Fundamentals of General Relativity

Christian Villandseie

December 2023

This is heavily inspired by [1].

1 The Equivalence Principle

The *equivalence principle* is one of the corners stones of general relativity. One formulation of the equivalence principle is that locally the laws of physics reduce to those of special relativity, and that it is impossible to detect the existence of a gravitational field locally. It is impossible to distinguish between uniform acceleration and a gravitational field. This means that locally the *metric* always reduces to the *Minkowski* metric, which is the metric describing special relativity.

2 Special Relativity

Before moving on to general relativity (GR) we should discuss the special case of GR in the absence of matter and energy, aptly named special relativity (SR). SR is based on two postulates:

1. The laws of physics take the same form in all inertial reference frames.
2. The speed of light measured in all inertial frames is always the same.

A consequence of these postulates is that there exists no inertial frames moving at the speed of light relative any other inertial frames, because then light would be at rest in this frame, and so the second postulate would be broken.

As we will see more formally later, we define the *line element* ds^2 as $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ where $g_{\mu\nu}$ is a general metric tensor and dx^μ is an infinitesimal increment in the x^μ coordinate. In special relativity we use the *Minkowski metric* $\eta_{\mu\nu} = \mathrm{diag}\,(-1, 1, 1, 1)$. The line element for Minkowski is therefore

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{1}$$

As we will show later this line element is coordinate invariant, meaning it takes the same value no matter which coordinate system it is evaluated in. Consider the time between two ticks on some clock. In the rest frame of the clock these two events, the ticks, happen at the same position. So $dx = dy = dz = 0$. Then $ds^2 = -dt^2$. The time measured in the rest frame of some object is called the object's *proper time* and is usually denoted by τ. Two events which have $ds^2 < 0$, like the two ticks on a clock, are said to have a *timelike* separation. Events with $ds^2 > 0$ have a *spacelike* separation, and events with $ds^2 = 0$ have a *lightlike* separation. Since the line element is coordinate invariant, this means that for any two events with timelike separation there exists a coordinate system in which these happen at the same position. So for timelike events we can always set $ds^2 = -d\tau^2$, where $d\tau$ is the proper time between the events. This means that for timelike events we can always write

$$
\begin{aligned}
-d\tau^2 &= -dt^2 + dx^2 + dy^2 + dz^2 \\
&= -dt^2 + (d\mathbf{x})^2 \\
&= -dt^2 \left(1 - \left(\frac{d\mathbf{x}}{dt}\right)^2\right) \\
&= -dt^2 \left(1 - \mathbf{v}^2\right),
\end{aligned}
$$

where we define $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ as usual. So $dt = \frac{d\tau}{\sqrt{1-\mathbf{v}^2}} = \gamma d\tau$, where we define $\gamma \equiv \frac{1}{\sqrt{1-\mathbf{v}^2}}$. The equation $dt = \gamma d\tau$ is the famous equation for the time dilation between observers in relative motion.

3 Four-vectors

The concept of *four-vectors* is related to the *vectors* we will encounter in the next section (which are not just the standard spatial vectors like $\mathbf{x}$). An example of a four-vector is $x^\mu = (t, x, y, z) = (t, \mathbf{x})$, sometimes called the *four-position*. We can define the *four-velocity* of some massive particle as $u^\mu = \frac{dx^\mu}{d\tau} = \left(\frac{dt}{d\tau}, \frac{dt}{d\tau}\frac{d\mathbf{x}}{dt}\right) \equiv \frac{dt}{d\tau}(1, \mathbf{v})$. The four-velocity has norm

$$\eta_{\mu\nu} u^\mu u^\nu = \left(\frac{dt}{d\tau}\right)^2 \left(-1 + \mathbf{v}^2\right) = -\gamma^2\gamma^{-2} = -1 \tag{2}$$

We can also define the *four-momentum* $p^\mu = \frac{dx^\mu}{d\lambda}$, where λ is an *affine parameter*, meaning it is related to the proper time τ through a linear transformation.

For massive particles we usually choose $\lambda = \frac{\tau}{m}$, with m being the mass of the particle, in which case we get

$$p^\mu = m\frac{dx^\mu}{d\tau} = mu^\mu = (\gamma m, \gamma m\mathbf{v})\,. \tag{3}$$

The norm of the four-momentum is $\eta_{\mu\nu}p^\mu p^\nu = m^2\eta_{\mu\nu}u^\mu u^\nu = -m^2$. These norms are a special case of a *contraction*, which is defined as $\eta_{\mu\nu}a^\mu b^\nu$ for arbitrary a^μ and b^ν. We define the relativistic energy to be $E = \gamma m$ and the relativistic momentum $\mathbf{p} = \gamma m\mathbf{v}$. The four-momentum of the massive particle can then be written as $p^\mu = (E, \mathbf{p})$. For massless particles we similarly define the four-momentum to be $p^\mu = (E, \mathbf{p})$, and choose λ correspondingly. Note that even in curved spacetime the norms of p^μ and u^μ are still $-m^2$ and -1 respectively. This is due to the equivalence principle - spacetime is always locally Minkowski, so we can always evaluate the four-momentum and four-velocity in frames which are locally Minkowski - and because contractions between vectors are coordinate invariant.

4 Differential geometry

4.1 Vectors

Consider a point p in spacetime. The set of *vectors* at p is denoted by T_p, and is called the *tangent space* at p. T_p is a vector space, meaning that

$$\forall a, b \in \mathbb{R} \quad V, W \in T_p$$

$$(a+b)(V+W) = aV + bV + aW + bW \in T_p$$

Any element V of T_p can be written as a linear combination of a set of basis vectors for T_p. Assume that we have one such set of basis vectors $\hat{e}_{(\mu)}$. Then we can write

$$V = V^\mu e_{(\mu)}, \tag{4}$$

where V^μ are the *components* of V in the basis $e_{(\mu)}$. Consider now a passive coordinate transformation where we simply transform our basis vectors. Our new basis vectors become

$$e'_{(\mu)} = \frac{\partial x^\nu}{\partial x'^\mu} e_{(\nu)}. \tag{5}$$

Since the vector V itself is just an object living in T_p, independent of the basis we choose to express it in, the change of basis can not change the vector itslf. So we need

$$V = V^{\mu} e_{(\mu)}$$

$$\stackrel{!}{=} V' = V'^{\mu} e_{(\mu)'}$$

$$= V'^{\mu} \frac{\partial x^{\nu}}{\partial x'^{\mu}} e_{(\nu)}$$

$$= V'^{\nu} \frac{\partial x^{\mu}}{\partial x'^{\nu}} e_{(\mu)}.$$

For these equalities to hold we need $V^{\mu} = V'^{\nu} \frac{\partial x^{\mu}}{\partial x'^{\nu}}$. Multiply this relation by $\frac{\partial x'^{\rho}}{\partial x^{\sigma}}$. The partial derivatives on the right hand side are equal to $\delta^{\rho}_{\nu} \delta^{\mu}_{\sigma}$ by the chain rule, where δ^{ν}_{μ} is the Kronecker-delta symbol. We get

$$V'^{\rho} = V^{\mu} \frac{\partial x'^{\rho}}{\partial x^{\mu}}. \tag{6}$$

So (6) specifies how the components of an element of T_p transform under a change of basis

$$e_{(\mu)} \to e'_{(\mu)} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} e_{(\nu)}. \tag{7}$$

4.2 Dual vectors

In addition to the tangent space T_p at each point in spacetime we also have the *cotangent space* T^*_p. Dual vectors are linear maps from the tangent space T_p to the real numbers $\mathbb{R}$, or stated more formally

$$\omega \in T^*_p : V \in T_p \to \omega(V) \in \mathbb{R} \tag{8}$$

In analogy with the basis vectors for the tangent space T_p we can introduce basis dual vectors $\theta^{(\nu)}$ and require that $\theta^{(\nu)}\left(e_{(\mu)}\right) = \delta^{\nu}_{\mu}$. The dual vector ω can then be written as a linear combination of the basis dual vectors $\theta^{(\mu)}$

$$\omega = \omega_{\mu} \theta^{(\mu)}, \tag{9}$$

where ω_μ are the components of the dual vector ω in the basis $\theta^{(\mu)}$. The linear map from T_p to $\mathbb{R}$ can then be written as

$$\omega(V) = \omega_\mu V^\nu \theta^{(\mu)}\left(e_{(\nu)}\right)$$

$$= \omega_\mu V^\nu \delta^\mu_\nu$$

$$= \omega_\mu V^\mu \in \mathbb{R}$$

Similarly we can also view vectors as being linear maps from the cotangent space T_p^* to $\mathbb{R}$. Let us now again consider a passive coordinate transformation $\theta'^{(\mu)} = \frac{\partial x'^\mu}{\partial x^\nu}\theta^{(\nu)}$. Then, since the dual vector ω is invariant under change of basis we have

$$\omega = \omega_\mu \theta^{(\mu)}$$

$$\stackrel{!}{=} \omega' = \omega'_\mu \theta'^{(\mu)}$$

$$= \omega'_\mu \frac{\partial x'^\mu}{\partial x^\nu}\theta^{(\nu)}$$

$$= \omega'_\nu \frac{\partial x'^\nu}{\partial x^\mu}\theta^{(\mu)}.$$

Again, in oreder for ω to be invariant we need $\omega_\mu = \omega'_\nu \frac{\partial x'^\nu}{\partial x^\mu}$. The components of dual vectors therefore transform as

$$\omega'_\rho = \omega_\mu \frac{\partial x^\mu}{\partial x'^\rho} \tag{10}$$

4.3 Tensors

The concept of a tensor is very closely related to vectors and dual vectors. A tensor of rank (k, l) is a multilinear map from k copies of T_p^* and l copies of T_p to $\mathbb{R}$. Or stated more formally

$$T : T_p^* \times ... \times T_p^* \times T_p \times ... \times T_p \to \mathbb{R} \tag{11}$$

with T_p^* appearing k times and T_p appearing l times. Informally we can therefore think of a rank (k, l) tensor as behaving like a collection of k vectors and l dual vectors.

