pseudo)-Riemannian Geometry: Introduction of a Metric Tensor, \mathbf{g} ; Its Relation to the Affine Connection

1. Functions and Properties of a metric tensor, possibly indefinite

The metric is originally introduced as a (real-valued) measure for scalar products of tangent vectors with themselves; this is quite analogous to the metric, $\boldsymbol{\eta}_{\mu\nu}$ used in special relativity.

This tensor should therefore have rank $\left\{ \begin{smallmatrix} 0 \\ 2 \end{smallmatrix} \right\}$, i.e., it should be an element of the space $\Lambda \otimes \Lambda$, and should be symmetric in its effect on these two arguments:

$$\mathbf{g} \in \Lambda^1 \otimes \Lambda^1 \text{ or } \mathbf{g} : \mathcal{T} \otimes \mathcal{T} \longrightarrow \mathcal{F} \equiv \Lambda^0 : \quad \text{i.e., } \mathbf{g} = g_{\mu\nu} \varpi^\mu \otimes \varpi^\nu , \quad (5.1)$$

$$\text{and the definitions } g_{\mu\nu} = \mathbf{g}(\tilde{e}_\mu, \tilde{e}_\nu) \implies \mathbf{g}(\tilde{u}, \tilde{v}) = g_{\mu\nu} u^\mu v^\nu . \quad (5.2)$$

It is very useful to often view the set of components $g_{\mu\nu}$ as constituting the elements of a (symmetric) matrix, $G \equiv ((g_{\mu\nu}))$, that of course describes the metric relative to the (previously-chosen) basis of 1-forms.

In almost all cases, it is desirable that this matrix be invertible, so that the matrix G^{-1} exists; it is customary to denote its elements by the symbol $g^{\mu\nu}$; i.e., $G^{-1} = ((g^{\mu\nu}))$, and to treat it as the components of a tensor of rank $\left\{ \begin{smallmatrix} 2 \\ 0 \end{smallmatrix} \right\}$, i.e., an element of $\mathcal{T} \otimes \mathcal{T}$, which can generate scalar products of 1-forms.

$$(G^{-1}G)^\mu{}_\lambda = g^{\mu\nu} g_{\lambda\nu} = \delta_\lambda^\mu = g_{\lambda\nu} g^{\nu\mu} = (GG^{-1})_\lambda{}^\mu , \quad (5.3)$$

$$\mathbf{g}^{-1} \in \mathcal{T} \otimes \mathcal{T} \text{ or } \mathbf{g}^{-1} : \Lambda^1 \otimes \Lambda^1 \longrightarrow \mathcal{F} \equiv \Lambda^0 \text{ so that } \mathbf{g}^{-1}(\alpha, \beta) \equiv g^{\mu\nu} \alpha_\mu \beta_\nu .$$

The existence of an invertible metric induces various other important mappings, which map tangent vectors into 1-forms, and vice versa. These mappings are often referred to as “raising” and “lowering” of indices; thus, the existence of an invertible metric “obscures” the differences between the two kinds of vectors that we use, namely tangent vectors and 1-forms. We describe the most fundamental of these induced mappings below; where we take $\tilde{u} = u^\mu \tilde{e}_\mu$, $\tilde{v} = v^\lambda \tilde{e}_\lambda \in \mathcal{T}$ and $\alpha = \alpha_\nu \varpi^\nu$, $\beta = \beta_\rho \varpi^\rho \in \Lambda^1$ as arbitrary tangent vectors, and 1-forms, respectively.

$$\mathbf{g}_* : \mathcal{T} \longrightarrow \Lambda^1 \text{ so that } \{\mathbf{g}_*(\tilde{u})\}(\tilde{v}) \equiv \mathbf{g}(\tilde{u}, \tilde{v}) , \text{ or } \mathbf{g}_*(\tilde{u}) = (g_{\mu\nu} u^\nu) \varpi^\mu \equiv u_\nu \varpi^\nu \in \Lambda^1 , \quad (5.4)$$

$$\mathbf{g}^* : \Lambda^1 \longrightarrow \mathcal{T} \text{ so that } \beta\{\mathbf{g}^*(\alpha)\} \equiv \mathbf{g}(\beta, \alpha) , \text{ or } \mathbf{g}^*(\alpha) = (g^{\mu\nu} \alpha_\mu) \tilde{e}_\nu \equiv \alpha^\nu \tilde{e}_\nu \in \mathcal{T} .$$

This is then the general process of “raising” and “lowering” indices on a vector and on a 1-form. The upper- and/or lower-stars on g just remind us that we are looking at the metric tensor as a mapping different from its original intention, as a metric.

It should be clear that the process can easily be extended to operate on any other sorts of tensors one might desire. In particular, because we use the symbols $g^{\mu\nu}$ for the components of G^{-1} , it turns out that the components that one might think of as “the metric with one index up and one index down” are just the components of the identity matrix, i.e., $g^{\mu\nu} g_{\nu\lambda} = \delta_{\lambda}^{\mu}$.

2. Determination of the Affine Connection

Having introduced a metric into our system, we may now use the affine connection that we already have to ask how that metric varies as we proceed from one vector space to some other nearby one, i.e., we should consider the rank $\left\{ \begin{smallmatrix} 0 \\ 3 \end{smallmatrix} \right\}$ tensor, $\nabla \mathbf{g}$. In the standard version of Einstein’s theory of general relativity, one assumes that physical reasons cause us to require this tensor to be zero. For the moment, I will go ahead and simply give this extra tensor a name, and see how it would enter into our calculations, being aware that there are in fact alternative theories of gravity where it is presumed to be an interesting physical quantity. As before, since this is a minor variation, I will put this section in small print, and label the few important results on the left.

We define the set of 1-forms, referred to as the “**metrizability coefficients**”, $\underset{\sim}{Q}^{\mu\nu}$:

$$\nabla g^{\mu\nu} \equiv \underset{\sim}{Q}^{\mu\nu} = Q_{\lambda}^{\mu\nu} \underset{\sim}{\omega}^{\lambda}, \implies \nabla g_{\mu\nu} = -\underset{\sim}{Q}_{\mu\nu}, \quad (5.5)$$

$$\implies g_{\mu\nu ;\lambda} = \partial_{\lambda} g_{\mu\nu} - \Gamma^{\eta}_{\mu\lambda} g_{\eta\nu} - \Gamma^{\eta}_{\nu\lambda} g_{\mu\eta} = -Q_{\lambda\mu\nu} \quad (5.6a)$$

$$\text{or } g_{\mu\nu ,\lambda} = \Gamma_{\nu\mu\lambda} + \Gamma_{\mu\nu\lambda} - Q_{\lambda\mu\nu} = \Gamma_{(\nu\mu)\lambda} - Q_{\lambda\mu\nu} \quad (5.6b)$$

They are symmetric on their indices, since $g_{\mu\nu}$ is symmetric on its indices, and the other relations follow from the usual (inverse) relation for the metric, $g^{\mu\nu} g_{\nu\lambda} = \delta_{\lambda}^{\mu}$, and the expression for the covariant derivative in terms of the components of the (affine) connection.

This finally defines all the necessary geometrical quantities needed to determine an equation that gives the components of the connection in terms of physically more relevant quantities; i.e., we may now contemplate justifying our “choice” for the connection 1-forms, $\Gamma^{\lambda}_{\mu\nu}$. They can be determined in terms of

- (1) the ordinary derivatives of the components of the metric, i.e., $g_{\mu\nu ,\lambda}$,

- (2) the metrizable coefficients, $Q_{\lambda\mu\nu}$,
- (3) the torsion coefficients $T^\lambda_{\mu\nu}$, and
- (4) the commutativity coefficients, $C_{\mu\nu}{}^\lambda$.

Since these various objects come, *a priori*, with indices in different sorts of places, it is most useful to first use the metric to “raise and/or lower” indices so that we have all of them on the same level. We now suppose that this has been done—with the ordering of the indices **definitely unchanged** by this process. The algebraic process for solving for the connection coefficients is begun by re-writing our last equation, Eq. (5.6), three times, one under the other, each time permuting the names of the indices (cyclicly), and multiplying the last copy by -1:

$$\begin{aligned} g_{\mu\nu, \lambda} &= \Gamma_{\nu\mu\lambda} + \Gamma_{\mu\nu\lambda} - Q_{\lambda\mu\nu} \\ g_{\nu\lambda, \mu} &= \Gamma_{\lambda\nu\mu} + \Gamma_{\nu\lambda\mu} - Q_{\mu\nu\lambda} \\ -g_{\lambda\mu, \nu} &= -\Gamma_{\mu\lambda\nu} - \Gamma_{\lambda\mu\nu} + Q_{\nu\lambda\mu} \end{aligned}$$

Adding these three equations gives the rather lengthy, but quite useful, equation

$$\begin{aligned} g_{\mu\nu, \lambda} + g_{\nu\lambda, \mu} - g_{\lambda\mu, \nu} \\ &= \Gamma_{\nu(\mu\lambda)} + \Gamma_{\mu[\nu\lambda]} + \Gamma_{\lambda[\nu\mu]} + Q_{\nu\lambda\mu} - Q_{\mu\nu\lambda} - Q_{\lambda\mu\nu} \\ &= 2\Gamma_{\nu\mu\lambda} - \Gamma_{\nu[\mu\lambda]} + \Gamma_{\mu[\nu\lambda]} + \Gamma_{\lambda[\nu\mu]} + Q_{\nu\lambda\mu} - Q_{\mu\nu\lambda} - Q_{\lambda\mu\nu} . \end{aligned} \tag{5.7}$$

This equation can be solved for the desired connection coefficients, namely $\Gamma_{\nu\mu\lambda}$, provided we know their skew part. However, one form of Cartan’s first structure equations, Eqs. (4.8), gives that part of the connection coefficients in terms of the torsion and the commutativity coefficients. It is sufficiently important that I now re-write it here, with all indices lowered:

$$2\Gamma_{\lambda[\mu\nu]} = -T_{\lambda\mu\nu} - C_{\mu\nu\lambda} . \tag{5.8}$$

At this point the algebra in question is clearly straight-forward, if lengthy, and gives the following result. Various special cases of this will be used often:

$$\begin{aligned} \Gamma_{\mu\nu\lambda} &= \frac{1}{2}(-g_{\nu\lambda, \mu} + g_{\mu\nu, \lambda} + g_{\mu\lambda, \nu}) \\ &+ \frac{1}{2}(-Q_{\mu\nu\lambda} + Q_{\lambda\mu\nu} + Q_{\nu\mu\lambda}) \\ &+ \frac{1}{2}(-T_{\mu\nu\lambda} + T_{\lambda\mu\nu} + T_{\nu\mu\lambda}) \\ &+ \frac{1}{2}(-C_{\nu\lambda\mu} + C_{\mu\nu\lambda} + C_{\mu\lambda\nu}) , \end{aligned} \tag{5.9}$$

where the four distinct entries in the equation come from the four distinct sorts of geometric contributions already mentioned, above. Of course, once we have the connection coefficients, they can be used to create the connection 1-forms, $\mathfrak{L}^\lambda{}_\mu$, and, from there, the curvature 2-forms, $\mathcal{Q}^\lambda{}_\mu$, or the (equivalent) Riemann curvature tensor components, $R^\lambda{}_{\mu\rho\sigma}$, from Eqs. (4.11). Our “choice” for the connection coefficients, which connect quantities in “adjoining” vector spaces, has now been rephrased in terms of more physical quantities.

3. the Levi-Civita Connection, the one we will be using for our studies

It is unlikely that one would ever want to use all parts of the general formula for an affine connection, given in Eqs. (5.9). Nonetheless, one may use different portions of it in different places.

We will now concentrate **only** on the version that corresponds to Einstein’s “official” theory of general relativity, which uses a **metric-compatible, torsion-free connection**; i.e., Einstein’s general relativity assumes explicitly that the metrization coefficients and the torsion components are exactly zero!

We have already discussed why it is plausible, at least, to set the torsion tensor to zero. At this point, let me note two very important things that happen, to our notation. The first is that **when we set the torsion tensor to zero**, as already noted in Eqs. (4.10), we acquire quite useful content from **Cartan’s First Structure Equations**. They now tell us explicitly a relationship between the components of the affine connection and the commutation coefficients that describe the lack of commutativity of the (non-holonomic) basis we are using for tangent vectors:

$$\begin{aligned} d\varpi^\lambda &= \varpi^\mu \wedge \mathfrak{L}^\lambda{}_\mu = \Gamma^\lambda{}_{\mu\nu} \varpi^\mu \wedge \varpi^\nu = \frac{1}{2} \Gamma^\lambda{}_{[\mu\nu]} \varpi^\mu \wedge \varpi^\nu \\ &\implies \Gamma^\lambda{}_{[\mu\nu]} = -C_{\mu\nu}{}^\lambda . \end{aligned} \tag{5.10 = 4.9}$$

There is however an additional, very pleasant thing that happens to the notation when the torsion vanishes, since we now have an explicit relation between the commutation coefficients and the skew-symmetric part of the connection 1-forms. We consider the exterior derivative of

a 1-form, and also a 2-form, in an arbitrary basis (while an arbitrary p-form follows in exactly the analogous way):

$$\begin{aligned}
d\alpha &= (d\alpha_\mu) \wedge \varpi^\mu + \alpha_\mu d\varpi^\mu = (d\alpha_\nu - \alpha_\mu \Gamma^\mu{}_\nu) \wedge \varpi^\nu \\
&= (\alpha_{\nu;\lambda} - \alpha_\mu \Gamma^\mu{}_{\nu\lambda}) \varpi^\lambda \wedge \varpi^\nu = (\alpha_{\nu;\lambda}) \varpi^\lambda \wedge \varpi^\nu ; \\
d\beta &= (d\beta_{\mu\nu} \wedge \varpi^\mu \wedge \varpi^\nu + \beta_{\mu\nu} d\varpi^\mu \wedge \varpi^\nu - \beta_{\mu\nu} \varpi^\mu \wedge d\varpi^\nu) = (d\beta_{\lambda\eta} - \beta_{\mu\eta} \Gamma^\mu{}_\lambda - \beta_{\lambda\nu} \Gamma^\nu{}_\eta) \wedge \varpi^\lambda \wedge \varpi^\eta \\
&= (\beta_{\lambda\eta\sigma} - \beta_{\mu\eta} \Gamma^\mu{}_{\lambda\sigma} - \beta_{\lambda\nu} \Gamma^\nu{}_{\eta\sigma}) \varpi^\sigma \wedge \varpi^\lambda \wedge \varpi^\eta = (\beta_{\lambda\eta;\sigma}) \varpi^\sigma \wedge \varpi^\lambda \wedge \varpi^\eta .
\end{aligned} \tag{5.11}$$

What this says is that

if we insert the components of the covariant derivative into an exterior derivative, of a p-form, then this automatically includes effects from the exterior derivatives of the basis forms.

