11 Tensor Formalism for General Relativity

- This chapter may be viewed as the mathematical appendix of the book. While some important GR results have been stated previously without proof, we now introduce the basics of the tensor formalism needed to properly formulate GR. Specific topics are introduced with an emphasis on their mathematical content, so the reader should refer back to the previous chapters for their physics context.
- While the tensors used in GR are basically the same as those in SR, differentiation of tensor components in a curved space must be handled with extra care, because basis vectors in a curved spacetime are position-dependent.
- By adding extra terms (involving a combination of the metric's first derivatives called Christoffel symbols) to the ordinary derivative operator, we can form a covariant derivative, which acts on tensor components to yield components of a new tensor. Covariant differentiation has a clear geometric meaning in terms of parallel transport of tensors.
- The Riemann tensor reflects multiple aspects of curvature. Its expression (11.40) can be derived from
  - the change of a vector parallel-transported around a closed path (which is related to the noncommutivity of covariant derivatives of a vector);
  - the deviation of geodesics (tidal forces).
- We use the Bianchi identity to show that the Einstein tensor has no covariant divergence, qualifying it to be the geometric term in the GR field equation. The metric tensor itself also satisfies this criterion, thereby allowing the cosmological constant term.
- The approach to Einstein field equation via the principle of least action is sketched. The relevant mathematics of its Schwarzschild solution is outlined.

General relativity requires that physics equations be covariant under any general coordinate transformation that leaves invariant the infinitesimal interval

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu. \tag{11.1}$$

Just as SR requires physics equations to be tensor equations with respect to Lorentz transformations, GR equations must be tensor equations with respect to general coordinate transformations. In this way, the principle of GR can be fulfilled automatically.

General coordinate transformations

Recall our discussion in Section 3.2 that tensor components are the expansion coefficients of a tensor in terms of the basis vectors. Under a coordinate transformation, a tensor does not itself change, but its components transform because of the changed bases. The transformation rules of tensor components are listed in (3.25)–(3.27). Because repeated reference to tensor components can be cumbersome, we often simply refer to tensor components as tensors.

In Chapter 5 and in particular (5.10), we suggested that coordinate transformations can be written in terms of partial derivatives.[1] We now discuss these general coordinate transformations further. From the basic chain rule of differentiation, we have

[1] This also applies to position-independent coordinate transformations such as ordinary rotations and Lorentz transformations; see Exercise 2.4.

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}\, dx^\nu, \qquad \partial'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}\, \partial_\nu. \tag{11.2}$$

We can interpret these relations as the transformations $(dx^\nu, \partial_\nu) \rightarrow (dx'^\mu, \partial'_\mu)$ by the respective transformation matrices $(\partial x'^\mu/\partial x^\nu, \partial x^\nu/\partial x'^\mu)$. Recall our Chapter 3 definitions of contravariant and covariant vector components (A^μ, A_μ); they transform in the same way as (dx^ν, ∂_ν). Thus we can write the respective transformations of the contravariant and covariant components of a vector as

$$A^\mu \longrightarrow A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu, \tag{11.3}$$

$$A_\mu \longrightarrow A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} A_\nu. \tag{11.4}$$

We display the contravariant transformation in 4D spacetime:

$$\begin{pmatrix} A'^0 \\ A'^1 \\ A'^2 \\ A'^3 \end{pmatrix} = \begin{pmatrix} \dfrac{\partial x'^0}{\partial x^0} & \dfrac{\partial x'^0}{\partial x^1} & \dfrac{\partial x'^0}{\partial x^2} & \dfrac{\partial x'^0}{\partial x^3} \\ \dfrac{\partial x'^1}{\partial x^0} & \dfrac{\partial x'^1}{\partial x^1} & \dfrac{\partial x'^1}{\partial x^2} & \dfrac{\partial x'^1}{\partial x^3} \\ \dfrac{\partial x'^2}{\partial x^0} & \dfrac{\partial x'^2}{\partial x^1} & \dfrac{\partial x'^2}{\partial x^2} & \dfrac{\partial x'^2}{\partial x^3} \\ \dfrac{\partial x'^3}{\partial x^0} & \dfrac{\partial x'^3}{\partial x^1} & \dfrac{\partial x'^3}{\partial x^2} & \dfrac{\partial x'^3}{\partial x^3} \end{pmatrix} \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix}. \tag{11.5}$$

This way of writing a transformation also has the advantage of preventing us from misidentifying the transformation $[...]^{\mu}_{\nu}$ as a tensor.

Because any vector $\mathbf{A}$ is coordinate-independent and may be expanded as $\mathbf{A} = A^{\mu}\mathbf{e}_{\mu} = A_{\mu}\mathbf{e}^{\mu}$, the transformations of its expansion coefficients A^{μ} and A_{μ} must cancel out (i.e., be the inverse of) those of the corresponding bases:

$$\mathbf{e}'_{\mu} = \frac{\partial x^{\nu}}{\partial x'^{\mu}}\mathbf{e}_{\nu} \quad \text{and} \quad \mathbf{e}'^{\nu} = \frac{\partial x'^{\nu}}{\partial x^{\rho}}\mathbf{e}^{\rho}. \tag{11.6}$$

This is the reason why $\{A_{\mu}\}$ are called the **co**variant components: they transform in the same way as the basis vectors $\{\mathbf{e}_{\mu}\}$, while the **contra**variant components $\{A^{\mu}\}$ transform oppositely, like the inverse bases $\{\mathbf{e}^{\mu}\}$.

A tensor $\mathbf{T}$ of higher rank is likewise coordinate-independent and can be similarly expanded in terms of basis elements that are (direct) products of the vector bases:

$$\mathbf{T}{=}T^{\mu\nu\ldots}{}_{\lambda\ldots}\mathbf{e}_{\mu}\otimes\mathbf{e}_{\nu}\cdots\otimes\mathbf{e}^{\lambda}\otimes\cdots. \tag{11.7}$$

Therefore, tensor components $T^{\mu\nu\ldots}{}_{\lambda\ldots}$ transform like products of vector components $A^{\mu}B^{\nu}\cdots C_{\lambda}\cdots$. For example, mixed tensor components $T_{\nu}{}^{\mu}$ transform as

$$T_{\nu}{}^{\mu} \longrightarrow T'_{\nu}{}^{\mu} = \frac{\partial x^{\lambda}}{\partial x'^{\nu}}\frac{\partial x'^{\mu}}{\partial x^{\rho}}T_{\lambda}{}^{\rho}. \tag{11.8}$$

In particular, the metric tensor components change as

$$g_{\mu\nu} \longrightarrow g'_{\mu\nu} = \frac{\partial x^{\lambda}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}g_{\lambda\rho}. \tag{11.9}$$

Recall that the metric tensor components are related to the basis vectors $\{\mathbf{e}_{\mu}\}$ (and components of the inverse metric to the inverse basis vectors $\{\mathbf{e}^{\mu}\}$) by

$$g_{\mu\nu} = \mathbf{e}_{\mu}\cdot\mathbf{e}_{\nu} \qquad \text{and} \qquad g^{\mu\nu} = \mathbf{e}^{\mu}\cdot\mathbf{e}^{\nu}. \tag{11.10}$$

Since the basis vectors of a curved space are position-dependent, so must be the associated metric. Relations such as (11.9) with position-dependent $g_{\lambda\rho}$ and $g'_{\mu\nu}$ imply that general coordinate transformations must themselves vary over spacetime.

11.1 Covariant derivatives and parallel transport

Physics equations usually involve differentiation. While tensors in GR are basically the same as SR tensors, the derivative operators in a curved space require considerable care. General coordinate transformations are position-dependent, so ordinary derivatives of tensor components, except for the trivial case of a scalar tensor, are not components of tensors. Nevertheless, we shall construct covariant differentiation operations that do result in tensor component derivatives.

11.1.1 Derivatives in a curved space and Christoffel symbols

We first demonstrate that ordinary derivatives spoil the transformation properties of tensor components. We then construct covariant derivatives that correct this problem.

Ordinary derivatives of tensor components are not tensors

In a curved space, the derivative $\partial_\nu A^\mu$ is not a tensor. Namely, even though A^μ and ∂_ν transform like vectors, as indicated by (11.3) and (11.2), their combination $\partial_\nu A^\mu$ does not transform as required by (11.8):

$$\partial_\nu A^\mu \longrightarrow \partial'_\nu A'^\mu \neq \frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial x'^\mu}{\partial x^\rho}\partial_\lambda A^\rho. \tag{11.11}$$

We can find the full expression for $\partial'_\nu A'^\mu$ by differentiating ($\partial'_\nu \equiv \partial/\partial x'^\nu$) both sides of (11.3):

$$\begin{aligned}\partial'_\nu A'^\mu &= \frac{\partial}{\partial x'^\nu}\left(\frac{\partial x'^\mu}{\partial x^\rho}A^\rho\right)\\ &= \frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial x'^\mu}{\partial x^\rho}(\partial_\lambda A^\rho) + \frac{\partial^2 x'^\mu}{\partial x'^\nu \partial x^\rho}A^\rho,\end{aligned} \tag{11.12}$$

where the chain rule (11.2) has been used. Compared with the right-hand side of (11.11), there is an extra term, the second term on the right-hand side, because

$$\frac{\partial}{\partial x'^\nu}\left(\frac{\partial x'^\mu}{\partial x^\rho}\right) \neq 0; \tag{11.13}$$

the transformations are position-dependent, which follows from the position dependence of the metric. We see that the fundamental problem lies in the changing bases, $\mathbf{e}^\mu = \mathbf{e}^\mu(x)$, of the curved space. More explicitly, because the vector components are the projections of the vector onto the basis vectors $A^\mu = \mathbf{e}^\mu \cdot \mathbf{A}$, the changing bases $\partial_\nu \mathbf{e}^\mu \neq 0$ produce an extra (second) term in the derivative:

$$\partial_\nu A^\mu = \mathbf{e}^\mu \cdot (\partial_\nu \mathbf{A}) + \mathbf{A} \cdot (\partial_\nu \mathbf{e}^\mu). \tag{11.14}$$

The properties of the two terms on the right-hand side will be studied separately below.

