The Physical Geometry of Spacetime

General Relativity Lecture Notes

Jericho Cain, PhD

Fall 2025

Contents

Preface: From Pure Geometry to Spacetime

Modern general relativity stands at the crossroads of geometry and physics. Its equations describe how the geometry of spacetime determines the motion of matter, and how matter in turn shapes geometry. To understand this interplay, one must first learn the mathematical language that expresses geometry precisely: the theory of smooth manifolds.

This language provides the setting for all that follows. A spacetime is modeled as a smooth four-dimensional manifold M, and the objects of physics, fields, momenta, and curvature, are smooth tensor fields defined on M. The fundamental laws of physics are statements about how these geometric objects transform and interact under smooth maps between manifolds.

Mathematically, the framework of smooth manifolds is developed in works such as John M. Lee's *Introduction to Smooth Manifolds*, while the physicist's perspective is presented in Robert M. Wald's *General Relativity*. The two texts share the same geometric foundation, but speak in different dialects: Lee emphasizes rigor and abstraction, while Wald focuses on physical interpretation and tensor calculus. The aim of these notes is to bridge the two.

Here we reinterpret the core ideas of differential geometry in the language and intuition of spacetime physics:

- how tangent spaces become the spaces of physical directions at each event,
- how the differential encodes coordinate transformations and embeddings,
- and how tensor bundles carry the physical quantities that define geometry.

Our goal is not to replace mathematical rigor with physical intuition, but to reveal the unity between them. By translating the mathematics of smooth manifolds into the language of general relativity, we arrive at a single coherent picture: spacetime as a geometric manifold whose curvature embodies gravity itself.

Once this foundation is established, the transition to Wald's Chapter 2 becomes seamless: the abstract structures of differential geometry reappear as the concrete geometric tools of spacetime physics, leading naturally toward curvature, dynamics, and the Einstein field equations.

Chapter 2

Spacetime as a Differentiable Manifold

Imagine a civilization of beings who live in a one–dimensional universe. They can move only forward or backward along a single line. One day, one of them notices that walking through a certain region requires more effort—motion feels heavier there. To us, higher–dimensional observers, the explanation is obvious: the line is bending through space, and the being is walking uphill.

But for the one–dimensional inhabitants, there is no concept of "up" or "curvature." They cannot appeal to a surrounding $\mathbb{R}^3$ to explain their experience. Their entire physics must be described *within* their own line. To do so, they would have to invent a calculus native to their world: a way to describe change and motion using only information intrinsic to the line itself.

This is precisely what differential geometry accomplishes. We do not imagine that our spacetime manifold M is embedded in some higher Euclidean space and borrow the derivative from there. Instead, we construct calculus *on* M directly. Every point of M carries its own local coordinate system, its own tangent space, and its own family of derivative operators. Together these local versions of calculus vary smoothly across M, defining the geometry of spacetime from within.

What feels like "resistance" to the one–dimensional beings is, to us, curvature.

Overview of Chapter 2: Foundations of the Differential Geometry of Spacetime

In this chapter we construct the mathematical framework that allows spacetime to have a geometry of its own. Our goal is to define what it means to speak of "direction," "change," and "structure" on a smooth manifold without appealing to any external space.

We begin by formalizing the notion of a *tangent vector*: an intrinsic way to describe motion or differentiation at a single point. From this idea we build the tangent space T_pM, which collects all possible directions of change at p, and the cotangent space T_p^*M, its dual, which contains linear functionals that act on vectors. These two spaces—the tangent and cotangent spaces—are the atoms from which all other geometric quantities are assembled.

With these pieces in hand we define tensors as multilinear maps built from vectors and covectors, and tensor fields as smooth assignments of such objects to each point of the manifold. Through them we can express every geometric or physical field that lives on spacetime—scalars, vectors, metrics, and beyond— within one unified algebraic language.

Finally, we learn to treat tangent vectors as *derivations* on the space of smooth functions, giving them the power to express directional derivatives and flows. This interpretation is the seed of calculus on manifolds: it will allow us, in Chapter 3, to define the covariant derivative, the connection, and ultimately curvature itself.

2.1 Manifolds and Coordinate Systems

In general relativity, spacetime is modeled as a smooth four-dimensional manifold M. Locally, it can always be described by a set of coordinates $x^\mu = (x^0, x^1, x^2, x^3)$, but no single coordinate system covers all of spacetime. Instead, spacetime is assembled from overlapping coordinate patches, each smoothly related to its neighbors.

Definition 2.1. Smooth Manifold: A *topological n-manifold* is a topological space M satisfying:

1. M is *Hausdorff* (any two distinct points have disjoint neighborhoods);
2. M is *second-countable* (it possesses a countable basis for its topology); and
3. every point $p \in M$ has a neighborhood U that is homeomorphic to an open subset of $\mathbb{R}^n$.

A *smooth structure* on M is a collection of charts $(U_\alpha, \varphi_\alpha)$ whose domains cover M and whose *transition maps* $\varphi_\beta \circ \varphi_\alpha^{-1}$ are smooth wherever defined. A topological manifold equipped with such a maximal collection of compatible charts is called a *smooth manifold*.

Each chart (U, φ) identifies an open subset $U \subset M$ with an open region $\hat{U} = \varphi(U) \subset \mathbb{R}^4$, whose standard coordinates (x^0, x^1, x^2, x^3) serve as the local coordinate system on U. Figure 2.1 illustrates this structure: each region of the manifold is mapped smoothly into a Euclidean patch, and the overlaps between charts are connected by smooth *transition maps*. This framework provides the mathematical foundation for expressing physics in a coordinate-independent way.

Wald's perspective. A spacetime is a smooth four-dimensional manifold. Every point $p \in M$ lies in some neighborhood U that can be labeled by smooth coordinates x^μ, and any smooth change of coordinates $x^{\mu'} = x^{\mu'}(x^0, x^1, x^2, x^3)$ preserves the differentiable structure. In Lee's terminology, these coordinates arise as the component functions of the chart map φ.

Lee's formal framework.

- A *chart* on M is a pair (U, φ), where $U \subseteq M$ is open and $\varphi : U \to \mathbb{R}^4$ is a homeomorphism onto its image. The image $\hat{U} = \varphi(U) \subset \mathbb{R}^4$ represents the local coordinate domain.
- The components of $\varphi(p)$ define the *coordinate functions* $x^\mu(p) = \pi^\mu \circ \varphi(p)$, so that $\varphi(p) = (x^0(p), x^1(p), x^2(p), x^3(p))$.
- The components of $\varphi(p)$ define the *coordinate functions* $x^\mu(p) = \pi^\mu \circ \varphi(p)$, where $\pi^\mu : \mathbb{R}^4 \to \mathbb{R}$ denotes the projection onto the μth coordinate. Thus $\varphi(p) = (x^0(p), x^1(p), x^2(p), x^3(p))$.

- If (U, φ) and (V, ψ) are charts with $U \cap V \neq \varnothing$, the *transition map* $\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V)$ must be smooth.
- A collection of mutually compatible charts forms an *atlas*; its maximal extension defines the smooth structure on M.

Common Terminology.

- **Bijective.** A map $f : X \to Y$ is *bijective* if it is both injective (one-to-one) and surjective (onto), so that each $y \in Y$ corresponds to exactly one $x \in X$.
- **Homeomorphism.** A bijective map $f : X \to Y$ between topological spaces is a *homeomorphism* if both f and its inverse f^{-1} are continuous. Homeomorphic spaces share the same topological structure.
- **Diffeomorphism.** A bijective map $f : M \to N$ between smooth manifolds is a *diffeomorphism* if both f and f^{-1} are smooth. Diffeomorphic manifolds have identical differentiable structures.
- **Isomorphism (general use).** The term *isomorphism* means "structure-preserving bijection." Its precise meaning depends on context: a homeomorphism is an isomorphism in the category of topological spaces, and a diffeomorphism is an isomorphism in the category of smooth manifolds.

Bridging the viewpoints. Wald's coordinates x^μ are precisely the component functions of the chart map φ in Lee's formalism. A "coordinate system" in physics is therefore a chart (U, φ), and the requirement that coordinate transformations be smooth is the mathematical statement that the transition maps $\psi \circ \varphi^{-1}$ are smooth diffeomorphisms. The compatibility of these maps ensures that all observers— each using their own coordinates—describe overlapping regions of spacetime consistently (see Fig. 2.1).

Physical interpretation. The manifold M represents the set of all possible spacetime events. Each chart (U, φ) corresponds to a local observer assigning coordinates x^μ to those events within their accessible region. Smooth transition maps express the physical requirement that when two observers describe the same region of spacetime, their coordinate systems are related by smooth, differentiable transformations.

Example: Minkowski spacetime. In special relativity, $M = \mathbb{R}^4$ equipped with the standard Cartesian coordinates

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

Here a single chart covers all of M because spacetime is globally Euclidean as a manifold (though not as a metric space), and the transition functions are simply the Lorentz transformations, which are smooth diffeomorphisms of $\mathbb{R}^4$ (Fig 2.2).

Here the minus sign that distinguishes the time direction appears in the metric tensor,

$$g = -dt^2 + dx^2 + dy^2 + dz^2,$$

not in the coordinate labels themselves.

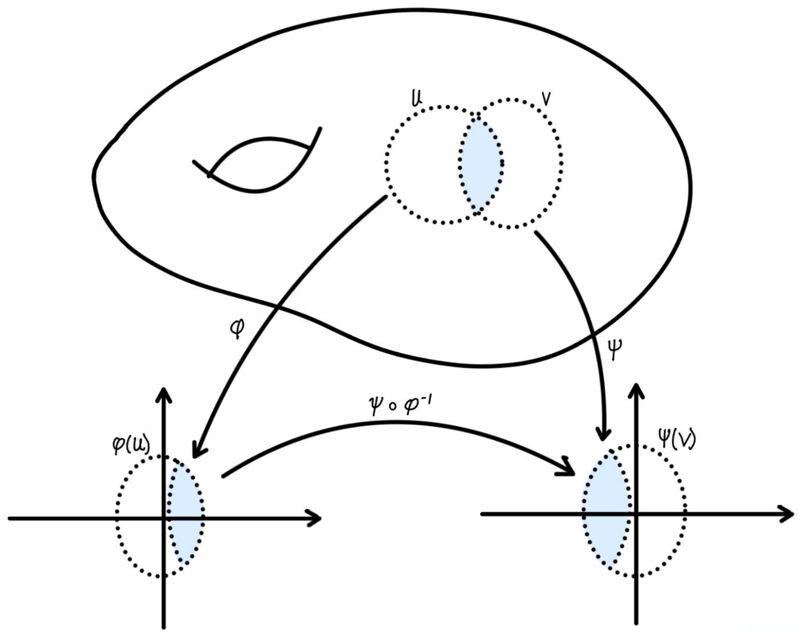

Figure 2.1: Overlapping coordinate charts (U, φ) and (V, ψ) on a manifold M. The transition map $\psi \circ \varphi^{-1}$ relates the coordinates assigned to the same events in the overlap $U \cap V$. The smoothness of these maps is what makes M a smooth manifold.

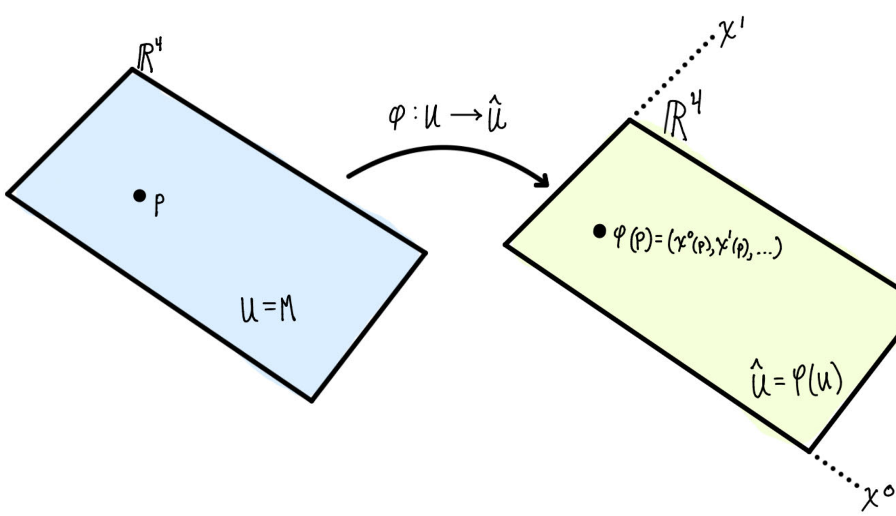

Figure 2.2: A schematic representation of the global chart for Minkowski spacetime. A point $p \in M$ on the manifold is mapped by $\varphi : M \to \mathbb{R}^4$ to its coordinate image $\varphi(p) = (x^0(p), x^1(p), x^2(p), x^3(p))$. Only two coordinate directions (x^0, x^1) are shown; the remaining components (x^2, x^3) are suppressed. The diagram represents the abstract correspondence between the manifold and its coordinate space, not a literal embedding in higher dimensions.

Note 2.1. In the notation $x^\mu(p)$, each $x^\mu : U \to \mathbb{R}$ is a coordinate function that assigns to the point $p \in M$ its corresponding coordinate value. In Minkowski spacetime, these are simply the familiar coordinates $x^0(p) = t,\ x^1(p) = x,\ x^2(p) = y,\ x^3(p) = z$, or $x^0(p) = ct$ if we use time units of length. This notation does not denote an algebraic operation, but rather the coordinate value of the point p under the chart map.

Remark. A chart (U, φ) does not embed the manifold into $\mathbb{R}^n$; it installs a local coordinate system on the patch U. The image $\hat{U} = \varphi(U)$ is a Euclidean model of that region, not a geometric copy. Depicting $\varphi : U \to \hat{U}$ in diagrams emphasizes that coordinates are assigned values in $\mathbb{R}^n$, not that the manifold itself lives there.

Example: S^2

This example serves as background for Problem 1

(a). Charts on the* 2*-sphere. Let

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

For $i \in \{1, 2, 3\}$ and a sign $\pm$, set

$$U_1^\pm = \{(x, y, z) \in S^2 :\ \pm x > 0\},$$
$$U_2^\pm = \{(x, y, z) \in S^2 :\ \pm y > 0\},$$
$$U_3^\pm = \{(x, y, z) \in S^2 :\ \pm z > 0\}.$$

These are the six open hemispheres as shown in Fig. 2.3. On each $U_i^\pm$ we use the two

remaining Cartesian coordinates as chart coordinates:

$$f_1^\pm(x, y, z) = (y, z), \qquad f_2^\pm(x, y, z) = (x, z), \qquad f_3^\pm(x, y, z) = (x, y).$$

Using our terminology,

- $S^2 \subset \mathbb{R}^3$ is the manifold.
- $p = (x, y, z) \in S^2$ is a point on the manifold.
- $U_i^\pm \subset S^2$ is an open hemisphere containing p (our open subsets).
- $f_i^\pm : U_i^\pm \to \mathbb{R}^2$ is a coordinate chart.

So when $p \in U_i^\pm$, the chart $f_i^\pm$ assigns Euclidean coordinates in $\mathbb{R}^2$ to that point. Here (x, y, z) is the geometric point in the ambient space $\mathbb{R}^3$. With the constraint $x^2 + y^2 + z^2 = 1$, it represents a point on the manifold S^2. These coordinates are not local coordinates on the manifold; they come from the embedding $S^2 \subset \mathbb{R}^3$. More on embeddings later.

The local coordinates on the manifold are $(u, v) \in \mathbb{R}^2$. They depend on the chosen chart and parametrize a neighborhood of the point on S^2. Formally,

$$(u, v) = f_i^\pm(x, y, z)$$

Important: Even when (u, v) happens to equal (x, y) numerically (as in the chart f_3^+), they play different conceptual roles.

For example, suppose $p = (x, y, z) \in U_3^+ \quad (z > 0)$. Then the chart is, $f_3^+(x, y, z) = (x, y) = (u, v)$. So (x, y, z) = geometric point on the sphere and $(u, v) = (x, y)$ = local coordinates in $\mathbb{R}^2$. Recovering the geometric point from its local coordinates gives $(x, y, z) = (u, v, \sqrt{1 - u^2 - v^2})$.

The image of each chart is the open unit disk $\mathbb{D} = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 < 1\}$.

Why an open disk? Consider a point $p = (x, y, z) \in U_1^- \subset S^2$. The chart $f_1^- : U_1^- \to \mathbb{R}^2$ assigns coordinates $f_1^-(p) = (y, z)$. Here we identify (y, z) with the local coordinates (u, v) - that is the coordinates on $\mathbb{R}^2$. The image, $f_1^-(p)$ is the local Euclidean coordinate representation of the manifold point p. Because $p \in S^2$ it is constrained by $x^2 + y^2 + z^2 = 1$, therefore $y^2 + z^2 = 1 - x^2$. Since $p \in U_1^-$, we know $x < 0$, but more importantly we know that $x \neq 0$, hence $x^2 > 0$, therefore, $y^2 + z^2 < 1$. Thus,

$$(y, z) \in \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 < 1\}.$$

This is exactly an open unit disk. The openness of the disk reflects the fact that U_1^- is an open hemisphere; using a closed hemisphere would produce a closed disk, which is not allowed for a chart.

Each chart $f_i^\pm : U_i^\pm \to \mathbb{R}^2$ “forgets" one coordinate. The inverse chart recovers the omitted coordinate using the sphere constraint $x^2 + y^2 + z^2 = 1$, with the sign determined by the chosen hemisphere. Explicitly, for $(u, v) \in \mathbb{D}$,

$$\begin{aligned}
(f_1^\pm)^{-1}(u, v) &= (\ \pm\sqrt{1 - u^2 - v^2},\ u,\ v), \\
(f_2^\pm)^{-1}(u, v) &= (\ u,\ \pm\sqrt{1 - u^2 - v^2},\ v), \\
(f_3^\pm)^{-1}(u, v) &= (\ u,\ v,\ \pm\sqrt{1 - u^2 - v^2}).
\end{aligned}$$

What does the inverse chart do? Given local coordinates $(u, v) \in \mathbb{R}^2$, the inverse map $f_i^\pm$ returns the unique geometric point, $p = (x, y, z) \in U_i^\pm \subset S^2$ having those coordinates. In other words, $p = (f_i^\pm)^{-1}(f_i^\pm(p))$, so the inverse chart reconstructs the original geometric point on the sphere from its local Euclidean coordinates.

What you will check in 1(a). For any pair of charts with nonempty overlap, the *overlap map*

$$f_i^\pm \circ (f_j^\pm)^{-1} :\ f_j^\pm(U_i^\pm \cap U_j^\pm) \subset \mathbb{D} \longrightarrow \mathbb{D}$$

is a smooth map. For example, between U_3^+ and U_1^+,

$$(f_1^+ \circ (f_3^+)^{-1})(u, v) = (v,\ \sqrt{1 - u^2 - v^2}),$$

which is C^∞ on its domain (where $1 - u^2 - v^2 > 0$).

Anchor: Consider the geometric point $p = (1, 0, 0)$.

- $1^2 + 0^2 + 0^2 = 1$, so $p \in S^2$.
- $x = 1 > 0$, so $p \in U_1^+$.

- $y = 0$, so $p \notin U_2^{\pm}$.
- $z = 0$, so $p \notin U_3^{\pm}$.
- The chart $f_1^+ : U_1^+ \to \mathbb{R}^2$ assigns the local coordinates

$$f_1^+(p) = (y, z) = (0, 0).$$

- Applying the inverse chart recovers the geometric point:

$$(f_1^+)^{-1}(0, 0) = \big(\sqrt{1 - 0^2 - 0^2}, 0, 0\big) = (1, 0, 0).$$

(b). Two-chart cover: stereographic projections.

Figures 2.4, 2.5 illustrate the idea behind the two stereographic projections used to construct a two–chart atlas for S^2. In the first figure, the sphere is projected onto the equatorial plane $z = 0$ from the north pole $N = (0, 0, 1)$; every point $x \in S^2 \setminus \{N\}$ is joined to N by a straight line, which intersects the plane at a point $y \in \mathbb{R}^2$. This defines the *northern stereographic projection*, $\sigma_N(x, y, x)$. The second figure shows the same construction in a meridional cross–section: for each point x on the circle (representing a point on the sphere), a line through the north pole meets the plane at the image point y. Performing the analogous construction from the south pole $S = (0, 0, -1)$ gives the *southern stereographic projection*, $\sigma_S(x, y, z)$,

$$\sigma_N : S^2 \backslash \{N\} \longrightarrow \mathbb{R}^2, \qquad \sigma_N(x, y, z) = \Big(\frac{x}{1 - z}, \frac{y}{1 - z}\Big),$$
$$\sigma_S : S^2 \backslash \{S\} \longrightarrow \mathbb{R}^2, \qquad \sigma_S(x, y, z) = \Big(\frac{x}{1 + z}, \frac{y}{1 + z}\Big).$$

Their inverses are smooth on all of $\mathbb{R}^2$:

$$\sigma_N^{-1}(u, v) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right),$$
$$\sigma_S^{-1}(u, v) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, -\frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right).$$

Thus $\{(S^2 \backslash \{N\}, \sigma_N),\ (S^2 \backslash \{S\}, \sigma_S)\}$ is a two-chart atlas covering S^2.