We now want to note the “(physical) advantages” to setting the metrizable coefficients to zero. (The usual “language” for such a process is referred to as insisting that the connection be “metric compatible.”) Metric compatibility has the great advantage that we don’t have to worry particularly when raising and lowering indices, or when calculating scalar products. If, contrariwise, the connection were not metric compatible, then we would have the following very strange behavior for an “ordinary” scalar product as it was moved from one vector space to a nearby one, using the usual product rule:

$$\nabla_{\tilde{w}} \{ \mathbf{g}(\tilde{u}, \tilde{v}) \} = \{ \nabla_{\tilde{w}} \mathbf{g} \}(\tilde{u}, \tilde{v}) + \mathbf{g}(\nabla_{\tilde{w}} \tilde{u}, \tilde{v}) + \mathbf{g}(\tilde{u}, \nabla_{\tilde{w}} \tilde{v}) \quad . \tag{5.5'}$$

The first term on the right-hand side of the equation is just the metrizable tensor; by setting it equal to zero, we arrange our theory so that the “scalar product,” i.e., the metric, just commutes through the covariant derivative operation, so that it then seems to be the same—in functional form—at every point on our manifold, thereby making the physical meaning of the scalar product much clearer!

This particular choice of affine connection was actually first made by Levi-Civita, the person who also first invented “tensor calculus,” back in the latter parts of the 19th century;

therefore it is usually referred to as **the Levi-Civita connection**. From now on, we will not worry further about any other connection. However, we need to consider, in some little detail, how to actually calculate the Levi-Civita connection for a given choice of (1) a metric, and (2) a basis of 1-forms (or of tangent vectors). Therefore we now re-write the very general equation for the affine connection, given as Eqs. (5.9), but now specialized for the Levi-Civita connection, and with the use of the metric to lower the contravariant index, so that we can look at both indices—both covariant—on the same footing:

the Levi-Civita Connection

$$\begin{aligned} \Gamma_{\mu\nu} \equiv g_{\mu\eta} \Gamma^{\eta}_{\nu} = \Gamma_{\mu\nu\lambda} \omega^{\lambda} = & \left\{ \frac{1}{2}(-g_{\nu\lambda, \mu} + g_{\mu\nu, \lambda} + g_{\mu\lambda, \nu}) \right. \\ & \left. + \frac{1}{2}(C_{\lambda\nu\mu} + C_{\mu\nu\lambda} + C_{\mu\lambda\nu}) \right\} \omega^{\lambda}, \end{aligned} \quad (5.10LC)$$

Also note that the triplet of terms involving derivatives of the metric is obviously symmetric under the interchange of the indices ν and λ , and therefore does **not contribute** to that part of the connection that is skew-symmetric on those indices. As well, $C_{\lambda\nu\mu}$ is skew-symmetric on those indices, so that this equation is completely consistent—as it surely must be—with Eqs. (5.10), as well as Eqs. (5.8) and Eqs. (4.9b).

The first triplet of terms, involving the partial derivatives of the components of the metric, comes with its own name, the **Christoffel symbol**; it is said to be of the *first kind* when all of its indices are lowered, and of the *second kind* when the “first” index is raised. The following notation is very common, although not quite universally used, for the two kinds of Christoffel symbols, where we also present the form for the Riemann tensor, taken from Eqs. (4.12), remembering that the commutation coefficients are zero for a holonomic basis:

$$\begin{aligned} [\mu; \nu\lambda] & \equiv \frac{1}{2}(-g_{\nu\lambda, \mu} + g_{\mu\nu, \lambda} + g_{\mu\lambda, \nu}) \\ \left\{ \begin{matrix} \mu \\ \nu\lambda \end{matrix} \right\} & \equiv g^{\mu\eta}[\eta; \nu\lambda] = \frac{1}{2} g^{\mu\eta}(-g_{\nu\lambda, \eta} + g_{\eta\nu, \lambda} + g_{\eta\lambda, \nu}), \\ R^{\mu}_{\nu\sigma\tau} & = \left\{ \begin{matrix} \mu \\ \nu[\tau] \end{matrix} \right\}_{, \sigma} + \left\{ \begin{matrix} \mu \\ \lambda[\sigma] \end{matrix} \right\} \left\{ \begin{matrix} \lambda \\ \nu\tau \end{matrix} \right\}. \end{aligned} \quad (5.10Cr)$$

By its definition, one easily notes that the Christöffel symbol is symmetric in its second pair of indices, implying of course similar symmetry for that part of the connection coefficients that are created from this contribution. Secondly, we recall that if one makes the choice to use a holonomic basis set for our vector spaces—one where the basis vectors for tangent vectors are just the ordinary partial derivatives of the coordinates—then the commutation coefficients $C_{\mu\nu\lambda}$ would vanish, and this last choice, of vector basis, would have now completely determined the connection as simply that given by the Christöffel symbols. This approach was once used in all books by physicists on this subject. It is still being used by, for instance, the texts by Stephani and by Carroll.

On the other hand, there is quite a different approach, originally invented by Elie Cartan, for the determination of the Levi-Civita connection that has become much more common and popular in these days, and is in fact used by most working relativists. In this mode, one chooses a non-holonomic basis set for the tangent vector spaces with the property that the components of the metric are constants, just as one might in flat space. (It is a characteristic of curved spaces that one may **not** require both that the metric coefficients be constant and that the basis set be holonomic!) Having made such a choice, of course the partial derivatives of the metric coefficients are all zero, thereby eliminating that “half” of the contributions to the connection. In this case all the contributions come from the commutation coefficients, causing the connection coefficients to have quite different symmetry properties. If we re-write the equation for the metrizable coefficients, Eqs. (5.5), one last time, we have the following completely general form for it:

$$\nabla g_{\mu\nu} = dg_{\mu\nu} - \mathfrak{L}^\lambda_{(\mu} g_{\lambda\nu)} = dg_{\mu\nu} - \mathfrak{L}_{(\nu\mu)}. \quad (5.5'')$$

Since the Levi-Civita connection is metric compatible, the left-hand side of this equation is zero, so that we have a very simple form for the symmetric part of the connection 1-forms:

$$\text{for metric compatible connection:} \quad dg_{\mu\nu} = \mathfrak{L}_{(\mu\nu)}. \quad (5.11LC)$$

Therefore, having chosen a (non-holonomic) basis so that the metric components are constant, it becomes immediately apparent that the various connection 1-forms are skew-symmetric. In our 4-dimensional spacetime, this indicates that there are only 6 independent 1-forms to be determined. Looking at the defining equation for the curvature 2-forms, $\mathcal{Q}_{\lambda\mu}$, we see that the same skew-symmetry condition applies there, as well, giving us only 6 independent curvature 2-forms, also! Therefore while the general Levi-Civita connection forms require considerable calculation, these two distinct modes of calculating them simplify the process greatly:

$$\text{a choice of } \{\tilde{e}_a\} \text{ as } \left\{ \begin{array}{ll} \text{holonomic} \Rightarrow \Gamma_{abc} = \frac{1}{2}\Gamma_{a(bc)}, & \text{since } C_{ab}{}^c \equiv 0, \\ \text{for } \frac{1}{2}N^2(N+1) & \\ \text{ind. components} & \xrightarrow{\text{for } N=4} \text{40 independent components,} \\ \\ \text{a tetrad} \Rightarrow \Gamma_{abc} = \frac{1}{2}\Gamma_{[ab]c}, & \text{since } g_{ab,c} \equiv 0, \\ \text{for } \frac{1}{2}N^2(N-1) & \\ \text{ind. components} & \xrightarrow{\text{for } N=4} \text{24 independent components} \end{array} \right. \quad (5.12a)$$

I almost always use the approach that uses a non-holonomic choice of vector basis with constant metric coefficients. The most important reason for this is that the basis vectors are orthonormal, in the same way that they are in ordinary, flat-space physics, which makes an understanding of vector components much more intuitive. It is of course also true that it reduces the number of independent connection and curvature components

There is, however, another very valuable set of features that are acquired by using the Levi-Civita connection, and its associated metric. Let us return to Cartan's second structure equation, as given in Eq. (4.15), and use the metric to lower the one upper index:

$$d\mathfrak{L}_{\lambda\mu} + \mathfrak{L}_{\lambda\nu} \wedge \mathfrak{L}^{\nu}{}_{\mu} = \mathcal{Q}_{\lambda\mu} = \frac{1}{2}R_{\lambda\mu\eta\kappa}\omega^{\eta} \wedge \omega^{\kappa}. \quad (5.12b)$$

Studying the second term in the expression on the left above, we easily see that it is skew symmetric in the free indices, i.e., in λ and μ :

$$\mathfrak{L}_{\lambda\nu} \wedge \mathfrak{L}^{\nu}{}_{\mu} = -\mathfrak{L}^{\nu}{}_{\mu} \wedge \mathfrak{L}_{\lambda\nu} = +\mathfrak{L}^{\nu}{}_{\mu} \wedge \mathfrak{L}_{\nu\lambda} = +\mathfrak{L}_{\nu\mu} \wedge \mathfrak{L}^{\nu}{}_{\lambda} = -\mathfrak{L}_{\mu\nu} \wedge \mathfrak{L}^{\nu}{}_{\lambda}.$$

Since our connection 1-forms are skew-symmetric in those two indices, it follows that $\Omega_{\lambda\nu}$ is skew-symmetric in those indices, so that we also can see that the Riemann tensor components are skew symmetric in the first two indices as well as the last two, i.e., $R_{\mu\nu\lambda\eta} = -R_{\nu\mu\lambda\eta}$. **It is very important that this is a tensor equation, and is therefore true in all coordinate systems, even though we derived it in a particular orthonormal one.**

Workers in the field usually characterize that case by using the special word **tetrad**—from the Greek, more or less, to mean something like “4-legged thing”—to indicate a choice of non-holonomic basis set such that the components of the metric are constant. While any constant choice would satisfy the criteria for being a tetrad, it turns out there are really only two further choices that occupy much space in the research literature on the 4-dimensional space-times for general relativity, which I now review. (They were already discussed in the manifolds, etc. handout)

a. **Orthonormal tetrads** are those that are commonly used by true, physical observers, since they are the “obvious” generalization of the simplest sort of basis set in our ordinary 3-dimensional space; therefore, they are often referred to in the literature as “physical basis sets,” or “physical tetrads.” For our choice of signature, such a tetrad would appear as follows:

$$\begin{aligned} \text{an orthonormal tetrad : } \mathbf{g} &= \hat{\omega}^1 \otimes \hat{\omega}^1 + \hat{\omega}^2 \otimes \hat{\omega}^2 + \hat{\omega}^3 \otimes \hat{\omega}^3 - \hat{\omega}^4 \otimes \hat{\omega}^4 \\ \text{or } \mathbf{g} &\approx (\hat{\omega}^1)^2 + (\hat{\omega}^2)^2 + (\hat{\omega}^3)^2 - (\hat{\omega}^4)^2, \\ \text{with } ((g_{\mu\nu})) &\implies \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \end{aligned} \quad (5.13)$$

where the symbol \approx is used to “copy” a very common mode of writing in the literature that doesn’t “bother” to write the actual tensor product symbols, and “presumes” that the metric is symmetric. More precisely, when considering a metric tensor, which is actually an element of $\Lambda^1 \otimes \Lambda^1$, it is very common to just write $dx dy$, when what is really meant is $\frac{1}{2}\{dx \otimes dy + dy \otimes dx\}$. This form of writing is “sloppy” but very common, and of course convenient.

b. **Null tetrads** determine a different form for a constant tetrad, which comes from the fact that one often studies physical fields moving with the speed of light, such as electromagnetic or gravitational radiation. These have directions associated with them which are “null,” i.e., of zero length; therefore it is also quite common to use a system of 4 null vectors as a choice of tetrad. In special relativity, it is easy to see, for instance, that $\hat{z} \pm \hat{t}$ constitute a pair of (linearly-independent) null vectors that describe “light rays” either outgoing and incoming along the \hat{z} -direction. On the other hand, another linearly-independent pair of null rays does not exist; nonetheless, one should never let a simple bothersome fact like nonexistence deter one from doing what “needs to be done.” Therefore, the standard approach to resolving this difficulty is to introduce complex, null-length basis vectors in, for instance, the plane of the wave-front. Again, in special relativity, this would correspond to the pair of basis vectors, $\hat{x} \pm i\hat{y}$. I note that it is “somewhat” customary to use the symbols $\{\varrho^\alpha\}_1^4$ for the elements of a null basis, and also $\nu_{\alpha\beta}$ for the components of a metric made from a null tetrad, just as it is customary to use the symbols $\eta_{\mu\nu}$ for the components of a metric made from an orthonormal tetrad:

$$\begin{aligned}
\text{a null tetrad : } \mathbf{g} \equiv \boldsymbol{\nu} &= \varrho^1 \otimes \varrho^2 + \varrho^2 \otimes \varrho^1 + \varrho^3 \otimes \varrho^4 + \varrho^4 \otimes \varrho^3 \\
\text{or } \boldsymbol{\nu} &\approx 2\varrho^1\varrho^2 + 2\varrho^3\varrho^4, \\
\text{with } ((\nu_{\alpha\beta})) &\implies \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{5.14}
\end{aligned}$$

(Other relative signs are used in the literature as well.)