Covariant derivatives as expansion coefficients of $\partial_\nu \mathbf{A}$

In order for an equation to be manifestly relativistic, we must be able to cast it as a tensor equation, whose form is unchanged under coordinate transformations.

Thus, we seek a covariant derivative D_ν to be used in covariant physics equations. Such a differentiation is constructed so that it acts on tensor components to yield a new tensor of rank one greater, which transforms per (11.8):

$$D_\nu A^\mu \longrightarrow D'_\nu A'^\mu = \frac{\partial x^\lambda}{\partial x'^\nu}\frac{\partial x'^\mu}{\partial x^\rho} D_\lambda A^\rho. \tag{11.15}$$

As will be demonstrated below, the first term on the right-hand side of (11.14) is just this desired covariant derivative term.

We have suggested that the difficulty with differentiating vector components A^μ is their coordinate dependence. By this reasoning, derivatives of a scalar function Φ should not have this complication, because a scalar tensor does not depend on the bases: $\Phi' = \Phi$, so

$$\partial_\mu \Phi \longrightarrow \partial'_\mu \Phi' = \frac{\partial x^\lambda}{\partial x'^\mu}\partial_\lambda \Phi. \tag{11.16}$$

Similarly, the derivatives of the vector **A** itself (not its components) transform properly, because **A** is coordinate-independent:

$$\partial_\mu \mathbf{A} \longrightarrow \partial'_\mu \mathbf{A} = \frac{\partial x^\lambda}{\partial x'^\mu}\partial_\lambda \mathbf{A}. \tag{11.17}$$

Both (11.16) and (11.17) merely reflect the transformation of the del operator (11.2). If we dot both sides of (11.17) by the inverse basis vectors, $\mathbf{e}'^\nu = (\partial x'^\nu/\partial x^\rho)\mathbf{e}^\rho$, we obtain

$$\mathbf{e}'^\nu \cdot \partial'_\mu \mathbf{A} = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\rho}\mathbf{e}^\rho \cdot \partial_\lambda \mathbf{A}. \tag{11.18}$$

This shows that $\mathbf{e}^\nu \cdot \partial_\mu \mathbf{A}$ is a proper mixed tensor[2] as required by (11.8), and this is the covariant derivative (11.15) we have been seeking:

$$D_\mu A^\nu = \mathbf{e}^\nu \cdot \partial_\mu \mathbf{A}. \tag{11.19}$$

This relation implies that $D_\mu A^\nu$ can be viewed as the projection[3] of the vectors $\{\partial_\mu \mathbf{A}\}$ along the direction of $\mathbf{e}^\nu$; we can then interpret $D_\mu A^\nu$ as the coefficient of expansion of $\{\partial_\mu \mathbf{A}\}$ in terms of the basis vectors:

$$\partial_\mu \mathbf{A} = (D_\mu A^\nu)\mathbf{e}_\nu. \tag{11.20}$$

Christoffel symbols as expansion coefficients of $\partial_\nu e^\mu$

On the other hand, we do not have a similarly simple transformation relation like (11.17) when the coordinate-independent **A** is replaced by one of the coordinate basis vectors $\mathbf{e}_\mu$, which by definition change under coordinate transformations. Still, by mimicking (11.20), we can expand $\partial_\nu \mathbf{e}^\mu$ as

[2] We can reach the same conclusion by applying the quotient theorem (see Exercise 3.2) to (11.20), with the observation that since both $\partial_\mu \mathbf{A}$ and $\mathbf{e}_\nu$ are good tensors, so must be their quotient $(D_\mu A^\nu)$.

[3] We are treating $\{\partial_\mu \mathbf{A}\}$ as a set of vectors, each labeled by an index μ. The combination $D_\mu A^\nu$ is a projection of $\partial_\mu \mathbf{A}$, in the same way that $A^\nu = \mathbf{e}^\nu \cdot \mathbf{A}$ is a projection of the vector **A**.

$$\partial_\nu \mathbf{e}^\mu = -\Gamma^\mu_{\nu\lambda}\mathbf{e}^\lambda \qquad \text{or} \qquad \mathbf{A}\cdot(\partial_\nu \mathbf{e}^\mu) = -\Gamma^\mu_{\nu\lambda}A^\lambda. \tag{11.21}$$

Similarly, we have[4]

$$\partial_\nu \mathbf{e}_\mu = +\Gamma^\lambda_{\nu\mu}\mathbf{e}_\lambda \qquad \text{or} \qquad \mathbf{A}\cdot(\partial_\nu \mathbf{e}_\mu) = +\Gamma^\lambda_{\nu\mu}A_\lambda. \tag{11.22}$$

[4] $\mathbf{e}_\mu \cdot \mathbf{e}^\nu = [\mathbb{I}]^\nu_\mu$, so $\partial_\lambda(\mathbf{e}_\mu \cdot \mathbf{e}^\nu) = 0$. One can then apply the derivative product rule and plug in the expansion of $\partial_\nu \mathbf{e}^\mu$ to solve for $\partial_\nu \mathbf{e}_\mu$.

But the expansion coefficients $\{\Gamma^\mu_{\nu\lambda}\}$ are not tensors. Anticipating the result, we have here used the same notation for these expansion coefficients as for the Christoffel symbols introduced in Chapter 5 (cf. (5.30))—also called the affine connection (connection, for short).

Plugging (11.19) and (11.21) into (11.14), we have

$$D_\nu A^\mu = \partial_\nu A^\mu + \Gamma^\mu_{\nu\lambda}A^\lambda. \tag{11.23}$$

Thus, in order to produce the covariant derivative, the ordinary derivative $\partial_\nu A^\mu$ must be supplemented by another term. This second term directly reflects the position dependence of the basis vectors, shown in (11.21). Even though neither $\partial_\nu A^\mu$ nor $\Gamma^\mu_{\nu\lambda}A^\lambda$ has the correct tensor transformation properties, the transformation of $\Gamma^\mu_{\nu\lambda}A^\lambda$ cancels the unwanted term in the transformation of $\partial_\nu A^\mu$ (11.12), so that their sum $D_\nu A^\mu$ is a good tensor. Further insight into the structure of the covariant derivative can be gleaned by invoking the basic geometric concept of parallel displacement of a vector, to be presented in Section 11.1.2.

One can easily show that the covariant derivative of a covariant vector A_μ takes on a form similar to (11.23) for the contravariant vector A^μ:

$$D_\nu A_\mu = \partial_\nu A_\mu - \Gamma^\lambda_{\nu\mu}A_\lambda. \tag{11.24}$$

A mixed tensor such as T^μ_ν, which transforms in the same way as the product $A^\mu B_\nu$, will have a covariant derivative

$$D_\nu T^\rho_\mu = \partial_\nu T^\rho_\mu - \Gamma^\lambda_{\nu\mu}T^\rho_\lambda + \Gamma^\rho_{\nu\sigma}T^\sigma_\mu. \tag{11.25}$$

There should be a Christoffel term for each index of the tensor—a $(+\Gamma T)$ for each contravariant index and a $(-\Gamma T)$ for each covariant index. A specific example is the covariant differentiation of the (covariant) metric tensor $g_{\mu\nu}$:

$$D_\lambda g_{\mu\nu} = \partial_\lambda g_{\mu\nu} - \Gamma^\rho_{\lambda\mu}g_{\rho\nu} - \Gamma^\rho_{\lambda\nu}g_{\mu\rho}. \tag{11.26}$$

Christoffel symbols and metric tensor

We have introduced the Christoffel symbols $\Gamma^\mu_{\nu\lambda}$ as the coefficients of expansion of $\partial_\nu \mathbf{e}^\mu$ as in (11.21). In this section, we shall relate these $\Gamma^\mu_{\nu\lambda}$ to the first derivatives of the metric tensor. This will justify the identification with the symbols first defined in (5.30).