Summary. A smooth manifold provides the coordinate-independent setting for spacetime geometry:

- Locally, each point has coordinates x^μ, defined by a chart (U, φ).
- Globally, the manifold is patched together from overlapping coordinate systems related by smooth maps.
- Physical fields must transform smoothly under such changes of coordinates, ensuring that their meaning is intrinsic to spacetime, not tied to any single chart.

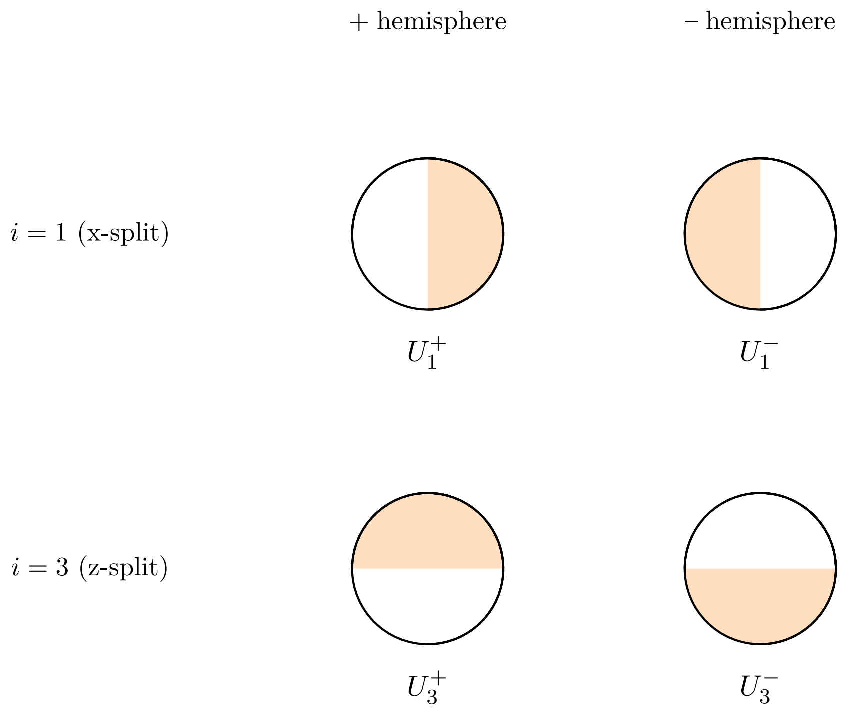

Figure 2.3: The six open hemispheres $U_i^{\pm} = \{(x, y, z) \in S^2 : \pm x_i > 0\}$ for $i = 1, 2, 3$ (with $x_1 = x,\ x_2 = y,\ x_3 = z$). Each chart $f_i^{\pm}$ uses the other two Cartesian coordinates as local coordinates on $U_i^{\pm}$; e.g. on $U_3^{\pm}$, $f_3^{\pm}(x, y, z) = (x, y)$. Shading indicates the hemisphere where the indicated coordinate is positive (left column) or negative (right column). Only the x- and z-splits $(i = 1, 3)$ are shown here for clarity—these can be drawn without obscuring the geometry. The y-split hemispheres $(i = 2)$ are analogous but lie in planes that would appear edge-on in this projection and are omitted.

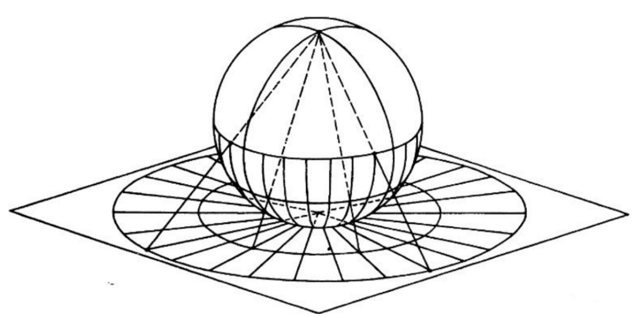

Figure 2.4: Stereographic projection from the north pole $N = (0, 0, 1)$ of the sphere onto the equatorial plane $z = 0$. Each line through N and a point $(x, y, z) \in S^2 \setminus \{N\}$ intersects the plane at a unique point $(u, v) \in \mathbb{R}^2$, defining the stereographic chart $\sigma_N(x, y, z) = (x/(1 - z),\ y/(1 - z))$. The entire sphere minus N is thus mapped smoothly onto the plane.

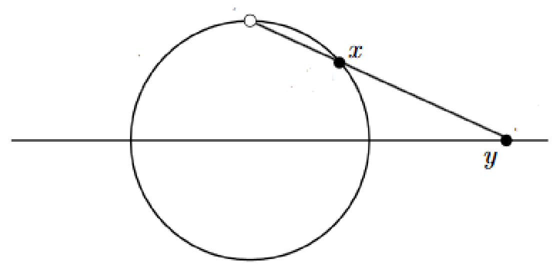

Figure 2.5: Cross-sectional view of stereographic projection. The point x on the circle represents a point on S^2, and the line through the north pole (open circle) meets the plane at the image point y. This geometric construction yields the coordinate map $\sigma_N : S^2 \setminus \{N\} \to \mathbb{R}^2$, and a similar projection from the south pole S defines σ_S.

2.2 Smooth Maps Between Manifolds

Up to now, we have described how coordinate systems are installed on regions of a manifold M by smooth maps $\varphi : U \to \mathbb{R}^n$. We now generalize this idea: a *smooth map* between manifolds $F : M \to N$ relates points of one manifold to points of another in a way that is compatible with their differentiable structures. The guiding principle is simple: *a map between manifolds is smooth if, after choosing coordinates on the domain and codomain, it becomes an ordinary smooth map between open subsets of Euclidean space.* This viewpoint is the common language behind coordinate transformations, embeddings, and physical fields.

Definition (Lee). Let $F : M \to N$ be a map between smooth manifolds. We say that F is *smooth at a point* $p \in M$ if there exist charts (U, φ) around p in M and (V, ψ) around $F(p)$ in N such that the coordinate representative

$$\psi \circ F \circ \varphi^{-1} : \ \varphi(U \cap F^{-1}(V)) \ \longrightarrow \ \psi(V)$$

is a smooth map between open subsets of Euclidean space. If F is smooth at every point of M, we call F a *smooth map*. (See Fig. 2.6.)

In physics language, this says that a coordinate transformation $x^{\mu'} = x^{\mu'}(x^\nu)$ is smooth precisely when its coordinate expression is a smooth function, and similarly a field $f : M \to \mathbb{R}$ (a scalar field) is smooth when it varies differentiably in every smooth coordinate system. Smoothness is what guarantees that derivatives—and therefore local dynamics—are well defined.

Wald's perspective. A smooth function $f : M \to \mathbb{R}$ assigns a real number (a scalar field) to each event in spacetime. More generally, a smooth map $F : M \to N$ can represent a coordinate transformation, a projection, or another relationship between manifolds. Smoothness ensures that physical quantities vary continuously and differentiably throughout spacetime.

Finally, smooth maps are stable under composition: if $F : M \to N$ and $G : N \to P$ are smooth, then $G \circ F : M \to P$ is smooth. In coordinates this is exactly the multivariable chain rule, and it

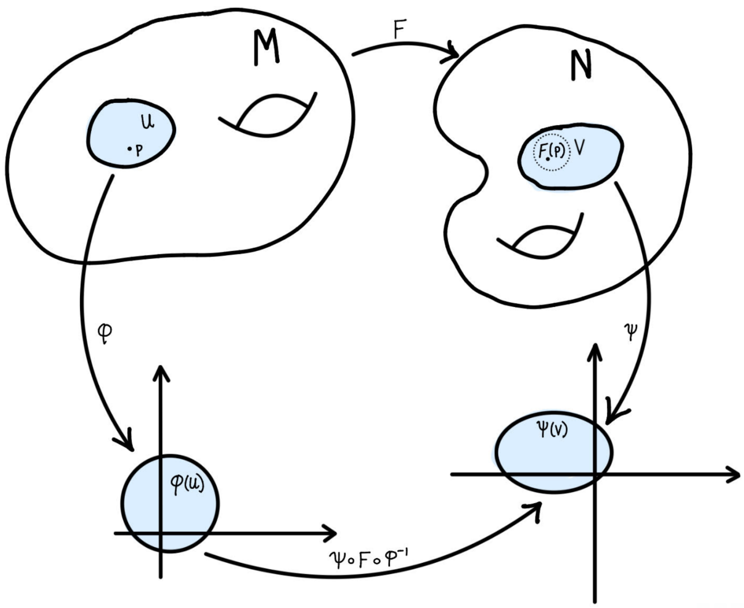

Figure 2.6: A smooth map $F : M \to N$ between manifolds. In local coordinates, the composition $\psi \circ F \circ \varphi^{-1}$ is an ordinary smooth map between regions of $\mathbb{R}^m$ and $\mathbb{R}^n$.

formalizes the idea that successive coordinate changes preserve differentiability.

We now illustrate the definition of a smooth map between manifolds with two concrete examples. In each case, both the domain and codomain are smooth manifolds, and smoothness is verified by expressing the map in local coordinates.

The first example is a map $S^2 \to \mathbb{R}$, which represents a scalar field on the sphere. This is the simplest and most important class of manifold maps in both mathematics and physics: functions that assign a real number to each point of a manifold. Such maps appear throughout geometry and spacetime physics, and they provide a direct bridge between manifolds and ordinary calculus.

The second example involves maps between coordinate charts on the same manifold S^2. These transition maps encode changes of coordinates and are the prototype for coordinate transformations in differential geometry and general relativity. Understanding how the same geometric point is described by different coordinate systems is essential for working with tensorial quantities later on.

We deliberately avoid more exotic examples, such as maps from S^2 to a torus, at this stage. While such maps are mathematically valid, they emphasize global topological features that are not central to our immediate goals. The examples chosen here are meant to anchor the notation and definitions in situations that students will encounter repeatedly, especially in the study of spacetime geometry, where local coordinate descriptions and scalar fields play a foundational role.

Example I (a scalar field on S^2): Define the *height function*

$$h : S^2 \to \mathbb{R}, \qquad h(x, y, z) = z.$$

- If $p = (x, y, z) \in S^2$, then $h(p) = z \in \mathbb{R}$. Thus h assigns a real number (a *scalar*) to each point on the sphere.

- The function h is smooth because it is the restriction to $S^2 \subset \mathbb{R}^3$ of the smooth ambient function
$$H : \mathbb{R}^3 \to \mathbb{R}, \qquad H(x, y, z) = z.$$
- In local coordinates this smoothness is visible directly. For example, on the chart (U_3^+, f_3^+) where
$$f_3^+(x, y, z) = (x, y) = (u, v), \qquad (f_3^+)^{-1}(u, v) = (u, v, \sqrt{1 - u^2 - v^2}),$$
the coordinate expression of h is the map
$$h \circ (f_3^+)^{-1} : \mathbb{D} \to \mathbb{R}, \qquad (h \circ (f_3^+)^{-1})(u, v) = \sqrt{1 - u^2 - v^2},$$
which is smooth on the open disk $\mathbb{D} = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 < 1\}$.
- Similarly, on the chart (U_3^-, f_3^-) one obtains
$$(h \circ (f_3^-)^{-1})(u, v) = -\sqrt{1 - u^2 - v^2},$$
again smooth on $\mathbb{D}$. Hence $h : S^2 \to \mathbb{R}$ is a smooth map between manifolds.

This construction is illustrated in Fig. 2.7, where the smoothness of $h : S^2 \to \mathbb{R}$ is determined by expressing the map in local coordinates as the ordinary function $\psi \circ h \circ (f_i^\pm)^{-1}$ between Euclidean spaces; in this case the chart ψ on $\mathbb{R}$ is the identity and is therefore suppressed.

Note: The function $H : \mathbb{R}^3 \to \mathbb{R}$ is an ambient extension used to prove smoothness, but the map we study geometrically is its restriction $h : S^2 \to \mathbb{R}$, which is the arrow shown in the diagram.

Example II (a map between coordinate charts on S^2): Let S^2 be covered by the charts $(U_i^\pm, f_i^\pm)$ defined previously. Suppose a point $p \in S^2$ lies in the overlap
$$U_1^+ \cap U_3^+ = \{(x, y, z) \in S^2 : x > 0,\ z > 0\}.$$

- The chart $f_3^+ : U_3^+ \to \mathbb{R}^2$ assigns coordinates
$$f_3^+(x, y, z) = (x, y) = (u, v).$$
- The inverse chart recovers the geometric point from these coordinates:
$$(f_3^+)^{-1}(u, v) = (u,\ v,\ \sqrt{1 - u^2 - v^2}).$$
- Applying the chart f_1^+ to this point gives new coordinates:
$$f_1^+((f_3^+)^{-1}(u, v)) = f_1^+(u,\ v,\ \sqrt{1 - u^2 - v^2}) = (v,\ \sqrt{1 - u^2 - v^2}).$$
- The map
$$f_1^+ \circ (f_3^+)^{-1} : \mathbb{D} \to \mathbb{R}^2, \qquad (u, v) \mapsto (v, \sqrt{1 - u^2 - v^2}),$$
is called a *transition map* (or *change of coordinates*) between the two charts.

This example shows how the *same geometric point* on the manifold can be described by different coordinate pairs, depending on the chosen chart. Smoothness of the manifold is encoded in the requirement that all such transition maps between overlapping charts are smooth functions between

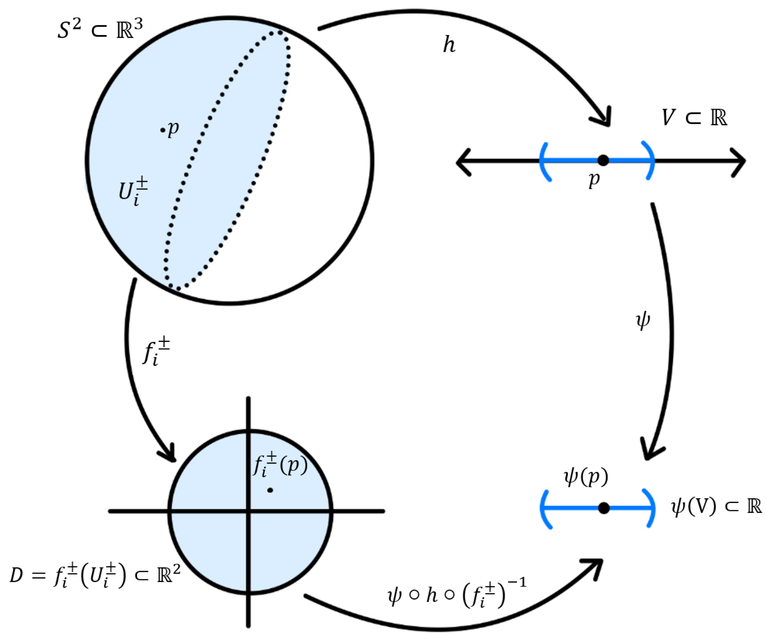

Figure 2.7: Example I: A smooth map $h : S^2 \to \mathbb{R}$ expressed in local coordinates. An open set $U_i^\pm \subset S^2$ containing p is mapped by the chart $f_i^\pm$ to its coordinate image $D = f_i^\pm(U_i^\pm) \subset \mathbb{R}^2$. An open interval $V \subset \mathbb{R}$ containing $h(p)$ serves as a chart domain on the target manifold, with coordinate map ψ. The smoothness of h at p is determined by the coordinate representative $\psi \circ h \circ (f_i^\pm)^{-1} : D \to \psi(V) \subset \mathbb{R}$, which is an ordinary smooth function between Euclidean spaces.

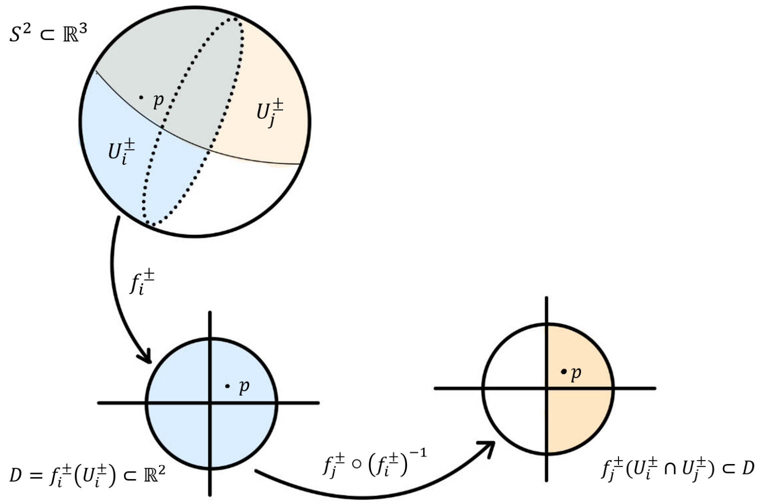

Figure 2.8: A transition map between overlapping coordinate charts on S^2. Two chart domains $U_i^\pm$ and $U_j^\pm$ overlap on the sphere, and a point p lies in their intersection. The chart $f_i^\pm$ maps $U_i^\pm$ to its coordinate image $D_i = f_i^\pm(U_i^\pm) \subset \mathbb{R}^2$. The transition map $f_j^\pm \circ (f_i^\pm)^{-1}$ is defined only on the coordinate image of the overlap $f_i^\pm(U_i^\pm \cap U_j^\pm) \subset D_i$ and takes values in $f_j^\pm(U_i^\pm \cap U_j^\pm) \subset D_j$. Smoothness of the manifold is encoded in the requirement that all such transition maps are smooth functions between open subsets of Euclidean space.

open subsets of Euclidean space. This relationship between overlapping chart domains and their coordinate images is illustrated in Fig. 2.8, where the transition map is shown acting only on the coordinate image of the overlap.

In practice, smooth maps are always analyzed locally: by choosing coordinates on the domain and codomain, smoothness reduces to ordinary multivariable calculus.

2.3 Tangent Vectors and Tangent Spaces

Up to this point, we have focused on smooth functions and smooth maps between manifolds, emphasizing that all notions of smoothness are ultimately defined using local coordinates. We now turn to a new type of geometric object: tangent vectors. Rather than assigning numbers to points of a manifold, tangent vectors encode the possible *directions of change* at a point. Although they will soon be defined in an intrinsic, coordinate-free way, tangent vectors are constructed using the same local-coordinate ideas developed above. In particular, their behavior is determined by how they act on smooth functions, whose smoothness is itself defined via charts.

At each point p on a smooth manifold M, we can define the possible *directions* in which one can move away from p. These directions form the *tangent space* T_pM, which provides the local linear approximation to the manifold. Tangent vectors are the building blocks of all tensorial objects in general relativity: velocities, momenta, and directional derivatives of fields. First, some definitions:

Definition 2.2 (Inner and Outer Leibniz Rules)**.** Let X and Y be smooth vector fields on a manifold M, acting as derivations on $C^\infty(M)$.

- **Inner Leibniz rule:** When Y acts on a product of functions,

$$Y(fg) = f\,Y(g) + g\,Y(f),$$

  expressing that Y itself satisfies the Leibniz (product) rule.
- **Outer Leibniz rule:** When X acts on a product that already contains the result of another derivation (for instance, $X(f\,Yg)$), the Leibniz rule applies again:

$$X(f\,Yg) = (Xf)\,Yg + f\,X(Yg).$$

  The same holds with X and Y interchanged. This second application of the Leibniz rule is what produces the full expansion in the proof that $[X, Y]$ is a derivation.

Definition 2.3 (Linearity of derivations)**.** Let X and Y be smooth vector fields on a manifold M, acting as derivations on the algebra of smooth functions $C^\infty(M)$. A derivation is a linear operator with respect to real scalars and addition of functions:

$$X(af + bg) = a\,X(f) + b\,X(g), \qquad a, b \in \mathbb{R},\ f, g \in C^\infty(M).$$

Consequently, for any two derivations X and Y, their commutator

$$[X, Y](f) := X(Yf) - Y(Xf)$$

is also linear:

$$[X, Y](af + bg) = a\,[X, Y](f) + b\,[X, Y](g).$$

This property, together with the Leibniz rule, ensures that $[X, Y]$ defines another derivation—hence a vector field on M.

Wald's perspective: A tangent vector at $p \in M$ is a *derivation*, that is, a linear operator X_p acting on smooth functions $f \in C^\infty(M)$ and satisfying the Leibniz rule:

$$X_p(fg) = f(p)X_p(g) + g(p)X_p(f).$$

These tangent vectors form a vector space T_pM, and in coordinates (x^μ), the natural basis is given by $\{\frac{\partial}{\partial x^\mu}\big|_p\}$.