Let T be a rank (k,l) tensor and S a rank (m,n) tensor. We can form a new tensor $T \otimes S$, which is a rank $(k+m, l+n)$ tensor. And

$$T \otimes S(\omega^{(1)}, ..., \omega^{(k+m)}, V^{(1)}, ..., V^{(l+n)})$$

$$= T(\omega^{(1)}, ..., \omega^{(k)}, V^{(1)}, ..., V^{(l)}) \cdot S(\omega^{(k+1)}, ..., \omega^{(k+m)}, V^{(l+1)}, ..., V^{(l+n)}).$$

Tensors of rank (m,n) form a vector space, meaning that we can write any rank (m,n) tensor as a linear combination of the basis "vectors" (meaning the (m,n) basis tensors). So for any rank (m,n) tensor we can write

$$T = T^{\mu_1 ... \mu_m}_{\nu_1 ... \nu_n} e_{(\mu_1)} \otimes ... \otimes e_{(\mu_m)} \otimes \theta^{(\nu_1)} \otimes ... \otimes \theta^{(\nu_n)}, \tag{12}$$

where $e_{(\mu_1)} \otimes ... \otimes e_{(\mu_m)} \otimes \theta^{(\nu_1)} \otimes ... \otimes \theta^{(\nu_n)}$ are these basis tensors, and $T^{\mu_1 ... \mu_m}_{\nu_1 ... \nu_n}$ are the components of T in that basis.

Keeping in mind the symmetries of a given tensor is often very important. We say that a tensor $T_{\mu\nu}$ is symmetric in μ and ν if $T_{\mu\nu} = T_{\nu\mu}$. The same holds for a tensor with upper indices, and for tensors of higher rank where the tensor is unchanged under the exchange of a pair of indices. Similarly tensor is antisymmetric in μ and ν if $T_{\mu\nu} = -T_{\nu\mu}$. We can form the symmetric part of any tensor $T_{\mu_1 ... \mu_n}$ in the following way

$$T_{(\mu_1 ... \mu_n)} = \frac{1}{n!} \left(T_{\mu_1 ... \mu_n} + T_{\mu_n ... \mu_1} + ... \right), \tag{13}$$

where the ... represents a sum over all the permutations of $\mu_1, ..., \mu_n$. And we denote symmetrized indices with round brackets. Similarly we can from the antisymmetric part of the same tensor

$$T_{[\mu_1 ... \mu_n]} = \frac{1}{n!} \left(T_{\mu_1 ... \mu_n} - T_{\mu_n ... \mu_1} + ... \right), \tag{14}$$

where the ... represents a sum over the permutations of the indices, but with alternating signs. And we denote antisymmetrized indices with square brackets. As an example, consider $T_{\mu\nu\rho}$. The symmetric part of this tensor is

$$T_{(\mu\nu\rho)} = \frac{1}{6} \left(T_{\mu\nu\rho} + T_{\mu\rho\nu} + T_{\rho\nu\mu} + T_{\rho\mu\nu} + T_{\nu\rho\mu} + T_{\nu\mu\rho} \right), \tag{15}$$

and the antisymmetric part is

$$T_{[\mu\nu\rho]} = \frac{1}{6} \left(T_{\mu\nu\rho} - T_{\mu\rho\nu} + T_{\rho\mu\nu} - T_{\rho\nu\mu} + T_{\nu\rho\mu} - T_{\nu\mu\rho} \right). \tag{16}$$

An important property of symmetric and antisymmetric tensors is that the contraction of a symmetric tensor with an antisymmetric tesnor is zero. Here we just show this for tensors with two indices, but the proof is really the same for an arbitrary number of indices, just more tedious. Assume that $T_{\mu\nu} = T_{\nu\mu}$ and $S^{\mu\nu} = -S^{\nu\mu}$. Then

$$T_{\mu\nu}S^{\mu\nu} = T_{\nu\mu}S^{\mu\nu} = T_{\mu\nu}S^{\nu\mu}. \tag{17}$$

In the first step we swap the indices in $T_{\mu\nu}$, and because the tensor is symmetric the contraction is left unchanged. Then in the next step I rename the indices $\mu \to \nu$ and $\nu \to \mu$. But the first and last expressions cannot be equal - since $S^{\mu\nu}$ is antisymmetric - unless the contraction is equal to zero. Now assume that we have some new $S^{\mu\nu} = A\mu\nu + B^{\mu\nu}$, where $A^{\mu\nu}$ is antisymmetric and $B^{\mu\nu}$ is symmetric. And consider what happens if we contract this with the symmetric $T_{\mu\nu}$. Only the symmetric part $B^{\mu\nu}$ survives the contraction. Now note that any tensor $S^{\mu\nu}$ can be written as a sum of a symmetric tensor and antisymmetric tensor, in the following way for tensors of two indices

$$S^{\mu\nu} = \frac{1}{2}\left(S^{\mu\nu} + S^{\nu\mu}\right) + \frac{1}{2}\left(S^{\mu\nu} - S^{\nu\mu}\right). \tag{18}$$

Combined with the previous observation this means that any time we contract a general tensor $S^{\mu\nu}$ with no special symmetry properties with a symmetric tensor $T_{\mu\nu}$ the result is always $T_{\mu\nu}$ contracted with the symmetric part of $S^{\mu\nu}$. So the contraction with a symmetric tensor picks out the symmetric part of the other tensor.

4.4 Bases for the tangent and cotangent spaces

Spacetime can be represented as an n-dimensional manifold M. And at each point p on M we have a tangent space T_p, which is the space of all vectors at p. We need some basis for the tangent spaces, and we can find this by considering the *directional derivative*. Consider curves $\gamma : \mathbb{R} \to M$, parameterized by λ, which pass through p. Each such curve through p defines an operator $\frac{d}{d\lambda}$, called the directional derivative. And it turns out that the space of directional derivative operators which pass through p form the tangent space T_p. We need to show that the space of directional derivatives is a vector space and that it can be identified with the tangent space.

Showing that the space of directional derivatives is a vector space is straight forward. Consider two operators $\frac{d}{d\lambda}$ and $\frac{d}{d\eta}$. We can form a linear combination of these to form the new operator $a\frac{d}{d\lambda} + b\frac{d}{d\eta}$ with $a, b \in \mathbb{R}$. We need to show that the linear combination is itself a derivative operator, meaning that it needs to act linearly and to obey the product rule.

$$\left(a\frac{d}{d\lambda}+b\frac{d}{d\eta}\right)(fg)=af\frac{dg}{d\lambda}+ag\frac{df}{d\lambda}+bf\frac{dg}{d\eta}+bg\frac{df}{d\eta}$$

$$=\left(a\frac{df}{d\lambda}+b\frac{df}{d\eta}\right)g+\left(a\frac{dg}{d\lambda}+b\frac{dg}{d\eta}\right)f.$$

Evidently the linear combination does have these two properties, and so the space of directional derivatives is a vector space. Now we need to show that the space of directional derivatives can be identified with the tangent space. For this we want to show that the vector space of directional derivatives has the same dimensionality as the manifold M and that the idea of a vector pointing in a certain direction naturally arises. Consider an n-dimensional manifold M and assume that we have a map from $\varphi : M \to \mathbb{R}^n$, a map $\gamma : \mathbb{R} \to M$ and a function $f : M \to \mathbb{R}$. γ can be thought of as a parameterized curve on M that takes a single real value as input and spits out a point on the manifold. φ is a coordinate chart, and can be thought of as a map that assigns coordinates to points on the manifold. And f is a function that maps points on the manifold to the real numbers. The following figure shows how one can move between spaces using γ, ϕ and f and their inverses.