The metric tensor is covariantly constant While the metric tensor is position-dependent, $\partial[g] \neq 0$, its components are constant with respect to covariant differentiation, $D[g] = 0$ (we say that $g_{\mu\nu}$ is covariantly constant):

$$D_\lambda g_{\mu\nu} = 0. \tag{11.27}$$

One way to prove this is to express the metric in terms of the basis vectors, $g_{\mu\nu} = \mathbf{e}_\mu \cdot \mathbf{e}_\nu$, and apply the definition of the affine connection, $\partial_\nu \mathbf{e}_\mu = +\Gamma^\rho_{\nu\mu}\mathbf{e}_\rho$, given in (11.22):

$$\begin{aligned}\partial_\lambda(\mathbf{e}_\mu \cdot \mathbf{e}_\nu) &= (\partial_\lambda \mathbf{e}_\mu) \cdot \mathbf{e}_\nu + \mathbf{e}_\mu \cdot (\partial_\lambda \mathbf{e}_\nu) \\ &= \Gamma^\rho_{\lambda\mu}\mathbf{e}_\rho \cdot \mathbf{e}_\nu + \Gamma^\rho_{\lambda\nu}\mathbf{e}_\mu \cdot \mathbf{e}_\rho. \end{aligned} \tag{11.28}$$

Reverting back to the metric tensors, this relation becomes

$$\partial_\lambda g_{\mu\nu} - \Gamma^\rho_{\lambda\mu} g_{\rho\nu} - \Gamma^\rho_{\lambda\nu} g_{\mu\rho} = D_\lambda g_{\mu\nu} = 0, \tag{11.29}$$

where we have applied the definition of the covariant derivative of a covariant tensor $g_{\mu\nu}$ as in (11.26). As we shall see, the covariant constancy of the metric tensor is the key property that allowed Einstein to introduce his cosmological constant term in the GR field equation.

Exercise 11.1 Christoffel symbols as the metric tensor derivative

(a) The geometry in which we are working has the property that two covariant differentiation operations on a scalar tensor commute: $D_\mu D_\nu \Phi = D_\nu D_\mu \Phi$ (we call such derivatives torsion-free). From this, prove that Christoffel symbols are symmetric with respect to interchange of their lower indices: $\Gamma^\mu_{\nu\lambda} = \Gamma^\mu_{\lambda\nu}$.
(b) Using the definition (11.22) of Christoffel symbols as the coefficients of expansion of the derivative $\partial_\nu \mathbf{e}_\mu$, we showed that the metric is covariantly constant as in (11.29). After this, derive the expression for Christoffel symbols, as the first derivatives of the metric tensor, shown in (5.30). To signify its importance, this relation is called the fundamental theorem of Riemannian geometry.

***Suggestion**: One can obtain the result by taking the linear combination of three equations expressing ($Dg = 0$) with indices cyclically permuted and by using $\Gamma^\mu_{\nu\lambda} = \Gamma^\mu_{\lambda\nu}$ as shown in (a).*

Once the connection of the first derivative of the metric and the Christoffel symbols is established, we can better understand the result in (11.29). Since Christoffel symbols vanish in the local Euclidean frame, $0 = \partial_\mu g_{\nu\lambda} = D_\mu g_{\nu\lambda}$. The first equality follows from the flatness theorem (discussed in Section 5.1.3); the second follows from $\Gamma^\mu_{\nu\lambda} = 0$, hence $\partial = D$, in the local Euclidean frame. The last expression is covariant, so it must equal zero in every frame of reference, thus proving (11.27).

11.1.2 Parallel transport and geodesics as straight lines

Parallel transport is a fundamental concept in differential geometry. It illuminates the meaning of covariant differentiation and the associated Christoffel symbols. Furthermore, we can use this operation to clearly portray the geodesic as the straightest possible curve,[5] the curve traced out by the parallel transport of its tangent vector. In Section 11.3, we shall derive the Riemann curvature tensor by way of parallel-transporting a vector around a closed path.

[5] This is to be compared with our previous discussion in Box 5.2, using only ordinary derivatives.

Component changes under parallel transport

Equation (11.23) follows from (11.14). It expresses the relation between ordinary and covariant derivatives. Writing $DA^\mu = (D_\nu A^\mu)\,dx^\nu$ and $dA^\mu = (\partial_\nu A^\mu)\,dx^\nu$, (11.14) becomes

$$dA^\mu = DA^\mu - \Gamma^\mu_{\nu\lambda} A^\nu\, dx^\lambda. \tag{11.30}$$

We will show that the Christoffel symbols in the derivatives of vector components reflect the effects of parallel transport of a vector by a displacement of $d\mathbf{x}$. First, what is a parallel transport? Why does one need to perform such an operation? Recall the definition of the derivative of a scalar function $\Phi(\mathbf{x})$,

$$\partial_\mu \Phi = \frac{d\Phi(\mathbf{x})}{dx^\mu} = \lim_{h\to 0} \frac{\Phi(\mathbf{x} + h\hat{\mathbf{e}}_\mu) - \Phi(\mathbf{x})}{h}. \tag{11.31}$$

Its numerator involves the difference of the function's values at two different positions. Evaluating the coordinate-independent scalar function $\Phi(\mathbf{x})$ at two locations does not introduce any complication. This is not so for vector components. The differential dA^μ on the left-hand side of (11.30) is the difference $A^\mu(\mathbf{x} + d\mathbf{x}) - A^\mu(\mathbf{x}) \equiv A^\mu_{(2)} - A^\mu_{(1)}$ between the vector components $A^\mu = \mathbf{e}^\mu \cdot \mathbf{A}$, evaluated at two nearby positions, (1) and (2), separated by $d\mathbf{x}$. There are two sources of this difference: the change in the vector itself, $\mathbf{A}_{(2)} \neq \mathbf{A}_{(1)}$, and a coordinate change, $\mathbf{e}^\mu_{(2)} \neq \mathbf{e}^\mu_{(1)}$; they correspond to the two terms on the right-hand side of (11.14). Thus the total change is the sum of two terms:

$$dA^\mu = [dA^\mu]_{\text{total}} = [dA^\mu]_{\text{true}} + [dA^\mu]_{\text{coord}}. \tag{11.32}$$

The term representing the change in the vector itself may be called the true change,

$$[dA^\mu]_{\text{true}} = \mathbf{e}^\mu \cdot d\mathbf{A} = DA^\mu. \tag{11.33}$$

The other term represents the projection of $\mathbf{A}$ onto the change in the (inverse) basis vector between the two points separated by $d\mathbf{x}$. This change is a linear combination of the products of components of the vector A^ν with the separation dx^λ, with the Christoffel symbols as coefficients:

$$[dA^\mu]_{\text{coord}} = \mathbf{A} \cdot d\mathbf{e}^\mu = -\Gamma^\mu_{\nu\lambda} A^\nu\, dx^\lambda. \tag{11.34}$$

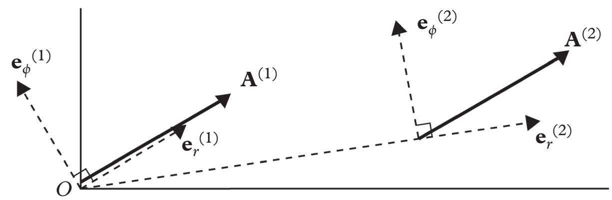

Figure 11.1 *Parallel transport of a vector $\boldsymbol{A}$ in a flat plane with polar coordinates: from position-1 at the origin, where $A^{(1)} = (A_\phi^{(1)}, A_r^{(1)})$, to position-2, $A^{(2)} = (A_\phi^{(2)}, A_r^{(2)})$. The differences in the basis vectors at these two positions, $(\boldsymbol{e}_\phi^{(1)}, \boldsymbol{e}_r^{(1)}) \neq (\boldsymbol{e}_\phi^{(2)}, \boldsymbol{e}_r^{(2)})$, bring about component changes. In particular, $A_\phi^{(1)} = 0$ while $A_\phi^{(2)} \neq 0$, and $A_r^{(1)} = \{[A_\phi^{(2)}]^2 + [A_r^{(2)}]^2\}^{1/2}$.*

This discussion motivates us to introduce the geometric concept of parallel transport. It is the process of moving a tensor without changing the tensor itself (in its local tangent space).[6] The only change in the tensor components under parallel displacement is due to coordinate changes, $dA^\mu = [dA^\mu]_{\text{coord}}$. In a flat space with a Cartesian coordinate system, this is trivial, since there is no coordinate change from point to point. But in a flat space with a curvilinear coordinate system such as polar coordinates, this parallel transport itself induces component changes, as shown in Fig. 11.1.

For the vector example discussed here, we have $[dA^\mu]_{\text{true}} = \mathbf{e}^\mu \cdot d\mathbf{A} = DA^\mu = 0$. Thus the mathematical expression for a parallel transport of vector components is

$$DA^\mu = dA^\mu + \Gamma^\mu_{\nu\lambda} A^\nu \, dx^\lambda = 0. \tag{11.35}$$

Recall that we have shown that the metric tensor is covariantly constant: $D_\mu g_{\nu\lambda} = 0$. We now understand that covariant constancy of a tensor means that any change in its tensor components is due to coordinate change only. But a change in the metric, by definition, is a pure coordinate change. Hence, it must have a vanishing covariant derivative.