Taking $\{\tilde{\varrho}^\mu\}_1^4$ as a representative orthonormal tetrad, and $\{\varrho^\alpha\}_1^4$ as the associated null tetrad, we may write explicitly the transformation matrix between them:

$$\begin{aligned}
\varrho^\alpha &= A^\alpha{}_\mu \tilde{\varrho}^\mu, \\
\begin{pmatrix} \varrho^1 \\ \varrho^2 \\ \varrho^3 \\ \varrho^4 \end{pmatrix} &= \frac{1}{\sqrt{2}} \begin{pmatrix} +1 & +i & 0 & 0 \\ +1 & -i & 0 & 0 \\ 0 & 0 & +1 & -1 \\ 0 & 0 & +1 & +1 \end{pmatrix} \begin{pmatrix} \tilde{\varrho}^1 \\ \tilde{\varrho}^2 \\ \tilde{\varrho}^3 \\ \tilde{\varrho}^4 \end{pmatrix}. \tag{5.15}
\end{aligned}$$

The matrix A then has determinant $-i$, consistent with the fact that the determinant of $\nu_{\alpha\beta}$ is $+1$ while the determinant of $\eta_{\mu\nu}$ is -1 . We could then show that the value of the metric, in this basis is indeed $\nu_{\alpha\beta}$ as stated above, by using the inverse of this matrix A to transform $\boldsymbol{\eta}$ into $\boldsymbol{\nu}$, i.e., the matrix

$$\nu_{\alpha\beta} = (A^{-1})^\mu{}_\alpha (A^{-1})^\nu{}_\beta \eta_{\mu\nu}, \quad A^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ -i & +i & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \quad (5.16)$$

■ Local Inertial Reference Frames Possible:

Gaussian Normal Coordinates

A significant portion of the equivalence principle is understood mathematically by the following proof of the (local) existence of very special coordinates in a neighborhood of whatever point on a manifold one chooses. Such coordinates are also often referred to as determining the existence of (local) inertial reference frames. To understand the construction of these coordinates, let us choose an arbitrary point $P \in U_P \subseteq \mathcal{M}$, where U_P is sufficiently small, and we are given an arbitrary coordinate system in U_P , which we denote by $\{x^\rho\}$. We define a new coordinate system, $\{\bar{x}^\mu\}$, which will be structured so that it has its zero values at P , will have the standard orthonormal metric of special relativity, $\eta_{\mu\nu}$, when evaluated at P , and, **also**, will have vanishing connection values at P . We write an equation for these coordinates in U_P , defining them quadratically in terms of the new, desired coordinates, that is to be solved, locally, for those new coordinates:

$$x^\rho \Big|_{U_P} \equiv x^\rho(P) + Q^\rho{}_\mu \bar{x}^\mu + \frac{1}{2} B^\rho{}_{\mu\nu} \bar{x}^\mu \bar{x}^\nu, \quad (5.17)$$

where the matrix $Q \equiv ((Q^\rho{}_\mu))$ and the quantities $B^\rho{}_{\mu\nu} = B^\rho{}_{\nu\mu}$ are to be determined in order to cause the new coordinates $\{\bar{x}^\mu\}$ to have the desired properties. We see here the reason for the requirement that the neighborhood U_P be sufficiently small, namely so that these quadratic equations for \bar{x}^μ will have unique solutions. Note that the expansion is just a form of a finite Taylor series, which always can be inverted in a sufficiently small neighborhood, provided that the quantities $Q^\rho{}_\mu$ have a non-zero determinant.

- a. It is obvious that $\bar{x}^\mu(P) = 0$, by construction.
- b. To arrange for the values of the metric in this new coordinate system, we recall its transformation laws:

$$\bar{g}_{\mu\nu} = X^\rho{}_\mu X^\sigma{}_\nu g_{\rho\sigma} , \quad X^\rho{}_\mu \equiv \frac{\partial x^\rho}{\partial \bar{x}^\mu} . \quad (5.18a)$$

We return to the defining equation for the new coordinates, and easily calculate that $X^\rho{}_\mu = Q^\rho{}_\mu + B^\rho{}_{\mu\nu} \bar{x}^\nu$, so that we have the simple statement that $X^\rho{}_{\mu|P} = Q^\rho{}_\mu$. This tells us that the metric, evaluated exactly at the point P , is just

$$\bar{g}_{\mu\nu|P} = Q^\rho{}_\mu Q^\sigma{}_\nu g_{\rho\sigma} \quad \text{or, in matrix form,} \quad \bar{G}|_P = Q^T G|_P Q . \quad (5.18b)$$

Writing it in the matrix format is a “cue” that the desired property is easily arranged. This is because of Sylvester’s theorem about any symmetric matrix. The theorem tells us that there is always a *congruency transformation*, i.e., one of the kind in this equation, which reduces any given symmetric matrix to a diagonal matrix which has only ± 1 ’s and 0’s along that diagonal. (This arrangement of ± 1 ’s and 0’s, ignoring order, is often referred to as the signature of the given symmetric matrix, and is an invariant under congruency transformations.) Obviously, since our metric is invertible, there are no zero’s in its presentation, and this provides the desired $\eta_{\mu\nu}$. Note that the quantities $Q^\rho{}_\mu = X^\rho{}_{\mu|P}$ are, in general 16 independent quantities; we only need 10 of them to solve the equation to reduce the symmetric $g_{\mu\nu}$ —at the single point P —to its Sylvester form, $\eta_{\mu\nu}$; therefore, there are still 6 available degrees of freedom. These form exactly the 6-parameter family of such Q ’s, i.e., the full Lorentz group, that of course preserves $\eta_{\mu\nu}$.

- c. We now want to determine the form of the connection, in these new coordinates, at the point P . We recall that the transformation behavior is given by

$$\bar{\Gamma}^\lambda{}_{\mu\nu} = (X^{-1})^\lambda{}_\sigma \left\{ X^\rho{}_\mu X^\tau{}_\nu \Gamma^\sigma{}_{\rho\tau} - \frac{\partial}{\partial \bar{x}^\nu} X^\sigma{}_\mu \right\} . \quad (5.19)$$

Evaluating this general equation at the point P , for our coordinate transformation, we have

$$\bar{\Gamma}^\lambda{}_{\mu\nu|P} = \frac{\partial \bar{x}^\lambda}{\partial x^\tau} \{ Q^\rho{}_\mu Q^\sigma{}_\nu \Gamma^\tau{}_{\rho\sigma} - B^\tau{}_{\mu\nu} \} , \quad (5.20)$$

where of course all the quantities on the right-hand side above are to be evaluated at the point P . As the matrix Q has already been determined, and the connections in the original coordinate system are already known, it is straight-forward to just use this equation to define the quantities $B^\tau{}_{\mu\nu}$ so that the left-hand side is arranged to vanish—at the point P —as was desired! It is of course necessary to note that the desired values for $B^\rho{}_{\mu\nu}$ must be symmetric in those two lower indices; however, this is the case because of connection—in our current coordinate gauge—is symmetric in the two lower indices, and it is summed on both sides by the same set of quantities. Notice, in particular, that there are 40 different, independent Christöffel symbols, and there are $10 \times 4 = 40$ available values for $B^\rho{}_{\mu\nu}$, so that works out fine. Since the Christöffel symbols are determined via the first (coordinate) derivatives of the metric, as shown in the equations just before Eqs. (5.7), this tells us that the metric’s first derivatives all vanish at P , i.e., that the Taylor series expansion of the metric in the near neighborhood of P , namely U_P , is quadratic in the coordinates there.

All this is very promising, and sounds just exactly what we want, since this means that the geodesics passing through P simply satisfy $d^2\bar{x}^\mu/d\lambda^2 = 0$ at the point P , where λ is the parameter along the curve, i.e., they look like straight lines. In particular this means that they travel along straight lines there, as the principle of equivalence wants. However, now we should ask if we can do better yet. Could we also make the curvature zero at this point? Since the connections and the first derivatives of the metric vanish there, this is the same as asking whether or not we can choose coordinates so that the second derivatives of the metric, $g_{\mu\nu,\lambda,\eta}$, would vanish at P . Since these are symmetric in each pair of indices, there are 100 distinct values that we want to vanish. In order to do that, this would require some differently-defined new coordinates such that their third derivatives, $\partial x^\alpha/\partial\bar{x}^\mu\partial\bar{x}^\nu\partial\bar{x}^\lambda$, would be used in order to cause this to happen. However, there are only $4 \times 20 = 80$ distinct quantities here, because of the symmetry of the different partial derivatives.

It is worth noting somewhere some useful facts about the numbers of symmetries of this sort. The number of ways to choose r things from a total

of n things, when all must be different, and the order is irrelevant is given by the binomial coefficient, $\binom{n}{r}$. This is appropriate when counting skew-symmetric things. Alternatively the number of ways to choose r things from a total of n things when they are allowed to all be the same, i.e., you can repeatedly draw out the same thing if you so desire, is given by the binomial coefficient $\binom{n+r-1}{r}$, which is appropriate for objects with symmetric indices.

Therefore, we could only eliminate 80 of the 100 2nd derivatives, leaving 20 behind. These are in fact the 20 distinct components of the Riemann tensor. A slightly different way of looking at this last determination comes from writing out the curvature tensor at P , where we know the Christoffel symbols vanish, but their derivatives apparently do not. We begin by evaluating expressions in Eqs. (5.10Cr) at the point P , remembering that the first derivatives of the metric and the Christoffel symbols both vanish there, but the second derivatives of the metric need not:

$$\begin{aligned} \overline{\left\{ \begin{matrix} \mu \\ \nu[\tau] \end{matrix} \right\}}_{,\sigma}|_P &= \frac{1}{2} \eta^{\mu\eta} \left(-\bar{g}_{\nu[\tau,\eta\sigma]} + \bar{g}_{\eta[\tau,\nu\sigma]} \right) |_P , \\ \bar{R}_{\mu\nu\sigma\tau}|_P &= \frac{1}{2} \left(-\bar{g}_{\nu[\tau,\mu\sigma]} + \bar{g}_{\mu[\tau,\nu\sigma]} \right) . \end{aligned} \tag{5.21}$$

While the above approach is highly desirable in some “equivalence-principle neighborhood” of some point, it is true that many of our observations are not done at a single point but, rather, along the observer’s geodesic. Therefore, it is also useful to realize that this sort of very special coordinates may also be created in some neighborhood around points that lie on some geodesic. The construction of these coordinates is really quite similar, but does require the solution of implicit equations of higher order than just the quadratic ones that we just described. These are called geodesic coordinates, and we explain them below:

Geodesic Coordinates

Suppose that there is a geodesic that passes through some special point P on the manifold, and let us have some (arbitrary) set of coordinates in a neighborhood containing at least some significant region of this geodesic. In those coordinates the differential equations satisfied by

the geodesic are the usual ones, where we now are using the Christöffel symbols to describe the affine connection in these coordinates:

$$\frac{d^2 x^\alpha}{d\tau^2} + \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} \frac{dx^\beta}{d\tau} \frac{dx^\gamma}{d\tau} = 0 \equiv x''^\alpha + \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} x'^\beta x'^\gamma . \quad (5.22)$$

Now let P be a given point on the geodesic, and take $\tau = 0$ there, specify the value of the coordinates there as $x^\mu(0; P)$, and take this and also $x'^\mu(0; P)$ as initial conditions for the curve. For points elsewhere on the geodesic, at some value, say τ of the parameter we may write out the Taylor series:

$$x^\mu(s; Q) = x^\mu(0; P) + x'^\mu(0; P)s + \frac{1}{2}x''^\mu(0; P)s^2 + \frac{1}{6}x'''^\mu(0; P)s^3 + O(s^4) . \quad (5.23)$$

However, we may differentiate the equation for the geodesic itself, Eq. (5.22) and determine

$$\begin{aligned} x'''^\mu &= - \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\}_{,\delta} x'^\beta x'^\gamma x'^\delta - \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} x''^\beta x'^\gamma - \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} x'^\beta x''^\gamma \\ &= - \left(- \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\}_{,\delta} - \left\{ \begin{array}{c} \alpha \\ \beta\eta \end{array} \right\} \left\{ \begin{array}{c} \eta \\ \gamma\delta \end{array} \right\} \right) x'^\beta x'^\gamma x'^\delta \equiv - \left\{ \begin{array}{c} \alpha \\ \beta\gamma\delta \end{array} \right\} x'^\beta x'^\gamma x'^\delta , \end{aligned} \quad (5.24)$$

where the last equality is a definition of the Christöffel-like symbol with three indices in the bottom row, but one should notice that it is actually defined as only the symmetric part—on those bottom three indices—of the expression in the larger parentheses on the left, since only that part would contribute to the summation with the three first derivatives, which is surely symmetric in those indices.

We may now insert all this back into the equation for points along the geodesic, Eqs. (5.23), where I have simplified the notation by using the symbol $a^\alpha \equiv x'^\alpha(0; P)$:

$$x^\alpha(s; Q) = x^\alpha(0; Q) + a^\alpha s - \frac{1}{2} \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} a^\beta a^\gamma s^2 - \frac{1}{6} \left\{ \begin{array}{c} \alpha \\ \beta\gamma\delta \end{array} \right\} a^\beta a^\gamma a^\delta s^3 + O(s^4) . \quad (5.25)$$

Having this formulation, we now define a new coordinate system, $\{y^\mu\}$, in the neighborhood of the point P , but actually for a similar neighborhood along any (nearby) points on the geodesic, by the following implicit equations, to be solved for the $\{y^\alpha\}$:

$$x^\alpha \equiv x^\alpha(0; P) + y^\alpha - \frac{1}{2} \left\{ \begin{array}{c} \alpha \\ \beta\gamma \end{array} \right\} y^\beta y^\gamma - \frac{1}{6} \left\{ \begin{array}{c} \alpha \\ \beta\gamma\delta \end{array} \right\} y^\beta y^\gamma y^\delta + O(y^4) , \quad (5.26)$$

where one can ignore the higher-order terms, and solve the cubic equations for the coordinates. Using this approach one can see that for points on the geodesic we have just $y^\alpha(s; Q) = a^\alpha s + O(s^4)$.