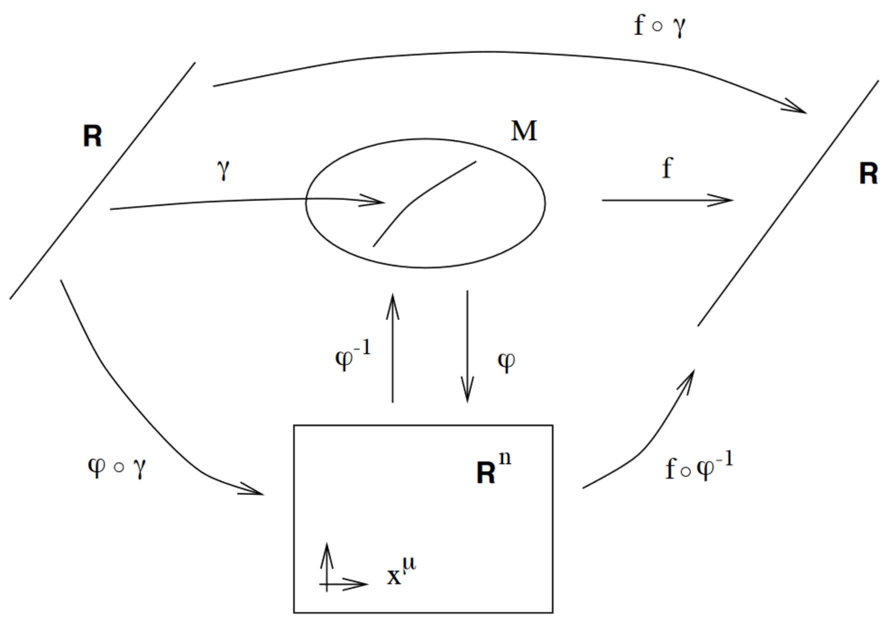

Figure 1: Sketch showing how γ, φ and f map between M, $\mathbb{R}^n$ and $\mathbb{R}$. The image is taken from [1].

Consider now a curve γ parameterized by λ. Let us compute the directional derivative

$$\begin{aligned} \frac{d}{d\lambda}(f \circ \gamma) &= \frac{d}{d\lambda}\left[\left(f \circ \varphi^{-1}\right) \circ (\varphi \circ \gamma))\right] \\ &= \frac{d\,(\varphi \circ \gamma)^{\mu}}{d\lambda} \frac{\partial\left(f \circ \varphi^{-1}\right)}{\partial x^{\mu}} \\ &= \frac{dx^{\mu}}{d\lambda} \partial_{\mu}\left(f \circ \varphi^{-1}\right). \end{aligned}$$

In the first equality we use the fact that there exists an inverse of φ. And in the second line we use the chain rule, remembering that $(\varphi \circ \gamma)^{\mu}$ represents coordinates in $\mathbb{R}^n$. And the last line is just inserting this fact and using the notation $\partial_{\mu} = \frac{\partial}{\partial x^{\mu}}$. Since the function f is arbitrary we can therefore write that

$$\frac{d}{d\lambda} = \frac{dx^{\mu}}{d\lambda}\partial_{\mu}. \tag{19}$$

This means that we can think of the set of partials ∂_{μ} as a basis for the vector space of directional derivatives. And of course these partials have the same dimensionality as M and very naturally lends themselves to describing vectors pointing in certain directions. So we can therefore identify the vector space of directional derivatives as the tangent space. This is to say that we have a basis consisting of partials ∂_{μ} for the tangent space. This basis is called a *coordinate basis* for T_p.

Now we want to form a basis for the cotangent space T_p^*. Remember that a dual vector is a linear map $\omega : T_p \to \mathbb{R}$. So we want a map from the directional derivatives to $\mathbb{R}$. Consider the action of the gradient df on the directional derivative. The gradient of a function f is defined as $df(V) = V(f)$, where $V \in T_p$. So with $V = \frac{d}{d\lambda}$ we get

$$df\left(\frac{d}{d\lambda}\right) = \frac{d}{d\lambda}(f) = \frac{df}{d\lambda} \in \mathbb{R},$$

meaning that df is a valid dual vector. And we have

$$dx^{\mu}\left(\partial_{\nu}\right) = \frac{\partial x^{\mu}}{\partial x^{\nu}} = \delta^{\mu}_{\nu}. \tag{20}$$

So the basis of gradients forms an appropriate basis for the cotangent space. A dual vector can therefore be written as $\omega = \omega_{\mu} dx^{\mu}$.

4.5 The metric tensor

The metric tensor is a non-degenerate, symmetric rank $(0,2)$ tensor, $g : T_p \times T_p \to \mathbb{R}$. The non-degeneracy means that $\det g \neq 0$, which in turn means that an inverse metric g^{-1} exists. We define the *inner product* between two vectors V and W as the action of the metric on V and W

$$g(V, W) = g(W, V) = g_{\mu\nu} V^\rho W^\sigma dx^\mu (\partial_\rho) \, dx^\nu (\partial_\sigma)$$

$$= g_{\mu\nu} V^\rho W^\sigma \delta^\mu_\rho \delta^\nu_\sigma$$

$$= g_{\mu\nu} V^\mu W^\nu.$$

Here we have used that in coordinate basis we have $g = g_{\mu\nu} dx^\mu \otimes dx^\nu \equiv ds^2$, where we also defined the *line element* ds^2. The inverse metric is defined through $g^{\mu\nu} g_{\nu\rho} = \delta^\mu_\rho$, so the inverse metric is a map $g^{-1} : T^*_p \times T^*_p \to \mathbb{R}$. Now consider what happens when acting with the metric on a single vector

$$g(V, \cdot) = g_{\mu\nu} V^\rho dx^\mu (\partial_\rho) \, dx^\nu$$

$$= g_{\mu\nu} V^\rho \delta^\mu_\rho dx^\nu$$

$$= g_{\mu\nu} V^\mu dx^\nu$$

Clearly this is a map from T_p to $\mathbb{R}$, which is the definition of a dual vector. So $g(V, \cdot)$ is a dual vector. we can define $g_{\mu\nu} V^\mu \equiv V_\nu$, in which case it is clear that the result is a dual vector. Similarly it can be shown that hte action of g^{-1} on a single dual vector is a vector. We can therefore say that the metric lowers indices, transforming vectors to dual vectors, while the inverse metric raises indices, transforming dual vectors to vectors.

5 Covariant derivatives and Christoffel symbols

We want to formulate our laws of physics in covariant forms, meaning that they are invariant under coordinate transformations. Tensorial equations are manifestly covariant, so we want to write our equations using tensors. Ideally we would want to use the partial derivative in our equations. Consider the partial derivative $\partial_\mu A^\nu$ of some arbitrary rank $(1,0)$ tensor A^ν. For this to be a tensor we need it to transform as $\partial_\mu A^\nu \to \partial'_\mu A'^\nu = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x'^\nu}{\partial x^\sigma} \partial_\rho A^\sigma$. But rather inconveniently it instead transforms as

$$\partial_\mu A^\nu \to \partial'_\mu A'^\nu = \frac{\partial x^\rho}{\partial x'^\mu}\partial_\rho \frac{\partial x'^\nu}{\partial x^\sigma}A^\sigma$$

$$= \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\sigma}\partial_\rho A^\sigma + \frac{\partial x^\rho}{\partial x'^\mu}A^\sigma \frac{\partial^2 x'^\nu}{\partial \rho \partial \sigma}.$$

So we get an extra term, meaning that $\partial_\mu A^\nu$ is not a tensor. There is a gemetric reason for this. Consider the vector $A = A^\mu e_{(\mu)}$, where we briefly go back to writing the basis vectors for T_p as $e_{(\mu)}$ for this discussion. Now

$$\frac{\partial A}{\partial x^\mu} = \frac{\partial A^\nu}{\partial x^\mu} e_{(\nu)} + A^\nu \frac{\partial e_{(\nu)}}{\partial x^\mu}. \tag{21}$$

The first term describes how the vector field A^μ changes from point to point, while the second term describes how the basis vectors $e_{(\mu)}$ change. It is this second term which prevents $\partial_\mu A^\nu$ from being a tensor. Now comes a small "trick". Define the basis vectors in a *local inertial reference frame* as $\varepsilon_{(\alpha)}$ and call the coordinates in this frame ξ^α. The defining property of these inertial basis vectors is that $\frac{\partial \varepsilon_{(\alpha)}}{\partial x^\nu} = 0$. The basis vectors $e_{(\mu)}$ in an arbitrary coordinate system are related to the inertial ones through

$$e_{(\mu)} = \frac{\partial \xi^\alpha}{\partial x^\mu}\varepsilon_{(\alpha)}. \tag{22}$$