[6] Strictly speaking, this statement is meaningful only in flat spaces. A tensor at one point in a curved space cannot be compared with a tensor at another point; they are different entities. For an obvious example, consider what a north-pointing vector on the equator would equal at the north pole—it is not defined. However, a curved space is locally flat (to first order), so we can parallel-transport a vector through a curved space, while keeping it constant in its local tangent space. We will see that this does induce second-order changes to a tensor after it has been parallel-transported in a closed path to its starting point, where it can be compared with its original state.

The geodesic as the straightest possible curve

The process of parallel-transporting a vector A^μ along a curve $x^\mu(\sigma)$ can be expressed, according to (11.35), as

$$\frac{DA^\mu}{d\sigma} = \frac{dA^\mu}{d\sigma} + \Gamma^\mu_{\nu\lambda} A^\nu \frac{dx^\lambda}{d\sigma} = 0. \tag{11.36}$$

From this, we can define the geodesic line as the straightest possible curve, because it is the line constructed by parallel transport of its tangent vector. See Fig. 11.2(a) for an illustration of such an operation in flat space. In this way, the geodesic condition can be formulated by setting $A^\mu = dx^\mu/d\sigma$ in (11.36):

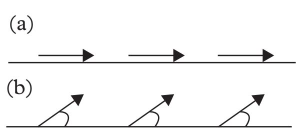

Figure 11.2 *(a) A straight line in a flat plane is a geodesic, the curve traced out by parallel transport of its tangents. (b) When a vector is parallel-transported along a straight line, the angle between the vector and the line is unchanged.*

$$\frac{D}{d\sigma}\left(\frac{dx^\mu}{d\sigma}\right) = 0, \tag{11.37}$$

or, more explicitly,

$$\frac{d}{d\sigma}\frac{dx^\mu}{d\sigma} + \Gamma^\mu_{\nu\lambda}\frac{dx^\nu}{d\sigma}\frac{dx^\lambda}{d\sigma} = 0. \tag{11.38}$$

This agrees with the geodesic equation (5.29).

Exercise 11.2 Parallel transport of a vector along a geodesic

Show that when a vector A_μ is parallel-transported along a geodesic, the angle between the vector and the geodesic (i.e., the tangent of the geodesic) is unchanged as in Fig. 11.2(b). Namely, prove the following relation:

$$\frac{D}{d\sigma}\left(A_\mu \frac{dx^\mu}{d\sigma}\right) = 0. \tag{11.39}$$

11.2 Riemann curvature tensor

As stated in Chapter 6, the Riemann curvature (6.20) is a tensor of rank 4:

$$R^\mu_{\lambda\alpha\beta} = \partial_\alpha \Gamma^\mu_{\lambda\beta} - \partial_\beta \Gamma^\mu_{\lambda\alpha} + \Gamma^\mu_{\nu\alpha}\Gamma^\nu_{\lambda\beta} - \Gamma^\mu_{\nu\beta}\Gamma^\nu_{\lambda\alpha}. \tag{11.40}$$

We shall demonstrate below that although the terms on the right-hand side are ordinary derivatives and Christoffel symbols, and thus are not tensors, their combination is nevertheless a proper tensor.

$R^\mu_{\lambda\alpha\beta}$ determines, independently of the coordinate choice, whether a space is curved. At any point in any curved space, one can always find a coordinate system (the local Euclidean frame) in which the metric's first derivatives vanish, $\partial g = 0$. However, the second derivatives of the metric, $\partial^2 g = 0$, vanish only for a flat space. Hence, in a flat space, $\partial^2 g + (\partial g)^2 \propto R^\mu_{\lambda\alpha\beta} = 0$. Since the Riemann curvature is a good tensor, if it vanishes for one set of coordinates, it vanishes for all coordinates.[7] In fact, we can also show that this is a sufficient condition for a space to be flat; i.e., $R^\mu_{\lambda\alpha\beta} = 0$ implies a flat space.

[7] This is exactly like the simpler 2D situation with the Gaussian curvature K of (6.7). The problem of reducing the Riemann curvature tensor of (11.40) in 2D space to the Gaussian curvature of (6.7) is worked out in Problem 13.11 in (Cheng 2010).

Two separate derivations of the Riemann tensor Having learned the formalism of covariant derivatives and the concept of parallel transport, we are now ready to derive the curvature tensor expression (11.40) in a space with arbitrary dimensions.

- The first derivation uses the feature that curvature measures the deviation of geometric relations from their corresponding Euclidean versions. We discussed in Section 6.1 the particular relation (6.19) that for a 2D curved surface the angular excess ϵ of an infinitesimal polygon (the sum of the interior angles over its Euclidean value) is proportional to the Gaussian curvature K at its location:

$$\epsilon = K\sigma, \tag{11.41}$$

  where σ is the area of the polygon. We will generalize this relation (11.41) for a 2D curvature K to an n-dimensional curved space. In this extension (Section 11.2.1), the concept of parallel transport plays a central role.

- The second derivation uses the feature that curvature causes relativistic tidal forces; we generalize the Newtonian deviation equation (6.24) to the GR equation of geodesic deviation. In this generalization (Section 11.2.2), covariant differentiation plays an indispensable role in finding the covariant equation that describes the relative motion of geodesic paths.

11.2.1 Parallel transport of a vector around a closed path

To extend the relation (11.41) to higher dimensions, one must first generalize the 2D quantities of angular excess ϵ and area σ to a higher-dimensional space.

Angular excess ϵ and directional change of a vector

How can an angular excess be measured in a higher-dimensional space? We first use the concept of parallel transport to cast the relation (11.41) in a form that can be generalized to n dimensions. The angular excess ϵ of a polygon is equal to the directional change in a vector after it has been parallel-transported around the perimeter. The simplest example of such a polygon is a spherical triangle with three 90° interior angles. Figure 6.2 shows that a vector parallel-transported around the triangle changes its direction by 90°, which is the angular excess. The generalization of (11.41) to an arbitrary triangle, and hence to any polygon, can be found in Sections 5.3.2 and 13.3 in (Cheng 2010). The key observation is that when a vector is parallel-transported along a geodesic, the angle it forms with the geodesic is unchanged; cf. Exercise 11.2. Recall the definition that an angle is the ratio of arclength to radius as shown in Fig. 11.3(a). Thus, the directional angular change, and hence the angular excess, can be written as the ratio of the change in a vector to its magnitude: $d\theta = \epsilon = dA/A$. Substituting this into (11.41), we obtain

$$dA = KA\sigma. \tag{11.42}$$

Namely, the change in a vector after a round-trip parallel transport is proportional to the vector itself and the area of the closed path. The coefficient of proportionality is identified as the curvature.

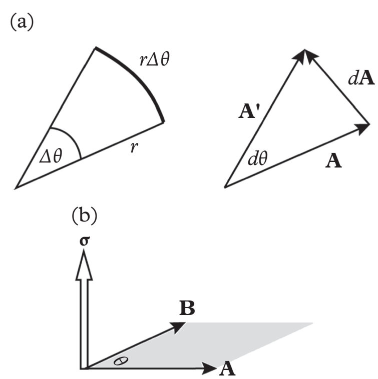

Figure 11.3 *(a) The directional change in a vector can be expressed as a fractional change in the vector: $d\theta = dA/A$. (b) The area vector of a parallelogram is the cross product of its two sides, $\boldsymbol{\sigma} = \boldsymbol{A} \times \boldsymbol{B}$.*

The area tensor

We will use (11.42) as a model for the curvature relation for a higher-dimensional curved space. We first need to write the 2D equation (11.42) in a proper index form that can be generalized to an n-dimensional space. Recall that the 2D area of a parallelogram spanned by two vectors **A** and **B** can be expressed as a vector product as in Fig. 11.3(b): $\boldsymbol{\sigma} = \mathbf{A} \times \mathbf{B}$. Using the antisymmetric Levi-Civita symbol in index notation,[8] we can write this as

$$\sigma_k = \epsilon_{ijk} A^i B^j. \tag{11.43}$$

[8] Levi-Civita symbols are discussed in Sidenote 19 in Chapter 3.

The area vector $\boldsymbol{\sigma}$ has magnitude $AB\sin\theta$ and direction given by the right-hand rule. But (11.43) is not a convenient form to use in a higher-dimensional space: (i) It refers to a 3D embedding space, even though the parallelogram resides in a 2D space. (ii) For a different number of dimensions, we would need to use the antisymmetric tensor with a different number of indices—e.g., in a 4D embedding space, $\epsilon_{\mu\nu\lambda\rho}$. We will instead use a two-index object σ^{ij} to represent the area:[9]

$$\sigma^{ij} \equiv \epsilon^{ijk}\sigma_k = \epsilon^{ijk}\epsilon_{mnk}A^m B^n = A^i B^j - A^j B^i. \tag{11.44}$$

In a 2D space, we can write σ^{ij} entirely with 2D indices, without reference to any embedding space. (Recall the distinction between intrinsic vs. extrinsic geometric descriptions discussed in Chapter 5.) In an n-dimensional space, we can represent the area spanned by a^λ and b^ρ by the antisymmetric combination

$$\sigma^{\lambda\rho} = a^\lambda b^\rho - b^\lambda a^\rho, \tag{11.45}$$

with the indices ranging over the dimensions of the space: $\lambda, \rho = \{1, 2, ..., n\}$.