For ideas of consistency it should be noted that there are exactly 80 different independent elements of these 4-index Christöffel-like objects, just as we could determine earlier, so that there are still $100-80=20$ free second derivatives of the metric. It should, however, also be noticed that one can push this method further by continuing with it, determining the fourth derivatives in terms of 5-index objects and extending the problem to solving quartic equations for the new coordinates, etc. Obviously the particular regions of independent solutions, with the given initial conditions, limit the size of the neighborhoods, and the distance along the geodesic which the neighborhoods may be pushed, and they doubtless become smaller and smaller as the order of the equations goes up, because degenerate solutions must clearly be avoided, i.e., places where solutions meet. To see this approach in considerable more detail see the paper by Agapitos Hatzinikitas, arXiv:hep-th/0001078v1 12 Jan 2000. Another quite interesting paper is by James Nester: *J. Phys. A.* **40**, 2751-2754 (2007), where he shows that as one proceeds upward in defining the new coordinates by higher- and higher-order equations the number of degrees of freedom one has left at each order is exactly the number of degrees of freedom in the appropriate level of covariant derivatives of the curvature tensor at that order.

■ Symmetries of the Curvature Tensor

This is a very good point to show the last symmetry of the Riemann tensor. We plan to show that it is symmetric under the interchange of the first pair with the second pair, i.e., that $R_{\mu\nu\sigma\tau} = R_{\sigma\tau\mu\nu}$. As this relationship is a tensor equation, we may use any coordinate system we like to show it, and it will remain true in every coordinate system, and choice of basis. Therefore, we will use Eqs. (5.21) above, in geodesic coordinates, and all evaluated at P :

$$\begin{aligned} 2\bar{R}_{\mu\nu\sigma\tau} &= -\bar{g}_{\nu[\tau,\mu\sigma]} + \bar{g}_{\mu[\tau,\nu\sigma]} = \bar{g}_{\mu\tau,\nu\sigma} + \bar{g}_{\nu\sigma,\mu\tau} - \bar{g}_{\mu\sigma,\nu\tau} - \bar{g}_{\nu\tau,\mu\sigma} , \\ 2\bar{R}_{\sigma\tau\mu\nu} &= \bar{g}_{\sigma[\nu,\tau\mu]} - \bar{g}_{\tau[\nu,\sigma\mu]} = \bar{g}_{\sigma\nu,\tau\mu} + \bar{g}_{\tau\mu,\sigma\nu} - \bar{g}_{\sigma\mu,\tau\nu} - \bar{g}_{\tau\nu,\sigma\mu} . \end{aligned} \tag{5.27}$$

Using the fact that the metric is symmetric in its two indices, and that partial differentiation is symmetric, one can quickly see that the two expressions on the right, just above in Eqs. (5.27), are equal; therefore, the two expressions on the left are equal. However, that equality is a tensor expression and must therefore be true in any coordinate system, and basis for tangent vectors, which is what was to be shown.

It is probably therefore useful to consider now the complete set of symmetries of the Riemann curvature tensor, induced by the Levi-Civita connection and its associated metric. $R_{\mu\nu\sigma\tau}$ is skew-symmetric under the interchange of the 1st and 2nd index, and also under the interchange of the 3rd and 4th index, while it is symmetric under the interchange of that first pair with the second pair. Since there are 6 possible values—in 4 dimensions—for independent values of a skew pair of indices, we can conceive of this as a 6×6 matrix—since it has 2 pair of such skew indices. In addition the matrix in question is symmetric. Such a matrix has $\frac{1}{2}(6 * 7) = 21$ independent elements. However, the Bianchi identity removes one of those, leaving us with 20. We will discuss this approach to the curvature tensor in more detail later.

However, we should also note that the symmetry of the Riemann tensor under the interchange of the two pairs of (skew-symmetric) indices causes the Ricci tensor to be a symmetric tensor—see Eq. (4.18)—so that it then has exactly 10 independent components; i.e., it takes on exactly half of the total independent components of the Riemann tensor, the remaining ones staying on, independently, in the Weyl tensor (or the conformal tensor).

■ Curvature Tensor Components as Tidal Gravitational Accelerations

Continuing onward with the structure of the special inertial reference frame created above, we have already seen that while the connection components vanish at the point P , their first derivatives do not vanish. Therefore, Taylor’s series tells us that the connection components do not vanish in the near neighborhood of P . In fact, this says that they are of first order in an appropriate affine parameter along any geodesic away from P . We return now to the thought “experiment” where the observer whose worldline goes through P is measuring the behavior of particles near him, where “near” really means that they are only first-order in

small quantities away. So in our new reference frame, a nearby object at location $\bar{x}^\mu(\tau)$ moves on its own trajectory, and we may think about its motion in the same way as we already did, for an object falling near the surface of the earth, in §II of the introductory lectures:

$$\ddot{\bar{x}}_{|\text{near } P}^m = g_{|\text{near } P}^m - g_{|P}^m = g_{,n|P}^m \bar{x}^n = -\bar{\Gamma}_{44,n|P}^m \bar{x}^n = -\bar{R}^m{}_{4n4|P} \bar{x}^n . \quad (5.28)$$

We began the equation with the tidal gravitational acceleration, i.e., the difference in two nearby gravitational accelerations, used our relationship between gravitational accelerations and connection components, and then used the equation relating derivatives of the connection to introduce the appropriate components of the curvature tensor, noting that those parts of the curvature tensor that are quadratic in the connections just vanish at P . So, re-iterating the equation in words, we say that the $\bar{R}^m{}_{4n4}$ components of the curvature tensor—measured in the local geodesic coordinate frame—tell us the (first-order) change—in the \tilde{e}_n -direction—of the \tilde{e}_m -component of the gravitational acceleration (or field). The reason for the two 4’s is because we are evaluating all this in the frame where the observer’s world velocity is simply \tilde{e}_4 . It is also true that we could write this tensor quantity as $\bar{R}^\mu{}_{4\nu 4}$, since if either μ or ν were to have the value 4 the component would automatically be zero; this method of writing it brings it back into a more plausible form, appropriate for 4-dimensional spacetime. On the other hand, in a more general frame—as, for instance, one where we were being watched by some other very nearby observer, who measures us to have 4-velocity u^γ , we would have $R^\alpha{}_{\beta\gamma\delta} u^\beta u^\delta$.

■ “Derivation” of Einstein’s equations

The discussion above gives us a very good physical basis for understanding these components of the curvature tensor. However, let us return a bit more closely to thinking about curvature as being caused by a weak gravitational field of the sort that we do have here on the earth, one which can be conceived of as being determined by a potential. In situations like ours where the earth is the only effective source of gravity, we can always determine a (scalar) gravitational potential, $\Phi(\vec{r})$, such that the acceleration is given by $\bar{g}^i = -\partial_{\bar{x}_i} \bar{\Phi}$. This then gives us the relation, coming from Eq. (5.28):

$$\partial_j \partial^i \bar{\Phi} = \bar{R}^i{}_{4j4} , \quad (5.29)$$

a yet more direct approach to the understanding of these components of the Riemann tensor, in those cases where the acceleration can be described by a single potential function. However, we can proceed a bit further in this regard because we remember that the gravitational potential is of course caused by some source of gravity, which is a local density of matter, ρ , so that it is quite general that Φ should satisfy the follow pde:

$$\partial^i \partial_i \Phi \equiv \nabla^2 \Phi = k\rho , \quad (5.30)$$

where k is some constant, depending on one's choice of units. (In our units it is 4π , the number of solid angles (steradians) in all possible directions in 3-dimensional space.) Thinking back to Eqs. (5.29), we can set $i = j$ and sum on their values, to re-create that relationship:

$$\bar{R}^j_{4j4} = \partial_j \partial^j \Phi = k\bar{\rho} = k\bar{T}_{44} , \quad (5.31)$$

where the last equality reminds us of the energy-momentum tensor of matter, which has its 4, 4-component as the energy density. There is a sum on j , from 1 to 3, in these relationships. However, we know that the curvature tensor is skew-symmetric in each of its 2 pairs of indices. Therefore, $R^4_{4j4} = 0 = R^j_{444}$, so that the sum may be replaced by a sum on μ , from 1 to 4: following may also be written

$$k\bar{T}_{44} = \bar{R}^j_{4j4} = \bar{R}^\mu_{4\mu4} = \bar{\mathcal{R}}_{44} , \quad (5.32)$$

the last entry being the 4, 4-component of the Ricci tensor, as defined in Eqs. (4.18). Of course this relationship is only expected to be true in our own local frame. In this frame, our velocity is just $u^\mu = \delta_4^\mu$; therefore, a tensorial approach to the equation above would be the following, where we note that the energy density is much larger than any other term in the energy-momentum tensor:

$$kT_{\mu\nu}u^\mu u^\nu = \mathcal{R}_{\mu\nu}u^\mu u^\nu . \quad (5.33)$$

However, this last should be true for any 4-velocity whatsoever; therefore, if we could divide out the two copies of the 4-velocity, provided that the tensors in question are symmetric, giving

us the following “guess” as to the relationship between the two tensors, that will agree with this weak gravity approach we have followed above:

$$\mathcal{R}_{\mu\nu} = kT_{\mu\nu} . \quad (5.34)$$

There is, however, a problem with this suggested result. We know that the energy-momentum tensor is divergenceless. More precisely, we know that the divergence of that tensor is the density of overall net force acting on the system at that point. If in fact there were some other force acting on the system, that would appear in the divergence of our tensor, we should go back and pull that force over to the other side, including its energy and momentum densities into the previous tensor, until we have made our **total energy-momentum tensor** as divergenceless.

But it turns out that the divergence of the Ricci tensor is not zero, so that we will have to modify this equation in some way, that will remain consistent with the earlier result concerning $\overline{\mathcal{R}}_{44}$. Preparatory to calculating the divergence of the Ricci tensor, it is convenient to first define the trace of the Ricci tensor, usually referred to as the Ricci scalar, which is a quantity which will help us understand the (covariant) divergence of the Ricci tensor:

$$\mathcal{R} \equiv g^{\mu\nu} \mathcal{R}_{\mu\nu} . \quad (5.35)$$

We may now proceed to the calculation of the (covariant) divergence of the entire Ricci tensor, beginning with the 2nd Bianchi identity, a cyclic sum of covariant derivatives of the Riemann tensor which vanishes, as stated in component form in Eqs. (4.17b), we raise the first lower index and re-write it as follows:

$$R^{\eta\lambda}{}_{\mu\nu;\sigma} + R^{\eta\lambda}{}_{\nu\sigma;\mu} + R^{\eta\lambda}{}_{\sigma\mu;\nu} = 0 . \quad (5.36)$$

We now sum on $\eta = \mu$, which acquires two differing copies of the Ricci tensor, remembering its definition:

$$\mathcal{R}^{\lambda}{}_{\nu;\sigma} + \mathcal{R}^{\mu\lambda}{}_{\nu\sigma;\mu} - \mathcal{R}^{\lambda}{}_{\sigma;\nu} = 0 , \quad (5.37)$$

and then we set $\lambda = \nu$ and sum one more time, and see that the Ricci scalar comes into the calculations:

$$\mathcal{R}_{;\sigma} - \mathcal{R}^\lambda_{\sigma;\lambda} - \mathcal{R}^\eta_{\sigma;\eta} = 0, \quad (5.38)$$

which we re-write slightly to tell us that the divergence of the Ricci tensor is half the (ordinary) derivative of the Ricci scalar, rather than zero:

$$\mathcal{R}^\lambda_{\sigma;\lambda} = \frac{1}{2}\mathcal{R}_{;\sigma}. \quad (5.39)$$

This tells us that the quantity $\mathcal{R}^\lambda_{\mu} - \frac{1}{2}\delta^\lambda_{\mu}\mathcal{R}$ is actually divergenceless. We therefore define a tensor closely-related to the Ricci tensor, namely the (divergenceless) Einstein tensor, which we define with both indices contravariant, noting that it is symmetric, just as is the Ricci tensor, and we attempt to set it equal to the energy-momentum tensor for local matter and energy, along with a constant κ , which we will need to determine:

$$\kappa T^{\mu\nu} = \mathcal{G}^{\mu\nu} \equiv \mathcal{R}^{\mu\nu} - \frac{1}{2}g^{\mu\nu}\mathcal{R} \quad \text{and} \quad \mathcal{G}^{\mu\nu}{}_{;\mu} = 0 = T^{\mu\nu}{}_{;\mu}. \quad (5.40)$$

Although there is more than one route to the desired answer, i.e., the value of the constant κ , let us take the following approach. We first take the trace of both sides of the equation:

$$\kappa T \equiv \kappa g_{\mu\nu}T^{\mu\nu} = \mathcal{R} - \frac{1}{2}(4)\mathcal{R} = -\mathcal{R}. \quad (5.41)$$

We may then use this equation to replace \mathcal{R} in the equation above, which gives us

$$\mathcal{R}^{\mu\nu} = \kappa(T^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T). \quad (5.42)$$

To determine T we note that $T^{44} = \rho$ is very much larger—in our weak field, slow-moving approximation—than any of the other components, so that a good approximation is $T = -T^{44} + g_{ij}T^{ij} \approx -T^{44}$. Then our equation becomes

$$\mathcal{R}^{\mu\nu} = \kappa(T^{\mu\nu} + \frac{1}{2}g^{\mu\nu}T^{44}) \implies \mathcal{R}^{44} = \kappa(T^{44} - \frac{1}{2}T^{44}) = \frac{1}{2}\kappa\rho. \quad (5.43)$$

However, we know that—in this specialized case—that $\mathcal{R}^{44} = 4\pi\rho$. Therefore, it follows that $\kappa = 8\pi$.