Lee's formal framework:

- A *tangent vector at* p is a derivation $v : C^\infty(M) \to \mathbb{R}$ satisfying linearity and the Leibniz rule:

$$v(fg) = f(p)v(g) + g(p)v(f).$$

- The collection of all such derivations forms a real vector space, the *tangent space* T_pM.
- In a coordinate chart $(U, x^1, \ldots, x^n)$ containing p, we define a natural basis

$$\left\{\frac{\partial}{\partial x^i}\bigg|_p\right\}, \qquad \text{where} \quad \frac{\partial}{\partial x^i}\bigg|_p(f) = \frac{\partial(f \circ x^{-1})}{\partial x^i}\bigg|_{x(p)}.$$

- Any tangent vector can thus be written as $v = v^i \frac{\partial}{\partial x^i}\big|_p$, with components $v^i \in \mathbb{R}$.

Bridging the viewpoints. Wald's derivation-based definition of tangent vectors is equivalent to Lee's construction using coordinate charts. In practice, we think of a tangent vector $v \in T_pM$ as an infinitesimal displacement at p or as a directional derivative acting on scalar fields. If $\gamma(t)$ is a smooth curve on M with $\gamma(0) = p$, then

$$v(f) = \frac{d}{dt}\big[f(\gamma(t))\big]_{t=0}$$

defines a tangent vector $v = \dot{\gamma}(0)$, called the *velocity vector* of the curve at p. Every tangent vector can be realized in this way.

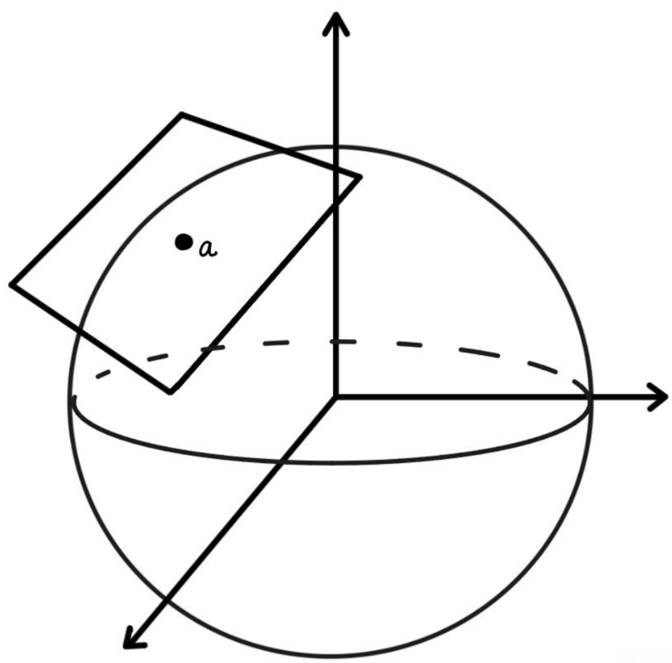

Figure 2.9: Tangent plane T_aM to a two-dimensional manifold M (the sphere) at a point a. The tangent vector $\dot{\gamma}(0)$ represents the velocity of a curve $\gamma(t)$ passing through a, showing how directions on M correspond to vectors in the tangent space.

Physical interpretation. In spacetime, a tangent vector represents the *four-velocity* of a particle at an event p:

$$u^a = \frac{dx^a}{d\tau}.$$

The tangent space T_pM is thus the space of all possible velocity vectors an observer at p could have. A field of tangent vectors over M defines a *vector field*, such as a velocity or momentum field.

Commutator (Lie bracket) of vector fields. Given two smooth vector fields v and w on M, we can form a new vector field $[v, w]$, called their *commutator* or *Lie bracket*, defined by

$$[v, w](f) := v[w(f)] - w[v(f)], \qquad f \in C^\infty(M).$$

This measures the failure of the two directional derivatives to commute when acting on smooth functions. It is straightforward to verify that $[v, w]$ again satisfies the Leibniz rule, so it is itself a smooth vector field. The space of smooth vector fields $\mathfrak{X}(M)$ is therefore closed under the commutator operation, and this operation satisfies

$$\begin{aligned} [v, w] &= -[w, v], \\ [v,\, aw_1 + bw_2] &= a[v, w_1] + b[v, w_2], \\ [[v, w], z] + [[w, z], v] + [[z, v], w] &= 0, \end{aligned}$$

making $\mathfrak{X}(M)$ into a *Lie algebra*. In local coordinates x^μ,

$$[v, w] = \left(v^\nu \partial_\nu w^\mu - w^\nu \partial_\nu v^\mu\right) \frac{\partial}{\partial x^\mu}.$$

Summary.

- Tangent vectors are derivations acting on smooth functions.
- The tangent space T_pM is a real vector space spanned by the coordinate basis $\{\frac{\partial}{\partial x^i}|_p\}$.
- Each tangent vector corresponds to a velocity vector of a curve through p.
- Physically, tangent vectors represent possible directions of motion or rates of change at a spacetime event.

2.4 The Differential and Pushforward

Overview. Up to this point, we have described smooth manifolds M, N and smooth maps $F : M \to N$. At each point $p \in M$, the map F induces a natural correspondence between directions at p and directions at its image $F(p)$. This correspondence is the *differential* of F at p, denoted

$$dF_p : T_pM \longrightarrow T_{F(p)}N.$$

In physics, this same object is called the *pushforward* of F and written F_*.

Definition (Lee). Let $F : M \to N$ be a smooth map and let $p \in M$. For each tangent vector $X_p \in T_pM$, the differential $dF_p(X_p) \in T_{F(p)}N$ is the tangent vector that acts on smooth functions $f \in C^\infty(N)$ by

$$(dF_p(X_p))(f) = X_p(f \circ F).$$

That is, the new vector differentiates a function f on the target manifold N by first pulling f back along F and then applying X_p. This definition guarantees that dF_p is linear. (Fig. 2.10)

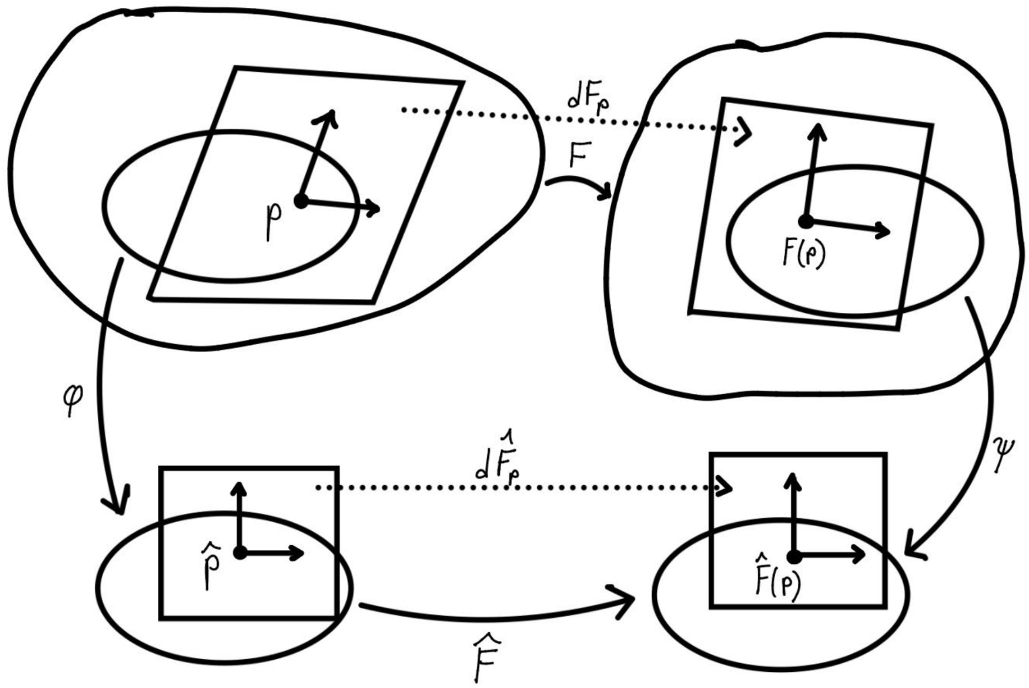

Figure 2.10: The differential dF_p relates tangent spaces of two manifolds M and N under a smooth map $F : M \to N$. In local coordinates (U, φ) and (V, ψ), it corresponds to the ordinary Jacobian $d\widehat{F}_{\hat{p}}$ of the coordinate representation $\widehat{F} = \psi \circ F \circ \varphi^{-1}$.

Coordinate expression. Suppose (x^i) are local coordinates on M near p and (y^α) are local coordinates on N near $F(p)$. Then F can be written in components as

$$y^\alpha = F^\alpha(x^1, \dots, x^m).$$

The differential acts on the coordinate basis vectors by

$$dF_p\left(\frac{\partial}{\partial x^i}\Big|_p\right) = \frac{\partial F^\alpha}{\partial x^i}(p)\frac{\partial}{\partial y^\alpha}\Big|_{F(p)}.$$

Thus, in coordinates, dF_p is represented by the *Jacobian matrix*

$$\left[\frac{\partial F^\alpha}{\partial x^i}(p)\right].$$

This Jacobian encodes how the coordinate components of a tangent vector transform under the map F.

Example (for intuition). If $F : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $F(x, y) = (r, \theta) = (\sqrt{x^2 + y^2}, \tan^{-1}(y/x))$, then

$$dF_{(x,y)} = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix}.$$

This Jacobian is precisely the matrix that converts Cartesian basis vectors (∂_x, ∂_y) into polar basis vectors $(\partial_r, \partial_\theta)$.

Geometric interpretation. The map dF_p carries tangent vectors at p to their "images" under F. If a curve $\gamma(t)$ passes through p with velocity $v_p = \dot{\gamma}(0)$, then the image curve $F \circ \gamma$ passes through $F(p)$ with velocity

$$(dF_p)(v_p) = \frac{d}{dt}(F \circ \gamma(t))\Big|_{t=0}.$$

Thus the differential describes how directions—velocities of curves— transform under smooth maps. In spacetime language, the pushforward F_* tells how a coordinate transformation or embedding moves tangent vectors from one manifold to another.

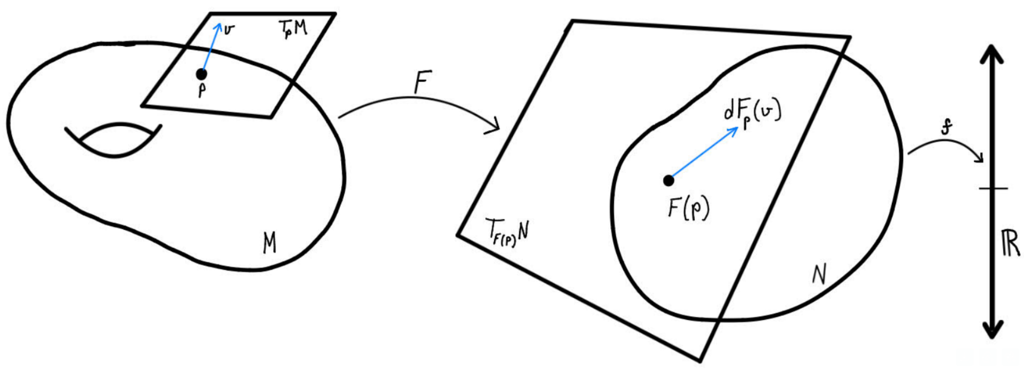

Figure 2.11: Geometric picture of the pushforward dF_p: a tangent vector $v \in T_pM$ to a curve γ at p is mapped to the tangent vector $dF_p(v) \in T_{F(p)}N$ of the image curve $F \circ \gamma$ at $F(p)$. Acting on a function $f : N \to \mathbb{R}$, the pushforward satisfies $(dF_p v)(f) = v(f \circ F)$.

Physical viewpoint. In general relativity, F_* appears constantly:

- A coordinate transformation $x^\mu \mapsto x^{\mu'}(x)$ acts on tangent vectors by the Jacobian $\frac{\partial x^{\mu'}}{\partial x^\nu}$.
- The inclusion $i : \Sigma \hookrightarrow \mathcal{M}$ of a spatial hypersurface pushes tangent vectors from Σ into the ambient spacetime $\mathcal{M}$.
- Along a worldline $\gamma : \mathbb{R} \to M$, the pushforward relates derivatives with respect to proper time τ to vectors in $T_{\gamma(\tau)}M$.

Summary. The differential dF_p (or pushforward F_*) is:

- a linear map between tangent spaces,
- represented in coordinates by the Jacobian matrix,
- and the geometric mechanism by which smooth maps carry directions and velocities from one manifold to another.

2.5 Covectors and the Pullback

Overview. Tangent vectors describe *directions of change* on a manifold. Their duals—called *covectors* or *one-forms*—measure *rates of change* along those directions. Together they form the natural algebraic pair that underlies tensor calculus in general relativity.

Mathematical definition (Lee). At a point $p \in M$, the *cotangent space* to M is defined as

$$T_p^*M = \mathrm{Hom}(T_pM, \mathbb{R}),$$

the vector space of all linear maps from the tangent space T_pM to the real numbers. Elements of T_p^*M are called **covectors** or **one-forms**. If $v_p \in T_pM$ and $\omega_p \in T_p^*M$, then their pairing is a real number

$$\omega_p(v_p) \in \mathbb{R}.$$

This operation is bilinear: $\omega_p(av_1 + bv_2) = a\,\omega_p(v_1) + b\,\omega_p(v_2)$.

Coordinate representation. Given local coordinates $(x^1, \ldots, x^n)$, the basis for T_pM is

$$\left\{ \frac{\partial}{\partial x^i}\Big|_p \right\},$$

and the dual basis for T_p^*M is

$$\{(dx^i)_p\},$$

defined by

$$(dx^i)_p\left(\frac{\partial}{\partial x^j}\Big|_p\right) = \delta^i_j.$$

Hence any covector at p can be written as a linear combination

$$\omega_p = \omega_i (dx^i)_p,$$

and acts on a vector $v_p = v^j \frac{\partial}{\partial x^j}\big|_p$ via

$$\omega_p(v_p) = \omega_i v^i.$$

Differentials of functions. For a smooth scalar field $f : M \to \mathbb{R}$, its *differential* at p, denoted df_p, is the covector that acts on a tangent vector v_p by

$$df_p(v_p) = v_p(f) = \frac{d}{dt} f(\gamma(t))\Big|_{t=0},$$

where $\gamma(t)$ is any curve in M with $\gamma(0) = p$ and velocity $\dot{\gamma}(0) = v_p$. In coordinates,

$$df = \frac{\partial f}{\partial x^i}\, dx^i.$$

Thus df is the one-form corresponding to the gradient of f: it maps each vector to the directional derivative of f along that vector.

Example (gradient on $\mathbb{R}^2$). Let $f(x, y) = x^2 + y^2$. Then

$$df = 2x\,dx + 2y\,dy.$$

For a vector $v = a\,\partial_x + b\,\partial_y$, we have

$$df(v) = 2x\,a + 2y\,b,$$

which equals the directional derivative of f in the direction v.

Pullback of covectors. Given a smooth map $F : M \to N$, the *pullback* of a covector $\omega_{F(p)} \in T^*_{F(p)}N$ is the covector $F^*\omega_{F(p)} \in T^*_p M$ defined by

$$(F^*\omega_{F(p)})(v_p) = \omega_{F(p)}(dF_p(v_p)).$$

That is, the pullback acts by first pushing forward the vector v_p and then applying the covector ω. In coordinates, if $\omega = \omega_\alpha\,dy^\alpha$ and $y^\alpha = F^\alpha(x^i)$, then

$$F^*\omega = \omega_\alpha \frac{\partial F^\alpha}{\partial x^i}\,dx^i.$$

Physical interpretation. Covectors are the mathematical form of gradients, momenta, and field differentials.

- The gradient of a scalar field f is the one-form df, assigning a real number (rate of change) to each direction.
- Under a coordinate transformation $x^\mu \mapsto x^{\mu'}(x)$, the components of a one-form transform with the inverse Jacobian:
$$\omega_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}}\omega_\mu.$$
- The pullback F^* tells how to compare measurements made in different manifolds—e.g. how a potential or flux density on one space induces one on another.

Summary.

- The cotangent space $T^*_p M$ consists of all linear maps $T_p M \to \mathbb{R}$.
- Covectors (one-forms) act on vectors to produce real numbers.
- The differential df of a function is the one-form corresponding to its gradient.
- A smooth map $F : M \to N$ induces a pullback $F^* : T^*_{F(p)}N \to T^*_p M$ that transports covectors "backward" along F.

Summary of relationships between f, v, df, and $df_p(v)$:

- $f : M \to \mathbb{R}$ — a **smooth scalar field** on the manifold. It assigns a real number to each point $p \in M$ (e.g., temperature, potential).

- $v_p \in T_pM$ — a **tangent vector** at p. It represents a direction of motion or infinitesimal displacement through p.
- $df_p \in T_p^*M$ — the **differential** (or covector) of f at p. It acts linearly on tangent vectors and encodes the local gradient of f.
- $df_p(v_p) = v_p(f)$ — the **directional derivative** of f along v_p. It gives the rate of change of the scalar field f as one moves in direction v_p.
- Geometrically:
  - f is the landscape (scalar field),
  - df_p is the local slope (gradient covector),
  - v_p is a direction you choose to walk,
  - $df_p(v_p)$ is how steep the slope feels in that direction.

2.6 Tensors and Tensor Fields

Overview. Once tangent and cotangent vectors are in hand, we can construct objects that combine them in linear ways. A *tensor* at a point $p \in M$ is simply a multilinear map that takes some number of tangent and cotangent vectors as inputs and returns a real number.

Wald's perspective: Tensors are multilinear functions built from vectors and covectors. A tensor of type (r, s) takes r covectors and s vectors as input:

$$T : (T_p^*M)^r \times (T_pM)^s \longrightarrow \mathbb{R}.$$

Their components $T^{a_1 \dots a_r}{}_{b_1 \dots b_s}$ transform under coordinate changes by one factor of the Jacobian for each upper index and one inverse Jacobian for each lower index. Tensor *fields* assign such an object to every point $p \in M$.

Lee's formal framework:
- For any vector space V, its dual is $V^* = \{\text{linear maps } V \to \mathbb{R}\}$.
- The tensor product $V^{\otimes r} \otimes (V^*)^{\otimes s}$ is the space of all multilinear maps $(V^*)^r \times V^s \to \mathbb{R}$, called tensors of type (r, s).
- On a manifold M, this defines the tensor space $T^r_s(M) = \bigcup_{p \in M} T^r_s(T_pM)$, and smooth sections of this bundle are the tensor fields.

Bridging the viewpoints. Wald's index notation and Lee's multilinear definition describe the same object. At each point $p \in M$, a tensor T of type (r, s) acts on r covectors and s vectors to yield a real number:

$$T(\omega_1, \dots, \omega_r, v_1, \dots, v_s) = T^{a_1 \dots a_r}{}_{b_1 \dots b_s}\, \omega_{1a_1} \cdots \omega_{ra_r} v_1^{b_1} \cdots v_s^{b_s}.$$

The upper (contravariant) indices correspond to the vector slots, and the lower (covariant) indices correspond to the covector slots.

Physical interpretation. Tensors are the language of field theory:

- The *metric tensor* g_{ab} measures spacetime intervals.
- The *electromagnetic field* F_{ab} is an antisymmetric tensor field.
- The *stress–energy tensor* T_{ab} encodes energy, momentum, and pressure.

Each of these is a smooth assignment of a multilinear map to every event in spacetime.

Summary.

- A tensor of type (r, s) acts on r covectors and s vectors.
- A tensor field assigns such an object smoothly to each point in M.
- Components transform by Jacobian factors under coordinate changes, ensuring the physical laws they encode are coordinate-independent.

2.7 The Metric Tensor

Overview. Up to this point, we have introduced tangent and cotangent spaces, and the way covectors act on vectors to produce scalars. The next structure, the *metric tensor*, provides a systematic way to compare vectors at a point: it measures lengths, angles, and, in relativity, causal relationships.

Wald's perspective: A spacetime is a smooth four-dimensional manifold M equipped with a *metric tensor* g_{ab}, a symmetric, non-degenerate tensor field of type $(0, 2)$. For any two tangent vectors $v^a, w^a \in T_pM$, the metric assigns a scalar

$$g_{ab}v^a w^b = g_p(v, w),$$

interpreted as the inner product between the two vectors. Its signature $(- + ++)$ encodes the distinction between timelike, null, and spacelike directions.