The curvature tensor in an n-dimensional space

Now we have the proper area tensor (11.45), we can cast (11.42) in tensor form to represent[10] the change dA^μ in a vector due to parallel transport around a parallelogram spanned by two infinitesimal vectors a^λ and b^ρ:

$$dA^\mu = -R^\mu_{\ \nu\lambda\rho}A^\nu a^\lambda b^\rho. \tag{11.46}$$

Namely, the change is proportional to the vector A^ν itself and to the two vectors a^λ and b^ρ spanning the parallelogram. The coefficient of proportionality $R^\mu_{\ \nu\lambda\rho}$ is a quantity with four indices, antisymmetric in λ and ρ so as to pick up both terms on the right-hand side of (11.45). We shall take this to be the definition of the curvature (called the Riemann curvature tensor) of an n-dimensional space.[11] Explicit calculation in Box 11.1 of the change in a vector parallel-transported around an infinitesimal parallelogram then leads to the expression (11.40).

[9] In the second equality of (11.44), we use the identity $\epsilon^{ijk}\epsilon_{mnk} = \delta^i_m\delta^j_n - \delta^i_n\delta^j_m$.

[10] The minus sign is required so as to be compatible with the curvature definition given in (11.40), if the direction of the parallel-transport loop is in accord with the area direction (11.45), i.e., given by the right-hand rule around σ in the 2D case (counterclockwise in Fig. 11.4).

[11] We can plausibly expect this coefficient $R^\mu_{\ \nu\lambda\rho}$ to be a tensor, because the differential dA^μ (taken at a given position), a^λ, b^ρ, and A^ν are tensors, so the quotient theorem (Exercise 3.2) tells us that $R^\mu_{\ \nu\lambda\rho}$ should be a good tensor.

Box 11.1 Deriving the Riemann tensor by parallel-transporting a vector around a closed path

Here we shall parallel-transport a vector around an infinitesimal parallelogram $PQP'Q'$ spanned by two infinitesimal vectors a^α and b^β, shown in Fig. 11.4. Recall that under parallel transport of a vector, $DA^\mu = 0$, so the total vectorial change in (11.35) is due entirely to coordinate change:

$$dA^\mu = -\Gamma^\mu_{\nu\lambda}A^\nu\, dx^\lambda. \tag{11.47}$$

continued

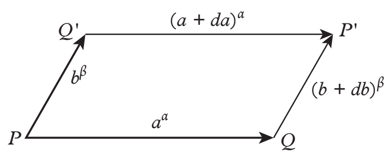

Figure 11.4 *The parallelogram $PQP'Q'$ is spanned by two vectors a^α and b^β. The opposite sides $(a+da)^\alpha$ and $(b+db)^\beta$ are obtained by parallel transport of a^α and b^β by b^ν and a^μ, respectively.*

Box 11.1 *continued*

The opposite sides of the parallelogram in Fig. 11.4, $(a+da)^\alpha$ and $(b+db)^\beta$, are obtained by parallel transport of a^α and b^β by b^ν and a^μ, respectively. The expression for parallel transport (11.47) gives the relations

$$\begin{aligned}(a+da)^\alpha &= a^\alpha - \Gamma^\alpha_{\mu\nu} a^\mu b^\nu,\\ (b+db)^\beta &= b^\beta - \Gamma^\beta_{\mu\nu} a^\mu b^\nu. \end{aligned} \tag{11.48}$$

Using (11.47) again, we now calculate the change in a vector A^μ due to parallel transport from $P \to Q \to P'$:

$$\begin{aligned} dA^\mu_{PQP'} &= dA^\mu_{PQ} + dA^\mu_{QP'} \\ &= -(\Gamma^\mu_{\nu\alpha}A^\nu)_P a^\alpha - (\Gamma^\mu_{\nu\beta}A^\nu)_Q (b+db)^\beta. \end{aligned} \tag{11.49}$$

The subscripts P and Q on the last line denote the respective positions where these functions are to be evaluated. Since eventually we shall compare all quantities at one position, say P, we will Taylor-expand the quantities $(...)_Q$ around the point P:

$$\begin{aligned} (\Gamma^\mu_{\nu\beta})_Q &= (\Gamma^\mu_{\nu\beta})_P + a^\alpha(\partial_\alpha \Gamma^\mu_{\nu\beta})_P, \\ (A^\nu)_Q &= (A^\nu)_P + a^\alpha(\partial_\alpha A^\nu)_P = (A^\nu)_P - a^\alpha(\Gamma^\nu_{\lambda\alpha}A^\lambda)_P, \end{aligned} \tag{11.50}$$

where we have used (11.47) to reach the last expression. From now on, we shall drop the subscript P. We substitute into (11.49) the expansions (11.48) and (11.50):

$$\begin{aligned} dA^\mu_{PQP'} &= -\Gamma^\mu_{\nu\alpha}A^\nu a^\alpha \\ &\quad -(\Gamma^\mu_{\nu\beta} + a^\alpha\partial_\alpha\Gamma^\mu_{\nu\beta})(A^\nu - a^\alpha\Gamma^\nu_{\lambda\alpha}A^\lambda)(b^\beta - \Gamma^\beta_{\rho\sigma}a^\rho b^\sigma). \end{aligned} \tag{11.51}$$

We multiply this out and keep terms up to $O(ab)$:

$$\begin{aligned} dA^\mu_{PQP'} &= -\Gamma^\mu_{\nu\alpha}A^\nu a^\alpha - \Gamma^\mu_{\nu\beta}A^\nu b^\beta + A^\nu\Gamma^\mu_{\nu\beta}\Gamma^\beta_{\rho\sigma}a^\rho b^\sigma \\ &\quad -\partial_\alpha\Gamma^\mu_{\lambda\beta}A^\lambda a^\alpha b^\beta + \Gamma^\mu_{\nu\beta}\Gamma^\nu_{\lambda\alpha}A^\lambda a^\alpha b^\beta. \end{aligned} \tag{11.52}$$

The vectorial change due to parallel transport along the other sides, $P \to Q' \to P'$, can be obtained from this expression by the simple interchange $a \leftrightarrow b$:

$$\begin{aligned} dA^\mu_{PQ'P'} &= -\Gamma^\mu_{\nu\alpha}A^\nu b^\alpha - \Gamma^\mu_{\nu\beta}A^\nu a^\beta + A^\nu\Gamma^\mu_{\nu\beta}\Gamma^\beta_{\rho\sigma}a^\rho b^\sigma \\ &\quad -\partial_\beta\Gamma^\mu_{\lambda\alpha}A^\lambda a^\alpha b^\beta + \Gamma^\mu_{\nu\alpha}\Gamma^\nu_{\lambda\beta}A^\lambda a^\alpha b^\beta. \end{aligned} \tag{11.53}$$

For a round-trip parallel transport[12] from P back to P, the vectorial change dA^μ corresponds to the difference of the above two equations (which results in cancellation of the first three terms on the right-hand sides):

$$\begin{aligned} dA^\mu &= dA^\mu_{PQP'} - dA^\mu_{PQ'P'} \\ &= -(\partial_\alpha \Gamma^\mu_{\lambda\beta} - \partial_\beta \Gamma^\mu_{\lambda\alpha} + \Gamma^\mu_{\nu\alpha}\Gamma^\nu_{\lambda\beta} - \Gamma^\mu_{\nu\beta}\Gamma^\nu_{\lambda\alpha})A^\lambda a^\alpha b^\beta. \end{aligned} \tag{11.54}$$

We conclude, after comparing (11.54) with (11.46), that the sought-after Riemann curvature tensor in terms of Christoffel symbols is just the quoted result (11.40).

[12] The order of the difference in (11.54) corresponds to parallel transport in the counterclockwise direction (on Fig. 11.4) in accordance with the area direction as defined in (11.43) and (11.45).

Exercise 11.3 Riemann curvature tensor as the commutator of covariant derivatives

We can obtain the same result as in Box 11.1 somewhat more efficiently by calculating the double covariant derivative

$$D_\alpha D_\beta A^\mu = D_\alpha(\partial_\beta A^\mu + \Gamma^\mu_{\beta\lambda}A^\lambda) = \ldots, \tag{11.55}$$

as well as the reverse order $D_\beta D_\alpha A^\mu = D_\beta(\partial_\alpha A^\mu + \Gamma^\mu_{\alpha\lambda}A^\lambda) = \ldots$. Show that their difference (expressed here as a commutator) is just the expression for the Riemann tensor given by (11.40):

$$[D_\alpha, D_\beta]A^\mu = R^\mu{}_{\lambda\alpha\beta}A^\lambda. \tag{11.56}$$

Comments: *(i) At first sight, one may question this approach to the problem of parallel transport of a vector around a closed path—wouldn't parallel transport mean that $DA = 0$? But the calculation in Box 11.1 shows that $D(DA) \neq 0$; calculating the vectorial change requires consistency in keeping the higher-order terms in Taylor expansions. See Sidenote 6 for a related comment.*

(ii) It is also straightforward to show that the covariant derivative commutator acting on a mixed tensor (instead of on a contravariant vector) will lead to

$$[D_\alpha, D_\beta]T^\mu_\nu = R^\mu{}_{\lambda\alpha\beta}T^\lambda{}_\nu - R^\lambda{}_{\nu\alpha\beta}T^\mu{}_\lambda; \tag{11.57}$$

i.e., for each contravariant index, there will be a $+RT$ term on the right-hand side, and, for each covariant index, a $-RT$ term.