This would seem to be the end of all this, and, perhaps it was. However, as I'm sure most of you already know, Einstein decided somewhat later that there would be no “harm” in adding yet an additional term to the left-hand side of this equation, which was **a constant** multiplying the metric, $g_{\mu\nu}$. This constant would need to be sufficiently small to fit the observed data; however, that term is also divergenceless, so it doesn't fault the above argument. We name the constant as *the cosmological constant*, denote it by Λ and add it to the equations so that the final form of Einstein's equations relating half the curvature, i.e., the Ricci (or Einstein) tensor with the local density of energy and momentum is given by

$$\mathcal{G}_{\mu\nu} + \Lambda g_{\mu\nu} \equiv \mathcal{R}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \quad \kappa = 8\pi. \quad (5.44)$$

It is worth noting that this is in fact the most general sort of equation that one could have to describe this physics: the assumptions are simply that there should be a function of the metric, its first, and its second derivatives—only—that equals the energy-momentum tensor, and therefore symmetric and divergenceless, and also has the correct Newtonian limits. To see a proof of this, see the paper by Lovelock, *Journ. Math. Phys.*, **12**, p. 498 (1971).

■ The Riemann Curvature Tensor, its Mathematical Meaning

The theorem is rather straightforward, and actually quite nice:

Let W be a simply connected region in some (smooth) manifold. The necessary and sufficient condition that at every point in W there exists a neighborhood, and a coordinate system defined there, such that everywhere in that neighborhood, in that coordinate system, $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$ is precisely that $R_{\mu\nu\lambda\eta} = 0$ throughout W .

In one direction, i.e., if we begin with the statement that the manifold is flat throughout W then it is clear that we can create such a coordinate system anywhere in W . Therefore,

the portion that needs proving begins with the statement that the curvature tensor vanishes throughout W . With that information we want to create a set of 4 linearly-independent, covariantly constant vector fields defined throughout W . We do this by first establishing such a set at some initial point $P \in U_P \subseteq W$. A reasonable example would be the set $\tilde{V}_{(\alpha)}$ such that $V_{(\alpha)|P}^\mu = \delta_\alpha^\mu$, which of course arranges that at the point P we have these 4 vectors as an orthonormal set, i.e.,

$$g_{\mu\nu} V_{(\alpha)}^\mu V_{(\beta)|P}^\nu = \eta_{\alpha\beta} .$$

We then insist that these 4 are parallel transported along some curve extending from P to some other point in W ; i.e., we insist that $\nabla_{\tilde{u}} \tilde{V}_{(\alpha)} = 0$, for all tangent vectors \tilde{u} . (This is of course the same as saying the these vectors are extended to other points by solving the differential equations that say that they are covariantly constant. In order for this to be a legitimate method to extend them it must be true that the extension via two different paths (or curves) should give the same result. This is of course the same as asking for the integrability conditions for the integration process just described to give consistent results. Those conditions are simply Eqs. (4.14) which tells us that the commutator of two covariant derivatives acting on an arbitrary vector equals the curvature tensor summed into that vector, i.e.,

$$V^\lambda_{[:\rho;\sigma]} = R^\lambda_{\mu\rho\sigma} V^\mu . \quad (4.14)$$

Since we have assumed the Riemann tensor vanishes, this guarantees that such covariantly constant vector fields exist in some (local) neighborhood. Having then their existence, continuing on the route of the proof is simpler if we now use the metric to map them to a set of 4 covariantly-constant 1-form fields. For each value of α —which we recall is NOT a tensor index but simply a label to distinguish the 4 distinct fields—we define the associated 1-form in the standard way:

$$\mathcal{V}^{(\alpha)} \equiv g_{\mu\nu} V_{(\alpha)}^\mu \omega^\nu \equiv V_\nu^{(\alpha)} \omega^\nu .$$

The statement that these 4 1-forms are covariant constant is simply that $d\mathcal{V}^{(\alpha)} = 0$, which means there exist—since we have already insisted that W is simply connected—a neighborhood

around P in which there are functions, $f^{(\alpha)}(x)$ such that $\mathcal{V}^{(\alpha)} = df^{(\alpha)}$, and we may take a new coordinate system in this neighborhood as $\bar{x}^\alpha \equiv f^{(\alpha)}(x^\mu)$, and it then follows that the components of our original vector fields in this new coordinate system are just δ_β^α throughout. As well it is obvious that they are constant throughout this neighborhood, so that the metric, established by their scalar product as above, in the new coordinate system is just $\eta_{\alpha\beta}$ as we desire; i.e., it is constant throughout the neighborhood.

However, now suppose we have two such points in W , i.e., $P_1 \in U_{P_1}$ and $P_2 \in U_{P_2}$, and at the first one we have found the new coordinate system, $\bar{x}^\alpha(x^\mu)$ such that the metric is as desired throughout, i.e.,

$$g_{\mu\nu} \frac{\partial x^\mu}{\partial \bar{x}^\alpha} \frac{\partial x^\nu}{\partial \bar{x}^\beta} = \bar{g}_{\alpha\beta} = \eta_{\alpha\beta} , \text{ throughout } U_{P_1} ,$$

and, at $P_2 \in U_{P_2}$ we have found a new coordinate system, $\tilde{x}^\sigma(x^\mu)$ such that the metric is as desired throughout:

$$g_{\mu\nu} \frac{\partial x^\mu}{\partial \tilde{x}^\sigma} \frac{\partial x^\nu}{\partial \tilde{x}^\rho} = \tilde{g}_{\sigma\rho} = \eta_{\sigma\rho} , \text{ throughout } U_{P_2} .$$

Can we be sure that these are consistent sets, so that we can use this method to extend the coordinates throughout the entirety of W ? The correct, positive answer is more easily seen by writing out the transformation equations above in matrix format, where we present the two sets of Jacobian matrices as

$$\left(\left(\frac{\partial x^\mu}{\partial \bar{x}^\alpha} \right) \right) \implies A , \quad \left(\left(\frac{\partial x^\mu}{\partial \tilde{x}^\sigma} \right) \right) \implies B ,$$

which then gives the transformations above in the form

$$\begin{aligned} A^T G A &= \bar{G} = \boldsymbol{\eta} = \tilde{G} = B^T G B \\ \implies \boldsymbol{\eta} &= \bar{G} = A^T G A = A^T (B^{-1})^T \tilde{G} B^{-1} A = A^T (B^{-1})^T \boldsymbol{\eta} B^{-1} A , \end{aligned}$$

so that it can be seen that the matrix $B^{-1}A$ is a Lorentz transformation, surely allowed at this point. Moreover, we can see that

$$B^{-1}A = \frac{\partial \tilde{x}^\sigma}{\partial x^\mu} \frac{\partial x^\mu}{\partial \bar{x}^\alpha} = \frac{\partial \tilde{x}^\sigma}{\partial \bar{x}^\alpha} ,$$

which is just the Jacobian matrix for the transformation from the one set of coordinates, defined in one of the neighborhoods, to the set defined in the other neighborhood, so that the transformation just above works fine, i.e., is a proper transition function, in the overlap of the neighborhoods. This allows us to extend this process throughout the entirety of W , and we are done!

■ What the Red Shift Might Say about the Metric,

We return to our very early discussion of the measured red shift in the gravitational field of the earth, which can surely be described by the gravitational potential, Φ , such that $\vec{g} = -\nabla\Phi$. We found there that the ratio of frequencies of an emitted light ray, at two different locations in a gravitational field was proportional to the difference in the gravitational potential at those two places. However, **now that we want to describe our system via a metric**, we can say that the proper time measured for any pair of events, such as two successive crests of a waveform—emitted by an atom at rest—would just be given by $(\delta\tau)^2 = -g_{44}(\Delta t)^2$. If we want this at two different places, then g_{44} will probably have a different value at these two places. So we can write

$$\frac{\nu_2}{\nu_1} = \frac{(\Delta t)_1}{(\Delta t)_2} = \frac{\Delta\tau/\sqrt{-g_{44}(x_1)}}{\Delta\tau/\sqrt{-g_{44}(x_2)}} = \sqrt{\frac{g_{44}(x_2)}{g_{44}(x_1)}}. \quad (5.45)$$

The Pound-Rebka experiment can then be interpreted as a measurement of the ratio of g_{44} at different places in the gravitational field. To determine exactly how it depends on the gravitational potential, we return to Eq. (3.4) of these notes, where we showed that in one's own (instantaneous) rest frame, in a gravitational field, that $\vec{g}^i = -\bar{\Gamma}^i_{44}$. However, since we are using a very simple approach to the situation, for a weak field, we now suppose that our metric—while NOT the one from special relativity—is not too much different. As well, since the gravitational potential with which we are concerned is not time dependent, we agree that our metric components should be independent of time. With those provisos we return to the formula for the connection viewed as a Christöffel symbol, namely

$$\partial^i\Phi = -g^i = \left\{ \begin{matrix} i \\ 44 \end{matrix} \right\} = \frac{1}{2}g^{i\eta}(-g_{44,\eta} + 2g_{\eta 4,4}) = -\frac{1}{2}\partial^i g_{44}. \quad (5.46)$$

Remembering that the special relativity value of g_{44} is just -1 , we may integrate this equation to give the desired value for this component of the metric in the presence of a weak gravitational field, described by a potential, Φ :

$$g_{44} = -1 - 2\Phi = -(1 + 2\Phi) \approx -(1 + \Phi)^2, \quad (5.47)$$

where the last approximation is valid to second order in Φ , and will make the algebra easier both now and later on. This tells us that

$$\frac{\nu_2}{\nu_1} \approx \frac{1 + \Phi(x_2)}{1 + \Phi(x_1)} \implies \frac{\nu_2 - \nu_1}{\nu_1} \approx \Delta\Phi, \quad (5.48)$$

which is completely consistent with the calculations made there, coming from a different point of view. We see that the very weakest approach to a metric consistent with the Pound-Rebka experiment is given by

$$\mathbf{g} = (dx)^2 + (dy)^2 + (dz)^2 - [(1 + \Phi)dt]^2. \quad (5.49)$$

We then want to see how all this agrees with the form of Einstein's equations determined above, via the requirement that $\mathcal{R}_{44} = 4\pi\rho = \nabla^2\Phi$. To do this we will use the following orthonormal tetrad, where we also note below its dual, non-holonomic basis for tangent vectors, which we will use a bit later:

$$\begin{aligned} \varpi^i &= dx^i, \quad i = 1, 2, 3, & \varpi^4 &= (1 + \Phi)dt, \\ \tilde{e}_i &= \partial_{x^i}, \quad i = 1, 2, 3, & \tilde{e}_4 &= \frac{1}{1 + \Phi}\partial_t, \\ \varpi^\mu(\tilde{e}_\nu) &= \delta_\nu^\mu. \end{aligned} \quad (5.50a)$$

We then use the so-called ‘‘guess’’ method of determining the connections:

$$\begin{aligned} 0 &= d(dx^i) = d\varpi^i = \varpi^\mu \wedge \Gamma_{\mu}^i & \implies & \Gamma_{\mu}^i \propto \varpi^\mu, \\ d\varpi^4 &= (\partial_i\Phi) dx^i \wedge dt = \frac{\partial_i\Phi}{1 + \Phi} \varpi^i \wedge \varpi^4 & \implies & \Gamma_{i4} = (\partial_i\Phi)dt = \frac{\partial_i\Phi}{1 + \Phi} \varpi^4, \end{aligned} \quad (5.50b)$$

with all other connections forms vanishing. Since all quadratic products of the connections are therefore zero, it is simple to determine the Cartan curvature 2-forms:

$$\Omega_{i4} = d\Gamma_{i4} = (\partial_i\partial_j\Phi) dx^j \wedge dt = \frac{\partial_i\partial_j\Phi}{1 + \Phi} \varpi^j \wedge \varpi^4 \implies R_{i4j4} = \frac{\partial_i\partial_j\Phi}{1 + \Phi}. \quad (5.50c)$$

Knowing that $R_{4444} = 0$, we can calculate the Ricci tensor corresponding to this curvature:

$$\mathcal{R}_{44} = g^{ij} \frac{\partial_i \partial_j \Phi}{1 + \Phi} = \frac{\nabla^2 \Phi}{1 + \Phi} = \frac{4\pi\rho}{1 + \Phi} \approx 4\pi\rho, \quad (5.50d)$$

which was, of course, the expected result!

As another quick result from this very simple case, we can retreat back to the equations of motion for a particle moving in this (weak) field, as given earlier in Eqs. (3.3-4), where we had described the motion relative to a coordinate basis. However, we have calculated the connections just above in a non-holonomic basis; therefore, having them in hand, let's re-describe these motions in that basis, beginning with the 4-velocity of the particle:

$$\tilde{u} = \frac{dx^\mu}{d\tau} \partial_{x^\mu} = \frac{dx^i}{d\tau} \partial_{x^i} + \frac{dt}{d\tau} \partial_t = \frac{dx^i}{d\tau} \tilde{e}_i + (1 + \Phi) \frac{dt}{d\tau} \tilde{e}_4 \equiv u^\alpha \tilde{e}_\alpha, \quad (5.51a)$$

where we have used this method to define the components of the 4-velocity relative to this orthonormal, non-holonomic reference frame. We now consider the various components of the statement that this 4-velocity is tangent to a geodesic, using the just-determined connections given above in Eq. (5.50b):

$$\begin{aligned} \frac{d}{d\tau} u^\alpha &= -\Gamma^\alpha_{\beta\gamma} u^\beta u^\gamma, \\ \implies \frac{d}{d\tau} \frac{dx^i}{d\tau} &= \frac{du^i}{d\tau} = -\Gamma^i_{44} (u^4)^2, \quad \frac{du^4}{d\tau} = -\Gamma^4_{i4} u^i u^4, \end{aligned} \quad (5.51b)$$

where we note that our orthonormal frame tells us that $\Gamma^i_{44} = +\Gamma_{i44} = -\Gamma_{4i4} = +\Gamma^4_{i4}$, while $\Gamma_{44\mu} = 0$. Therefore the equation involving $u^4 = (1 + \Phi) dt/d\tau$ tells us

$$\frac{d}{d\tau} \left[(1 + \Phi) \frac{dt}{d\tau} \right] = \frac{du^4}{d\tau} = -\frac{\partial_i \Phi}{1 + \Phi} \frac{dx^i}{d\tau} (1 + \Phi) \frac{dt}{d\tau} = -\frac{\partial \Phi}{d\tau} \frac{dt}{d\tau} = -\frac{\partial(1 + \Phi)}{d\tau} \frac{dt}{d\tau}. \quad (5.51c)$$

This equation is easily integrated to say that

$$(1 + \Phi)^2 \frac{dt}{d\tau} = A, \quad A \text{ an arbitrary constant}, \quad (5.51d)$$

where it is worth noting—see solutions to HW4—that this was expected because the metric components do not depend explicitly on the time coordinate, t .