Lee's formal framework: A *Riemannian metric* on a smooth manifold M is a smooth assignment

$$g_p : T_pM \times T_pM \to \mathbb{R},$$

such that for every point $p \in M$:

(a) g_p is *bilinear*: $g_p(av_1 + bw_1, v_2) = a\, g_p(v_1, v_2) + b\, g_p(w_1, v_2)$, and linear in each argument.
(b) g_p is *symmetric*: $g_p(v, w) = g_p(w, v)$.
(c) g_p is *non-degenerate*: if $g_p(v, w) = 0$ for all w, then $v = 0$.

When $g_p(v, v) > 0$ for all nonzero v, g is called *Riemannian.* If instead g has signature $(- + ++)$ or another indefinite form, it is called a *pseudo-Riemannian metric*—the Lorentzian case of relativity.

Geometric meaning. The metric defines an inner product on each tangent space T_pM, allowing us to measure:

- the *length* of a vector, $\|v\|^2 = g_p(v, v)$;
- the *angle* between vectors, via $\cos\theta = \frac{g_p(v,w)}{\|v\|\,\|w\|}$;
- and the *orthogonality condition*, $g_p(v, w) = 0$.

In Lorentzian signature, $g_p(v, v)$ can be positive, negative, or zero: timelike, spacelike, or null. Thus the metric is not merely a notion of "distance," but the structure that defines *causality* in spacetime.

Raising and lowering indices. Because g_p is non-degenerate, it provides a natural isomorphism between tangent and cotangent spaces:

$$\flat : T_pM \to T_p^*M, \qquad v_p \mapsto v_p^\flat,$$

where

$$v_p^\flat(w_p) = g_p(v_p, w_p) \quad \text{for all } w_p \in T_pM.$$

This "lowering" map allows us to represent a vector as a covector. The inverse map,

$$\sharp : T_p^*M \to T_pM, \qquad \omega_p \mapsto \omega_p^\sharp,$$

is the corresponding "raising" operation, defined so that

$$g_p(\omega_p^\sharp, w_p) = \omega_p(w_p).$$

Together these define the familiar index operations $v_a = g_{ab}v^b$ and $v^a = g^{ab}v_b$ in tensor notation.

Physical interpretation. The metric provides the geometric scaffolding of spacetime:

- It defines proper time and proper distance: $ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu$.
- It establishes light cones and causal structure.
- It allows us to convert between vectors and covectors, enabling the formulation of physical laws in covariant form.

Example (Minkowski metric). In flat spacetime with global coordinates $(x^0, x^1, x^2, x^3) = (t, x, y, z)$, the metric components are

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

For two tangent vectors v^μ, w^μ, their inner product is

$$g_{\mu\nu}v^\mu w^\nu = -v^0w^0 + v^1w^1 + v^2w^2 + v^3w^3.$$

This distinguishes timelike, spacelike, and null directions:

$$\begin{cases} g(v, v) < 0, & \text{timelike,} \\ g(v, v) = 0, & \text{null,} \\ g(v, v) > 0, & \text{spacelike.} \end{cases}$$

But wait... I thought a tensor takes in r covectors and s vectors and outputs a real number?
That's absolutely right — by definition, a tensor at a point p is a multilinear map

$$T_p : (T_p^*M)^r \times (T_pM)^s \to \mathbb{R}.$$

So the *metric at a point* g_p is indeed a $(0,2)$-tensor: it takes in two tangent vectors v_p, w_p and returns a number $g_p(v_p, w_p)$.
When we write

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

we're no longer talking about just one g_p, but about the entire *metric tensor field*: a smooth assignment $p \mapsto g_p$ over the manifold. In flat (Minkowski) spacetime the components are constant everywhere, so the field looks identical at every point.
That's why physicists often blur the language and say "the metric tensor" when they really mean "the metric tensor field." The distinction only becomes important once the components $g_{\mu\nu}(x)$ start varying with position in a curved spacetime.

Summary. The metric tensor g is a symmetric, non-degenerate bilinear form on each tangent space, smoothly varying from point to point. It enables:

- measurement of lengths and angles;
- conversion between vectors and covectors;
- and, in relativity, the distinction between causal types of vectors.

It is the geometric heart of both Riemannian geometry and spacetime physics.

2.8 Orientation and Volume Forms (Optional)

Overview. Orientation gives a manifold a consistent notion of "handedness," while a *volume form* provides the geometric tool needed to define integration independently of coordinates. These concepts become essential later for flux integrals, the divergence theorem, and the Einstein–Hilbert action.

Wald's perspective: Spacetime possesses a smooth, nowhere-vanishing volume element ϵ_{abcd}, a completely antisymmetric tensor field of type $(0,4)$. In any coordinate system with positive orientation,

$$\epsilon_{0123} = +\sqrt{|\det g|},$$

and it transforms with the sign of the Jacobian determinant under coordinate changes. This tensor defines oriented volume elements used in integration and in expressing physical conservation laws such as $\nabla_a T^{ab} = 0$.

Lee's formal framework: A smooth manifold M is said to be *orientable* if it admits an atlas whose transition maps all have positive Jacobian determinant. An *orientation* is a choice of such equivalence class of atlases.
A *volume form* on an n-dimensional oriented manifold is a smooth, nowhere-vanishing differential form of top degree:

$$\omega \in \Omega^n(M), \qquad \omega_p(v_1, \dots, v_n) = (\text{signed volume spanned by the } v_i).$$

In the Riemannian or Lorentzian case, the metric naturally induces a canonical volume form

$$\omega = \sqrt{|\det g|}\, dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n.$$

Bridging the viewpoints. For physicists, the symbol ϵ_{abcd} plays two roles: it encodes the manifold's orientation (sign convention), and it acts as the metric-compatible volume measure. Formally, it is the component representation of the volume form ω with respect to a positively oriented coordinate basis:

$$\epsilon_{abcd} = \omega\left(\frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c}, \frac{\partial}{\partial x^d}\right).$$

Integrals over spacetime regions or hypersurfaces use this form to ensure coordinate invariance:

$$\int_M f\, \epsilon = \int f\, \sqrt{|\det g|}\, d^4x.$$

Physical interpretation. Orientation distinguishes "future" from "past," or "right-handed" from "left-handed" coordinate systems. The volume form defines the invariant measure of spacetime, so that physical quantities such as energy or charge are independent of the coordinate chart. In general relativity, ϵ_{abcd} enters directly into:

- flux integrals, e.g. $\int_\Sigma J^a \epsilon_{abcd}$;
- definitions of dual tensors via the Hodge star;
- and the Einstein–Hilbert action, $S = \int R \sqrt{|\det g|}\, d^4x$.

Example (Euclidean 3-space). On $\mathbb{R}^3$ with standard coordinates (x, y, z), the canonical orientation is given by the right-hand rule, and

$$\omega = dx \wedge dy \wedge dz.$$

For vectors v_1, v_2, v_3, $\omega(v_1, v_2, v_3)$ equals the signed volume of the parallelepiped they span. In curved 3-space with metric g, this becomes $\omega = \sqrt{\det g}\, dx \wedge dy \wedge dz$.

Summary. Orientation and the volume form together provide:

- a consistent notion of handedness on M;
- a metric-independent definition of signed volume;
- and the geometric foundation for integration on manifolds.

In physics, these ideas appear through ϵ_{abcd} and $\sqrt{|\det g|}\, d^4x$, which ensure that integrals and conservation laws are invariant under smooth coordinate transformations.

2.9 Summary and Outlook

Summary. We have built the full geometric foundation needed for general relativity in purely differential-geometric terms:

- Spacetime M is a smooth 4-dimensional manifold equipped with a Lorentzian metric g.
- Tangent vectors, covectors, and general tensors arise naturally from the tangent and cotangent bundles.
- Smooth maps between manifolds induce pushforwards and pullbacks that describe how physical quantities transform under coordinate changes.
- The metric tensor g provides an inner product on each tangent space T_pM, defines distances and causal structure, and allows indices to be raised and lowered.

Physical picture. At this point we have a complete mathematical language for describing spacetime as a differentiable manifold: points are events, tangent vectors are directions of motion, and tensor fields represent physical quantities such as fields, fluxes, and stresses. All of this structure is purely kinematical— it describes *how* objects live and transform on spacetime, but not yet *how they evolve or curve.*

Interdisciplinary Insight: Geometry in Modern Machine Learning
The mathematical structures developed in this chapter—manifolds, tangent spaces, covectors, and tensors—form not only the language of general relativity but also a growing foundation for modern machine learning.
Most real-world data do not occupy the full Euclidean space in which they are embedded. Instead, they lie on smooth *manifolds* of much lower intrinsic dimension. Understanding how to analyze and compare data on such curved spaces requires precisely the tools of differential geometry.

- **Manifold Learning:** Algorithms such as Isomap, LLE, and t-SNE attempt to uncover the manifold structure underlying high-dimensional data.
- **Geometric Deep Learning:** Extends neural networks to non-Euclidean domains—graphs, surfaces, and manifolds—using ideas of symmetry, invariance, and connection.
- **Riemannian Optimization:** Many learning problems involve constraints that make parameter spaces curved (e.g. spheres, orthogonal matrices, covariance manifolds). Gradients and geodesics on these spaces rely on Riemannian geometry.
- **Information Geometry:** Views probability distributions as points on a manifold endowed with the Fisher information metric. The resulting "natural gradient" follows geodesics in information space.
- **Topological Data Analysis:** Uses persistent homology to capture the global shape and connectivity of data sets in a way that is robust to noise.

In all these areas, the central questions mirror those of spacetime geometry:

What is the shape of the space in which we live—and how do objects change and relate within it?

Differential geometry provides the language for answering this question, whether the manifold is the universe itself or the high-dimensional space of data.

Problems

The following problems come from Chapter 2 of Wald's General Relativity.

1. (a) Show that the overlap functions $f_i^{\pm} \circ (f_j^{\pm})^{-1}$ are C^{∞}, thus completing the demonstration given in section 2.1 of Wald that S^2 is a manifold.
   (b) Show by explicit construction that two coordinate systems (as opposed to the six used in the text) suffice to cover S^2. (It is impossible to cover S^2 with a single chart, as follows from the fact that S^2 is compact, but every open subset of $\mathbb{R}^2$ is noncompact; see Appendix A.)

Solution

Solution.
(a). Let

$$U_i^{\pm} = \{(x^1, x^2, x^3) \in S^2 : \pm x^i > 0\}, \qquad (x^1, x^2, x^3) = (x, y, z).$$

Each $U_i^{\pm}$ is mapped diffeomorphically onto the open unit disk $\mathbb{D} = \{(u,v) \in \mathbb{R}^2 : u^2 + v^2 < 1\}$ by the "projection" charts

$$f_1^{\pm}(x,y,z) = (y,z), \qquad f_2^{\pm}(x,y,z) = (x,z), \qquad f_3^{\pm}(x,y,z) = (x,y),$$

whose inverses (using $x^2 + y^2 + z^2 = 1$) are

$$\begin{aligned}
(f_1^{\pm})^{-1}(u,v) &= (\ \pm\sqrt{1-u^2-v^2},\ u,\ v),\\
(f_2^{\pm})^{-1}(u,v) &= (\ u,\ \pm\sqrt{1-u^2-v^2},\ v),\\
(f_3^{\pm})^{-1}(u,v) &= (\ u,\ v,\ \pm\sqrt{1-u^2-v^2}).
\end{aligned}$$

For any pair of charts with nonempty overlap, the *overlap map*

$$f_i^{\pm} \circ (f_j^{\pm})^{-1} :\ f_j^{\pm}(U_i^{\pm} \cap U_j^{\pm}) \subset \mathbb{D} \longrightarrow \mathbb{D}$$

is smooth. (The only disjoint pair is U_i^+ and U_i^-.) Each component of $f_i^{\pm} \circ (f_j^{\pm})^{-1}$ is one of u, v, or $\pm\sqrt{1-u^2-v^2}$, hence C^{∞} on its (open) domain $u^2 + v^2 < 1$ with the relevant sign inequality. For example, on $U_3^+ \cap U_1^+$,

$$(f_1^+ \circ (f_3^+)^{-1})(u,v) = (v,\ \sqrt{1-u^2-v^2}),$$

which is C^{∞} where $1 - u^2 - v^2 > 0$. Thus all transition maps are smooth, completing the verification that this atlas makes S^2 a smooth manifold.

(b). Define the northern and southern stereographic projections

$$\sigma_N : S^2 \setminus \{N\} \longrightarrow \mathbb{R}^2, \qquad \sigma_N(x,y,z) = \Big(\frac{x}{1-z}, \frac{y}{1-z}\Big),$$
$$\sigma_S : S^2 \setminus \{S\} \longrightarrow \mathbb{R}^2, \qquad \sigma_S(x,y,z) = \Big(\frac{x}{1+z}, \frac{y}{1+z}\Big),$$

with inverses (smooth on all of $\mathbb{R}^2$)

$$\sigma_N^{-1}(u,v) = \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\right),$$
$$\sigma_S^{-1}(u,v) = \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, -\frac{u^2+v^2-1}{u^2+v^2+1}\right).$$

Hence σ_N and σ_S are diffeomorphisms from $S^2 \setminus \{N\}$ and $S^2 \setminus \{S\}$ onto $\mathbb{R}^2$, respectively. On the overlap, the transition maps are

$$\sigma_N \circ \sigma_S^{-1}(u,v) = \frac{(u,v)}{u^2+v^2}, \qquad \sigma_S \circ \sigma_N^{-1}(u,v) = \frac{(u,v)}{u^2+v^2},$$

which are smooth on $\mathbb{R}^2 \setminus \{(0,0)\}$. Therefore the two charts $(S^2 \setminus \{N\}, \sigma_N)$ and $(S^2 \setminus \{S\}, \sigma_S)$ form a smooth atlas covering S^2.

2. Prove that any smooth function $F : \mathbb{R}^n \to \mathbb{R}$ can be written in the form equation (2.2.2). (Hint: For $n = 1$, use the identity

$$F(x) - F(a) = (x-a) \int_0^1 F'[\,t(x-a) + a\,]\, dt;$$

then prove it for general n by induction.)

Solution

Solution (2). Let $F : \mathbb{R}^n \to \mathbb{R}$ be smooth and fix $a \in \mathbb{R}^n$. Consider the line segment $\gamma : [0,1] \to \mathbb{R}^n$,

$$\gamma(t) = a + t(x-a), \qquad \phi(t) := F(\gamma(t)).$$

By the multivariable chain rule,

$$\phi'(t) = \sum_{\mu=1}^{n} \partial_\mu F(\gamma(t))\, \frac{d\gamma^\mu}{dt} = \sum_{\mu=1}^{n} \partial_\mu F(a + t(x-a))\,(x^\mu - a^\mu),$$

since $\dfrac{d\gamma^\mu}{dt} = x^\mu - a^\mu$. Integrating from 0 to 1 and using the Fundamental Theorem of Calculus,

$$F(x) - F(a) = \int_0^1 \phi'(t)\, dt = \sum_{\mu=1}^{n} (x^\mu - a^\mu) \int_0^1 \partial_\mu F(a + t(x-a))\, dt.$$

Thus

$$F(x) = F(a) + \sum_{\mu=1}^{n} (x^\mu - a^\mu)\, H_\mu(x), \qquad H_\mu(x) := \int_0^1 \partial_\mu F(a + t(x-a))\, dt.$$

Each H_μ is smooth (integral of a smooth function depending smoothly on x over a compact interval), and

$$H_\mu(a) = \int_0^1 \partial_\mu F(a)\, dt = \partial_\mu F(a).$$

This is equation (2.2.2).

3. (a) Verify that the commutator, defined by equation (2.2.14), satisfies the linearity and Leibnitz properties, and hence defines a vector field.
(b) Let X, Y, Z be smooth vector fields on a manifold M. Verify that their commutator satisfies the Jacobi identity:

$$[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0.$$

(c) Let $Y_1, \dots, Y_n$ be smooth vector fields on an n-dimensional manifold M such that at each $p \in M$ they form a basis of the tangent space V_p. Then, at each point, we may expand each commutator $[Y_\alpha, Y_\beta]$ in this basis, thereby defining the functions $C^\gamma_{\alpha\beta} = -C^\gamma_{\beta\alpha}$ by

$$[Y_\alpha, Y_\beta] = \sum_\gamma C^\gamma_{\alpha\beta} Y_\gamma.$$

Use the Jacobi identity to derive an equation satisfied by $C^\gamma_{\alpha\beta}$. (This equation is a useful algebraic relation if the $C^\gamma_{\alpha\beta}$ are constants, as will be the case if $Y_1, \dots, Y_n$ are left [or right] invariant vector fields on a Lie group [see section 7.2].)

Solution

Problem. For smooth vector fields X, Y, Z on a manifold M, with the commutator (Lie bracket)

$$[X,Y](f) := X(Yf) - Y(Xf), \qquad f \in C^\infty(M),$$

(a) Linearity and Leibniz. Linearity in f is immediate from linearity of X and Y:

$$[X,Y](af + bg) = a[X,Y]f + b[X,Y]g \quad (a, b \in \mathbb{R}).$$

For the Leibniz rule, use that X and Y are derivations:

$$\begin{aligned}
[X,Y](fg) &= X\big(Y(fg)\big) - Y\big(X(fg)\big) \\
&\overset{\text{(inner Leibniz)}}{=} X\big(f\,Yg + g\,Yf\big) - Y\big(f\,Xg + g\,Xf\big) \\
&= X(fYg) + X(gYf) - Y(fXg) - Y(gXf) \\
&\overset{\text{(outer Leibniz)}}{=} (Xf)\,Yg + f\,X(Yg) + (Xg)\,Yf + g\,X(Yf) \\
&\quad - (Yf)\,Xg - f\,Y(Xg) - (Yg)\,Xf - g\,Y(Xf) \\
&= f\big(X(Yg) - Y(Xg)\big) + g\big(X(Yf) - Y(Xf)\big) \\
&= f\,[X,Y](g) + g\,[X,Y](f).
\end{aligned}$$

Thus $[X,Y]$ is again a derivation, hence a (smooth) vector field.

(b) Jacobi identity. We verify explicitly that the commutator of vector fields satisfies the Jacobi identity. For any smooth function $f \in C^\infty(M)$,

$$[X,Y](f) = X(Yf) - Y(Xf).$$

Then

$$\begin{aligned}
[X,[Y,Z]](f) &= X\big([Y,Z]f\big) - [Y,Z](Xf) \\
&= X(YZf - ZYf) - (YZXf - ZYXf) \\
&= \boxed{XYZf} - \boxed{XZYf} - \boxed{YZXf} + \boxed{ZYXf},
\end{aligned}$$

$$\begin{aligned}
[Y,[Z,X]](f) &= Y\big([Z,X]f\big) - [Z,X](Yf) \\
&= Y(ZXf - XZf) - (ZXYf - XZYf) \\
&= \boxed{YZXf} - \boxed{YXZf} - \boxed{ZXYf} + \boxed{XZYf},
\end{aligned}$$

$$\begin{aligned}
[Z,[X,Y]](f) &= Z\big([X,Y]f\big) - [X,Y](Zf) \\
&= Z(XYf - YXf) - (XYZf - YXZf) \\
&= \boxed{ZXYf} - \boxed{ZYXf} - \boxed{XYZf} + \boxed{YXZf}.
\end{aligned}$$

Adding the three expressions gives

$$[X,[Y,Z]](f) + [Y,[Z,X]](f) + [Z,[X,Y]](f) = 0,$$

since each term cancels with an identical term of opposite sign:

$$\begin{aligned}
&+XYZf \text{ (from the first) cancels with } -XYZf \text{ (from the third)}, \\
&+XZYf \text{ (from the second) cancels with } -XZYf \text{ (from the first)}, \\
&+YZXf \text{ (from the second) cancels with } -YZXf \text{ (from the first)}, \\
&+ZYXf \text{ (from the first) cancels with } -ZYXf \text{ (from the third)}, \\
&+ZXYf \text{ (from the third) cancels with } -ZXYf \text{ (from the second)}, \\
&+YXZf \text{ (from the third) cancels with } -YXZf \text{ (from the second)}.
\end{aligned}$$

Therefore, for all $f \in C^\infty(M)$,

$$[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0,$$

which is the Jacobi identity for the Lie bracket of vector fields.