11.2.2 Equation of geodesic deviation

In Section 11.2.1, we have derived an expression for the curvature (11.40) by a purely geometric method. A more physical approach would be to seek the GR generalization of the tidal forces discussed in Section 6.2.2. Following exactly the same steps used to derive the Newtonian deviation equation (6.24), let us consider two particles: one follows the spacetime trajectory $x^\mu(\tau)$, and the other follows

$x^\mu(\tau) + s^\mu(\tau)$. These two particles, separated by the displacement vector s^μ, obey their respective GR equations of motion, the geodesic equations, cf. (5.29):

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}(x)\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0 \tag{11.58}$$

and

$$\left(\frac{d^2x^\mu}{d\tau^2} + \frac{d^2s^\mu}{d\tau^2}\right) + \Gamma^\mu_{\alpha\beta}(x+s)\left(\frac{dx^\alpha}{d\tau} + \frac{ds^\alpha}{d\tau}\right)\left(\frac{dx^\beta}{d\tau} + \frac{ds^\beta}{d\tau}\right) = 0. \tag{11.59}$$

When the separation s^μ is small, we can approximate the Christoffel symbols $\Gamma^\mu_{\alpha\beta}(x+s)$ by a Taylor expansion

$$\Gamma^\mu_{\alpha\beta}(x+s) = \Gamma^\mu_{\alpha\beta}(x) + \partial_\lambda\Gamma^\mu_{\alpha\beta}s^\lambda + \cdots. \tag{11.60}$$

From the difference of the two geodesic equations, we obtain, to first order in s^μ,

$$\frac{d^2s^\mu}{d\tau^2} = -2\Gamma^\mu_{\alpha\beta}\frac{ds^\alpha}{d\tau}\frac{dx^\beta}{d\tau} - \partial_\lambda\Gamma^\mu_{\alpha\beta}s^\lambda\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}. \tag{11.61}$$

We are seeking the relative acceleration (the second derivative of the separation s^μ) along the worldline. So far we have only written down ordinary derivatives. In GR equations, we must use covariant derivatives. From (11.36), we have the first covariant derivative,

$$\frac{Ds^\mu}{d\tau} = \frac{ds^\mu}{d\tau} + \Gamma^\mu_{\alpha\beta}s^\alpha\frac{dx^\beta}{d\tau}, \tag{11.62}$$

and the second covariant derivative,

$$\begin{aligned}
\frac{D^2s^\mu}{d\tau^2} &= \frac{D}{d\tau}\left(\frac{Ds^\mu}{d\tau}\right) = \frac{d}{d\tau}\left(\frac{Ds^\mu}{d\tau}\right) + \Gamma^\mu_{\alpha\beta}\left(\frac{Ds^\alpha}{d\tau}\right)\frac{dx^\beta}{d\tau}\\
&= \frac{d}{d\tau}\left(\frac{ds^\mu}{d\tau} + \Gamma^\mu_{\alpha\beta}s^\alpha\frac{dx^\beta}{d\tau}\right) + \Gamma^\mu_{\alpha\beta}\left(\frac{ds^\alpha}{d\tau} + \Gamma^\alpha_{\lambda\rho}s^\lambda\frac{dx^\rho}{d\tau}\right)\frac{dx^\beta}{d\tau}\\
&= \frac{d^2s^\mu}{d\tau^2} + \partial_\lambda\Gamma^\mu_{\alpha\beta}\frac{dx^\lambda}{d\tau}s^\alpha\frac{dx^\beta}{d\tau} + \Gamma^\mu_{\alpha\beta}\frac{ds^\alpha}{d\tau}\frac{dx^\beta}{d\tau} + \Gamma^\mu_{\alpha\beta}s^\alpha\frac{d^2x^\beta}{d\tau^2}\\
&\quad + \Gamma^\mu_{\alpha\beta}\frac{ds^\alpha}{d\tau}\frac{dx^\beta}{d\tau} + \Gamma^\mu_{\alpha\beta}\Gamma^\alpha_{\lambda\rho}s^\lambda\frac{dx^\rho}{d\tau}\frac{dx^\beta}{d\tau}.
\end{aligned} \tag{11.63}$$

For the first term on the right-hand side ($d^2s^\mu/d\tau^2$), we apply (11.61); for $d^2x^\beta/d\tau^2$ in the fourth term we use the geodesic equation (11.58):

$$\frac{d^2x^\beta}{d\tau^2} = -\Gamma^\beta_{\lambda\rho}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}. \tag{11.64}$$

In this way, we find

$$\begin{aligned}\frac{D^2 s^\mu}{d\tau^2} &= -2\Gamma^\mu_{\alpha\beta}\frac{ds^\alpha}{d\tau}\frac{dx^\beta}{d\tau} - \partial_\lambda \Gamma^\mu_{\alpha\beta} s^\lambda \frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} + \partial_\lambda \Gamma^\mu_{\alpha\beta}\frac{dx^\lambda}{d\tau} s^\alpha \frac{dx^\beta}{d\tau}\\ &+2\Gamma^\mu_{\alpha\beta}\frac{ds^\alpha}{d\tau}\frac{dx^\beta}{d\tau} - \Gamma^\mu_{\alpha\beta} s^\alpha \Gamma^\beta_{\lambda\rho}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}\\ &+\Gamma^\mu_{\alpha\beta}\Gamma^\alpha_{\lambda\rho} s^\lambda \frac{dx^\rho}{d\tau}\frac{dx^\beta}{d\tau}.\end{aligned} \tag{11.65}$$

After a cancellation of two terms and the relabeling of several dummy indices, this becomes

$$\begin{aligned}\frac{D^2 s^\mu}{d\tau^2} &= -\partial_\lambda \Gamma^\mu_{\alpha\beta} s^\lambda \frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} + \partial_\beta \Gamma^\mu_{\lambda\alpha}\frac{dx^\alpha}{d\tau} s^\lambda \frac{dx^\beta}{d\tau} \\ &-\Gamma^\mu_{\lambda\rho} s^\lambda \Gamma^\rho_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} + \Gamma^\mu_{\rho\beta}\Gamma^\rho_{\lambda\alpha} s^\lambda \frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}.\end{aligned} \tag{11.66}$$

Factoring out the common $(dx^\alpha/d\tau)(dx^\beta/d\tau)$ yields the equation of geodesic deviation:

$$\frac{D^2 s^\mu}{D\tau^2} = -R^\mu{}_{\alpha\lambda\beta} s^\lambda \frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}, \tag{11.67}$$

where $R^\mu{}_{\alpha\lambda\beta}$ is just the Riemann curvature given in (11.40). Namely, the tensor of the gravitational potential's second derivatives (tidal gravity) in (6.24) is replaced in GR by the Riemann curvature tensor (11.40).

Exercise 11.4 From geodesic deviation to nonrelativistic tidal forces

Show that the equation of geodesic deviation (11.67) reduces to the Newtonian deviation equation (6.24) in the Newtonian limit. In the nonrelativistic limit of slow-moving particles with 4-velocity $dx^\alpha/d\tau \simeq (c,0,0,0)$, the GR equation (11.67) is reduced to

$$\frac{d^2 s^i}{dt^2} = -c^2 R^i{}_{0j0} s^j. \tag{11.68}$$

We have also set $s^0 = 0$, because we are comparing the two particles' accelerations at the same time. Thus (6.24) can be recovered by showing the relation

$$R^i{}_{0j0} = \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial x^i \partial x^j} \tag{11.69}$$

in the Newtonian limit. You are asked to prove that this expression is the limit of Riemann curvature in (11.40).