■ The General Equation for Geodesic Deviation,

We used the simple versions of gravitational forces created by weak fields, above, to understand the physical meanings of both the (Levi-Civita) connection and the Riemann curvature tensor. However, there is much more to say about that, in even the strongest of fields. We want now to derive the generic form of the equation describing how the curvature of the space-time in which they “live” causes two neighboring geodesics to deviate, one from the other—an equation also often referred to by the name the Jacobi equation.

As in the weak field case, we consider two neighboring geodesics, with the same tangent vector, $\tilde{u} \equiv d/d\lambda$, but with slightly different initial locations, so that there is a vector, $\tilde{\xi}$, connecting the points on these two geodesics that correspond to the same value of λ , and construe a coordinate system so that $\tilde{\xi} = \partial/\partial n$, with (λ, n) being two of the coordinates locally. This means that we have required that

$$[\tilde{u}, \tilde{\xi}] = 0 \implies \nabla_{\tilde{u}} \tilde{\xi} = \nabla_{\tilde{\xi}} \tilde{u}, \quad (5.52)$$

where this latter implication comes from the calculation we made for the torsion of two (nearby) tangent vectors, which we have now set equal to zero.

Having this construction we may now ask our basic question, namely, how does this vector vary as we move along the geodesics, i.e., how does it vary as λ varies. To answer this question we refer back to the effect of the curvature tensor on 3 vectors:

$$R(\tilde{\xi}, \tilde{u})\tilde{u} \equiv \nabla_{\tilde{\xi}} \nabla_{\tilde{u}} \tilde{u} - \nabla_{\tilde{u}} \nabla_{\tilde{\xi}} \tilde{u} - \nabla_{[\tilde{u}, \tilde{\xi}]} \tilde{u} = -\nabla_{\tilde{u}} \nabla_{\tilde{\xi}} \tilde{u}, \quad (5.53)$$

where the first term on the left vanished because \tilde{u} was a geodesic, and the last term on the left vanished because the two sets of curves commute, one with the other. However, as already noted above, the vanishing of the torsion requires the commutativity shown in Eq. (5.52).

Therefore, we may re-write the last equation as

$$\nabla_{\tilde{u}}^2 \tilde{\xi} = +R(\tilde{u}, \tilde{\xi})\tilde{u} = -R(\tilde{\xi}, \tilde{u})\tilde{u}, \quad (5.54)$$

which tells us that, even in the most general of situations, the first set of measurements, local but not so local that only special relativity is involved, can determine the values of the curvature tensor.

VI. More discussion on p-forms in 4 dimensions, and the Hodge duality mapping between them

1. Over 4-dimensional manifolds, there are 5 distinct spaces of p -forms, Λ^p :
 - i. Λ^0 is just the space of continuous (C^∞) functions, also denoted by \mathcal{F} . We say that it has dimension 1, since no true “directions” are involved.
 - ii. Λ^1 is the space of 1-forms, already considered; it has as many dimensions as the manifold, so for 4-dimensional spacetime, it has dimension 4.
 - iii. Λ^2 is the space of 2-forms, i.e., skew-symmetric tensors, or linear combinations of wedge products of 1-forms; therefore in general it has dimension $\frac{1}{2}n(n-1)$, which becomes 6 for 4-dimensional spacetime. A basis can be created by taking all wedge products of the basis set for 1-forms: $\{\omega^\alpha \wedge \omega^\beta \mid \alpha, \beta = 1, \dots, 4; \alpha < \beta\}$. Below we will see that we can use the ideas of Hodge duality to divide this set of 6 basis vectors into two triplets of basis vectors—in a Lorentz invariant way—although we will have to allow complex-valued expressions in order to do so. We will refer to those two triplets as a basis set for the self-dual and anti-self-dual sectors of Λ^2 , respectively.
 - iv. Λ^3 is the space of 3-forms, i.e., linear combinations of wedge products of 1-forms, three at a time; in general it has dimension $\binom{n}{3} = \frac{1}{6}n(n-1)(n-2)$, which becomes 4 for 4-dimensional spacetime. Once again, Hodge duality will provide a mapping between these 4 dimensions and the 4 dimensions of Λ^1 .
 - v. Λ^4 is the space of 4-forms; in general it has dimension $\binom{n}{4}$. For 4-dimensional spacetime, this is a 1-dimensional space; i.e., every 4-form is proportional to every other. We make a particular choice of basis for this space and refer to this as *the volume form*. Because of skew symmetry there are only two possible different choices, which differ only by a minus

sign; these two different choices are usually referred to as *a choice of orientation*. It is this that is referred to, in freshman physics, as *the right-hand rule*, i.e., that particular orientation is chosen, rather than the one based on the left hand. (In a more general m -dimensional space, the volume form is always an m -form.)

- vi. Over m -dimensional spacetime, it is impossible to have more than m things skew all at once; therefore, the volume form is always the last in the sequence of basis sets for p -forms. So, in 4 dimensions, there is no Λ^p for $p \geq 5$.
- vii. The union of all m of the non-zero vector spaces Λ^p is sometimes referred to as the entire Grassmann algebra of p -forms over a manifold, and is denoted simply by Λ . (It is troublesome, occasionally, that the superscript 1 on Λ^1 is sometimes dropped, allowing some possible confusion.)

2. Working in the usual (local) Minkowski coordinates, where it is reasonable to choose $\{dx, dy, dz, dt\}$ as a basis for 1-forms, we choose the particular 4-form

$$\mathcal{V} \equiv dx \wedge dy \wedge dz \wedge dt, \quad \text{the (standard) volume form} \quad (6.1)$$

as our choice of a volume form. Do notice that the alternate choice, where dt comes first in the sequence, has the opposite sign, i.e., the opposite orientation.

More generally, if $\{\varpi^\alpha\}_1^4$ is an arbitrary basis for 1-forms, in particular not necessarily orthonormal, we may define **the very important tensor quantity** $\eta^{\alpha\beta\gamma\delta}$, often named as the Levi-Civita tensor, which gives the “components” of the volume form relative to an arbitrary choice of basis:

$$\varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma \wedge \varpi^\delta \equiv \eta^{\alpha\beta\gamma\delta} \mathcal{V}, \quad (6.2)$$

$$\begin{aligned} \implies \mathcal{V} &= \frac{(-1)^s}{4!} \eta_{\alpha\beta\gamma\delta} \varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma \wedge \varpi^\delta \\ &\equiv \frac{(-1)^s}{4!} \{g_{\alpha\rho} g_{\beta\sigma} g_{\gamma\tau} g_{\delta\varphi} \eta^{\rho\sigma\tau\varphi}\} \varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma \wedge \varpi^\delta, \end{aligned} \quad (6.3)$$

where s is the number of linearly-independent timelike basis vectors available, either 1 for spacetime or 0 if we are restricting our studies to 3-dimensional, spatial-only manifolds.

This tensor is completely skew-symmetric, i.e., it changes sign when any two indices are interchanged, and so must be proportional to the *Levi-Civita symbol*, $\epsilon^{\alpha\beta\gamma\delta}$, used for determinants. One verifies that the following defines the Levi-Civita tensor of rank $\left\{ \begin{smallmatrix} 4 \\ 0 \end{smallmatrix} \right\}$, and its associated tensor of rank $\left\{ \begin{smallmatrix} 0 \\ 4 \end{smallmatrix} \right\}$, related as usual by the raising/lowering of indices via the metric tensor, where G is the matrix presentation for the components of the metric \mathbf{g} which goes along with our choice of basis, $\{\omega^\mu\}_1^4$, while $\boldsymbol{\eta}$ is the matrix presentation of the standard, Minkowski-basis metric from special relativity:

$$\eta^{\alpha\beta\gamma\delta} = \frac{1}{m} \epsilon^{\alpha\beta\gamma\delta} \quad , \quad \eta_{\alpha\beta\gamma\delta} = g_{\alpha\mu} g_{\beta\nu} g_{\gamma\rho} g_{\delta\tau} \eta^{\mu\nu\rho\tau} = (-1)^s m \epsilon_{\alpha\beta\gamma\delta} \quad ,$$

where $m \equiv \det(M)$, $G = M^T \boldsymbol{\eta} M$, (6.4)

and $g \equiv \det(G) = m^2 \det(\boldsymbol{\eta}) = (-1)^s m^2$, or $m = \sqrt{(-1)^s \det(G)}$

and one chooses $s = 0$ or 1 , as the number of timelike directions.

The matrix M^{-1} creates the congruency transformation that Sylvester's theorem asserts exists, that puts the metric into its normal form.

The values of $\eta^{\alpha\beta\gamma\delta}$ depend on the basis chosen; however, let us first consider briefly the problem for an orthonormal tetrad, or triad, in (4-dimensional) spacetime or 3-dimensional space where the matrix M , above, is just the identity matrix, so that $m = 1$:

- a. when the metric components are just $\eta_{\mu\nu}$ as they would be with Minkowski coordinates $\{x, y, z, t\}$, we must choose $s = 1$, which then implies that $\eta^{1234} = +1 = -\eta_{1234} = +\eta_{2341}$; however, if the arbitrary metric, $\{\omega^\mu\}_1^4$, is chosen to be **any orthonormal tetrad**, $\{\hat{\omega}^\mu\}_1^4$, then it is also true that the metric is given by $\eta_{\mu\nu}$, and therefore the values of $\eta^{\alpha\beta\gamma\delta}$ are the same as presented just above;
- b. or, when we are in ordinary, Cartesian coordinates, $\{x, y, z\}$, in 3-dimensional space, we choose $s = 0$, although $s = 2$ also has some advantages; then we simply have $\eta^{123} = +1 = \eta^{231} = \eta^{312} = \eta_{123}$.

3. the (Hodge) dual, $*$: $\Lambda^p \longrightarrow \Lambda^{m-p}$

Since we have already seen that the vector spaces Λ^p and Λ^{m-p} have the same dimension, when defined over an m -dimensional manifold, it is reasonable that there should be a useful mapping between the two. It was first studied by Hodge. We have not previously discussed it because—at least in this form about to be presented, which is the way we will always use it—it requires a metric to define it. Therefore, we now let α be an arbitrary p -form; then we denote the (Hodge) dual by $^*\alpha$, a p' -form, where we habitually use $p' \equiv m - p$ as a useful abbreviation. The two are related as follows:

$$\begin{aligned} \alpha &= \frac{1}{p!} \alpha_{\mu_1 \dots \mu_p} \omega^{\mu_1} \wedge \dots \wedge \omega^{\mu_p} \quad , \\ ^*\alpha &\equiv \frac{i^{pp'+s}}{p!p'} \alpha_{a_1 \dots a_p} g^{a_1 b_1} \dots g^{a_p b_p} \eta_{b_1 \dots b_p c_1 \dots c_{p'}} \omega^{c_1} \wedge \dots \wedge \omega^{c_{p'}} \quad (6.5) \\ \text{or } ^*\alpha_{c_1 \dots c_{p'}} &= i^{pp'+s} \left(\frac{1}{p!} \alpha_{a_1 \dots a_p} \right) g^{a_1 b_1} \dots g^{a_p b_p} \eta_{b_1 \dots b_p c_1 \dots c_{p'}} \quad , \end{aligned}$$

where the last line just shows explicitly the components of the dual $(m - p)$ -form.

The factors of $i \equiv \sqrt{-1}$ have been inserted in **just such a way that** the dual of the dual brings one back to where she started:

$$^*\{^*\alpha\} = \alpha \quad . \quad (6.6)$$

There are various conventions concerning the i 's in the definition. My convention, using the factors of i , allows for eigen-2-forms of the * operator, since Eq. (6.6) obviously tells us that the eigenvalues of the duality operator, * , are just ± 1 . Many authors omit this extra factor, which causes the eigenvalues to be $\pm i$ raised to a power that depends on p , but which then does not insert factors of i in the process of taking the dual of some tensor. As it turns out, later, I believe that there is some value in having such extra i 's when one wants to look at tensors as complex objects, but, as I say, there is considerable disagreement. My approach comes from Plebański.