Applying ∂_ν in (22) we get

$$\frac{\partial e_{(\mu)}}{\partial x^\nu} = \left[\frac{\partial}{\partial x^\nu}\left(\frac{\partial \xi^\alpha}{\partial x^\mu}\right)\right]\varepsilon_{(\alpha)}, \tag{23}$$

where we used the defining property $\frac{\partial \varepsilon_{(\alpha)}}{\partial x^\nu} = 0$. The right hand side of (23) is a linear combination of the inertial basis vectors. This linear combination must live in T_p, so we can alternatively write it as a linear combination of the basis vectors $e_{(\rho)}$. Write

$$\frac{\partial e_{(\mu)}}{\partial x^\nu} = \Gamma^\rho_{\mu\nu} e_{(\rho)}. \tag{24}$$

where we write the coefficients of the linear combination suggestively as $\Gamma^\rho_{\mu\nu}$. Then (21) can be written as

$$\frac{\partial A}{\partial x^\mu} = \frac{\partial A^\nu}{\partial x^\mu} e_{(\nu)} + A^\nu \Gamma^\rho_{\nu\mu} e_{(\rho)}$$

$$= \left(\frac{\partial A^\nu}{\partial x^\mu} + A^\rho \Gamma^\nu_{\rho\mu} \right) e_{(\nu)}$$

$$\equiv \left(\nabla_\mu A^\nu \right) e_{(\nu)}.$$

We define the *covariant derivative* of the components of a vector (called *contravariant* components) as

$$\nabla_\mu A^\nu = \partial_\mu A^\nu + \Gamma^\nu_{\rho\mu} A^\rho. \tag{25}$$

In order to find the expression for the covariant derivative of the components of a dual vector (called *covariant* components) we consider a scalar $\phi = V_\mu U^\mu$. The scalar is independent of the basis vectors, so

$$\nabla_\nu \phi = \partial_\nu \phi = \frac{\partial V_\mu}{\partial x^\nu} U^\mu + V_\mu \frac{\partial U^\mu}{\partial x^\nu}. \tag{26}$$

Use (25) to replace $\frac{\partial U^\mu}{\partial x^\nu}$. Then

$$\nabla_\nu \phi = \frac{\partial V_\mu}{\partial x^\nu} U^\mu + V_\mu \left[\nabla_\nu U^\mu - \Gamma^\mu_{\rho\nu} U^\rho \right]$$

$$= \left[\frac{\partial V_\mu}{\partial x^\nu} - \Gamma^\rho_{\mu\nu} V_\rho \right] U^\mu + V_\mu \nabla_\nu U^\mu.$$

The left hand side is a tensor, and so are U^μ and V_μ. This means that the terms in the square bracket must also form a tensor, and that $\nabla_\nu U^\mu$ is a tensor (One can also show explicitly that $\nabla_\mu U^\mu$ is a tensor from (25)). The covariant derivative of the covariant components of a dual vector is therefore

$$\nabla_\nu V_\mu = \partial_\nu V_\mu - \Gamma^\rho_{\mu\nu} V_\rho. \tag{27}$$

Note two facts about the covariant derivative: we require $\nabla_\rho g_{\mu\nu} = 0$, a property which is called *metric compatibility.* And as a consequence

$$\nabla_\nu V^\mu = \nabla_\nu \left(g^{\mu\sigma} V_\sigma \right) = g^{\mu\sigma} \nabla_\nu V_\sigma, \tag{28}$$

meaning that we can raise and lower indices inside covariant derivatives without worrying. This of course only holds due to metric compatibility.

In general $\Gamma^{\nu}_{\rho\mu}$ does not need to equal $\Gamma^{\nu}_{\mu\rho}$. If $\Gamma^{\nu}_{\rho\mu} \neq \Gamma^{\nu}_{\mu\rho}$ we say that the spacetime has torsion. But in general relativity we require a torsionless spacetime, meaning $\Gamma^{\nu}_{\mu\rho} = \Gamma^{\nu}_{\rho\mu}$. Now we want to relate these coefficients $\Gamma^{\nu}_{\mu\rho}$ to the metric tensor. Before we get there we take a short detour into discussing the line element ds^2. We know from earlier that $ds^2 = g_{\mu\nu} dx^{\mu} \otimes dx^{\nu}$. But notice that we can also express this as $ds^2 = \left(dx^{\mu} e_{(\mu)}\right) \cdot \left(dx^{\nu} e_{(\nu)}\right)$, interpreting the line element as a dot product between two infinitesimal vectors with components dx^{μ} and dx^{ν}. Then $g_{\mu\nu} = e_{(\mu)} \cdot e_{(\nu)}$, leading to

$$
\begin{aligned}
\partial_{\rho} g_{\mu\nu} &= \partial_{\rho} e_{(\mu)} \cdot e_{(\nu)} + e_{(\mu)} \cdot \partial_{\rho} e_{(\nu)} \\
&= \Gamma^{\sigma}_{\mu\rho} e_{(\sigma)} \cdot e_{(\nu)} + e_{(\mu)} \cdot \Gamma^{\sigma}_{\nu\rho} e_{(\sigma)} \\
&= \Gamma^{\sigma}_{\mu\rho} g_{\sigma\nu} + \Gamma^{\sigma}_{\nu\rho} g_{\mu\sigma}.
\end{aligned}
$$

In the second expression I used (24). By cyclically permuting the indices in the previous expression we can obtain

$$
\begin{aligned}
\partial_{\nu} g_{\rho\mu} &= \Gamma^{\sigma}_{\rho\nu} g_{\sigma\mu} + \Gamma^{\sigma}_{\mu\nu} g_{\rho\sigma} \\
\partial_{\mu} g_{\nu\rho} &= \Gamma^{\sigma}_{\nu\mu} g_{\rho\sigma} + \Gamma^{\sigma}_{\rho\mu} g_{\nu\sigma}
\end{aligned}
$$

If we combine these, together with the enforced property $\Gamma^{\sigma}_{\mu\nu} = \Gamma^{\sigma}_{\nu\mu}$, we get

$$
\partial_{\rho} g_{\mu\nu} + \partial_{\nu} g_{\rho\mu} - \partial_{\mu} g_{\nu\rho} = 2\Gamma^{\sigma}_{\rho\nu} g_{\mu\sigma}. \tag{29}
$$

By multiplying with the inverse metric $g^{\alpha\mu}$ and using the definition of the inverse $g^{\alpha\mu} g_{\mu\sigma} = \delta^{\alpha}_{\sigma}$ we get

$$
\Gamma^{\mu}_{\nu\rho} = \frac{1}{2} g^{\mu\sigma} \left(\partial_{\nu} g_{\sigma\rho} + \partial_{\rho} g_{\nu\sigma} - \partial_{\sigma} g_{\nu\rho}\right). \tag{30}
$$

This is called the *Christoffel connection* or just the *Christoffel symbol*.

5.1 Coordinate transformation of the Christoffel symbols

The Christoffel symbols are called *symbols* rather than tensor for a reason - they do not transform as tensor. Here we show this. Assume we have a coordinate system x^{μ} and transform to a new coordinate system y^{μ}. Then the metric and its inverse transform as

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial x^\beta}{\partial y^\nu}g_{\alpha\beta}$$

$$g'^{\mu\nu} = \frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\nu}{\partial x^\beta}g^{\alpha\beta}.$$

The Christoffel symbols then transform as

$$\Gamma'^{\mu}_{\rho\sigma} = \frac{1}{2}g'^{\mu\lambda}\left(\partial'_\rho g'_{\lambda\sigma} + \partial'_\sigma g'_{\lambda\rho} - \partial'_\lambda g'_{\rho\sigma}\right)$$

$$= \frac{1}{2}\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\left[\frac{\partial x^\nu}{\partial y^\rho}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right) + \frac{\partial x^\nu}{\partial y^\sigma}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\rho}g_{\theta\phi}\right) - \frac{\partial x^\nu}{\partial y^\lambda}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right)\right]$$

Let us first evaluate each of the terms inside the square bracket. We present the first and third terms in the bracket, because the second term can be obtained from the first by an exchange $\rho \leftrightarrow \sigma$.

$$\frac{\partial x^\nu}{\partial y^\rho}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right) = \frac{\partial x^\nu}{\partial y^\rho}\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\sigma}\partial_\nu g_{\theta\phi} + \frac{\partial x^\phi}{\partial y^\sigma}\frac{\partial^2 x^\theta}{\partial y^\sigma\partial y^\lambda}g_{\theta\phi} + \frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial^2 x^\phi}{\partial y^\rho\partial y^\sigma}g_{\theta\phi}$$

$$\frac{\partial x^\nu}{\partial y^\lambda}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right) = \frac{\partial x^\nu}{\partial y^\lambda}\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}\partial_\nu g_{\theta\phi} + \frac{\partial x^\phi}{\partial y^\sigma}\frac{\partial^2 x^\theta}{\partial y^\lambda\partial y^\rho}g_{\theta\phi} + \frac{\partial x^\theta}{\partial y^\rho}\frac{\partial^2 x^\phi}{\partial y^\lambda\partial y^\sigma}g_{\theta\phi}.$$