(c) Identity for structure functions. Let $\{Y_\alpha\}_{\alpha=1}^n$ be a local frame (basis of T_pM at each p), and define smooth functions $C^\gamma_{\alpha\beta} = -C^\gamma_{\beta\alpha}$ by

$$[Y_\alpha, Y_\beta] = C^\gamma_{\alpha\beta}\, Y_\gamma.$$

Compute, using $[Y_\alpha, fY_\mu] = Y_\alpha(f)\, Y_\mu + f\,[Y_\alpha, Y_\mu]$:

$$\begin{aligned}
[Y_\alpha, [Y_\beta, Y_\gamma]] &= [Y_\alpha,\, C^\mu_{\beta\gamma} Y_\mu] \\
&= Y_\alpha(C^\mu_{\beta\gamma})\, Y_\mu + C^\mu_{\beta\gamma}\, [Y_\alpha, Y_\mu] \\
&= \big(Y_\alpha(C^\delta_{\beta\gamma}) + C^\mu_{\beta\gamma} C^\delta_{\alpha\mu}\big)\, Y_\delta.
\end{aligned}$$

Summing cyclically in (α,β,γ) and using the Jacobi identity gives, for each δ,

$$\boxed{\sum_{\text{cyc}(\alpha\beta\gamma)} \Big(Y_\alpha(C^\delta_{\beta\gamma}) + C^\mu_{\beta\gamma}\, C^\delta_{\alpha\mu}\Big) = 0}\,.$$

This is the desired relation. When the $C^{\gamma}_{\alpha\beta}$ are constants (e.g. for left-invariant frames on a Lie group), the first (derivative) term vanishes and we obtain the purely algebraic condition

$$\sum_{\mathrm{cyc}(\alpha\beta\gamma)} C^{\mu}_{\beta\gamma}\, C^{\delta}_{\alpha\mu} = 0,$$

i.e. the Jacobi identity for the structure constants.

4. (a) Show that in any coordinate basis, the components of the commutator of two vector fields v and w are given by

$$[v,w]^{\mu} = \sum_{\nu} \left(v^{\nu} \frac{\partial w^{\mu}}{\partial x^{\nu}} - w^{\nu} \frac{\partial v^{\mu}}{\partial x^{\nu}} \right).$$

(b) Let $Y_1, \ldots, Y_n$ be as in problem 3(c). Let $Y^{1*}, \ldots, Y^{n*}$ be the dual basis. Show that the components $(Y^{\gamma*})_{\mu}$ of $Y^{\gamma*}$ in any coordinate basis satisfy

$$\frac{\partial (Y^{\gamma*})_{\mu}}{\partial x^{\nu}} - \frac{\partial (Y^{\gamma*})_{\nu}}{\partial x^{\mu}} = \sum_{\alpha,\beta} C^{\gamma}_{\alpha\beta} (Y^{\alpha*})_{\mu} (Y^{\beta*})_{\nu}.$$

(Hint: Contract both sides with $(Y_{\alpha})^{\mu}(Y_{\beta})^{\nu}$.)

Solution

(a) Show that in a coordinate basis,

$$[v,w]^{\mu} = v^{\nu} \frac{\partial w^{\mu}}{\partial x^{\nu}} - w^{\nu} \frac{\partial v^{\mu}}{\partial x^{\nu}}.$$

In a coordinate basis $\{\partial_{\mu}\}$ every vector field has the form $v = v^{\mu}\partial_{\mu}$ and $w = w^{\mu}\partial_{\mu}$. The commutator of vector fields is defined by

$$[v,w] := v(w) - w(v).$$

Compute the action of v on w:

$$v(w) = v^{\nu}\partial_{\nu}(w^{\mu}\partial_{\mu}) = v^{\nu}(\partial_{\nu}w^{\mu})\partial_{\mu} + v^{\nu}w^{\mu}\partial_{\nu}\partial_{\mu}.$$

Likewise,

$$w(v) = w^{\nu}(\partial_{\nu}v^{\mu})\partial_{\mu} + w^{\nu}v^{\mu}\partial_{\nu}\partial_{\mu}.$$

But partial derivatives commute in a coordinate basis:

$$[\partial_{\nu}, \partial_{\mu}] = 0.$$

Therefore the second terms cancel, leaving

$$[v,w] = \left(v^{\nu}\partial_{\nu}w^{\mu} - w^{\nu}\partial_{\nu}v^{\mu}\right)\partial_{\mu}.$$

Thus the components are

$$[v,w]^{\mu} = v^{\nu}\partial_{\nu}w^{\mu} - w^{\nu}\partial_{\nu}v^{\mu}.$$

(b) Show that

$$\partial_\nu (Y^{\gamma *})_\mu - \partial_\mu (Y^{\gamma *})_\nu = C^\gamma_{\alpha\beta} (Y^{\alpha *})_\mu (Y^{\beta *})_\nu.$$

Let $\{Y_\alpha\}$ be the vector fields of problem 3(c), which satisfy

$$[Y_\alpha, Y_\beta] = C^\gamma_{\alpha\beta} Y_\gamma.$$

Let $\{Y^{\gamma *}\}$ be the dual basis, satisfying

$$Y^{\gamma *}(Y_\alpha) = \delta^\gamma_\alpha.$$

Step 1: Express the dual basis in coordinates. Write

$$Y^{\gamma *} = (Y^{\gamma *})_\mu \, dx^\mu, \qquad Y_\alpha = (Y_\alpha)^\nu \, \partial_\nu.$$

The duality condition becomes

$$(Y^{\gamma *})_\mu (Y_\alpha)^\mu = \delta^\gamma_\alpha. \tag{1}$$

Step 2: Differentiate the duality identity. Apply ∂_ν to both sides of (1):

$$\partial_\nu \big((Y^{\gamma *})_\mu (Y_\alpha)^\mu \big) = 0.$$

Expand:

$$(\partial_\nu (Y^{\gamma *})_\mu)(Y_\alpha)^\mu + (Y^{\gamma *})_\mu (\partial_\nu (Y_\alpha)^\mu) = 0. \tag{2}$$

Step 3: Use the definition of the structure constants. From

$$[Y_\alpha, Y_\beta]^\mu = (Y_\alpha)^\nu \partial_\nu (Y_\beta)^\mu - (Y_\beta)^\nu \partial_\nu (Y_\alpha)^\mu = C^\gamma_{\alpha\beta} (Y_\gamma)^\mu,$$

rearrange to isolate $\partial_\nu (Y_\alpha)^\mu$ and use it in (2). Now antisymmetrize in μ and ν:

$$\partial_\nu (Y^{\gamma *})_\mu - \partial_\mu (Y^{\gamma *})_\nu = C^\gamma_{\alpha\beta} (Y^{\alpha *})_\mu (Y^{\beta *})_\nu.$$

This is exactly the desired identity.

5. Let $Y_1, \ldots, Y_n$ be smooth vector fields on an n-dimensional manifold M which form a basis of V_p at each $p \in M$. Suppose $[Y_\alpha, Y_\beta] = 0$ for all α, β. Prove that in a neighborhood of each $p \in M$ there exist coordinates $y_1, \ldots, y_n$ such that $Y_1, \ldots, Y_n$ are the coordinate vector fields, $Y_\mu = \partial/\partial y^\mu$. (Hint: In an open ball of $\mathbb{R}^n$, the equations $\partial f/\partial x^\mu = F_\mu$ with $\mu = 1, \ldots, n$ for the unknown function f have a solution if and only if $\partial F_\mu/\partial x^\nu = \partial F_\nu/\partial x^\mu$. [See the end of section B.1 of appendix B for a statement of generalizations of this result.] Use this fact together with the results of problem 4(b) to obtain the new coordinates.)

6. (a) Verify that the dual vectors $\{v^{\mu *}\}$ defined by equation (2.3.1) constitute a basis of V^*.
   (b) Let $v_1, \ldots, v_n$ be a basis of the vector space V, and let $v^{1*}, \ldots, v^{n*}$ be its dual basis. Let $w \in V$ and let $\omega \in V^*$. Show that

$$w = \sum_\alpha v^{\alpha *}(w) \, v_\alpha, \qquad \omega = \sum_\alpha \omega(v_\alpha) \, v^{\alpha *}.$$

(c) Prove that the operation of contraction, equation (2.3.2), is independent of the choice of basis.

7. Let V be an n-dimensional vector space and let g be a metric on V.
   (a) Show that one always can find an orthonormal basis $v_1, \ldots, v_n$ of V, i.e. a basis such that $g(v_\alpha, v_\beta) = \pm\delta_{\alpha\beta}$. (Hint: Use induction.)
   (b) Show that the signature of g is independent of the choice of orthonormal basis.

8. (a) The metric of flat, three-dimensional Euclidean space is

$$ds^2 = dx^2 + dy^2 + dz^2.$$

Show that the metric components $g_{\mu\nu}$ in spherical polar coordinates r, θ, ϕ, defined by

$$r = (x^2 + y^2 + z^2)^{1/2}, \quad \cos\theta = z/r, \quad \tan\phi = y/x,$$

are given by

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2.$$

(b) The spacetime metric of special relativity is

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2.$$

Find the components $g_{\mu\nu}$ and $g^{\mu\nu}$ of the metric and inverse metric in "rotating coordinates," defined by

$$t' = t, \quad x' = (x^2 + y^2)^{1/2} \cos(\phi - \omega t), \quad y' = (x^2 + y^2)^{1/2} \sin(\phi - \omega t), \quad z' = z,$$

where $\tan\phi = y/x$.

Chapter 3

Curvature

Imagine that our one–dimensional beings from the previous chapter have now learned to compare vectors at different points along their line. They can say whether the "direction" of a quantity changes as they move. Now suppose their universe were two–dimensional. They carry a little arrow (a vector) and drag it carefully around a closed path, always trying to keep it pointing "the same way." When they return to where they started, the arrow has rotated slightly. Something about the surface itself has twisted their notion of "sameness."

That twist is what we call *curvature*. It is not about how a surface bends in some higher space, but how parallelism itself behaves *within* the manifold.

Curvature measures the failure of vectors to return unchanged after parallel transport around a closed loop.

3.1 Chapter Summary

In this chapter, we finally give mathematical meaning to what we call the curvature of spacetime.

We start with a simple question: how do we tell whether a space is curved without looking at it from the outside? For a surface like a sphere, we can see it bending in three-dimensional space, but spacetime doesn't live inside anything higher-dimensional that we can peek at. So, we need a way to describe curvature using only information that lives within the manifold itself.

To do that, we first learn how to compare vectors at different points on a manifold. Each point has its own tangent space, its own little copy of "directions", and there is no built-in rule that tells us how to say whether a direction at one point is "the same" as a direction at another. So we invent such a rule. That rule is called a derivative operator or connection, and it tells us how to parallel transport vectors along curves so that they "stay pointing the same way" according to the geometry of the manifold.

Once we know how to transport vectors, we can ask what happens if we carry one all the way around a small closed loop and come back to the starting point. If the vector returns unchanged, the space is flat. If it comes back rotated, the space is curved. This failure to come back unchanged is the essence of curvature.

Mathematically, that failure shows up as the fact that if you take two covariant derivatives in different orders, they don't quite cancel each other out. The amount by which they fail to commute defines the curvature tensor, a precise, coordinate-free measure of how the geometry twists and turns.

We then connect this idea to geodesics, the straightest possible paths in the manifold. In curved space, two geodesics that start out parallel can begin to converge or diverge, this is encoded in the geodesic deviation equation, which expresses curvature in terms of the relative acceleration of nearby free-falling particles.

Finally, we learn practical ways to compute curvature once a metric is given: how to find the connection coefficients, build the Riemann curvature tensor, and contract it to get familiar quantities like the Ricci tensor and scalar curvature.

3.2 Derivative Operators and Parallel Transport

To talk about curvature, we must first learn how to compare tensors at different points on a manifold. The tangent spaces T_pM and T_qM are distinct vector spaces, so there is no natural way to say that a vector at p "equals" a vector at q. We therefore introduce additional structure: a *derivative operator* (or *connection*) ∇.

Geometrically, ∇ encodes how vectors and tensors *change* as we move infinitesimally along curves on M. Operationally, it defines how to take derivatives of tensor fields in a way that respects the tensor algebra of M.

Remark (Abstract vs. Component Indices). The symbols $a_1, \ldots, a_k$ and $b_1, \ldots, b_l$ are *abstract indices*—they label the tensor's type, not its components. Upper indices correspond to contravariant (vector) slots, lower indices to covariant (covector) slots. They tell us how the tensor contracts or transforms, but do not take numerical values until a coordinate basis is chosen.
For example, a rank $(3,2)$ tensor field $T^{a_1a_2a_3}{}_{b_1b_2} \in \mathscr{T}(3,2)$ assigns to each point $p \in M$ a tensor $T^{a_1a_2a_3}{}_{b_1b_2}(p) \in T_p^{(3,2)}M$, which is a multilinear map taking two vectors and three covectors at p and returning a real number.

Definition 3.1. Derivative Operator A **derivative operator** ∇ is a map that takes a smooth tensor field

$$T^{a_1\cdots a_k}{}_{b_1\cdots b_l} \in \mathscr{T}(k,l)$$

to another tensor field

$$\nabla_c T^{a_1\cdots a_k}{}_{b_1\cdots b_l} \in \mathscr{T}(k,l+1)$$

such that the following five properties hold.

Five Defining Properties of a Derivative Operator

1. **Linearity.** For all tensor fields A, B and scalars $\alpha, \beta \in \mathbb{R}$,

$$\nabla_c(\alpha A^{a_1\cdots a_k}{}_{b_1\cdots b_l} + \beta B^{a_1\cdots a_k}{}_{b_1\cdots b_l}) = \alpha\nabla_c A^{a_1\cdots a_k}{}_{b_1\cdots b_l} + \beta\nabla_c B^{a_1\cdots a_k}{}_{b_1\cdots b_l}.$$

2. **Leibniz Rule.** For $A \in \mathscr{T}(k,l)$ and $B \in \mathscr{T}(k',l')$,

$$\nabla_e(A^{a_1\cdots a_k}{}_{b_1\cdots b_l} B^{c_1\cdots c_{k'}}{}_{d_1\cdots d_{l'}}) = (\nabla_e A^{a_1\cdots a_k}{}_{b_1\cdots b_l}) B^{c_1\cdots c_{k'}}{}_{d_1\cdots d_{l'}} + A^{a_1\cdots a_k}{}_{b_1\cdots b_l}(\nabla_e B^{c_1\cdots c_{k'}}{}_{d_1\cdots d_{l'}}).$$

3. **Commutativity with Contraction.** Derivative operators commute with index contraction:

$$\nabla_d(A^{a_1\cdots a_k}{}_{b_1\cdots a_d\cdots b_l}) = \nabla_d(A^{a_1\cdots a_{k-1}}{}_{b_1\cdots b_l}).$$

4. **Consistency with Directional Derivatives.** For any smooth function $f \in \mathscr{F}(M)$ and any vector field t^a,

$$t(f) = t^a\nabla_a f.$$

   Thus, $\nabla_a f$ is the covariant generalization of the ordinary directional derivative of f.
5. **Torsion-Free Condition.** For all smooth functions f,

$$\nabla_a\nabla_b f = \nabla_b\nabla_a f.$$

   This states that the derivative operator has no torsion.

In general relativity, we *always* assume torsion-free derivative operators, unless stated otherwise.

Remark. Here $\mathscr{T}(k,l)$ denotes the space of smooth tensor fields of type (k,l) on M. That is, for each point $p \in M$, a field $T \in \mathscr{T}(k,l)$ assigns a tensor $T(p) \in T_p^{(k,l)}M$, smoothly varying with p.

Geometric Intuition for the Five Properties

- **(1) Linearity:** Differentiation is a linear operation, just as in calculus. The rate of change of a linear combination is the same linear combination of rates of change.
- **(2) Leibniz Rule:** The product rule ensures that the derivative respects tensor multiplication. Without this, the derivative would not be compatible with the tensor algebra of the manifold.
- **(3) Contraction:** Contraction is purely algebraic; differentiation should not interfere with it. This guarantees that derived tensors have the same index behavior as the originals.
- **(4) Directional Derivative:** On scalar fields, ∇ must reduce to the familiar notion of differentiation along a vector field. This anchors ∇ to our geometric intuition of change along flow lines.
- **(5) Torsion-Free:** Moving infinitesimally along two directions in different orders should lead to the same endpoint. The failure of this closure defines the *torsion tensor*. Setting torsion to zero ensures infinitesimal parallelograms on M close. Tangent spaces always exist at every point, but torsion-free means that the way we slide those tangent vectors along curves (via parallel transport) doesn't introduce any twist that would make the notion of "tangency" inconsistent.

Commutator of Vector Fields

Using properties (4) and (5) with the Leibniz rule, we can express the Lie bracket of two vector fields v^a and w^a in terms of the derivative operator ∇_a.

For any smooth function f,

$$\begin{aligned}[v, w](f) &= v\{w(f)\} - w\{v(f)\} \\ &= v^a \nabla_a (w^b \nabla_b f) - w^a \nabla_a (v^b \nabla_b f) \\ &= (v^a \nabla_a w^b - w^a \nabla_a v^b) \nabla_b f.\end{aligned}$$

Since this holds for all f, we obtain the component expression:

$$\boxed{[v, w]^b = v^a \nabla_a w^b - w^a \nabla_a v^b.}$$

This shows that the commutator of two vector fields can be written purely in terms of the derivative operator.

Torsion Tensor (Optional)

If condition (5) is not imposed, one can define the **torsion tensor**

$$T^c{}_{ab} = -(\nabla_a \nabla_b f - \nabla_b \nabla_a f)/(\nabla_c f),$$

which measures the antisymmetric part of the derivative operator. The torsion-free condition $T^c{}_{ab} = 0$ ensures that ∇_a is symmetric in its lower indices when acting on scalar functions.

Existence of Derivative Operators

Up to this point, we have defined what a derivative operator ∇ must do: it takes tensor fields to tensor fields, satisfies linearity, the Leibniz rule, commutes with contraction, reduces to directional derivatives on scalars, and is torsion-free. These conditions are axiomatic, they describe how a derivative *should* behave on a manifold. But we have not yet shown that such operators actually exist.

Step 1: Constructing a derivative in coordinates.

Given a coordinate system ψ on a region of the manifold M, with coordinates x^μ, we can always define a "derivative operator" using the ordinary partial derivatives of component functions. Let $\{\partial/\partial x^\mu\}$ and $\{dx^\mu\}$ be the associated coordinate bases. For any smooth tensor field $T^{a_1\cdots a_k}{}_{b_1\cdots b_l} \in \mathscr{T}(k, l)$, we define an operator ∂_a by taking the partial derivatives of its components in this coordinate basis:

$$(\partial_c T)^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} = \frac{\partial T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l}}{\partial x^\sigma}.$$

We call ∂_a the **ordinary derivative operator** associated with the coordinate system ψ.

Because partial derivatives in $\mathbb{R}^n$ already satisfy the standard rules of differentiation, this coordinate-based operator automatically satisfies the five defining properties of a derivative operator:

- Linearity follows from the linearity of partial differentiation.
- The Leibniz rule holds by the product rule for partial derivatives.
- Commutation with contraction is automatic because contraction is algebraic.
- Acting on scalars, $\partial_a f$ gives the ordinary directional derivative.
- Mixed partial derivatives commute, so the operator is torsion-free.

Thus, in every coordinate system, we can construct such a derivative operator.

Step 2: The catch – coordinate dependence.

The ordinary derivative operator ∂_a depends on the choice of coordinates used to define it. If we choose a different coordinate system ψ' with coordinates $x^{\mu'}$, then the same tensor field T will have new components $T^{\mu'_1\cdots\mu'_k}{}_{\nu'_1\cdots\nu'_l}$, and the corresponding partial derivatives $\partial'_\sigma T^{\mu'_1\cdots\mu'_k}{}_{\nu'_1\cdots\nu'_l}$ will generally *not* transform as the components of a tensor. In other words,

$$(\partial_a T)^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} \quad \text{and} \quad (\partial'_a T)^{\mu'_1\cdots\mu'_k}{}_{\nu'_1\cdots\nu'_l}$$

are not related by the tensor transformation law. This means that the operator ∂_a is not a geometric object; it depends on the coordinate system in which it was defined.

Step 3: The conceptual lesson.

This construction shows that derivative operators do exist in a purely formal sense: in any coordinate patch, we can always take partial derivatives of tensor components. However, this "ordinary derivative" is *coordinate dependent* and therefore not intrinsic to the manifold. The manifold itself does not come equipped with a canonical notion of how to compare vectors or tensors at different points. Partial derivatives depend on how we label points with coordinates.

To obtain a derivative operator that is truly geometric, one that produces tensors independent of any coordinate system, we must add additional structure to the manifold: a **connection**. The connection modifies the coordinate-based derivative by adding correction terms (connection coefficients) that transform in such a way that the full covariant derivative $\nabla_a T^{b_1\cdots b_k}{}_{c_1\cdots c_l}$ *does* transform tensorially.

Summary.

- Ordinary partial differentiation defines a derivative operator in any coordinate system.
- This operator satisfies the five conditions but depends on the chosen coordinates.
- Therefore, it is not naturally associated with the geometry of the manifold.
- To define a coordinate-independent derivative, we introduce a connection in the next section.

Uniqueness of Derivative Operators

Having shown that derivative operators exist (at least in coordinate systems), we now ask: *how unique are they?* In other words, can there be more than one derivative operator on the same manifold, and if so, how might they differ?

Step 1: Agreement on scalar fields.