As some examples for the use of the exterior derivative and the Hodge dual, let us begin by calculating the action of what will turn out to be the wave operator on a scalar function of the

coordinates, which is sometimes called the d'Alembert operator, or the de Rham Laplacian. We begin with the simplest case, which is a scalar field, such as the electric potential, $V = V(x^\beta)$, for some choice of coordinates on the manifold. Since we do not need more generality than necessary, we will do this calculation in the usual 4-dimensional spacetime, so that the sign factor in the definition of the Hodge dual, s , has value $s = 1$, and of course the metric has negative determinant, which we write out as $\det g = -m^2$, i.e., $m = \sqrt{-\det g}$. To begin, we suppose given a set of basis vectors, $\{\tilde{e}_\mu\}_1^4$ for the tangent vectors and a dual set of basis 1-forms, $\{\varpi^\nu\}_1^4$, for 1-forms. Then it is straightforward to write out its exterior derivative:

$$dV = (V_{,\mu})\varpi^\mu, \quad V_{,\mu} \equiv \tilde{e}_\mu(V) \equiv (\tilde{e}_\mu)^\alpha \frac{\partial}{\partial x^\alpha} V(x^\beta). \quad (6.7)$$

However, at this point we have a 1-form, and I want to determine its Hodge dual, which should be a 3-form. The formula in Eqs. (6.5) tells us that

$$*dV = \frac{i^4}{3!} (V_{,\mu}) g^{\mu\nu} \eta_{\nu\alpha\beta\gamma} \varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma = -\frac{m}{3!} V_{,\mu} g^{\mu\nu} \epsilon_{\nu\alpha\beta\gamma} \varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma. \quad (6.8)$$

Now we would like to take the exterior derivative a second time:

$$d^*dV = -\frac{1}{3!} (mV_{,\mu} g^{\mu\nu})_{,\lambda} \epsilon_{\nu\alpha\beta\gamma} \varpi^\lambda \wedge \varpi^\alpha \wedge \varpi^\beta \wedge \varpi^\gamma, \quad (6.9)$$

which gives us a 4-form, which we know will be dual to a scalar function. Therefore, we should now take the Hodge dual one more time:

$$*d^*dV = -\frac{i^1}{3!} (mV_{,\mu} g^{\mu\nu})_{,\lambda} \epsilon_{\nu\alpha\beta\gamma} g^{\lambda\rho} g^{\alpha\sigma} g^{\beta\tau} g^{\gamma\phi} \eta_{\rho\sigma\tau\phi}. \quad (6.10)$$

We now simplify this form by remembering to calculate the determinant of the (inverse) metric which is hidden in it, which gives us

$$*d^*dV = -\frac{i^1}{3! m} (mV_{,\mu} g^{\mu\nu})_{,\lambda} \epsilon_{\nu\alpha\beta\gamma} \epsilon^{\lambda\alpha\beta\gamma} = -\frac{i}{m} (mV_{,\mu} g^{\mu\nu})_{,\lambda} \delta_\nu^\lambda = -\frac{i}{m} (mV_{,\mu} g^{\mu\lambda})_{,\lambda}, \quad (6.11)$$

where we have used a form from Eqs. (4.19) to simplify the partially-summed product of two Levi-Civita symbols. One can see that, if we were using Cartesian coordinates, where $m = +1$,

and in flat space so that the metric is the usual diagonal one, $\eta^{\mu\nu}$, then the calculation above would simply be such that

$$i(*d^*dV + d^*d^*V) = \square^2V \equiv \nabla^2V - \frac{\partial^2}{\partial t^2}V, \quad (6.12)$$

where we are allowed to add the second term, on the left-hand side, since it has just the value zero, since d^*V is a 5-form, and of course all 5-forms are zero in 4-dimensional space. However, we have added it because it “provides symmetry” to the expression, and, much more importantly, because in the more general cases where we want this expression on an arbitrary p -form, it is needed in order to obtain the result on the right-hand side, although I do not here provide a proof of that. [Note that some texts, such as Stephani, for example, do NOT put a superscript 2 on his use of the symbol \square , although he means the same thing I have defined above.]

A different way to get over the fact that (Hodge) duality appears complicated is to write it down for **all** plausible exemplars that may occur, in our 4-dimensional spacetime, at least in the standard Minkowski tetrad, $\{dx, dy, dz, dt\} = \{\omega^\mu\}_1^4$, and the bases of each distinct space of p -forms, Λ^p , remembering that taking the dual twice simply gets you back to where you began:

$$\begin{aligned} \Lambda^1 \leftrightarrow \Lambda^3 : * \begin{pmatrix} dx \\ dy \\ dz \\ dt \end{pmatrix} &= - \begin{pmatrix} dy \wedge dz \wedge dt \\ dz \wedge dx \wedge dt \\ dx \wedge dy \wedge dt \\ dx \wedge dy \wedge dz \end{pmatrix}, \\ \Lambda^0 \leftrightarrow \Lambda^4 : *1 &= -i dx \wedge dy \wedge dz \wedge dt = -i\mathcal{V}, \\ \Lambda^2 \leftrightarrow \Lambda^2 : * \begin{pmatrix} dx \wedge dy \\ dy \wedge dz \\ dz \wedge dx \end{pmatrix} &= -i \begin{pmatrix} dz \wedge dt \\ dx \wedge dt \\ dy \wedge dt \end{pmatrix}, \quad . \end{aligned} \quad (6.13)$$

On the other hand, were we to perform the same calculation in 4-dimensional spacetime in spherical coordinates, in the coordinate basis $\{dr, d\theta, d\varphi, dt\}$, then we would obtain

$$* \begin{pmatrix} dr \\ d\theta \\ d\varphi \\ dt \end{pmatrix} = -r^2 \sin \theta \begin{pmatrix} d\theta \wedge d\varphi \wedge dt \\ \frac{1}{r^2} d\varphi \wedge dr \wedge dt \\ \frac{1}{r^2 \sin^2 \theta} dr \wedge d\theta \wedge dt \\ dr \wedge d\theta \wedge d\varphi \end{pmatrix}. \quad (6.14)$$

Since the Hodge dual maps 2-forms into 2-forms, it follows that there can be self-dual 2-forms, i.e., those which are mapped into themselves under the duality operation, and, of course, also those which are mapped into their own negative, under this operation. We can see this quite easily via the following calculation, where \underline{F} is an arbitrary 2-form:

$$\underline{F} = \frac{1}{2} (\underline{F} + *\underline{F}) + \frac{1}{2} (\underline{F} - *\underline{F}) .$$

The validity of the equality is obvious, but because the dual of the dual of a p -form is the same p -form back again, then it must be that the first full expression on the right-hand side is self-dual while the second expression on that side is anti-self-dual. As an example of the details of the calculation, consider the electromagnetic 2-form, or Faraday, which has the following form in terms of the Cartesian components of the electric and magnetic field 3-vectors:

$$\frac{1}{2} F_{\mu\nu} \varpi^\mu \wedge \varpi^\nu = \underline{F} \implies F_{\mu\nu} = \begin{pmatrix} 0 & B^z & -B^y & E_x \\ -B^z & 0 & B^x & E_y \\ B^y & -B^x & 0 & E_z \\ -E_x & -E_y & -E_z & 0 \end{pmatrix} , \quad (6.15)$$

$$\frac{1}{2} (*F_{\alpha\beta}) \varpi^\alpha \wedge \varpi^\beta = *\underline{F} \implies (*F)_{\alpha\beta} = -i \begin{pmatrix} 0 & -E^z & E^y & B_x \\ E^z & 0 & -E^x & B_y \\ -E^y & E^x & 0 & B_z \\ -B_x & 0 & -B_y & 0 \end{pmatrix} .$$

Note that the map that sends $-i\underline{F}$ to $*\underline{F}$ is accomplished by sending $\vec{B} \rightarrow -\vec{E}$ and $\vec{E} \rightarrow +\vec{B}$; this is a symmetry of Maxwell's equations originally discovered by Maxwell and Hertz. They saw this because of the intriguing properties of the *self-dual part* of this tensor, which can be completely characterized by a single, 3-dimensional but complex vector, $\vec{C} \equiv \vec{E} - i\vec{B}$:

$$\underline{W} \equiv \underline{F} + *\underline{F} = \frac{1}{2} W_{\mu\nu} \varpi^\mu \wedge \varpi^\nu , \quad W_{\mu\nu} \implies \begin{pmatrix} 0 & iC^z & -iC^y & C_x \\ -iC^z & 0 & iC^x & C_y \\ iC^y & -iC^x & 0 & C_z \\ -C_x & -C_y & -C_z & 0 \end{pmatrix} . \quad (6.16)$$

As yet a different aspect of the picture, we also give details for **3-dimensional, Euclidean space**, with Cartesian basis, $\{dx, dy, dz\} = \{\varpi^a\}_1^3$, and I insert a subscript 3 on the star for the Hodge dual in this 3-dimensional space:

$$\Lambda^1 \leftrightarrow \Lambda^2 : \begin{matrix} * \\ \text{\scriptsize 3} \end{matrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = - \begin{pmatrix} dy \wedge dz \\ dz \wedge dx \\ dx \wedge dy \end{pmatrix} , \quad \Lambda^0 \leftrightarrow \Lambda^3 : \begin{matrix} * \\ \text{\scriptsize 3} \end{matrix} 1 = dx \wedge dy \wedge dz \quad . \quad (6.17)$$

With our understanding of the Hodge dual in a 3-dimensional space, we may now look at that 3×3 submatrix of the electromagnetic 2-form that contains only spatial parts, i.e., that only involves the magnetic field, but oriented in perhaps an unexpected way, and consider it as a matrix presentation of the components of a 2-form in 3-dimensional space, with basis $\{dx, dy, dz\}$:

$$\begin{pmatrix} 0 & B^z & -B^y \\ -B^z & 0 & B^x \\ B^y & -B^x & 0 \end{pmatrix} \leftarrow \equiv \mathcal{B} = B^z dx \wedge dy + B^y dz \wedge dx + B^x dy \wedge dz \quad (6.18)$$

$$\implies \begin{matrix} - * \\ \text{\scriptsize 3} \end{matrix} \mathcal{B} = B^z dz + B^y dy + B^x dx .$$

This shows us that in order to properly move the usual, 3-dimensional, magnetic-field vector, \vec{B} , into a 4-dimensional spacetime, and make appropriate its relationship with the 3-dimensional electric-field vector, \vec{E} , we must first take its dual, making it part of a 2-form, instead of a 1-form. [This is what is sometimes stated as saying that \vec{B} is a different sort of vector than \vec{E} . In particular, when one considers their behavior under a parity transformation, \vec{E} changes sign, but \vec{B} does not.] It is of course also true that this particular way of uniting the two quantities that one thought were both 3-vectors, back in 3-dimensional space, causes the joint to transform in a simple, tensorial, way in the entire spacetime. As a slightly different way of looking at the same thing, we now notice that this formulation of the components of \vec{B} into the Faraday 2-form allows the matrix product to generate a cross product in the 3-dimensional vector space:

$$\begin{pmatrix} 0 & B^z & -B^y \\ -B^z & 0 & B^x \\ B^y & -B^x & 0 \end{pmatrix} \begin{pmatrix} A^x \\ A^y \\ A^z \end{pmatrix} = \begin{pmatrix} A^y B^z - A^z B^y \\ A^z B^x - A^x B^z \\ A^x B^y - A^y B^x \end{pmatrix} \leftarrow \equiv \vec{A} \times \vec{B} , \quad (6.19)$$

so that the 3-dimensional matrix portion itself may be thought of much as the operator $-\vec{B} \times$.

4. the (Hodge) dual in the Null Tetrad Basis

As a last example, let us see how all this looks in a null tetrad basis, of the sort described near Eqs. (5.14-16) above. We first determine the volume form for this basis, using the matrix A , from Eqs. (5.14), (which has determinant $-i$), that transforms an orthonormal basis to this null tetrad:

$$\eta^{\alpha\beta\gamma\delta} = A^\alpha{}_\mu A^\beta{}_\nu A^\gamma{}_\lambda A^\delta{}_\eta \epsilon^{\mu\nu\lambda\eta} \implies \det(A) = \eta^{1234} = -i = \eta_{1234} = -\det(A^{-1}). \quad (6.20)$$

We may then use this to determine the Hodge dual of a 2-form in the null tetrad basis:

$$\Lambda^2 \leftrightarrow \Lambda^2 : * \begin{pmatrix} \varrho^1 \wedge \varrho^2 \\ \varrho^2 \wedge \varrho^3 \\ \varrho^3 \wedge \varrho^1 \\ \varrho^1 \wedge \varrho^4 \\ \varrho^2 \wedge \varrho^4 \end{pmatrix} = \begin{pmatrix} -\varrho^3 \wedge \varrho^4 \\ +\varrho^2 \wedge \varrho^3 \\ -\varrho^3 \wedge \varrho^1 \\ +\varrho^1 \wedge \varrho^4 \\ -\varrho^2 \wedge \varrho^4 \end{pmatrix}. \quad (6.21)$$

One can see that this allows us to very easily pick out the (two) **3-complex dimensional** subspaces of Λ^2 that correspond to *self-dual* and *anti-self-dual* 2-forms, and describe an appropriate basis set for them, by taking half the sum (or difference) of each basis 2-form and its Hodge dual:

$$\begin{aligned} \Lambda_{SD}^2 : \quad \text{basis is} & \begin{pmatrix} \varrho^2 \wedge \varrho^3 \\ \frac{1}{\sqrt{2}}(-\varrho^1 \wedge \varrho^2 + \varrho^3 \wedge \varrho^4) \\ \varrho^1 \wedge \varrho^4 \end{pmatrix} \equiv \mathcal{Z}^a, \\ \Lambda_{aSD}^2 : \quad \text{basis is} & \begin{pmatrix} \varrho^1 \wedge \varrho^3 \\ \frac{1}{\sqrt{2}}(\varrho^1 \wedge \varrho^2 + \varrho^3 \wedge \varrho^4) \\ \varrho^2 \wedge \varrho^4 \end{pmatrix} \equiv \mathcal{Z}^{\dot{a}}, \end{aligned} \quad (6.22)$$

and remembering that ϱ^2 is the complex conjugate of ϱ^1 , we see that the anti-self-dual basis set is simply the complex conjugate of the self-dual basis set, which explains the “dot” over the index a on the anti-self-dual basis set. We intend that index to take on values of $+$, 0 , and $-$, going down the 3-vector defining them. With that choice the components of the self-dual part of the Faraday are simply the 3 (null) components of the vector \vec{C} defined above at Eqs. (6.16), while the self-dual part has the complex conjugate of the components of that vector.

Let us now recall the electromagnetic field tensor, \underline{F} , and (twice) its self-dual part, written as \underline{W} , at Eqs. (6.16), relative to an orthonormal basis. We may, however, now transform it to

a null basis using the matrix A^{-1} given at Eq. (5.16), denoting its components relative to this null basis by an overtilde as we have done with other tensors:

$$\widetilde{W}_{\alpha\beta} = (A^{-1})^\mu{}_\alpha (A^{-1})^\nu{}_\beta W_{\mu\nu} = \begin{pmatrix} 0 & -C_z & 0 & C_x - iC_y \\ C_z & 0 & -C_x - iC_y & 0 \\ 0 & C_x + iC_y & 0 & C_z \\ -C_x + iC_y & 0 & -C_z & 0 \end{pmatrix}, \quad (6.23)$$

where we see that there are entries of 0 in those places in the matrix where the anti-self-dual basis set would be, as of course should be the case.