The terms in the bracket multiplied by the factor in front then give

$$\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\nu}{\partial y^\rho}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right) = \frac{\partial y^\mu}{\partial x^\alpha}g^{\alpha\beta}\frac{\partial x^\nu}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}\partial_\nu g_{\beta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\phi}{\partial y^\sigma}\frac{\partial^2 x^\theta}{\partial y^\rho\partial y^\lambda}g_{\theta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial y^\rho\partial y^\sigma}$$

$$\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\nu}{\partial y^\sigma}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\lambda}\frac{\partial x^\phi}{\partial y^\rho}g_{\theta\phi}\right) = \frac{\partial y^\mu}{\partial x^\alpha}g^{\alpha\beta}\frac{\partial x^\nu}{\partial y^\sigma}\frac{\partial x^\phi}{\partial y^\rho}\partial_\nu g_{\beta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\phi}{\partial y^\rho}\frac{\partial^2 x^\theta}{\partial y^\sigma\partial y^\lambda}g_{\theta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial y^\rho\partial y^\sigma}$$

$$\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\nu}{\partial y^\lambda}\partial_\nu\left(\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}g_{\theta\phi}\right) = \frac{\partial y^\mu}{\partial x^\alpha}g^{\alpha\nu}\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}\partial_\nu g_{\theta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\phi}{\partial y^\sigma}\frac{\partial^2 x^\theta}{\partial y^\lambda\partial y^\rho}g_{\theta\phi}$$

$$+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\lambda}{\partial x^\beta}g^{\alpha\beta}\frac{\partial x^\theta}{\partial y^\rho}\frac{\partial^2 x^\phi}{\partial y^\lambda\partial y^\sigma}g_{\theta\phi}$$

And the transformation of the Christoffel symbol is therefore

$$\Gamma'^\mu_{\rho\sigma} = \frac{1}{2}\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial x^\nu}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}g^{\alpha\beta}\left[\partial_\nu g_{\beta\phi}+\partial_\phi g_{\beta\nu}-\partial_\beta g_{\nu\phi}\right]+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial y^\rho\partial y^\sigma}$$

$$= \frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial x^\nu}{\partial y^\rho}\frac{\partial x^\phi}{\partial y^\sigma}\Gamma^\alpha_{\nu\phi}+\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial y^\rho\partial y^\sigma}$$

The second term here is the reason why the Christoffel symbol is not a tensor.

6 Parallel transport and geodesics

In flat space we can compare vectors/tensors between different points p and q without having to worry about how one moves the vectors over to each other. We can keep one of the vectors/tensors constant while moving it from p to q or vice versa. The path that we follow from p to q does not matter. But in curved spacetime it does matter. When we say that a tensor T in flat space is kept constant along the curve from p to q we mean that $\frac{dT}{d\lambda} = \frac{dx^\mu}{d\lambda}\frac{\partial T}{\partial x^\mu} = 0$. In analogy to this directional derivative operator we can define the covariant directional derivative along a path in a general spacetime to be given by the operator

$$\frac{D}{d\lambda} = \frac{dx^\mu}{d\lambda}\nabla_\mu. \tag{31}$$

The concept of *parallel transport* generalizes this notion about keeping a tensor "constant" while moving it from point p to q to curved spacetime. Parallel transport of a tensor T along a path $x^\mu(\lambda)$ is defined to be the requirement that

$$\left(\frac{D}{d\lambda}T\right)^{\mu_1\dots\mu_k}_{\nu_1\dots\nu_l} \equiv \frac{dx^\sigma}{d\lambda}\nabla_\sigma T^{\mu_1\dots\mu_k}_{\nu_1\dots\nu_l} = 0 \tag{32}$$

along the path. This is the *equation of parallel transport*. For a vector V^μ we get

$$\frac{d}{d\lambda}V^\mu + \Gamma^\mu_{\sigma\rho}\frac{dx^\sigma}{d\lambda}V^\rho = 0. \tag{33}$$

The metric is always parallel transported, which follows directly from metric compatibility. We can use this to show that the inner product between two parallel transported vectors is also preserved.

$$\frac{D}{d\lambda}\left(g_{\mu\nu}V^\mu W^\nu\right) = \left(\frac{D}{d\lambda}g_{\mu\nu}\right)V^\mu W^\nu + g_{\mu\nu}\left(\frac{D}{d\lambda}V^\mu\right)W^\nu + g_{\mu\nu}V^\mu\left(\frac{D}{d\lambda}W^\nu\right) = 0.$$

The first term is zero due to metric compatibility, while the other two terms are zero because V^μ and W^ν are themselves parallel transported. So the norm and orthogonality of parallel transported vectors are preserved.

6.1 Geodesics

Geodesics are important concepts in general relativity, and they represent the generalization of straight lines in Euclidean space to curved spaces. But a geodesic is also a path which parallel transports its own tangent vector $x^\mu(\lambda)$ is $\frac{dx^\mu}{d\lambda}$

$$\frac{D}{d\lambda}\frac{dx^\mu}{d\lambda} = 0. \tag{34}$$

This can also be written as

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0. \tag{35}$$

And this is called the *geodesic equation*. The tangent vector $\frac{dx^\mu}{d\lambda}$ is related to the four-velocity of the particle following the geodesic. So the four-velocity of the particle is covariantly conserved. On a purely intuitive level this should feel reasonable, since geodesic motion is unaccelerated motion in general relativity. Geodesics are also the paths that maximize proper time. We saw earlier that the line element for timelike paths can always be given as $ds^2 = -d\tau^2$. So $d\tau = \sqrt{-ds^2} = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}$. So the proper time along a path parameterized by λ is

$$\tau = \int \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}d\lambda \tag{36}$$

We want to find the path that extremizes the proper time, which is the path that is such that a small variation in the path around the extremum induces no variation in the proper time, i.e. $\delta\tau = 0$. We want to vary the path $x^\mu \to x^\mu + \delta x^\mu$, which also induces a variation in the metric $g_{\mu\nu} \to g_{\mu\nu} + \delta x^\sigma \partial_\sigma g_{\mu\nu}$. The variation in the proper time is

$$\delta\tau = \int \frac{1}{2\sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}}\left(-\partial_\sigma g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\delta x^\sigma - g_{\mu\nu}\frac{dx^\nu}{d\lambda}\delta\left(\frac{dx^\mu}{d\lambda}\right) - g_{\mu\nu}\frac{dx^\mu}{d\lambda}\delta\left(\frac{dx^\nu}{d\lambda}\right)\right)d\lambda$$

$$= \int \frac{1}{2\sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}}\left[-\partial_\sigma g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\delta x^\sigma - 2g_{\mu\nu}\frac{dx^\nu}{d\lambda}\frac{d\left(\delta x^\mu\right)}{d\lambda},\right]d\lambda$$

where we have used the symmetry of $g_{\mu\nu}$ and the fact that $\delta\left(\frac{dx^\mu}{d\lambda}\right) = \frac{d}{d\lambda}\left(\delta x^\mu\right)$. We can now use that $d\tau = \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}d\lambda$ to replace λ with τ.

$$\delta\tau = \int \left[-\frac{1}{2}\partial_\sigma g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\delta x^\sigma - g_{\mu\nu}\frac{dx^\nu}{d\tau}\frac{d\left(\delta x^\mu\right)}{d\tau}\right] d\tau$$

$$= \int \left[-\frac{1}{2}\partial_\sigma g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\delta x^\sigma + \partial_\sigma g_{\mu\nu}\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau}\delta x^\mu + g_{\mu\nu}\frac{d^2 x^\nu}{d\tau^2}\delta x^\mu - \frac{d}{d\tau}\left(g_{\mu\nu}\frac{dx^\nu}{d\tau}\delta x^\mu \right)\right] d\tau$$

The last term is a total derivative, so the integral of that term produces boundary terms, also called *surface terms*. We assume that the variation δx^μ is zero at the endpoints, so the surface terms vanish. This can be thought of as requiring that the particle moves along a path from a known point p to another known point q, and we just vary the path between p and q. Then, requiring that this surface term vanishes gives

$$\delta\tau = \int \left(g_{\mu\nu}\frac{d^2 x^\nu}{d\tau^2} + \partial_\sigma g_{\mu\nu}\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau} - \frac{1}{2}\partial_\mu g_{\sigma\nu}\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau}\right)\delta x^\mu d\tau$$

The variation δx^μ is of course arbitrary, and since we want $\delta\tau = 0$ this means that the terms in the parenthesis must vanish. So

$$g_{\mu\nu}\frac{d^2 x^\nu}{d\tau^2} + \frac{1}{2}\left(2\partial_\sigma g_{\mu\nu} - \partial_\mu g_{\sigma\nu}\right)\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau} = 0. \tag{37}$$

Multiply by the inverse $g^{\mu\rho}$ to get $g^{\mu\rho}g_{\mu\nu} = \delta^\rho_\nu$ in the first term. Then

$$\frac{d^2 x^\rho}{d\tau^2} + \frac{1}{2}g^{\mu\rho}\left(\partial_\sigma g_{\mu\nu} + \partial_\nu g_{\mu\sigma} - \partial_\mu g_{\sigma\nu}\right)\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau} = 0$$

$$\frac{d^2 x^\rho}{d\tau^2} + \Gamma^\mu_{\sigma\nu}\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau} = 0,$$

where $\Gamma^\mu_{\sigma\nu} = \frac{1}{2}g^{\mu\rho}\left(\partial_\sigma g_{\mu\nu} + \partial_\nu g_{\mu\sigma} - \partial_\mu g_{\sigma\nu}\right)$ is of course the Christoffel symbol defined earlier. In the first step we used the symmetry of $\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau}$ to write $2\partial_\sigma g_{\mu\nu}$ as two different partial derivatives. What we have arrived at is of course the geodesic equation we saw earlier. This shows that the geodesic between two points extremizes the proper time between these points. This does not prove that the proper time is *maximized* along a geodesic, however. Only that it is extremized. But for any timelike curve we can approximate it as a series of infinitely many infinitesimal null (lightlike) curves. The series of null curves approaches the timelike curve, but it has zero proper time. Therefore timelike geodesics cannot be curves of minimum proper time, since the null curve has zero proper time. So the proper time along the geodesic must instead *maximize* proper time.