By condition (4), all derivative operators must act identically on scalar fields:

$$\nabla_a f = \tilde{\nabla}_a f$$

for any smooth function $f \in \mathscr{F}$. Thus, any difference between two derivative operators ∇_a and $\tilde{\nabla}_a$ can only appear when they act on tensors of rank one or higher.

Step 2: Investigating their difference on covector fields.

Let ω_b be a smooth covector field. Consider the difference

$$(\tilde{\nabla}_a - \nabla_a)\omega_b.$$

To understand what kind of object this is, let us multiply ω_b by an arbitrary scalar field f and apply both derivative operators. Using the Leibniz rule for each, we find

$$\begin{aligned}\tilde{\nabla}_a(f\omega_b) - \nabla_a(f\omega_b) &= (\tilde{\nabla}_a f)\,\omega_b + f\,\tilde{\nabla}_a\omega_b - (\nabla_a f)\,\omega_b - f\,\nabla_a\omega_b \\ &= f(\tilde{\nabla}_a\omega_b - \nabla_a\omega_b),\end{aligned}$$

since $\tilde{\nabla}_a f = \nabla_a f$ by condition (4). Hence,

$$\tilde{\nabla}_a(f\omega_b) - \nabla_a(f\omega_b) = f(\tilde{\nabla}_a\omega_b - \nabla_a\omega_b). \tag{3.1.3}$$

This shows that the difference depends only on the value of ω_b at the point where the derivative is taken, not on how ω_b varies nearby.

Note on Locality and the Cancellation in (3.1.3).
A derivative operator ∇_a always depends on how a field varies in a *neighborhood* of a point. For a scalar f,

$$\nabla_a f(p)$$

measures the infinitesimal change of f near p; it is not determined solely by the value of f at p. Thus, any expression containing $\nabla_a f$ includes information about *nearby* values of f. Whereas a covector $\omega_b(p)$ lives entirely in the cotangent space T_p^*M and depends only on its value at p. Thus, for $(\tilde{\nabla}_a - \nabla_a)$ to be a tensor, all terms involving $\nabla_a f$ must cancel, leaving a purely pointwise, linear operation on $\omega_b(p)$.

Step 3: Linearity and pointwise dependence.

At any point p, both $\nabla_a\omega_b$ and $\tilde{\nabla}_a\omega_b$ depend on the local behavior of ω_b, but their *difference* depends only on $\omega_b(p)$. To see this, suppose ω'_b is another covector field such that $\omega'_b(p) = \omega_b(p)$.

Then, following Wald's argument:

$$(\tilde{\nabla}_a - \nabla_a)(\omega'_b - \omega_b) = 0 \quad \text{at } p.$$

Hence,

$$(\tilde{\nabla}_a - \nabla_a)\omega'_b = (\tilde{\nabla}_a - \nabla_a)\omega_b \quad \text{at } p. \tag{3.1.6}$$

Therefore, the difference operator is linear and pointwise: it depends only on the value of the field at p, not its derivatives.

Step 4: The difference as a tensor field.

Any operation that is linear and depends only on the value of its argument at a single point is itself a tensor. Thus, we define a smooth tensor field $C^a{}_{bc}$ of type $(1, 2)$ by

$$(\tilde{\nabla}_b - \nabla_b)v^a = C^a{}_{bc}\, v^c,$$

for every vector field v^a. This tensor $C^a{}_{bc}$ describes how one derivative operator differs from another.

By construction, $C^a{}_{bc}$ is linear in v^a, and because the difference vanishes on scalar fields, it is purely tensorial—it transforms covariantly under coordinate changes.

Step 5: Conceptual meaning.

This result is extremely important. It tells us that:

- There is not just one derivative operator on a manifold. Infinitely many exist, and they can differ by a tensor $C^a{}_{bc}$.
- The space of all derivative operators forms an affine space: given any one operator ∇_a, adding a tensor $C^a{}_{bc}$ defines a new one $\tilde{\nabla}_a = \nabla_a + C^a{}_{bc}$.
- Later, when we choose coordinates, the components $C^a{}_{bc}$ will appear as the familiar *connection coefficients* or *Christoffel symbols*.

Step 6: Relation to coordinate derivatives.

The "ordinary" derivative operators ∂_a defined in different coordinate systems are examples of distinct derivative operators. The tensor $C^a{}_{bc}$ between them encodes how their definitions differ. When we later introduce the metric, we will use it to single out one particular connection—the *Levi-Civita connection*—by requiring it to be both torsion-free and compatible with the metric:

$$\nabla_a g_{bc} = 0.$$

Summary.

- All derivative operators agree on scalars but may differ on higher tensors.
- The difference between any two derivative operators is a tensor field $C^a{}_{bc}$ of type $(1, 2)$.
- This tensor represents how different coordinate or connection choices shift the notion of differentiation on the manifold.
- In coordinates, its components become the *Christoffel symbols*.

Difference of Derivative Operators on Dual Vector Fields

Previously we showed that the difference between any two derivative operators ∇_a and $\tilde{\nabla}_a$ is described by a tensor field $C^a{}_{bc}$ of type $(1, 2)$:

$$(\tilde{\nabla}_b - \nabla_b)v^a = C^a{}_{bc}v^c.$$

We now examine how this relation appears when the operators act on *covector* fields (dual vectors) rather than vector fields.

Step 1: Action on dual vectors.

Let ω_b be a smooth covector field. Because the derivative operator acts linearly and satisfies the Leibniz rule, the difference between $\tilde{\nabla}_a$ and ∇_a must again depend only on the value of ω_b at the point p. We therefore define

$$\nabla_a\omega_b = \tilde{\nabla}_a\omega_b - C^c{}_{ab}\,\omega_c. \tag{3.1.7}$$

This shows explicitly how two derivative operators can disagree in their action on dual vector fields: the tensor $C^c{}_{ab}$ measures that disagreement.

Because ∇_a lowers an index and $\tilde{\nabla}_a$ acts on the same covector, the correction term enters with a minus sign— a reflection of how dual vectors transform contragrediently to ordinary vectors.

Step 2: Symmetry from torsion-freeness.

Condition (5) (the torsion-free requirement) implies that for any smooth scalar field f, the second derivatives commute:

$$\nabla_a\nabla_b f = \nabla_b\nabla_a f.$$

Using this property, we can infer a symmetry of $C^c{}_{ab}$. Let $\omega_b = \nabla_b f = \tilde{\nabla}_b f$. Then, applying equation (3.1.7),

$$\nabla_a\nabla_b f = \tilde{\nabla}_a\tilde{\nabla}_b f - C^c{}_{ab}\,\nabla_c f. \tag{3.1.8}$$

Since both $\nabla_a\nabla_b f$ and $\tilde{\nabla}_a\tilde{\nabla}_b f$ are symmetric in a and b, the correction term must also be symmetric:

$$C^c{}_{ab} = C^c{}_{ba}. \tag{3.1.9}$$

Thus, the tensor $C^c{}_{ab}$ is symmetric in its lower indices whenever the derivative operators are torsion-free.

Step 3: Geometric meaning.

This symmetry expresses the same geometric condition that the Christoffel symbols later satisfy:

$$\Gamma^a{}_{bc} = \Gamma^a{}_{cb}.$$

Torsion-free means that parallel transport around an infinitesimal closed loop does not depend on the order of motion along the two directions. In a curved but torsion-free spacetime, basis vectors may rotate under transport, but they do not “twist” independently of the curvature.

Summary.

- For covector fields, the difference of two derivative operators is $\nabla_a \omega_b = \tilde{\nabla}_a \omega_b - C^c{}_{ab} \omega_c$.
- The tensor $C^c{}_{ab}$ measures their possible disagreement.
- Under the torsion-free condition, $C^c{}_{ab}$ is symmetric in its lower indices: $C^c{}_{ab} = C^c{}_{ba}$.
- This symmetry will reappear for the Christoffel symbols of the Levi-Civita connection.

Extending the Difference Tensor to All Tensor Fields

We have shown that for any two derivative operators ∇_a and $\tilde{\nabla}_a$, their difference is described by a tensor field $C^a{}_{bc}$ of type $(1, 2)$, symmetric in its lower indices. This tensor measures how the two operators "disagree" when acting on vector and covector fields. We now extend this to tensors of arbitrary rank.

Step 1: Acting on a vector field.

Let t^a be a smooth vector field and ω_a a one-form field. By property (4) of derivative operators (compatibility with contraction),

$$(\tilde{\nabla}_a - \nabla_a)(\omega_b t^b) = 0. \tag{3.1.10}$$

Using the Leibniz rule, we expand:

$$(\tilde{\nabla}_a - \nabla_a)(\omega_b t^b) = (C^c{}_{ab}\, \omega_c) t^b + \omega_b (\tilde{\nabla}_a - \nabla_a) t^b. \tag{3.1.11}$$

Since this must vanish for all ω_b, we find

$$\nabla_a t^b = \tilde{\nabla}_a t^b + C^b{}_{ac} t^c. \tag{3.1.13}$$

Thus, $C^b{}_{ac}$ provides the correction needed to translate between the two derivative operators on vector fields.

Step 2: Acting on general tensors.

Applying the same reasoning repeatedly, we can write the difference of ∇_a and $\tilde{\nabla}_a$ acting on any tensor field $T^{b_1 \cdots b_k}{}_{c_1 \cdots c_l}$ as

$$\boxed{\nabla_a T^{b_1 \cdots b_k}{}_{c_1 \cdots c_l} = \tilde{\nabla}_a T^{b_1 \cdots b_k}{}_{c_1 \cdots c_l} + \sum_{i=1}^{k} C^{b_i}{}_{ad}\, T^{b_1 \cdots d \cdots b_k}{}_{c_1 \cdots c_l} - \sum_{j=1}^{l} C^d{}_{ac_j}\, T^{b_1 \cdots b_k}{}_{c_1 \cdots d \cdots c_l}.} \tag{3.1.14}$$

Each contravariant index contributes a $+$ correction term, and each covariant index contributes a $-$ correction term, reflecting how vectors and covectors transform oppositely.

Step 3: Interpretation.

The tensor $C^a{}_{bc}$ completely characterizes the difference between any two derivative operators: if we know $\tilde{\nabla}_a$ and $C^a{}_{bc}$, we can reconstruct ∇_a. Conversely, for any smooth symmetric tensor field $C^a{}_{bc}$, the operator ∇_a defined by equation 3.1.14 satisfies all five defining properties of a derivative operator.

Geometric Meaning. The object $C^a{}_{bc}$ describes how the "rule of differentiation" is modified from one connection to another. It encodes how the notion of parallelism changes across the manifold. In coordinates, $C^a{}_{bc}$ will become the familiar *Christoffel symbol*, which corrects the partial derivative to account for the curvature or twisting of the coordinate basis.

Freedom in Choosing a Derivative Operator and the Christoffel Symbols

Equation (3.1.14) shows that the difference between two derivative operators ∇_a and $\tilde{\nabla}_a$ is completely determined by a smooth tensor field $C^a{}_{bc}$, symmetric in its lower indices. Conversely, given any such $C^a{}_{bc}$, we can define a new derivative operator ∇_a by adding the correction terms in equation (3.1.14). Thus, the manifold structure alone does not select a unique derivative operator—there is considerable freedom.

Degrees of freedom. On an n-dimensional manifold, a symmetric tensor $C^a{}_{bc}$ has

$$\frac{n^2(n+1)}{2}$$

independent components at each point. Each possible choice of these components defines a different derivative operator ∇_a. Hence, an ordinary manifold M admits infinitely many distinct derivative operators, none of which is naturally preferred unless additional structure, such as a metric, is introduced.

Ordinary derivative operator and Christoffel symbols. A particularly useful special case occurs when we take $\tilde{\nabla}_a$ to be the *ordinary derivative operator* ∂_a associated with some coordinate system. Then the difference tensor is denoted by

$$\Gamma^a{}_{bc} \equiv C^a{}_{bc},$$

and is called a *Christoffel symbol*. Equation (3.1.13) becomes

$$\nabla_a t^b = \partial_a t^b + \Gamma^b{}_{ac} t^c. \tag{3.1.15}$$

Here the partial derivative $\partial_a t^b$ captures the coordinate variation of the components of t^b, while the $\Gamma^b{}_{ac}$ term corrects for how the basis vectors themselves change from point to point.

Coordinate dependence. Because the coordinate basis $\{\partial/\partial x^\mu\}$ changes under coordinate transformations, the associated ordinary derivative operator ∂_a also changes. When we change coordinates from x^μ to $x^{\mu'}$, we replace ∂_a by $\partial_{a'}$ and the Christoffel symbols $\Gamma^a{}_{bc}$ by new symbols $\Gamma^{a'}{}_{b'c'}$ that are not related by the usual tensor transformation law. This shows that Christoffel symbols are *not tensors*; they depend on both the derivative operator and the coordinate system used to define it.

Summary.
- A manifold by itself admits many derivative operators.
- Each operator differs from any other by a symmetric tensor $C^a{}_{bc}$.
- Choosing $\tilde{\nabla}_a = \partial_a$ makes $C^a{}_{bc}$ the Christoffel symbols $\Gamma^a{}_{bc}$.

- The Christoffel symbols are not tensor components—they change inhomogeneously with the coordinates.

Metric Compatibility and the Levi–Civita Condition

So far, we have seen that a manifold admits infinitely many possible derivative operators, each corresponding to a different tensor $C^a{}_{bc}$ or, in coordinates, a different set of Christoffel symbols $\Gamma^a{}_{bc}$. No particular choice is preferred by the manifold itself. However, if the manifold is equipped with a metric g_{ab}, the metric provides a natural way to single out one distinguished derivative operator.

Preservation of inner products.

Suppose we parallel transport two vector fields v^a and w^a along a curve with tangent t^a. It is natural to require that their inner product $g_{ab}v^a w^b$ remain constant along the curve:

$$t^a\nabla_a(g_{bc}v^b w^c) = 0. \tag{3.1.20}$$

This expresses the idea that parallel transport should not change lengths or angles as measured by the metric.

Using the Leibniz rule and the condition that v^a and w^a are parallel transported ($t^a\nabla_a v^b = t^a\nabla_a w^b = 0$), we obtain

$$t^a v^b w^c \nabla_a g_{bc} = 0. \tag{3.1.21}$$

Since this must hold for all curves and for all parallely transported vectors, we conclude that the derivative operator must satisfy

$$\boxed{\nabla_a g_{bc} = 0.} \tag{3.1.22}$$

Metric compatibility. Equation 3.1.22 is called the *metric compatibility condition.* It ensures that the derivative operator ∇_a preserves the metric under parallel transport:

$$\nabla_a g_{bc} = 0 \quad \Longleftrightarrow \quad \nabla_a g^{bc} = 0.$$

In geometric terms, the connection "respects" the geometry defined by g_{ab}—the inner product of any two parallely transported vectors is invariant along a curve.

Physical and Geometric Meaning. Metric compatibility encodes the principle that spacetime has no preferred directions or distortions: parallel transport preserves the notions of length and angle defined by the metric. Together with the torsion-free condition ($C^a{}_{bc} = C^a{}_{cb}$), it uniquely determines the derivative operator. This unique connection is known as the *Levi–Civita connection.*

The Levi–Civita Connection: Existence, Uniqueness, and Formula

We now show that on a (pseudo-)Riemannian manifold (M, g) there exists a *unique* torsion-free, metric-compatible derivative operator. This distinguished connection is the *Levi–Civita connection.*

Theorem 3.1 (Levi–Civita)**.** Let (M, g) be a smooth manifold with metric g_{ab}. Then there exists a unique derivative operator ∇_a satisfying:

1. **Torsion-free:** $\nabla_a \nabla_b f = \nabla_b \nabla_a f$ for all smooth functions f (equivalently, $T^c{}_{ab} = 0$);
2. **Metric-compatible:** $\nabla_a g_{bc} = 0$.

Geometric meaning of torsion-freeness. The Christoffel symbols encode how neighboring tangent spaces are "tilted" relative to one another when a vector is carried from one point to the next. They describe the adjustment needed for a vector to remain "as constant as possible" on a curved manifold. The torsion-free condition removes the *twisting* or *curl-like* part of this adjustment: it requires that taking a small step in one coordinate direction and then in another leads to the same result as reversing the order of the steps, so infinitesimal parallelograms close. What remains after eliminating this twist is the purely *dot-product-preserving* part determined by the metric. Thus torsion-freeness removes the "swirl" in how tangent spaces relate, while metric-compatibility fixes the remaining "tilt," uniquely producing the Levi–Civita connection.

Uniqueness of the Levi–Civita Connection

Suppose we have two derivative operators (connections) on the same manifold with metric g_{ab}:

$$\nabla_a \qquad \text{and} \qquad \tilde{\nabla}_a,$$

and assume that both satisfy:

1. **Torsion-free:**
$$\nabla_a \nabla_b f = \nabla_b \nabla_a f, \qquad \tilde{\nabla}_a \tilde{\nabla}_b f = \tilde{\nabla}_b \tilde{\nabla}_a f,$$
equivalently
$$\Gamma^c{}_{ab} = \Gamma^c{}_{ba}, \qquad \tilde{\Gamma}^c{}_{ab} = \tilde{\Gamma}^c{}_{ba}.$$
2. **Metric-compatible:**
$$\nabla_a g_{bc} = 0, \qquad \tilde{\nabla}_a g_{bc} = 0.$$

We show that these two connections must be identical.

Step 1: Define the difference operator

For any vector field v^c, define

$$(\tilde{\nabla}_a - \nabla_a) v^c = C^c{}_{ab} \, v^b. \tag{3.1}$$

Since the left-hand side is linear in v^b and depends only on the value of v^b at a point, this defines a smooth tensor field $C^c{}_{ab}$ of type $(1, 2)$.

Step 2: Identify $C^c{}_{ab}$ with the difference of Christoffel symbols

In a coordinate chart,

$$\nabla_a v^c = \partial_a v^c + \Gamma^c{}_{ab} v^b, \qquad \tilde{\nabla}_a v^c = \partial_a v^c + \tilde{\Gamma}^c{}_{ab} v^b.$$

Subtracting gives

$$(\tilde{\nabla}_a - \nabla_a) v^c = (\tilde{\Gamma}^c{}_{ab} - \Gamma^c{}_{ab})\, v^b.$$

Comparing with (3.1) and using that this holds for all v^b, we obtain

$$\boxed{C^c{}_{ab} = \tilde{\Gamma}^c{}_{ab} - \Gamma^c{}_{ab}.} \tag{3.2}$$

Step 3: Torsion-free condition implies symmetry

Torsion-free means

$$\Gamma^c{}_{ab} = \Gamma^c{}_{ba}, \qquad \tilde{\Gamma}^c{}_{ab} = \tilde{\Gamma}^c{}_{ba}.$$

Using (3.2),

$$C^c{}_{ab} = \tilde{\Gamma}^c{}_{ab} - \Gamma^c{}_{ab} = \tilde{\Gamma}^c{}_{ba} - \Gamma^c{}_{ba} = C^c{}_{ba}.$$

Thus

$$\boxed{C^c{}_{ab} = C^c{}_{ba}.} \tag{3.3}$$

Step 4: Apply metric compatibility

Consider the difference of the two metric-compatibility conditions:

$$0 = (\tilde{\nabla}_a - \nabla_a) g_{bc}.$$

Using Eq. 3.1.14 the covariant derivative of a $(0, 2)$ tensor is

$$\nabla_a g_{bc} = \partial_a g_{bc} - \Gamma^d{}_{ab} g_{dc} - \Gamma^d{}_{ac} g_{bd},$$

and similarly for $\tilde{\nabla}_a$, subtracting yields

$$(\tilde{\nabla}_a - \nabla_a) g_{bc} = -C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd}.$$

Thus metric-compatibility implies

$$0 = -C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd}. \tag{3.4}$$

Step 5: Contract with arbitrary vectors and use symmetry

Contract (3.4) with arbitrary vectors X^a, Y^b, Z^c:

$$0 = -g_{dc}C^d{}_{ab}\,X^aY^bZ^c - g_{bd}C^d{}_{ac}\,X^aY^bZ^c.$$

Using the symmetry $C^d{}_{ab} = C^d{}_{ba}$ from (3.3), interchange of b and c leaves the expression unchanged. Because X, Y, Z are arbitrary, the only possible solution is

$$C^d{}_{ab} = 0 \quad \text{everywhere.}$$

Conclusion

Since the difference tensor vanishes,

$$(\tilde{\nabla}_a - \nabla_a)v^c = 0 \qquad \forall v^c,$$

we obtain

$$\boxed{\tilde{\nabla}_a = \nabla_a.}$$

Hence the torsion-free, metric-compatible connection is *unique*.

Existence of the Levi–Civita Connection (Fully Expanded)

To prove existence, we start with:

$$(M, g) \quad \text{a smooth (pseudo-)Riemannian manifold.}$$

Our goal: construct a derivative operator ∇_a that is

1. **torsion-free**, and
2. **metric-compatible**: $\nabla_a g_{bc} = 0$.