We now define an appropriate set of components for our 3-dimensional, complex-valued vector, \vec{C} , to go “nicely” with our set of basis self-dual 2-forms defined in Eqs. (6.22):

$$\vec{C} \implies \begin{pmatrix} C_+ \\ C_0 \\ C_- \end{pmatrix} \equiv \begin{pmatrix} -\frac{1}{\sqrt{2}}(C_x + iC_y) \\ C_z \\ +\frac{1}{\sqrt{2}}(C_x - iC_y) \end{pmatrix}, \quad (6.24)$$

which allows this self-dual electromagnetic field to be written as

$$\begin{aligned} \frac{1}{2}\widetilde{W}_{\alpha\beta}\varrho^\alpha \wedge \varrho^\beta &= \mathbb{W} = \sqrt{2}(C_+\mathcal{Z}^+ + C_0\mathcal{Z}^0 + C_-\mathcal{Z}^-) \equiv \sqrt{2}C_a\mathcal{Z}^a, \\ \widetilde{W}_{\alpha\beta} &= \begin{pmatrix} 0 & -C_0 & 0 & \sqrt{2}C_- \\ C_0 & 0 & \sqrt{2}C_+ & 0 \\ 0 & -\sqrt{2}C_+ & 0 & C_0 \\ -\sqrt{2}C_- & 0 & -C_0 & 0 \end{pmatrix}. \end{aligned} \quad (6.25)$$

5. Use of p-forms to describe Areas, Volumes, etc.

a. (2-dimensional) Surfaces and their Areas

Our usual notion for an area is that it is determined by two vectors that are not parallel; i.e., it is the area enclosed by the parallelogram created from the sides of the two vectors. The two vectors are not unique in the sense that if we add some fraction of the one vector to the other one it will not change the enclosed area. If we had a metric then we could determine the area enclosed by determining the lengths of the two vectors and the angle between them; nonetheless, even without a metric, we do have a geometric object that takes two vectors and gives us a number, and also which gives us zero if the two vectors are parallel. That object is the 2-form, i.e., a skew-symmetric rank $\left\{ \begin{smallmatrix} 0 \\ 2 \end{smallmatrix} \right\}$ tensor. Since it is skew symmetric it changes sign

if we change the order of the two vectors determining the area; however, this is just the usual situation that we want to distinguish the upward and downward normals, or the inner and outer normals. Therefore, this says that the operator that describes an area, locally, at some given point, is a 2-form, defined in an appropriate vector space at that point. To determine the actual area desired, we need to be given two vector fields in that neighborhood, and integrate over the region desired, **but also** we need to have an actual definition of length, i.e., some sort of scale factor that says whether we are using inches or centimeters or miles, and also whether that scale factor varies from point to point, which is to say that for that purpose we will need a **metric tensor**. The notion as to how that scale factor is actually introduced is most easily obtained by beginning from the overall scale factor for the manifold itself, i.e., with the volume form. We will discuss this below; however, now, ignoring that scaling problem for the moment, one may certainly say that a different way to describe all this discussion of area is to say that 2-forms are the objects which need to be under integral signs, and need to have their desired two vectors telling them over what domain to obtain the needed value for the integral.

For example the 2-form $dx \wedge dy$, where $\{x, y\}$ are coordinates near some point, gives us the object to integrate to determine an area of a flat, 2-dimensional plane where x and y may also be thought of as parameters to describe the region of the plane desired, and therefore to determine the ranges over which they should be varied to perform the integral. However, it is much more likely that some given area, in an arbitrarily-curved spacetime is not nice and flat like that. Instead, for a general 2-dimensional area, we will need to think of it in the same way that we thought about curves on our manifold, which were parametrized by some single real variable that gave a mapping into the manifold. Therefore, for a 2-dimensional subspace of the manifold, what we will call a surface there, we want a mapping

$$S : \mathbb{R} \times \mathbb{R} \rightarrow \mathcal{M}, \tag{6.26}$$

where we think of the two variables as “parameters” that describe points locally on the surface, i.e., for some real numbers, λ^1 and λ^2 , $S(\lambda^1, \lambda^2) = P \in \mathcal{M}$, and each of these parameters varies

over some range, so that one describes the entire surface. One may also think of the surface as described by a continuous family of curves on the surface where one of the parameters is maintained fixed while the other is allowed to vary, and different members of the family are labeled by the fixed one. In general such a family of curves is usually referred to as a *congruence of curves*. Obviously the tangent vectors to these spanning curves are

$$\partial_{\lambda^j} \equiv \frac{\partial}{\partial \lambda^j} = \frac{\partial x^\mu}{\partial \lambda^j} \frac{\partial}{\partial x^\mu}, \quad (6.27)$$

where we have used the fact that the general coordinates on the manifold, $\{x^\mu\}$, must be functions of this pair of parameters when they describe points that lie on the surface, i.e., one has $x^\mu = x^\mu(\lambda^1, \lambda^2)$, so that the expressions above for the tangent vectors make sense.

The 2-form that describes an infinitesimal portion of area on the surface, at some particular point, must then be

$$d\lambda^1 \wedge d\lambda^2 = \frac{\partial \lambda^1}{\partial x^\mu} \frac{\partial \lambda^2}{\partial x^\nu} dx^\mu \wedge dx^\nu. \quad (6.28)$$

When we look at how the various vector spaces vary as one moves around over a non-flat manifold, we will see that there is a direct correlation between the curvature of the manifold and the difference between a vector that has been moved around some closed path and the original vector at the beginning of the path. Because of this it is in general impossible to integrate geometric quantities that are not scalar functions. Therefore if we want to integrate something over a surface described by a 2-form, it would have to actually be that 2-form acting on the pair of vectors that define the surface locally. This says that if I need to integrate some 2-form \mathcal{Q} over some surface spanned as described above, then it is actually $\mathcal{Q}(\partial_\lambda^1, \partial_\lambda^2)$ that one integrates, as the parameters λ^j vary as desired to spell out the entire surface, i.e., the integral actually has the form

$$\int_S \mathcal{Q} \equiv \int_S d\lambda^1 \wedge d\lambda^2 \mathcal{Q}(\partial_\lambda^1, \partial_\lambda^2) = \int_{\lambda_0^1}^{\lambda_1^1} d\lambda^1 \int_{\lambda_0^2}^{\lambda_1^2} d\lambda^2 \mathcal{Q}(\partial_\lambda^1, \partial_\lambda^2), \quad (6.29)$$

where $\mathcal{Q}(\partial_\lambda^1, \partial_\lambda^2) = \frac{1}{2} \alpha_{\mu\nu} [x^\beta(\lambda^j)] \frac{\partial x^\mu}{\partial \lambda^m} \frac{\partial x^\nu}{\partial \lambda^n} d\lambda^m \wedge d\lambda^n$.

b. **Extension to higher-dimensional sub-manifolds**

In our 4-dimensional spacetime, in addition to the 1-dimensional curves and the 2-dimensional surfaces, we can have 3-dimensional “hypersurfaces,” and then infinitesimal pieces of the full 4-dimensional volume itself. The obvious extension of the above discussion for surfaces and their area integrals tells us that we should use 3-forms to describe hypersurfaces, locally, and 4-forms for a local description of volume itself. We have already discussed the volume form, \mathcal{V} , at Eqs. (6.2-3), although it is also worth writing it down in a somewhat simpler way, where I have inserted “hats” over the symbols for the 1-forms to indicate that here I certainly do intend that they denote an orthonormal basis here:

$$\mathcal{V} \equiv \hat{\omega}^1 \wedge \hat{\omega}^2 \wedge \hat{\omega}^3 \wedge \hat{\omega}^4 = \sqrt{-\det(G)} \, dx \wedge dy \wedge dz \wedge dt , \quad (6.30)$$

If we now choose some particular small volume—a 4-dimensional volume, in our spacetime—by choosing 4 linearly independent 4-vectors that determine the edges of the 4-dimensional parallelepiped they define—we may determine the numerical value of the volume defined by those four tangent vectors, say, $\{\tilde{\zeta}_\nu\}_1^4$, by determining $\mathcal{V}(\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3, \tilde{\zeta}_4)$ and integrating this as necessary over the 4-dimensional parameter space that defines it.

We now want to reduce this notion down to considerations of 3-dimensional volumes, but still in our 4-dimensional manifold. These we refer to as hypervolumes, or sometimes hypersurfaces, which in either case means a surface with one less dimension than that of the manifold. Such a surface can be defined by giving some single function of restraint, that constrains the 4 coordinates to lie on that surface. However, especially from the infinitesimal point of view, it can be described by giving either three, linearly-independent tangent vectors to some three-dimensional set of parametrizing curves, or by giving a 3-form that is waiting for those 3 vectors to determine its 3-volume. Choose the 3 parameters, say $\{\mu^i\}_1^3$, that define the surface, so that the manifold coordinates are then determined by $x^\alpha(\mu^i)$, and one could use the Hodge dual of the associated 3-form to define the surface with what amounts to its

“normal” 1-form:

$$\underline{\zeta}_\mu \equiv -\frac{1}{3!}\eta_{\mu\alpha\beta\gamma} dx^\alpha \wedge dx^\beta \wedge dx^\gamma \Big|_\Sigma, \quad (6.31)$$

where the minus sign comes from the fact that our metric, in spacetime, has negative determinant, and where the evaluation of the 1-forms on the surface requires re-thinking them in terms of the 3 parameters on the surface and re-expressing their differentials that way, i.e., $dx^\alpha(\lambda^j) = (\partial x^\alpha / \partial \lambda^j) d\lambda^j$.

This is also a good place to note that a 3-surface that one would/should call spacelike is one that would have a normal vector that is timelike, and, vice versa, a 3-surface with a spacelike normal is one that it is reasonable to refer to as a timelike surface since one of its tangent vectors is surely timelike.

Lastly, we return to the notion that a 3-surface is determined—on the manifold—by specifying a constant value for some single function of the 4 coordinates, i.e., $f(x^\mu) = 0$, determines a 3-surface, $f(x^\mu) = 1$ determines a related one, etc. We know that the associated 1-form for such a surface should have the form df , evaluated at whatever point on the manifold we like. We can think of this 1-form as lying in the vector space of 1-forms, and—under the influence of the metric—having an associated vector, locally in the tangent space of the manifold, that is made from this 1-form via the action of the metric, that maps 1-forms into tangent vectors (and of course vice versa). However, the normal vector is just a direction; therefore, for any function h on the manifold, the 1-form $h df$ would be just as good to specify the 3-surface in question. Now, given some arbitrary 1-form, say \mathcal{Q} , how do we know whether it can be “pushed down” onto the manifold to determine a single 3-surface, at least in some neighborhood. This is of course the same as asking whether or not there exist functions h and f such that $\mathcal{Q} = h df$. To answer this, we note that

$$d(h df) = dh \wedge df \implies (h df) \wedge d(h df) = h df \wedge dh \wedge df = 0. \quad (6.32)$$

Following this line of reasoning one can show that a given 1-form is *surface forming* if and only if it satisfies the property

$$\mathcal{Q} \wedge d(\mathcal{Q}) = 0 \iff \mathcal{Q} \text{ is surface-forming.} \quad (6.33)$$

One should note that the discussion above concerning 2-dimensional areas must be construed so that the correct scale factors introduced by a given metric definition of length require that we use orthonormal basis forms there, or, if you prefer, the appropriate reduction of the volume form to that 2-dimensional situation. It is rather more complicated since a 2-surface in a 4-space requires the specification of two linearly-independent normals in order to describe its orientation. Therefore, for some arbitrary surface S , with parameters $\{\lambda^i\}_1^2$, such that it is specified by a 2-form with choices of normals so that its 2-form may be given as

$$\mathcal{S}_{\mu\nu} \equiv -\frac{1}{2}\eta_{\mu\nu\alpha\beta} dx^\alpha \wedge dx^\beta \Big|_S. \quad (6.34)$$

As an example, to understand what the indices mean we consider the 2-form that would describe a simple surface with tangent vectors in the x, y -plane; it would therefore have normals proportional to the vectors ∂_z and ∂_t , so that we would want the 2-form \mathcal{S}_{34} .

c. Extrinsic Curvature

An important subject which has so far not been mentioned is how the curvature of a manifold looks from the point of view of some manifold with larger dimension from which we are viewing the original one. This view of the curvature of a manifold of dimension, say m , as measured from a manifold of dimension $n > m$ is referred to as the *extrinsic curvature* of the original manifold, while its internal curvature is of course measured by its Riemann tensor. An example of the importance of extrinsic curvature is given, for instance, by a cylinder which we might construct in 3-dimensional space. When the Riemann curvature of a cylinder is calculated the result is just zero. This result is not overly surprising—one hopes—since (internal) curvature is a locally-defined property of a manifold, available in any neighborhood of each and every point. At any given point on a cylinder one could imagine un-sealing the two edges of the cylinder which were joined and rolling it out and it does indeed appear flat. On the other hand, from our 3-dimensional perspective it is clear that the cylinder is (globally) curved. Therefore, we need a perspective of the curvature from a higher-dimensional point of view, at any time when we are considering subspaces of our 4-dimensional manifold. The ideas

are discussed at some length in Grøn's text, §7.2 and §7.4. Perhaps I will manage to write some more details concerning it for your use, but—at least for now—if you are interested you should consider those sections, and also Problem 7.4, on p. 173, which discusses the Gauss-Codazzi equations, which are the most important ingredients to begin a serious study of 3-dimensional (hyper)surfaces in 4-dimensional manifolds. They can also be used to answer questions such as can one have an $n > 4$ -dimensional, flat spacetime where one can view the 4-dimensional spacetime under study as a submanifold there?