7 Symmetries

There is yet another way of writing the geodesic equation. Write the four-momentum as $p^\mu = \frac{dx^\mu}{d\lambda}$ such that this holds for both massive and massless particles. Our claim now is that the geodesic equation can be written as

$$p^\nu \nabla_\nu p_\mu = 0. \tag{38}$$

This can be written as

$$\frac{dx^\nu}{d\lambda}\partial_\nu p_\mu - p^\nu \Gamma^\sigma_{\nu\mu} p_\sigma = 0. \tag{39}$$

This equation strongly resembles the geodesic equation (35), but it is not obvious that they should be exactly equivalent. To show this we use that $\frac{dx^\nu}{d\lambda}\partial_\nu = \frac{d}{d\lambda}$ to get

$$\frac{dp_\mu}{d\lambda} - g_{\sigma\rho}\Gamma^\sigma_{\nu\mu}p^\nu p^\rho = \partial_\sigma g_{\mu\nu} p^\nu \frac{dx^\sigma}{d\lambda} + g_{\mu\nu}\frac{dp^\nu}{d\lambda} - g_{\sigma\rho}\Gamma^\sigma_{\nu\mu}p^\nu p^\rho$$

$$= g_{\mu\nu}\frac{dp^\nu}{d\lambda} + \partial_\sigma g_{\mu\nu}p^\nu \frac{dx^\sigma}{d\lambda} - \frac{1}{2} g_{\sigma\rho} g^{\sigma\alpha}\left(\partial_\nu g_{\alpha\mu} + \partial_\mu g_{\alpha\nu} - \partial_\alpha g_{\mu\nu}\right) p^\nu p^\rho = 0.$$

Multiply by $g^{\mu\beta}$ to get

$$\frac{dp^\beta}{d\lambda} + g^{\mu\beta}\partial_\rho g_{\mu\nu} p^\nu p^\rho - \frac{1}{2} g^{\mu\beta}\left(\partial_\nu g_{\rho\mu} + \partial_\mu g_{\rho\nu} - \partial_\rho g_{\mu\nu}\right) p^\nu p^\rho$$

$$= \frac{dp^\beta}{d\lambda} + \frac{1}{2} g^{\mu\beta}\left(2\partial_\rho g_{\mu\nu} - \partial_\mu g_{\rho\nu}\right) p^\nu p^\rho$$

$$= \frac{dp^\beta}{d\lambda} + \frac{1}{2} g^{\mu\beta}\left(\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho} - \partial_\mu g_{\rho\nu}\right) p^\nu p^\rho$$

$$= \frac{dp^\beta}{d\lambda} + \Gamma^\beta_{\nu\rho} p^\nu p^\rho = 0,$$

which is again the geodesic equation. In the first step we used that $\partial_\mu g_{\rho\mu} - \partial_\rho g_{\nu\mu}$ is antisymmetric in ν and ρ while the factor $p^\nu p^\rho$ that it is multiplied by is symmetric. So these terms die. The form of the geodeisc equation in (38) is useful because it can very directly describe how symmetries lead to conservation laws. Consider

$$\frac{dp_\mu}{d\lambda} - \frac{1}{2} g^{\sigma\rho} \left(\partial_\mu g_{\rho\nu} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}\right) p^\nu p_\sigma = 0$$

$$\Longrightarrow \frac{dp_\mu}{d\lambda} - \frac{1}{2} \left(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\alpha\mu} - \partial_\alpha g_{\mu\nu}\right) p^\nu p^\alpha = 0.$$

The last two terms in the parenthesis are antisymmetric in α and ν, while the multiplying factor $p^\nu p^\alpha$ is symmetric. So these terms vanish. We are then left with

$$\frac{dp_\mu}{d\lambda} = \frac{1}{2} \partial_\mu g_{\alpha\nu} p^\nu p^\alpha = 0. \tag{40}$$

Notice that for every symmetry of the metric there is a conserved quantity along a geodesic. If $\partial_{\underline{\mu}} g_{\rho\sigma} = 0 \Leftrightarrow p_{\underline{\mu}} = \text{const.}$ along a geodesic. Here $\underline{\mu}$ means that this holds for one value of μ.

Now consider a dual vector K with the defining property $\nabla_{(\mu} K_{\nu)} = 0$. Such a vector is called a *Killing vector*. Then $K_\mu p^\mu$ is conserved along a geodesic. This is relatively simple to show

$$\frac{d}{d\lambda} \left(K_\mu p^\mu\right) = \frac{dx^\nu}{d\lambda} \partial_\nu \left(K_\mu p^\mu\right) = \frac{dx^\nu}{d\lambda} \nabla_\nu \left(K_\mu p^\mu\right)$$

$$= \frac{dx^\nu}{d\lambda} p^\mu \nabla_\nu K_\mu + K_\mu \frac{dx^\nu}{d\lambda} \nabla_\nu p^\mu$$

$$= p^\nu p^\mu \nabla_{(\nu} K_{\mu)} + K_\mu p^\nu \nabla_\nu p^\mu$$

$$= 0$$

The switch from ∂_ν to ∇_ν is valid because $K_\mu p^\mu$ is a scalar. We arrived at the last step by using the symmetry of $p^\nu p^\mu$. When multiplied by $p^\nu p^\mu$ only the symmetric part of $\nabla_\nu K_\mu$ survives. The first term in the last line is zero by definition of the Killing vector, and the last term is zero due to the geodesic equation. So the scalar $K_\mu p^\mu$ is indeed conserved along geodesics if $\nabla_{(\mu} K_{\nu)} = 0$.

8 Curvature

8.1 The Riemann tensor

In order to discuss curvature we need a proper way to quantify it. We noted earlier that in curved spacetime the path one takes when parallel transporting a vector matters. Therefore it should make sense that parallel transporting a vector around an infinitesimal loop provides a measure of the local curvature. The covariant derivative measures how much a tensor changes relative to if it had been parallel transported. This is because the covariant derivative in the direction along which it is parallel transported is zero. We can therefore find a measure of the curvature by computing the *commutator* of two covariant derivatives, namely $[\nabla_\mu, \nabla_\nu] = \nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$. This commutator measures the difference between parallel transport in one direction then in another direction, and vice versa. Consider the commutator between ∇_μ and ∇_ν acting on a vector V^ρ.

$$
\begin{aligned}
[\nabla_\mu, \nabla_\nu] V^\rho &= \nabla_\mu \nabla_\nu V^\rho - \nabla_\nu \nabla_\mu V^\rho \\
&= \partial_\mu \left(\nabla_\nu V^\rho\right) - \Gamma^\lambda_{\mu\nu} \nabla_\lambda V^\rho + \Gamma^\rho_{\mu\sigma} \nabla_\nu V^\sigma - \partial_\nu \left(\nabla_\mu V^\rho\right) + \Gamma^\lambda_{\nu\mu} \nabla_\lambda V^\rho - \Gamma^\rho_{\nu\sigma} \nabla_\mu V^\sigma \\
&= \partial_\mu \partial_\nu V^\rho + \left(\partial_\mu \Gamma^\rho_{\nu\sigma}\right) V^\sigma + \Gamma^\rho_{\nu\sigma} \partial_\mu V^\sigma - \Gamma^\lambda_{\mu\nu} \partial_\lambda V^\rho - \Gamma^\lambda_{\mu\nu} \Gamma^\rho_{\lambda\sigma} V^\sigma + \Gamma^\rho_{\mu\sigma} \partial_\nu V^\sigma + \Gamma^\rho_{\mu\sigma} \Gamma^\sigma_{\nu\lambda} V^\lambda \\
&- \partial_\nu \partial_\mu V^\rho - \left(\partial_\nu \Gamma^\rho_{\mu\sigma}\right) V^\sigma - \Gamma^\rho_{\mu\sigma} \partial_\nu V^\sigma + \Gamma^\lambda_{\nu\mu} \partial_\lambda V^\rho + \Gamma^\lambda_{\nu\mu} \Gamma^\rho_{\lambda\sigma} V^\sigma - \Gamma^\rho_{\nu\sigma} \partial_\mu V^\sigma - \Gamma^\rho_{\nu\sigma} \Gamma^\sigma_{\mu\lambda} V^\lambda \\
&= \left(\partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\lambda_{\nu\mu} \Gamma^\rho_{\lambda\sigma} - \Gamma^\lambda_{\mu\sigma} \Gamma^\rho_{\lambda\nu}\right) V^\sigma - \Gamma^\lambda_{\mu\nu} \nabla_\lambda V^\rho + \Gamma^\lambda_{\nu\mu} \nabla_\lambda V^\rho \\
&= \left(\partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\rho} + \Gamma^\lambda_{\nu\mu} \Gamma^\rho_{\lambda\sigma} - \Gamma^\lambda_{\mu\sigma} \Gamma^\rho_{\lambda\nu}\right) V^\sigma .
\end{aligned}
$$