11.2.3 Bianchi identity and the Einstein tensor

We have already displayed the symmetries and contractions of the Riemann curvature tensor in Section 6.3.1. In particular, we have the symmetric Ricci tensor and the Ricci scalar:

$$R_{\mu\nu} \equiv g^{\alpha\beta} R_{\alpha\mu\beta\nu}, \qquad R \equiv g^{\alpha\beta} R_{\alpha\beta}. \tag{11.70}$$

It was suggested that some linear combination $R_{\mu\nu} + aRg_{\mu\nu}$ will enter the GR field equation as both terms are rank-2 symmetric tensors composed of the metric and its derivatives. As it turns out, the constant a can be fixed by requiring the combination to be covariantly constant: $D^{\mu}(R_{\mu\nu} + aRg_{\mu\nu}) = 0$ (cf. the opening discussion in Section 11.3.2). An efficient way is to use the Bianchi identity,

$$D_{\lambda} R_{\gamma\alpha\mu\nu} + D_{\nu} R_{\gamma\alpha\lambda\mu} + D_{\mu} R_{\gamma\alpha\nu\lambda} = 0. \tag{11.71}$$

The structure of this identity (11.71) suggests that we derive it from the Jacobi identity for the double commutators of three operators—in this case, the operators are covariant derivatives:

$$[D_{\lambda}, [D_{\mu}, D_{\nu}]] + [D_{\nu}, [D_{\lambda}, D_{\mu}]] + [D_{\mu}, [D_{\nu}, D_{\lambda}]] = 0. \tag{11.72}$$

Exercise 11.5 Jacobi identity and double commutator of covariant derivatives

---

(a) Prove the Jacobi identity. Namely, demonstrate explicitly that the cyclic combination of three double commutators of any three operators (in particular the differential operators in (11.72)) vanishes.
(b) Use the expression for the Riemann tensor in terms of the double commutator in (11.57) to show that

$$[D_{\lambda}, [D_{\mu}, D_{\nu}]] A_{\alpha} = -D_{\lambda} R^{\gamma}{}_{\alpha\mu\nu} A_{\gamma} + R^{\gamma}{}_{\lambda\mu\nu} D_{\gamma} A_{\alpha}. \tag{11.73}$$

Applying (11.73) to every double commutator in (11.72) acting on the covariant vector A_{α}, we have

$$\begin{aligned}
0 &= ([D_{\lambda}, [D_{\mu}, D_{\nu}]] + [D_{\nu}, [D_{\lambda}, D_{\mu}]] + [D_{\mu}, [D_{\nu}, D_{\lambda}]]) A_{\alpha} \\
&= -D_{\lambda} R^{\gamma}{}_{\alpha\mu\nu} A_{\gamma} - D_{\nu} R^{\gamma}{}_{\alpha\lambda\mu} A_{\gamma} - D_{\mu} R^{\gamma}{}_{\alpha\nu\lambda} A_{\gamma} \\
&\quad + R^{\gamma}{}_{\lambda\mu\nu} D_{\gamma} A_{\alpha} + R^{\gamma}{}_{\nu\lambda\mu} D_{\gamma} A_{\alpha} + R^{\gamma}{}_{\mu\nu\lambda} D_{\gamma} A_{\alpha} \\
&= -(D_{\lambda} R^{\gamma}{}_{\alpha\mu\nu} + D_{\nu} R^{\gamma}{}_{\alpha\lambda\mu} + D_{\mu} R^{\gamma}{}_{\alpha\nu\lambda}) A_{\gamma} \\
&\quad + (R^{\gamma}{}_{\lambda\mu\nu} + R^{\gamma}{}_{\nu\lambda\mu} + R^{\gamma}{}_{\mu\nu\lambda}) D_{\gamma} A_{\alpha}.
\end{aligned} \tag{11.74}$$

The second term on the right-hand side vanishes because of the cyclic symmetry property (6.32). It then follows that the parenthesis in the first term must also vanish, leading to the Bianchi identity (11.71).

Our next objective is to use the Bianchi identity to find the covariantly constant linear combination of $R_{\mu\nu}$ and $Rg_{\mu\nu}$, which we need for the GR field equation.

First we contract (11.71) with $g^{\mu\alpha}$; because the metric tensor is covariantly constant, $D_\lambda g^{\alpha\beta} = 0$, this metric contraction can be pushed right through the covariant differentiation:

$$-D_\lambda R_{\gamma\nu} + D_\nu R_{\gamma\lambda} + D_\mu g^{\mu\alpha} R_{\gamma\alpha\nu\lambda} = 0. \tag{11.75}$$

Contracting another time with $g^{\gamma\nu}$ yields

$$-D_\lambda R + D_\nu g^{\gamma\nu} R_{\gamma\lambda} + D_\mu g^{\mu\alpha} R_{\alpha\lambda} = 0. \tag{11.76}$$

At the last two terms, the metric just raises the indices:

$$-D_\lambda R + D_\nu R^\nu_\lambda + D_\mu R^\mu_\lambda = -D_\lambda R + 2D_\nu R^\nu_\lambda = 0. \tag{11.77}$$

Pushing through yet another $g^{\mu\lambda}$ in order to raise the λ index at the last term gives us

$$D_\lambda(-Rg^{\mu\lambda} + 2R^{\mu\lambda}) = 0. \tag{11.78}$$

Thus this combination (6.36), called the Einstein tensor,

$$G^{\mu\nu} = R^{\mu\nu} - \tfrac{1}{2} R g^{\mu\nu}, \tag{11.79}$$

i.e., the combination $(R^{\mu\nu} + aRg^{\mu\nu})$ with the constant fixed $a = -1/2$ is covariantly constant:

$$D_\mu G^{\mu\nu} = 0. \tag{11.80}$$

$G^{\mu\nu}$ is a covariantly constant, rank-2, symmetric tensor involving the second derivatives of the metric $\partial^2 g$ as well as terms quadratic in ∂g—exactly what we were seeking for the GR field equation; cf. (6.5).

11.3 GR tensor equations

According to the strong EP, gravity can always be transformed away locally. As first discussed in Section 5.2, Einstein suggested an elegant formulation of the new theory of gravity based on a curved spacetime. The EP is thus fundamentally built into the theory. Local flatness (a metric structure of spacetime) means that SR (the theory of flat spacetime with no gravity) is automatically incorporated into the new theory. Gravity is modeled not as a force but as the structure of spacetime; free particles just follow geodesics in a curved spacetime.

11.3.1 The principle of general covariance

The field equation for the relativistic potential, the metric function $g_{\mu\nu}(x)$, must have the same form no matter what generalized coordinates are used to label worldpoints (events) in spacetime. One expresses this by the requirement that the physics equations must satisfy the principle of general covariance. This is a two-part statement:

1. Physics equations must be covariant under the general coordinate transformations that leave the infinitesimal spacetime line element interval ds^2 invariant.
2. Physics equations should reduce to the correct SR form in the local inertial frames. Namely, we must have the correct SR equations in free-fall frames, in which gravity is transformed away. Additionally, gravitational equations reduce to Newtonian equations in the limit of low-velocity particles in a weak and static field.

The minimal substitution rule

This general principle provides us with a well-defined path to go from SR equations (which are valid in local inertial frames with no gravity) to GR equations that are valid in every coordinate system in curved spacetime (curved because of the presence of gravity). GR equations must be covariant under general local transformations. The key feature of a general coordinate transformation, in contrast to the (Lorentz) transformation in flat spacetime, is its spacetime dependence—hence the requirement for covariant derivatives. To go from an SR equation to the corresponding GR equation is simple: we need to replace the ordinary derivatives (∂) in SR equations by covariant derivatives (D):

$$\partial \longrightarrow D \ \ (= \partial + \Gamma). \tag{11.81}$$

This is known as the minimal substitution rule, because we are assuming the absence of the (high-order) Riemann tensor $R^{\mu}_{\ \nu\lambda\rho}$ terms, which vanish in the flat-spacetime limit. Since the Christoffel symbols Γ are derivatives of the metric (i.e., they represent the gravitational field strength), the introduction of covariant derivatives naturally brings the gravitational field into the physics equations. In this way, we can, for example, find the equations that describe electromagnetism in the presence of a gravitational field:

$$\partial_\mu F^{\mu\nu} = -\frac{1}{c} j^\nu \quad \longrightarrow \quad D_\mu F^{\mu\nu} = -\frac{1}{c} j^\nu. \tag{11.82}$$

Namely, Gauss's and Ampère's laws of electromagnetism in the presence of a gravitational field take on the form

$$\partial_\mu F^{\mu\nu} + \Gamma^{\mu}_{\mu\lambda} F^{\lambda\nu} + \Gamma^{\nu}_{\mu\lambda} F^{\mu\lambda} = -\frac{1}{c}\, j^\nu, \tag{11.83}$$

with gravity entering through the Christoffel symbols.

GR equation of motion

Gravity in Einstein's formulation is not a force, but the structure of spacetime. Thus the motion of a particle under the sole influence of gravity is the motion of a free particle in spacetime. At this point, instead of arguing heuristically (as in Section 5.3) that the geodesic equation should be the equation of motion in GR, the minimal substitution rule actually provides us with a formal way to arrive at this conclusion. The SR equation of motion for a free particle is $d\dot{x}^\mu/d\tau = 0$, where τ is the proper time and $\dot{x}^\mu = dx^\mu/d\tau$ the 4-velocity of the particle. According to the minimal substitution rule, the corresponding GR equation should then be

$$\frac{D\dot{x}^\mu}{d\tau} = 0, \tag{11.84}$$

which we recognize from (11.37) and (11.38) to be simply the geodesic equation.

11.3.2 Einstein field equation

We note that the above minimal substitution procedure cannot be used to obtain the GR field equation, because there is no SR gravitational field equation. The search for the GR field equation must start all the way back at Newton's equation as discussed at the beginning of Chapter 6, cf. (6.4). Since the energy–momentum right-hand side of the field equation (6.4) must satisfy the GR conservation condition,

$$D_\mu T^{\mu\nu} = 0, \tag{11.85}$$

we must have for the geometric left-hand side a covariantly constant, symmetric, rank-2 tensor involving metric derivatives $(\partial^2 g), (\partial g)^2 \sim \partial\Gamma, \Gamma^2$, which are just the properties of the Einstein curvature tensor $G_{\mu\nu}$. This then leads to the Einstein field equation:

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = \kappa T_{\mu\nu}, \tag{11.86}$$

where κ is a proportionality constant. This in general is a set of six coupled partial differential equations. The number of independent elements of the symmetric tensors $G_{\mu\nu}$ and $T_{\mu\nu}$ are reduced from ten to six by the four component equations of (11.80) and (11.85), the covariant constancy conditions. This number of independent equations just matches the number of independent elements of the metric solution $g_{\mu\nu}(x)$—whose ten components, as we explained in Sidenote 12 in Chapter 6, are reduced to six by the requirement of general coordinate transformation $x^\mu \to x'^\mu$ so that the metric $g_{\mu\nu}(x')$ must still be a solution.