Step 1: Start with any torsion-free connection

Choose *any* torsion-free derivative operator $\tilde{\nabla}_a$. For example: the ordinary partial derivative operator ∂_a in a coordinate patch.

By definition of a covariant derivative on vector fields:

$$\tilde{\nabla}_a v^c = \partial_a v^c + \tilde{\Gamma}^c{}_{ab} v^b,$$

where $\tilde{\Gamma}$ is symmetric in a, b (torsion-free condition).

Step 2: Modify it by adding a $(1,2)$ tensor

Define a new operator ∇_a by

$$\nabla_a v^c := \tilde{\nabla}_a v^c + C^c{}_{ab} v^b, \tag{3.5}$$

where $C^c{}_{ab}$ is *not yet chosen.*

Since:

- $\tilde{\nabla}$ is a derivative operator, and - adding a $(1,2)$ tensor multiplied by v^b preserves linearity and Leibniz rules,

any such choice of $C^c{}_{ab}$ produces a valid derivative operator.

We now choose C so that ∇ is metric-compatible.

Step 3: Compute the metric derivative under the new ∇

Apply ∇_a to g_{bc}. Using definition (3.5) and the general formula for covariant derivatives of a $(0,2)$ tensor:

$$\nabla_a g_{bc} = \tilde{\nabla}_a g_{bc} - C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd}.$$

We want metric compatibility, so we impose

$$\nabla_a g_{bc} = 0.$$

Thus:

$$0 = \tilde{\nabla}_a g_{bc} - C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd}. \tag{3.6}$$

This is an algebraic equation for $C^c{}_{ab}$.

Step 4: Solve for $C^c{}_{ab}$

We will symmetrize and permute indices to isolate C.

First, write the metric-compatibility condition three times with permuted indices:

$$\begin{aligned}
&(1) \qquad 0 = \tilde{\nabla}_a g_{bc} - C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd},\\
&(2) \qquad 0 = \tilde{\nabla}_b g_{ca} - C^d{}_{bc} g_{da} - C^d{}_{ba} g_{cd},\\
&(3) \qquad 0 = \tilde{\nabla}_c g_{ab} - C^d{}_{ca} g_{db} - C^d{}_{cb} g_{ad}.
\end{aligned}$$

Now add equations (1) and (2), then subtract equation (3):

$$\tilde{\nabla}_a g_{bc} + \tilde{\nabla}_b g_{ca} - \tilde{\nabla}_c g_{ab}$$
$$= (C^d{}_{ab} g_{dc} + C^d{}_{ac} g_{bd} + C^d{}_{bc} g_{da} + C^d{}_{ba} g_{cd} - C^d{}_{ca} g_{db} - C^d{}_{cb} g_{ad}).$$

Use the symmetries

$$C^d{}_{ab} = C^d{}_{ba}, \qquad g_{ab} = g_{ba}.$$

Then the following pairs cancel:

$$C^d{}_{ac} g_{bd} - C^d{}_{ca} g_{db} = 0, \qquad C^d{}_{bc} g_{da} - C^d{}_{cb} g_{ad} = 0.$$

Thus the right-hand side reduces to

$$2\, C^d{}_{ab} g_{dc}.$$

Therefore we obtain the identity

$$\boxed{\tilde{\nabla}_a g_{bc} + \tilde{\nabla}_b g_{ac} - \tilde{\nabla}_c g_{ab} = 2C^d{}_{ab} g_{dc}.}$$

Multiply both sides by the inverse metric g^{ce} taking advantage of the fact that $g^{ce} g_{dc} = \delta^e_d$:

$$C^e{}_{ab} = \tfrac{1}{2}\, g^{ce} \left(\tilde{\nabla}_a g_{bc} + \tilde{\nabla}_b g_{ac} - \tilde{\nabla}_c g_{ab} \right).$$

Relabel index $e \to c$ to match conventions:

$$\boxed{C^c{}_{ab} = \tfrac{1}{2}\, g^{cd} \Big(\tilde{\nabla}_a g_{bd} + \tilde{\nabla}_b g_{ad} - \tilde{\nabla}_d g_{ab} \Big).} \qquad \text{(LC–abstract)}$$

This is the unique tensor that makes $\nabla g = 0$.

Step 5: Check torsion-freeness

Because $\tilde{\nabla}$ is torsion-free and $C^c{}_{ab} = C^c{}_{ba}$ is symmetric in (a, b),

$$\nabla_a v^c - \nabla_b v^c = (\tilde{\nabla}_a - \tilde{\nabla}_b) v^c + C^c{}_{ab} v^b - C^c{}_{ba} v^a = 0.$$

So ∇ is also torsion-free.

Thus ∇ satisfies:

- $\nabla g = 0$ - torsion-free

so it is the Levi–Civita connection.

3.3 Curvature

In the previous section we developed the notion of a derivative operator ∇_a (also called a *connection*) and showed how it determines the parallel transport of vectors and tensors along curves. We are now ready to use parallel transport to define curvature.

The starting point is the observation that parallel transport is generally path–dependent: transporting a vector from p to q along two different curves need not give the same result. Equivalently, a vector transported around a small closed loop will typically fail to return to its original direction. This failure is the geometric content of curvature.

The infinitesimal form of this idea is encoded in the failure of covariant derivatives to commute. Let ∇_a be any derivative operator and let ω_c be a smooth dual vector field. For a scalar field f we compute

$$\nabla_a\nabla_b(f\omega_c) = \nabla_a\big(\omega_c\nabla_b f + f\nabla_b\omega_c\big) \tag{3.7}$$
$$= (\nabla_a\nabla_b f)\,\omega_c + (\nabla_b f)\,\nabla_a\omega_c + (\nabla_a f)\,\nabla_b\omega_c + f\,\nabla_a\nabla_b\omega_c. \tag{3.2.1}$$

If we subtract from this the expression with a and b exchanged, the first three terms cancel pairwise, yielding

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)(f\omega_c) = f\,(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c. \tag{3.2.2}$$

As in our earlier analysis of derivative operators, this identity shows that the commutator $(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c$ at a point p depends only on the value of ω_c at p, not on its behavior nearby. Thus the map

$$\omega_c \longmapsto (\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c$$

is linear on T_p^*M and hence defines a tensor of type $(1,3)$ at p.

Definition 3.2 (Riemann curvature tensor)**.** There exists a unique tensor field $R_{abc}{}^d$ such that for all dual vector fields ω_c,

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\,\omega_c = R_{abc}{}^d\,\omega_d. \tag{3.2.3}$$

This tensor $R_{abc}{}^d$ is called the *Riemann curvature* associated with the derivative operator ∇_a.

3.3.1 Curvature and Parallel Transport Around a Loop

We now relate the curvature tensor $R_{abc}{}^d$ to the failure of a vector to return to its original value when parallel transported around a closed loop. Let $p \in M$ and choose a two–dimensional surface S through p. Introduce local coordinates (t,s) on S and choose these so that p corresponds to $(0,0)$.

Consider the small rectangular loop shown in Fig. 3.1:

$$(0,0) \;\to\; (\Delta t,0) \;\to\; (\Delta t,\Delta s) \;\to\; (0,\Delta s) \;\to\; (0,0).$$

Let v^a be a vector at p (not assumed tangent to S), and let us parallel transport v^a around the loop. We compute the resulting change in v^a by evaluating the change in the scalar $v^a\omega_a$ for an arbitrary dual vector field ω_a.

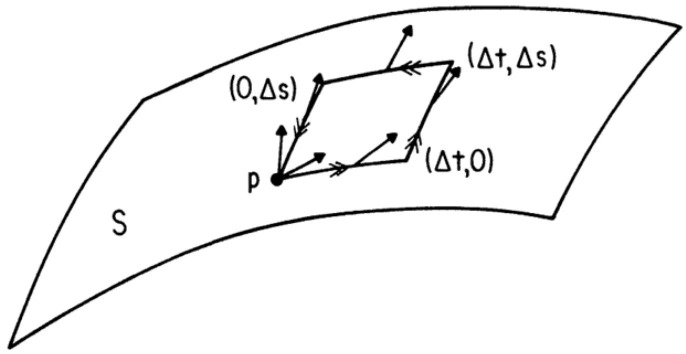

Figure 3.1: Parallel transport of a vector v^a around a small rectangular loop in the (t,s) surface through p. As derived in the text, the second–order change in v^a is governed by the Riemann tensor.

First–order variations (in detail). Along the first leg of the loop we move from $(0,0)$ to $(\Delta t,0)$, i.e. along the curve

$$\gamma(t) = (t,0), \qquad 0 \le t \le \Delta t.$$

We are interested in the change in the scalar field $v^a\omega_a$ between the endpoints of this segment. To first order in Δt, a Taylor expansion about the midpoint $t = \Delta t/2$ gives

$$\delta_1 = (v^a\omega_a)\big|_{(\Delta t,0)} - (v^a\omega_a)\big|_{(0,0)} \approx \Delta t\, \frac{\partial}{\partial t}(v^a\omega_a)\bigg|_{(\Delta t/2,0)}. \tag{3.2.4}$$

Now, the curve $\gamma(t)$ has tangent vector

$$T^b = \left(\frac{\partial}{\partial t}\right)^b,$$

which is simply the coordinate basis vector $\partial/\partial t$ restricted to the surface S along the line $s = 0$. Since $v^a\omega_a$ is a scalar field, its derivative along the curve is its *directional derivative* in the direction T^b:

$$\frac{\partial}{\partial t}(v^a\omega_a) = T^b\nabla_b(v^a\omega_a),$$

because for scalars f we have $\nabla_b f = \partial_b f$ by property (4) of the derivative operator.[1] Thus we may rewrite (3.2.4) as

$$\delta_1 = \Delta t\, T^b\nabla_b(v^a\omega_a)\big|_{(\Delta t/2,0)}.$$

[1]More explicitly, $\dfrac{d}{dt}f(\gamma(t)) = \dfrac{\partial x^b}{\partial t}\,\partial_b f = T^b\partial_b f = T^b\nabla_b f.$

Next we use the Leibniz rule for ∇_b acting on the product of a vector and a covector:

$$\nabla_b(v^a\omega_a) = (\nabla_b v^a)\,\omega_a + v^a\nabla_b\omega_a.$$

Contracting with T^b gives

$$T^b\nabla_b(v^a\omega_a) = (T^b\nabla_b v^a)\,\omega_a + v^a T^b\nabla_b\omega_a.$$

However, by construction v^a is being parallel transported along the t–curves, whose tangent is T^b. The definition of parallel transport (eq. (3.1.16)) therefore tells us

$$T^b\nabla_b v^a = 0$$

along this leg of the loop. Hence the first term above vanishes and we are left with

$$T^b\nabla_b(v^a\omega_a) = v^a T^b\nabla_b\omega_a.$$

Substituting back into the expression for δ_1, we obtain

$$\delta_1 = \Delta t\, v^a T^b\nabla_b\omega_a\big|_{(\Delta t/2,0)}. \tag{3.2.5}$$

Recap of this step.

- We first view $v^a\omega_a$ as a scalar field and expand its change along the curve $s=0$ using $\delta_1 \approx \Delta t\,\partial_t(v^a\omega_a)$.
- The scalar derivative ∂_t is the directional derivative in the direction of the tangent T^b, so $\partial_t(v^a\omega_a) = T^b\nabla_b(v^a\omega_a)$.
- Applying the Leibniz rule and using the fact that v^a is parallel transported ($T^b\nabla_b v^a = 0$) kills one term, leaving only $v^a T^b\nabla_b\omega_a$.

This is why the term involving $\nabla_b v^a$ disappears in eq. (3.2.5).

Second–order contributions. To evaluate the difference in brackets to first order in Δs, we move along the curve $s\mapsto s+\Delta s$ with $t=\Delta t/2$. Parallel transport is path–independent to first order, so v^a at $(\Delta t/2,\Delta s)$ equals the parallel transport of v^a at $(\Delta t/2,0)$. However,

$$T^b\nabla_b\omega_a\big|_{(\Delta t/2,\Delta s)} = T^b\nabla_b\omega_a\big|_{(\Delta t/2,0)} + \Delta s\,S^c\nabla_c(T^b\nabla_b\omega_a),$$

where $S^c = \partial/\partial s$ is the tangent to curves of constant t. Thus,

$$\delta_1 + \delta_3 = -\Delta t\,\Delta s\; v^a S^c\nabla_c(T^b\nabla_b\omega_a). \tag{3.2.7}$$

Repeating the calculation for $\delta_2+\delta_4$ and adding both contributions, we find the total change

$$\delta(v^a\omega_a) = \Delta t\,\Delta s\, v^a\{T^c S^b\nabla_c(\nabla_b\omega_a) - S^c T^b\nabla_c(\nabla_b\omega_a)\}.$$

Using the commutation of partial derivatives of coordinate basis vectors (§2.2) and the definition of the Riemann tensor (3.2.3), this becomes

$$\delta(v^a\omega_a) = \Delta t\,\Delta s\, v^a T^c S^b(\nabla_c\nabla_b - \nabla_b\nabla_c)\omega_a \tag{3.8}$$

$$= \Delta t\,\Delta s\, v^a T^c S^b R_{cba}{}^d\omega_d. \tag{3.2.8}$$

Since this must hold for all ω_a, the change in v^a itself is

$$\boxed{\delta v^a = \Delta t \, \Delta s \; v^d \, T^b S^c R_{cbd}{}^a.} \tag{3.2.9}$$

Thus the Riemann tensor directly measures the infinitesimal failure of a vector to return to its original value after parallel transport around a closed loop.

Curvature Acting on Vector Fields

We now derive the formula for the action of the commutator of covariant derivatives on a vector field t^c. Let ω_a be any dual vector field. Using property 5 (torsion–free), the Leibniz rule, and eq. (3.2.3), we compute

$$\begin{aligned} 0 &= (\nabla_a \nabla_b - \nabla_b \nabla_a)(t^c \omega_c) && (3.9)\\ &= \nabla_a(\omega_c \nabla_b t^c + t^c \nabla_b \omega_c) - \nabla_b(\omega_c \nabla_a t^c + t^c \nabla_a \omega_c) && (3.10)\\ &= \omega_c(\nabla_a \nabla_b - \nabla_b \nabla_a) t^c + t^c (\nabla_a \nabla_b - \nabla_b \nabla_a)\omega_c && (3.11)\\ &= \omega_c(\nabla_a \nabla_b - \nabla_b \nabla_a) t^c + t^c \omega_d R_{abc}{}^d. && (3.2.10) \end{aligned}$$

Since ω_c is arbitrary, we conclude

$$\boxed{(\nabla_a \nabla_b - \nabla_b \nabla_a) t^c = -R_{abd}{}^c \, t^d.} \tag{3.2.11}$$

This gives the action of the curvature on vector fields.

Curvature Acting on General Tensors

We have already seen how the commutator of covariant derivatives acts on dual vector fields and vector fields:

$$\begin{aligned} (\nabla_a \nabla_b - \nabla_b \nabla_a)\omega_c &= R_{abc}{}^d \, \omega_d, && (3.2.3)\\ (\nabla_a \nabla_b - \nabla_b \nabla_a) t^c &= -R_{abd}{}^c \, t^d. && (3.2.11) \end{aligned}$$

Using the Leibniz rule, these formulas extend to arbitrary tensor fields. By induction on rank one finds that for a tensor field $T^{c_1 \cdots c_k}{}_{d_1 \cdots d_l}$,

$$\boxed{(\nabla_a \nabla_b - \nabla_b \nabla_a) T^{c_1 \cdots c_k}{}_{d_1 \cdots d_l} = -\sum_{i=1}^{k} R_{abe}{}^{c_i} T^{c_1 \cdots e \cdots c_k}{}_{d_1 \cdots d_l} + \sum_{j=1}^{l} R_{abd_j}{}^{e} T^{c_1 \cdots c_k}{}_{d_1 \cdots e \cdots d_l}.} \tag{3.2.12}$$

Each upper index contributes a $-R$ term and each lower index a $+R$ term, exactly as in the simple vector and covector cases.

Algebraic Properties of the Riemann Tensor

Wald next establishes four key properties of $R_{abc}{}^{d}$:

1. Antisymmetry in the first two indices:
$$R_{abc}{}^{d} = -R_{bac}{}^{d}. \tag{3.2.13}$$
2. Cyclic (or "first Bianchi") identity in the lower three indices:
$$R_{[abc]}{}^{d} = 0. \tag{3.2.14}$$
3. For the Levi–Civita connection ∇_a associated with the metric ($\nabla_a g_{bc} = 0$), the Riemann tensor is antisymmetric on the last pair and obeys the pair-exchange symmetry
$$R_{abcd} = -R_{abdc}, \qquad R_{abcd} = R_{cdab}. \tag{3.2.15}$$
4. Differential Bianchi identity:
$$\nabla_{[a} R_{bc]d}{}^{e} = 0. \tag{3.2.16}$$

Sketch of the proofs. Property (1) follows directly from the definition (3.2.3) since

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)\omega_c = -(\nabla_b \nabla_a - \nabla_a \nabla_b)\omega_c.$$

To prove (2), we note that for any dual vector field ω_a and any connection ∇_a one has

$$\nabla_{[a} \nabla_{b]} \omega_c = 0, \tag{3.2.17}$$

when ∇_a is chosen to be an ordinary derivative operator ∂_a (see Eq. 3.1.14) and we use the symmetry $C^c{}_{ab} = C^c{}_{ba}$ of the difference tensor. In differential forms language this is just $d^2\omega = 0$. Thus

$$\begin{aligned} 0 &= 2\nabla_{[a} \nabla_{b]} \omega_c = \nabla_a \nabla_b \omega_c - \nabla_b \nabla_a \omega_c \\ &= (\nabla_a \nabla_b - \nabla_b \nabla_a)\omega_c = R_{abc}{}^{d}\, \omega_d. \end{aligned} \tag{3.2.18}$$

Antisymmetrizing over a, b, c in the last expression yields $R_{[abc]}{}^{d} = 0$, which is (2).

Property (3) makes use of metric-compatibility. Apply the general formula (3.2.12) to the metric tensor g_{ab}. Since $\nabla_a g_{bc} = 0$, we have

$$\begin{aligned} 0 &= (\nabla_a \nabla_b - \nabla_b \nabla_a) g_{cd} \\ &= R_{abc}{}^{e} g_{ed} + R_{abd}{}^{e} g_{ce} = R_{abcd} + R_{abdc}. \end{aligned} \tag{3.2.19}$$

This proves $R_{abcd} = -R_{abdc}$. Together with (1) and (2), one can show (see problem 3) that R_{abcd} also satisfies

$$R_{abcd} = R_{cdab}. \tag{3.2.20}$$

Finally, to establish (4), we apply the commutator of covariant derivatives to $\nabla_c \omega_d$ and use (3.2.12):

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)\nabla_c \omega_d = R_{abc}{}^{e}\, \nabla_e \omega_d + R_{abd}{}^{e}\, \nabla_c \omega_e. \tag{3.2.21}$$

On the other hand, we may first commute the covariant derivatives in the other order:

$$\begin{aligned}\nabla_a(\nabla_b\nabla_c\omega_d) - \nabla_b(\nabla_a\nabla_c\omega_d) &= \nabla_a(R_{bcd}{}^e\omega_e) - \nabla_b(R_{acd}{}^e\omega_e)\\ &= \omega_e\nabla_a R_{bcd}{}^e + R_{bcd}{}^e\nabla_a\omega_e - \omega_e\nabla_b R_{acd}{}^e - R_{acd}{}^e\nabla_b\omega_e. \end{aligned} \tag{3.2.22}$$

Antisymmetrizing over a, b, c in (3.2.21) and (3.2.22), the left-hand sides agree. Equality of the right-hand sides then yields

$$\omega_e\nabla_{[a}R_{bc]d}{}^e = 0 \tag{3.2.23}$$

for all ω_e, which implies

$$\nabla_{[a}R_{bc]d}{}^e = 0,$$

i.e. property (4), the differential Bianchi identity (3.2.16).

Ricci Tensor, Scalar Curvature, and Weyl Tensor

Because of the antisymmetries (1) and (3), some traces of the Riemann tensor vanish identically, but one nontrivial trace remains. Contracting the first and fourth indices gives the *Ricci tensor*:

$$\boxed{R_{ac} \equiv R_{abc}{}^b.} \tag{3.2.25}$$

Using the symmetry properties of R_{abcd}, one can show that R_{ab} is symmetric:

$$R_{ac} = R_{ca}. \tag{3.2.26}$$

Contracting the Ricci tensor once more with the metric produces the *scalar curvature*:

$$\boxed{R \equiv R_a{}^a = g^{ab}R_{ab}.} \tag{3.2.27}$$

It is often useful to decompose the Riemann tensor into a "trace part" and a "trace–free part." The trace–free part is called the *Weyl tensor*, C_{abcd}, and for $n \geq 3$ is defined by

$$\boxed{R_{abcd} = C_{abcd} + \frac{2}{n-2}\big(g_{a[c}R_{d]b} - g_{b[c}R_{d]a}\big) - \frac{2}{(n-1)(n-2)}\,R\,g_{a[c}g_{d]b}.} \tag{3.2.28}$$

The Weyl tensor satisfies the same algebraic symmetries (1), (2), and (3) as R_{abcd} and is trace-free on all index pairs. Geometrically, C_{abcd} encodes the "shape-changing" (conformal) part of curvature, while the Ricci tensor encodes the part related to volume change.