The last two terms in the second to last line cancel because we require a torsionless metric. The *Riemann tensor* is defined as these terms in the parenthesis

$$R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\lambda_{\nu\mu} \Gamma^\rho_{\lambda\sigma} - \Gamma^\lambda_{\mu\sigma} \Gamma^\rho_{\lambda\nu} . \tag{41}$$

Before we see an example of where this Riemann tensor appears we should quickly note that we can form the so-called *Ricci tensor* through a contraction of two of the indices of the Riemann tensor

$$R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu} . \tag{42}$$

And the *Ricci scalar* by contracting the indices of the Ricci tensor

$$R = g^{\mu\nu} R_{\mu\nu} \tag{43}$$

8.2 Geodesic deviation

We know that in flat spacetime two initially parallel lines stay parallel forever. In curved spacetime, however, geodesics do not stay parallel. This is called *geodesic deviation*, and we can quantify it by considering two different, initially infinitesimally separated geodesics. The derivation is somewhat lengthy, involving lots of index juggling, but it is interesting nonetheless. Consider two geodesics $x^\mu(\lambda)$ and $x^\mu(\lambda) + \xi^\mu(\lambda)$, where ξ^μ is assumed to be infinitesimal. Our task here is to find out how this ξ^μ evolves. Since these are geodesics we have

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}(x)\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0$$

$$\frac{d^2\left(x^\mu + \xi^\mu\right)}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}(x+\xi)\frac{d\left(x^\rho + \xi^\rho\right)}{d\lambda}\frac{d\left(x^\sigma + \xi^\sigma\right)}{d\lambda} = 0$$

Now we take the difference between the second and first equation

$$\frac{d^2\xi^\mu}{d\lambda^2} + \left(\Gamma^\mu_{\rho\sigma} + \partial_\nu\Gamma^\mu_{\rho\sigma}\xi^\nu\right)\left(\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} + \frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda} + \frac{d\xi^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\right) - \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0$$

$$\implies \frac{d^2\xi^\mu}{d\lambda^2} + 2\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda} + \partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu = 0 \tag{44}$$

As an intermediate step, let us compute $\frac{d}{d\lambda}\left(\frac{dx^\rho}{d\lambda}\nabla_\rho\xi^\mu\right) = \frac{d}{d\lambda}\left(\frac{d\xi^\mu}{d\lambda} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\xi^\sigma\right)$.

$$\frac{d}{d\lambda}\left(\nabla_\rho\xi^\mu\right) = \frac{d}{d\lambda}\left(\frac{d\xi^\mu}{d\lambda} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\xi^\sigma\right) = \frac{d^2\xi^\mu}{d\lambda^2} + \frac{dx^\nu}{d\lambda}\partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\xi^\sigma + \Gamma^\mu_{\rho\sigma}\frac{d^2x^\rho}{d\lambda^2}\xi^\sigma + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda}$$

$$= \frac{d^2\xi^\mu}{d\lambda^2} + \partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda}\xi^\sigma - \Gamma^\mu_{\rho\sigma}\Gamma^\rho_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}\xi^\sigma + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda}$$

Therefore (44) can be written as

$$\frac{d}{d\lambda}\left(\frac{d\xi^\mu}{d\lambda} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\xi^\sigma\right) + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda} + \partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu$$

$$-\partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda}\xi^\sigma + \Gamma^\mu_{\rho\sigma}\Gamma^\rho_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}\xi^\sigma = 0 \tag{45}$$

Now we define

$$\frac{D^2\xi^\mu}{d\lambda^2} = \frac{dx^\alpha}{d\lambda}\nabla_\alpha\left(\frac{dx^\rho}{d\lambda}\nabla_\rho\xi^\mu\right)$$

$$= \frac{dx^\alpha}{d\lambda}\nabla_\alpha\left(\frac{dx^\rho}{d\lambda}\partial_\rho\xi^\mu + \frac{dx^\rho}{d\lambda}\Gamma^\mu_{\rho\sigma}\xi^\sigma\right)$$

$$= \frac{dx^\alpha}{d\lambda}\nabla_\alpha\left(\frac{d\xi^\mu}{d\lambda} + \frac{dx^\rho}{d\lambda}\Gamma^\mu_{\rho\sigma}\xi^\sigma\right)$$

$$= \frac{d}{d\lambda}\left(\frac{d\xi^\mu}{d\lambda} + \frac{dx^\rho}{d\lambda}\Gamma^\mu_{\rho\sigma}\xi^\sigma\right) + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{d\xi^\sigma}{d\lambda} + \frac{dx^\alpha}{d\lambda}\frac{dx^\rho}{d\lambda}\Gamma^\beta_{\rho\sigma}\Gamma^\mu_{\alpha\beta}\xi^\sigma$$

By rearranging this expression for $\frac{d}{d\lambda}\left(\frac{d\xi^\mu}{d\lambda} + \frac{dx^\rho}{d\lambda}\Gamma^\mu_{\rho\sigma}\xi^\sigma\right)$ and plugging this into (45) we get

$$\frac{D^2\xi^\mu}{d\lambda^2} + \partial_\nu\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu - \partial_\sigma\Gamma^\mu_{\rho\nu}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu$$

$$+\Gamma^\mu_{\rho\nu}\Gamma^\rho_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}\xi^\nu - \Gamma^\beta_{\rho\nu}\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\rho}{d\lambda}\xi^\nu = 0$$

$$\implies \frac{D^2\xi^\mu}{d\lambda^2} + \left(\partial_\nu\Gamma^\mu_{\rho\sigma} - \partial_\sigma\Gamma^\mu_{\rho\nu} + \Gamma^\mu_{\beta\nu}\Gamma^\beta_{\rho\sigma} - \Gamma^\mu_{\rho\beta}\Gamma^\beta_{\sigma\nu}\right)\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu = 0$$

$$\implies \frac{D^2\xi^\mu}{d\lambda^2} + R^\mu{}_{\sigma\nu\rho}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}\xi^\nu = 0 \tag{30}$$

So the Riemann tensor is a measure of how quickly these geodesics deviate. Note that in flat space(time), where $\Gamma = 0 \implies R^\mu_{\sigma\nu\rho} = 0$, and hence $\frac{D^2\xi^\mu}{d\lambda^2}$ reduces to $\frac{d^2\xi^\mu}{d\lambda^2} = 0$. Which is to say that in flat space(time) initially parallel geodesics stay parallel.

9 Gravitation

9.1 The minimal coupling principle

We have seen how the covariant derivative is the generalization of the partial derivative to curved spacetime. It is reasonable then to think that the partial derivatives encountered in the physics of flat spacetime are just special cases of covariant derivatives. This leads us to the *minimal coupling principle*, the statements of which boil down to the fact that the laws which hold in inertial coordinates in flat spacetime can be generalized to curved spacetime simply by letting $\partial_\mu \to \nabla_\mu$ and $\eta_{\mu\nu} \to g_{\mu\nu}$. Consider as an example unaccelerated motion in flat spacetime, which can be expressed as

$$\frac{d^2 x^\mu}{d\lambda^2} = 0 \quad \Longrightarrow \frac{dx^\nu}{d\lambda}\partial_\nu \frac{dx^\mu}{d\lambda}. \tag{46}$$

The minimal coupling principle leads us to propose that the equivalent law in curved spacetime should be the same, just with $\partial_\nu \to \nabla_\nu$. We then get

$$\frac{dx^\nu}{d\lambda}\nabla_\nu \frac{dx^\mu}{d\lambda} = \frac{dx^\nu}{d\lambda}\partial_\nu \frac{dx^\mu}{d\lambda} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda} = 0.$$

This is of course the geodesic equation, which is indeed the equation for unaccelerated motion in curved spacetime. Another example of how this minimal coupling principle can be used can be explored by looking at the *energy-momentum tensor* $T^{\mu\nu}$. Stated with no rigour whatsoever the energy-momentum tensor describes the energy-content of the spacetime in question. We will be considering the energy-momentum tensors of *perfect fluids*, which are fluids that can be described only by their energy densities and isotropic pressure. For perfect fluids the energy-momentum tensor in flat spacetime is given by

$$T^{\mu\nu} = (\rho + p)\, U^\mu U^\nu + p\eta^{\mu\nu}, \tag{47}$$

where ρ is the energy density of the fluid and p its pressure. It can be shown that this tensor is conserved $\partial_\mu T^{\mu\nu} = 0$ in flat spacetime. When moving to curved spacetime we are then lead to propose that the energy-momentum tensor is now *covariantly conserved* $\nabla_\mu T^{\mu\nu} = 0$, where we also do $\eta^{\mu\nu} \to g^{\mu\nu}$ in (47).