This field equation can be written in an alternative form. Taking the trace of the above equation, we have

$$-R = \kappa T, \tag{11.87}$$

where T is the trace of the energy–momentum tensor, $T = g^{\mu\nu} T_{\mu\nu}$. In this way, we can rewrite (11.86) in an equivalent form by replacing $Rg_{\mu\nu}$ by $-\kappa\, Tg_{\mu\nu}$:

$$R_{\mu\nu} = \kappa(T_{\mu\nu} - \tfrac{1}{2} Tg_{\mu\nu}). \tag{11.88}$$

Exercise 11.6 Newtonian limit of Einstein's field equation

Show that Newton's gravitational field law, written in differential form (4.7), is the leading-order approximation to the Einstein field equation (11.88) in the Newtonian limit (cf. Section 5.3.1) for a slow ($v \ll c$) source particle producing a static and weak gravitational field. In this way, one can also establish the connection between the proportionality constant κ and Newton's constant as shown in (6.39).

Einstein equation via the principle of least action In the above derivation of the GR field equation, we worked directly with the rank-2 Einstein tensor, which must satisfy a bunch of necessary conditions. $T^{\mu\nu}$ and hence $G^{\mu\nu}$ had to be covariantly constant to satisfy the constraint of energy–momentum conservation in GR, which is rather complicated conceptually. The same field equation can be obtained more systematically via the principle of least action (cf. Box 5.1). The action for any field, instead of being a time integral of the Lagrangian L for a particle as in (5.18), is a 4D integral of the Lagrangian density $\mathcal{L}$. The invariant 4D volume differential product[13] is $\sqrt{-g}\, d^4x$, where g is the determinant of the metric tensor $g_{\mu\nu}$. The relevant action for the GR metric field in vacuum is called the Einstein–Hilbert action,

$$S_g = \int \mathcal{L}_g \sqrt{-g}\, d^4x = \int g^{\mu\nu} R_{\mu\nu} \sqrt{-g}\, d^4x. \tag{11.89}$$

For the invariant source-free GR Lagrangian density, one makes the natural identification with the Ricci scalar: $\mathcal{L}_g = R$. We can then derive the $T_{\mu\nu} = 0$ version of (11.86) as the Euler–Lagrange equation resulting from minimization of this action. The variation of the action δS_g has three pieces involving $\delta g^{\mu\nu}$, $\delta R_{\mu\nu}$, and $\delta\sqrt{-g}$. The integral containing the $\delta R_{\mu\nu}$ factor, after an integration by parts, turns into a vanishing surface term. The variation $\delta\sqrt{-g}$ is easiest to compute for a diagonal metric. The determinant is the product of its elements $g = \prod_\mu g_{\mu\mu}$, so its variation (using the inverse metric $g^{\mu\nu}$) is simplified to $\delta g = g\sum_\mu \delta g_{\mu\mu}/g_{\mu\mu} = gg^{\mu\nu}\,\delta g_{\mu\nu} = -gg_{\mu\nu}\,\delta g^{\mu\nu}$. As the metric matrix is symmetric, we can diagonalize it by coordinate transformation; this expression for δg must be valid for all coordinates. Thus $\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\,\delta g^{\mu\nu}$. Consequently, the variational principle requires

$$\delta S_g = \int \sqrt{-g}\, d^4x\, (R_{\mu\nu} - \tfrac{1}{2} Rg_{\mu\nu})\, \delta g^{\mu\nu} = 0, \tag{11.90}$$

which implies the Einstein equation (11.86).[14]

[13] Coordinate transformation involves the Jacobian $\mathcal{J}$ so that $d^4x' = d^4x\mathcal{J}$ and $g' = g\mathcal{J}^{-2}$ as the metric has two lower indices hence two inverse transformation factors. Thus $\sqrt{-g'}\, d^4x' = \sqrt{-g}\, d^4x$.

[14] It is possible to incorporate terms for other fields into the Lagrangian density to derive the field equations that describe the interaction of gravity with other sources.

The Schwarzschild solution The Einstein equation (11.88), in the exterior region (where $T_{\mu\nu} = 0$) of a spherical source, merely states that the Ricci tensor vanishes: $R_{\mu\nu} = 0$. From Box 6.2, we learnt that a spherically symmetric metric involves only two scalar functions, g_{00} and g_{rr}. We first express in terms of g_{00} and g_{rr} the Christoffel symbols $\Gamma^{\mu}_{\nu\lambda}$, and then the Riemann curvature tensor $R^{\mu}{}_{\nu\lambda\rho}$, which is then contracted to give the Ricci tensor $R_{\mu\nu}$. The field equations ($R_{00} = R_{rr} = 0$) are two coupled differential equations for g_{00} and g_{rr}. One of them simply yields the relation $g_{rr} = -1/g_{00}$ of (6.57); the other is an ordinary differential equation for g_{00},

$$\frac{dg_{00}}{dr} + \frac{g_{00}}{r} = -\frac{1}{r}. \tag{11.91}$$

The solution to the homogeneous part ($d\bar{g}_{00}/dr + \bar{g}_{00}/r = 0$) is $\bar{g}_{00} = r^*/r$ for some constant r^*. By adding to the easily checked particular solution, $g_{00,\text{part}} = -1$, we obtain the general result shown in (6.57): $g_{00} = -1 + r^*/r$.

Einstein's cosmological constant The energy–momentum tensor $T_{\mu\nu}$ and Einstein curvature tensor $G_{\mu\nu}$ are covariantly constant; clearly, any term to be added to the GR field equation must also have this property. This requirement for mathematical consistency allowed Einstein to insert another metric term as in (8.74),[15] because the metric tensor itself is covariantly constant, as shown in (11.27).

[15] The gravitational Lagrangian density is modified as $\mathcal{L}_g = R + 2\Lambda$; the variation of the resulting action then leads to (8.74).

Review questions

1. What is the fundamental difference between coordinate transformations in a curved space and those in flat space (e.g., Lorentz transformations in flat Minkowski space)?

2. Writing the coordinate transformation as a partial derivative matrix, give the transformation laws for a contravariant vector $A^{\mu} \longrightarrow A'^{\mu}$ and a covariant vector $A_{\mu} \longrightarrow A'_{\mu}$, as well as for a mixed tensor $T^{\mu}_{\nu} \longrightarrow T'^{\mu}_{\nu}$.

3. From the transformation of $A_{\mu} \longrightarrow A'_{\mu}$ in the answer to the previous question, work out the coordinate transformation of the derivatives $\partial_{\mu}A_{\nu}$. Why do we say that $\partial_{\mu}A_{\nu}$ is not a tensor? What is the underlying reason why $\partial_{\mu}A_{\nu}$ is not a tensor? How do the covariant derivatives $D_{\mu}A_{\nu}$ transform? Why is it important to have differentiations that result in tensors?

4. Write out the covariant derivative $D_{\mu}T^{\lambda\rho}_{\nu}$ (in terms of the connection symbols) of a mixed tensor $T^{\lambda\rho}_{\nu}$.

5. The relation between the Christoffel symbol and the metric tensor is called "the fundamental theorem of Riemannian geometry." Write out this relation.

6. What is the flatness theorem? Use this theorem to show that the metric tensor is covariantly constant: $D_{\mu}g_{\nu\lambda} = 0$.

7. As the Christoffel symbols $\Gamma^{\mu}_{\alpha\beta}$ are not components of a tensor (and ordinary derivatives generally do not yield tensors), how do we know that $R^{\mu}{}_{\lambda\alpha\beta} = \partial_{\alpha}\Gamma^{\mu}_{\lambda\beta} - \partial_{\beta}\Gamma^{\mu}_{\lambda\alpha} + \Gamma^{\mu}_{\nu\alpha}\Gamma^{\nu}_{\lambda\beta} - \Gamma^{\mu}_{\nu\beta}\Gamma^{\nu}_{\lambda\alpha}$ is really a tensor?

8. The quantitative description of tidal force is the Newtonian deviation equation. What is its GR analog?

What geometric quantity of a curved spacetime replaces the second derivatives of the gravitational potential?

9. In Einstein's search for a linear combination of $R^{\mu\nu}$ and $Rg^{\mu\nu}$ that is covariantly constant, why is such a constraint required? What is an efficient way to find this combination?

10. What is the principle of general covariance?

11. If a physics equation is known in the SR limit, how does one form its GR analog? Since SR equations are valid only in the absence of gravity, turning them into their GR versions implies the introduction of a gravitational field into the relativistic equations. How does gravity enter into this alteration of the equations?

12. How can one determine the GR equation of motion from that of SR? Why can we not find the GR field equation in the same way?