9.2 The Einstein field equations

One of the goals of general relativity is to relate the geometry of spacetime to its energy contents. The geometry is specified by the metric $g_{\mu\nu}$, while the energy contents are specified by the energy-momentum tensor $T^{\mu\nu}$. The equations that related these two are the *Einstein field equations*. There are multiple ways of proposing these equations, but arguably the most beautiful way of deriving them is through the principle of stationary action. We start by proposing the action

$$S = \int \left[\frac{1}{16\pi G} R + \mathcal{L}_M\right] \sqrt{-g}\; d^4 x. \tag{48}$$

Here $\mathcal{L}_M$ is the matter Lagrangian, and R is the Ricci scalar, while

$$g \equiv \det(g_{\mu\nu}). \tag{49}$$

The term $S = \frac{1}{16\pi G}\int \sqrt{-g}Rd^4x$ is called the *Einstein-Hilbert action*. We can obtain the Einstein field equations by using the stationary action principle - the variation of the action with respect to the metric is zero. So

$$\delta S = \int \left[\frac{1}{16\pi G}\left(\sqrt{-g}\delta R + R\delta\sqrt{-g}\right) + \frac{\partial\left(\sqrt{-g}\mathcal{L}_M\right)}{\partial g^{\mu\nu}}\delta g^{\mu\nu}\right] d^4x$$

$$= \int \left[\frac{1}{16\pi G}\left(\sqrt{-g}\frac{\partial}{\partial g^{\mu\nu}}\left(g^{\rho\sigma}R_{\rho\sigma}\right)\delta g^{\mu\nu} - R\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}\right) + \frac{\partial\left(\sqrt{-g}\mathcal{L}_M\right)}{\partial g^{\mu\nu}}\delta g^{\mu\nu}\right] d^4x$$

$$= \int \left[\frac{1}{16\pi G}\left(\sqrt{-g}R_{\mu\nu}\delta g^{\mu\nu} + \sqrt{-g}g^{\rho\sigma}\delta R_{\rho\sigma} - \frac{1}{2}\sqrt{-g}Rg_{\mu\nu}\delta g^{\mu\nu}\right) + \frac{\partial\left(\sqrt{-g}\mathcal{L}_M\right)}{\partial g^{\mu\nu}}\delta g^{\mu\nu}\right] d^4x$$

where we have used that $\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}$ (which can be found in [1], for instance). Now we need an expression for the variation of the Ricci tensor $\delta R_{\rho\sigma}$. Using (41) we can write the Ricci tensor as

$$R_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\sigma}\Gamma^\sigma_{\mu\nu} - \Gamma^\lambda_{\nu\sigma}\Gamma^\sigma_{\lambda\mu}. \tag{50}$$

The variation of the Ricci tensor is then

$$\delta R_{\mu\nu} = \partial_\lambda \delta\Gamma^\lambda_{\mu\nu} - \partial_\nu \delta\Gamma^\lambda_{\mu\lambda} + \delta\Gamma^\lambda_{\lambda\sigma}\Gamma^\sigma_{\mu\nu} + \Gamma^\lambda_{\lambda\sigma}\delta\Gamma^\sigma_{\mu\nu} - \delta\Gamma^\lambda_{\nu\sigma}\Gamma^\sigma_{\lambda\mu} - \Gamma^\lambda_{\nu\sigma}\delta\Gamma^\sigma_{\lambda\mu} \tag{35}$$

One might notice that this looks suspiciously reminiscent of the covariant derivative of the variation of Christoffel symbols. Consider the following two covariant derivatives

$$\nabla_\lambda \delta\Gamma^\lambda_{\mu\nu} = \partial_\lambda \delta\Gamma^\lambda_{\mu\nu} + \Gamma^\lambda_{\lambda\sigma}\delta\Gamma^\sigma_{\mu\nu} - \Gamma^\sigma_{\mu\lambda}\delta\Gamma^\lambda_{\sigma\nu} - \Gamma^\sigma_{\nu\lambda}\delta\Gamma^\lambda_{\mu\sigma}$$

$$\nabla_\nu \delta\Gamma^\lambda_{\lambda\mu\lambda} = \partial_\nu \delta\Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\nu\sigma}\delta\Gamma^\sigma_{\mu\lambda} - \Gamma^\sigma_{\mu\nu}\delta\Gamma^\lambda_{\sigma\lambda} - \Gamma^\sigma_{\lambda\nu}\delta\Gamma^\lambda_{\mu\sigma}$$

The difference between these two is

$$
\begin{aligned}
\nabla_\lambda \delta\Gamma^\lambda_{\mu\nu} - \nabla_\nu \delta\Gamma^\lambda_{\lambda\mu\lambda} &= \partial_\lambda \delta\Gamma^\lambda_{\mu\nu} - \partial_\nu \delta\Gamma^\lambda_{\mu\lambda} \\
&+ \Gamma^\lambda_{\lambda\sigma}\delta\Gamma^\sigma_{\mu\nu} - \Gamma^\sigma_{\mu\lambda}\delta\Gamma^\lambda_{\sigma\nu} - \Gamma^\sigma_{\nu\lambda}\delta\Gamma^\lambda_{\mu\sigma} \\
&- \Gamma^\lambda_{\nu\sigma}\delta\Gamma^\sigma_{\mu\lambda} + \Gamma^\sigma_{\mu\nu}\delta\Gamma^\lambda_{\sigma\lambda} + \Gamma^\sigma_{\lambda\nu}\delta\Gamma^\lambda_{\mu\sigma}
\end{aligned}
$$

Because we require a torsionless metric two of these terms cancel. And the result is what we have in (35). This means that

$$
\delta R_{\mu\nu} = \nabla_\lambda \delta\Gamma^\lambda_{\mu\nu} - \nabla_\nu \delta\Gamma^\lambda_{\mu\lambda} \tag{51}
$$

In the expression for the variation of the action we have $g^{\rho\sigma}\delta R_{\rho\sigma}$. We further have that

$$
\begin{aligned}
g^{\rho\sigma}\delta R_{\rho\sigma} &= g^{\rho\sigma}\nabla_\lambda \delta\Gamma^\lambda_{\rho\sigma} - g^{\rho\sigma}\nabla_\sigma \delta\Gamma^\lambda_{\rho\lambda} \\
&= \nabla_\lambda \left(g^{\rho\sigma}\delta\Gamma^\lambda_{\rho\sigma} - g^{\rho\lambda}\delta\Gamma^\sigma_{\rho\sigma} \right)
\end{aligned}
$$

Now, we saw earlier that the coordinate transformation law for Christoffel symbols contained a term which meant the Christoffel symbols are not tensors. But note that his non-tensorial term is only dependent on the coordinate transformation. So in the variations of the Christoffel symbols, which are essentially differences between Christoffel symbols, this term cancels out. This meant that the variation of a Christoffel symbol *is* a tensor, even if the Christoffel symbol itself is not. So $g^{\rho\sigma}\delta R_{\rho sigma}$ is the covariant derivative of some vector A^λ, where

$$
A^\lambda = g^{\rho\sigma}\delta\Gamma^\lambda_{\rho\sigma} - g^{\rho\lambda}\delta\Gamma^\sigma_{\rho\sigma}. \tag{52}
$$

When we plug this back into the variation of the action the term related to the variation of the Ricci tensor gives rise to a total derivative, which is again a surface term. And as usual we assume these go to zero at infinity, so we can safely neglect this term. The variation of the action then becomes

$$
\delta S = \int \left[\frac{1}{16\pi G} \left(\sqrt{-g} R_{\mu\nu} - \frac{1}{2}\sqrt{-g} R g_{\mu\nu} \right) + \frac{\partial \left(\sqrt{-g}\mathcal{L}_M \right)}{\partial g^{\mu\nu}} \right] \delta g^{\mu\nu} d^4 x
$$

Let us now write

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}}\frac{\partial\left(\sqrt{-g}\mathcal{L}_M\right)}{\partial g^{\mu\nu}}. \tag{53}$$

Inserting this into the variation of the action we get

$$\delta S = \int \left[\frac{1}{16\pi G}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right) - \frac{1}{2}T_{\mu\nu}\right]\sqrt{-g}\delta g^{\mu\nu} d^4x$$

This variation should equal zero, by the principle of stationary action. Since $\delta g^{\mu\nu}$ is arbitrary this means that the argument in the square bracket is zero ($\sqrt{-g}$ is part of the integral measure). So we have

$$\frac{1}{16\pi G}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right) - \frac{1}{2}T_{\mu\nu} = 0, \tag{54}$$

or written in a more well-known form

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{36}$$

These are the Einstein field equations.

References

[1] Sean M. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 2019.