3.4 Geometric Meaning of Riemann and Ricci Curvature

3.4.1 Riemann Curvature Acting on Vectors and Tensors

The Riemann curvature tensor

$$R^a{}_{bcd}$$

measures the failure of second covariant derivatives to commute:

$$(\nabla_c \nabla_d - \nabla_d \nabla_c)V^a = R^a{}_{bcd} V^b.$$

For a vector field V^a, this expresses the fundamental fact that parallel transport around an infinitesimal loop in the c–d directions produces a small "rotation" of V^a. In a flat manifold this commutator vanishes, and the vector returns to its original orientation.

This generalizes naturally to a tensor of type (k, l) as we have seen previously in Eq 3.2.12:

$$\begin{aligned}(\nabla_c \nabla_d - \nabla_d \nabla_c)T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l} = &-\sum_{i=1}^{k} R^{a_i}{}_{ecd}\, T^{a_1 \ldots e \ldots a_k}{}_{b_1 \ldots b_l} \\ &+\sum_{j=1}^{l} R^{e}{}_{b_j cd}\, T^{a_1 \ldots a_k}{}_{b_1 \ldots e \ldots b_l}.\end{aligned}$$

Each contravariant index of T picks up a curvature term with a *minus* sign, and each covariant index picks up a curvature term with a *plus* sign, reflecting how curvature acts separately on each vector or covector slot.

Interpretation. Riemann curvature tells you how each index of a tensor fails to return to its initial orientation after being parallel transported around a small loop. A rank–(k, l) tensor has $k + l$ independent "failure modes," one for each index slot.

3.4.2 Why Contracting Riemann Yields the Ricci Tensor

The Ricci tensor is obtained by contracting the first and third indices:

$$R_{bd} = R^a{}_{bad}.$$

To understand the meaning of this contraction, it is helpful to examine the index roles in $R^a{}_{bcd}$:

- c and d determine the infinitesimal loop,
- b labels the input vector being parallel transported,
- a labels the output direction after the loop.

Thus the map $V^b \mapsto R^a{}_{bcd} V^b$ is a linear transformation on vectors. Taking its trace over a and b extracts the overall "net" effect of this transformation.

Geometric meaning. Contracting the Riemann tensor traces over how curvature acts on all possible directions of the transported vector. The resulting Ricci tensor R_{bd} captures only the *volume-changing* part of the curvature, discarding the purely shape-changing (Weyl) part.

3.4.3 Ricci Curvature and the Volume of Geodesic Balls

Consider a small ball of freely falling test particles—a *geodesic ball*. Each particle follows a geodesic with tangent u^a. Their mutual separation vectors ξ^a obey the geodesic deviation equation:

$$\frac{D^2\xi^a}{d\tau^2} = -R^a{}_{bcd}\, u^b u^d\, \xi^c.$$

The *volume* $V(\tau)$ of the ball evolves according to

$$\frac{d^2V}{d\tau^2} = -R_{ab}u^a u^b\, V.$$

Meaning. The Riemann tensor governs how individual separation vectors between nearby geodesics bend. The Ricci tensor, being the trace of Riemann, governs how the *entire* ball expands or contracts. It measures the local focusing or defocusing of geodesic flows.

3.4.4 Index Interpretation Summary

The Riemann tensor has type $(1, 3)$:

$$R^a{}_{bcd} : T_pM \times T_p^*M \times T_p^*M \times T_p^*M \to \mathbb{R}.$$

Its indices have these roles:

- c, d: directions of the infinitesimal loop,
- b: the vector being transported,
- a: resulting change in vector after loop.

Contracting a and b yields

$$R_{cd} = R^a{}_{cad},$$

which eliminates the orientation-changing information and keeps only the volume-changing component of curvature.

Contracting the Ricci tensor gives the scalar curvature, which is the overall magnitude (or average strength) of volume distortion caused by curvature at a point.

Summary.
- Riemann curvature measures the full failure of parallel transport.
- Ricci curvature is its trace, measuring geodesic focusing and volume distortion.
- Scalar curvature R is the trace of Ricci, giving the average curvature in all directions.
- Contracting the Ricci tensor gives the scalar curvature, which is the overall magnitude (or average strength) of volume distortion caused by curvature at a point.

Bianchi Identity and Einstein Tensor

Contracting the Bianchi identity (3.2.16) over e and b yields an identity for the Ricci tensor:

$$\nabla_a R_{bcd}{}^a + \nabla_b R_{cd} - \nabla_c R_{bd} = 0. \tag{3.2.29}$$

Raising the index d with the metric and contracting again over b and d, we obtain

$$\nabla_a R^a{}_c + \nabla_b R^b{}_c - \nabla_c R = 0, \tag{3.2.30}$$

or, equivalently,

$$\nabla^a G_{ab} = 0, \tag{3.2.31}$$

where

$$\boxed{G_{ab} \equiv R_{ab} - \tfrac{1}{2} R\, g_{ab}} \tag{3.2.32}$$

is called the *Einstein tensor*.

The twice–contracted Bianchi identity $\nabla^a G_{ab} = 0$ will play a central role in general relativity. In Einstein's field equations, $G_{ab} = 8\pi T_{ab}$, it guarantees the consistency of the equations with local conservation of energy–momentum, $\nabla^a T_{ab} = 0$.

Interpretation: From Riemann to Ricci to Scalar Curvature and the Einstein Tensor

Curvature in general relativity is built hierarchically. Each level of contraction of the Riemann tensor discards some information and highlights a different geometric feature of spacetime. This culminates in the Einstein tensor, the unique combination of curvature that can act as the source of gravity.

Riemann Curvature: Loop Failure and Full Geometric Information

The Riemann tensor $R^a{}_{bcd}$ measures the failure of parallel transport around infinitesimal loops:

$$(\nabla_c \nabla_d - \nabla_d \nabla_c) V^a = R^a{}_{bcd} V^b.$$

It captures *all* aspects of curvature: rotations, shear, tidal effects, and changes in separation between geodesics. This is the most detailed notion of curvature.

Riemann meaning. Riemann curvature tells how vectors and tensors fail to return to their original orientation after transport around small loops. It encodes the full tidal and shape-distorting structure of spacetime.

Ricci Curvature: Focusing of Geodesic Flows (Volume Distortion)

Contracting Riemann on its first and third indices gives the Ricci tensor:

$$R_{bd} = R^a{}_{bad}.$$

This removes the purely shape-changing (Weyl) part of curvature and retains only the component that affects the *volume* of small geodesic balls.
If u^a is the tangent to a geodesic congruence, then

$$R_{ab}u^a u^b$$

governs whether nearby geodesics focus or defocus. This appears in the Raychaudhuri equation and determines the second derivative of the volume of a geodesic ball:

$$\frac{d^2V}{d\tau^2} = -R_{ab}u^a u^b V.$$

Ricci meaning. Ricci curvature measures the *volume-changing* part of curvature. It tells whether bundles of geodesics converge (positive R_{ab}) or diverge (negative R_{ab}).

Scalar Curvature: Net Magnitude of Volume Distortion

Tracing the Ricci tensor gives the scalar curvature:

$$R = g^{ab}R_{ab}.$$

This is the average of Ricci curvature over all orthonormal directions. It measures how the volume of an infinitesimally small geodesic ball compares to that of a ball in flat space:

$$\mathrm{Vol}(B_\varepsilon) = \omega_n \varepsilon^n \left(1 - \frac{R}{6(n+2)}\varepsilon^2 + \cdots\right).$$

Scalar curvature meaning. Scalar curvature measures the *overall magnitude* of volume distortion. It is the "average curvature" at a point, condensed into a single number.

Einstein Tensor: Divergence-Free Curvature Sourced by Matter

The Einstein tensor is defined by

$$G_{ab} = R_{ab} - \tfrac{1}{2}Rg_{ab}.$$

The Bianchi identity guarantees

$$\nabla^a G_{ab} = 0,$$

so G_{ab} is the unique rank–$(0,2)$ tensor built from the metric and its first two derivatives that is automatically divergence-free.
This property matches the physical requirement that the stress–energy tensor obeys the conservation law $\nabla^a T_{ab} = 0$, which is why Einstein's equation takes the form

$$G_{ab} = 8\pi T_{ab}.$$

Einstein tensor meaning. The Einstein tensor encodes precisely the part of curvature responsible for the focusing of geodesic flows, adjusted by its trace so that it satisfies a local conservation law. It is the geometric quantity that "responds" directly to matter and energy.

Hierarchy of Curvature (Summary)

Tensor	Construction	Interpretation
$R^a{}_{bcd}$	Full curvature	Loop failure, tidal effects
R_{ab}	$R^c{}_{acb}$	Volume distortion, geodesic focusing
R	$g^{ab}R_{ab}$	Net strength of curvature (average)
G_{ab}	$R_{ab} - \frac{1}{2}Rg_{ab}$	Curvature sourced by matter

3.5 Geodesics

3.5.1 Definition via Parallel Transport

Let C be a smooth curve on the manifold M, with affine parameter t. The tangent vector to the curve is

$$T^a = \left(\frac{d}{dt}\right)^a.$$

A curve is called a *geodesic* if its tangent vector parallel transports itself along C:

$$T^a \nabla_a T^b = 0. \tag{3.3.1}$$

This equation expresses the idea that as we move along the curve, the tangent vector does not "rotate" relative to the connection.

More generally one may allow

$$T^a \nabla_a T^b = \alpha\, T^b, \tag{3.3.2}$$

where α is a scalar function along the curve. This version permits reparameterizations of the curve, but the path followed in the manifold is unchanged. Any such curve can be reparameterized so that the right–hand side of (3.3.2) vanishes, yielding the affinely parametrized form (3.3.1). A parameter for which (3.3.1) holds is called an *affine parameter*.

Geometric meaning. The condition

$$T^a \nabla_a T^b = 0$$

states that the tangent vector is carried along the curve without any additional turning or twisting beyond what the connection itself dictates. In flat Euclidean space this reduces to $dT^i/dt = 0$, the condition that a straight line has constant direction. On a curved manifold the connection encodes how the basis of tangent spaces changes from point to point, and a geodesic is a curve whose tangent vector remains constant with respect to this changing basis. Thus a geodesic is "as straight as the geometry allows."

3.5.2 Coordinate Expression of the Geodesic Equation

Let $\gamma(t)$ be a geodesic, and let

$$x^\mu(t) := x^\mu(\gamma(t))$$

be its coordinate representation in some chart. The tangent vector to the curve is

$$T^a = \left(\frac{d}{dt}\right)^a, \qquad T^\mu = \frac{dx^\mu}{dt}.$$

To compute the geodesic equation in coordinates, we evaluate $T^a \nabla_a T^\mu$. Using the definition of the covariant derivative of a vector field,

$$\nabla_a T^\mu = \partial_a T^\mu + \Gamma^\mu{}_{a\lambda} T^\lambda,$$

and contracting with T^a gives

$$T^a \nabla_a T^\mu = T^a \partial_a T^\mu + \Gamma^\mu{}_{a\lambda} T^a T^\lambda.$$

In a coordinate basis, $T^a \partial_a = T^\nu \partial_\nu$, so the first term is

$$T^\nu \partial_\nu T^\mu = \frac{d}{dt}\left(\frac{dx^\mu}{dt}\right) = \frac{d^2 x^\mu}{dt^2}.$$

Substituting back, the geodesic condition $T^a \nabla_a T^b = 0$ becomes

$$\frac{d^2 x^\mu}{dt^2} + \Gamma^\mu{}_{\nu\lambda} \frac{dx^\nu}{dt} \frac{dx^\lambda}{dt} = 0. \tag{3.3.3}$$

Equation (3.3.3) is Wald's coordinate form of the geodesic equation: a system of n coupled second–order ordinary differential equations for the coordinate functions $x^\mu(t)$.

Existence and uniqueness. Given an initial point $p \in M$ and an initial tangent vector $T^a \in T_p M$, the standard theorems for ODEs guarantee a unique solution to (3.3.3). Thus, specifying initial position and initial velocity uniquely determines a geodesic.

Why this equation represents "straightest possible" motion.
In ordinary Euclidean space, a straight line satisfies $\frac{d^2 x^i}{dt^2} = 0$. The extra term involving $\Gamma^\mu{}_{\nu\lambda}$ in (3.3.3) is the correction needed because, on a curved manifold, the coordinate basis itself *changes* from point to point. The Christoffel symbols encode this change.
Thus the geodesic equation says:

"The acceleration is exactly what is required to keep the direction constant relative to the moving basis."

If the connection coefficients vanish at a point (as in Riemannian normal coordinates), then at that point the geodesic equation reduces to $d^2 x^\mu / dt^2 = 0$, mimicking a straight line in flat space.

3.5.3 The Exponential Map

Geodesics are uniquely determined by an initial point and an initial tangent vector. The exponential map packages this fact into a single, geometric operation that sends a tangent vector at a point into the manifold itself by "following the geodesic it generates." This construction becomes indispensable later—for defining normal coordinates, understanding curvature, and formalizing local inertial frames in general relativity.

Initial value formulation of geodesics. Fix a point $p \in M$ and let T_pM be its tangent space. Given any vector $T^a \in T_pM$, consider the geodesic equation

$$T^a \nabla_a T^b = 0,$$

with initial conditions

$$\gamma_T(0) = p, \qquad \dot{\gamma}_T(0) = T^a.$$

By the existence and uniqueness theorem for ODEs, this determines a unique geodesic

$$\gamma_T : I \to M,$$

defined on some open interval I containing 0. In other words, T^a fixes both the initial direction and the rate of change along the geodesic.

Smooth dependence on initial data. Because the geodesic equation is a smooth system of ODEs, solutions vary smoothly with respect to their initial data. Thus, for nearby vectors T^a and T'^a, the geodesics γ_T and $\gamma_{T'}$ differ smoothly.

Definition of the exponential map. We now follow Wald and define a map from the tangent space into the manifold by evaluating each geodesic at unit affine parameter:

$$\exp_p(T^a) := \gamma_T(1). \tag{3.3.6}$$

Thus: - start at p, - shoot out along the unique geodesic whose initial tangent is T^a, - travel for one unit of affine parameter, - arrive at the point $\exp_p(T^a)$.

The choice of parameter value 1 is conventional: changing it simply rescales the tangent vectors by a constant factor.

Local diffeomorphism. For sufficiently small T^a, the map $\exp_p$ is smooth and satisfies

$$(d\,\exp_p)_{0^a} = \mathrm{Id}_{T_pM}.$$

This follows from expanding $\gamma_T(t)$ for small t:

$$\gamma_T(t) = p + tT^a + O(t^2).$$

Thus the derivative of $\exp_p$ at the origin sends small vectors T^a to the corresponding initial displacement in M.

By the inverse function theorem, $\exp_p$ is therefore a diffeomorphism from a neighborhood of $0^a \in T_pM$ onto a neighborhood of p in M.

Geometric meaning. The exponential map provides the canonical way to move from p into the manifold in the direction indicated by a tangent vector. It is intrinsic and coordinate-free. In Euclidean space with the usual flat connection, $\exp_p(T^i)$ is simply $p + T^i$. On a curved manifold, $\exp_p$ accomplishes the analogous "straight motion" using geodesics instead of literal straight lines.

Why the exponential map matters.
The exponential map ties together:

- the linear structure of T_pM,
- the nonlinear geometry of the manifold M,
- and the geodesics determined by the connection ∇.

It allows us to:

- build coordinate systems whose axes *are* geodesics (Riemannian normal coordinates),
- express curvature through how $\exp_p$ deviates from linearity,
- treat small neighborhoods of p as if they were tangent-space vectors "exponentiated" into M,
- formalize the notion of freely falling observers and local inertial frames in general relativity.

In short:

$$\exp_p : T_pM \to M$$

is the bridge between the tangent space and the manifold's intrinsic geometry.

3.5.4 Riemannian Normal Coordinates

Riemannian normal coordinates (RNC) provide a coordinate system centered at a point $p \in M$ in which the geometry looks as flat as possible at that point:

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \qquad \Gamma^{\rho}{}_{\mu\nu}(p) = 0.$$

These coordinates arise directly from the exponential map.

Step 1: Use the exponential map to label nearby points. For each vector $T^a \in T_pM$, the exponential map sends T^a to the endpoint of the geodesic shot out from p with initial tangent T^a:

$$q = \exp_p(T^a).$$

When T^a is sufficiently small, this establishes a one-to-one correspondence between points q near p and tangent vectors T^a near the origin.

Step 2: Use components of T^a as coordinates. Choose a basis $e^a_{(\mu)}$ of T_pM. Every tangent vector can be uniquely expanded as

$$T^a = T^\mu e^a_{(\mu)}.$$

Definition: The Riemannian normal coordinates of q are defined by

$$x^\mu(q) := T^\mu \quad \text{where } q = \exp_p(T^a).$$

Thus each coordinate line x^μ corresponds to a geodesic starting from p in the direction of $e^a_{(\mu)}$.

Geometric meaning of coordinate axes. In RNC: - The coordinate lines through p are geodesics. - The coordinate basis vectors at p equal the original basis $e^a_{(\mu)}$. - Motion in a single coordinate direction corresponds to following a uniquely determined geodesic.

Consequences for the metric and the connection. Because coordinate axes are geodesics,

$$\Gamma^\mu{}_{\nu\lambda}(p) = 0.$$

To see this, consider the geodesic equation for a coordinate line parametrized by $t = x^\mu$:

$$\frac{d^2x^\mu}{dt^2} + \Gamma^\mu{}_{\nu\lambda}\frac{dx^\nu}{dt}\frac{dx^\lambda}{dt} = 0.$$

Along such a curve,

$$\frac{dx^\nu}{dt} = \delta^\nu{}_\mu, \qquad \frac{d^2x^\mu}{dt^2} = 0.$$

Hence the geodesic equation demands

$$\Gamma^\mu{}_{\mu\mu}(p) = 0,$$

and repeating this argument in different coordinate directions shows

$$\Gamma^\rho{}_{\mu\nu}(p) = 0 \quad \text{for all } \mu, \nu, \rho.$$

Next, at p, the metric can be arranged to equal the Minkowski/Euler metric $\eta_{\mu\nu}$ by choosing the basis $\{e^a_{(\mu)}\}$ orthonormally:

$$g_{\mu\nu}(p) = g_{ab}e^a_{(\mu)}e^b_{(\nu)} = \eta_{\mu\nu}.$$

Differentiating the identity $g_{\mu\nu} = \eta_{\mu\nu}$ at p and using $\Gamma^\rho{}_{\mu\nu}(p) = 0$, we also get

$$\partial_\lambda g_{\mu\nu}(p) = 0.$$

Thus, *RNC flatten the geometry at p to first order.* Curvature appears only in the second derivatives of the metric.

Higher–order behavior and curvature. Wald briefly states, and one can show, that:

$$\partial_\sigma\partial_\rho g_{\mu\nu}(p) = -\tfrac{1}{3}(R_{\mu\rho\nu\sigma} + R_{\mu\sigma\nu\rho})(p).$$

These second-derivative terms are the leading indicators of curvature in RNC and are responsible for tidal effects, geodesic deviation, and the non-Euclidean behavior of volumes and areas.

Why RNC are important.
Riemannian normal coordinates do not merely "look like" flat coordinates — they reproduce the geometry of flat space at a point p in the strongest possible sense:

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \qquad \partial_\lambda g_{\mu\nu}(p) = 0, \qquad \Gamma^\rho{}_{\mu\nu}(p) = 0.$$

Thus RNC give:

- a coordinate system in which gravity "vanishes" at a point, matching the physical intuition of a freely falling observer,
- the cleanest environment for expanding tensors in Taylor series,
- the natural coordinate system for analyzing curvature,
- the starting point for normal coordinate expansions used throughout general relativity, gauge theory, and Riemannian geometry.

They are the mathematical realization of the equivalence principle: *locally, spacetime is flat.*

3.5.5 Gaussian Normal Coordinates

Gaussian normal coordinates (GNC) generalize the idea of Riemannian normal coordinates from a single point to an entire hypersurface. Instead of shooting geodesics out of one point, we shoot geodesics *orthogonally* out of every point on a chosen hypersurface S.

Step 1: Start with a hypersurface and its normal. Let $S \subset M$ be a smooth hypersurface of codimension one. At each point $p \in S$, choose a unit normal vector $n^a(p)$ satisfying

$$g_{ab}n^a n^b = \epsilon, \qquad \epsilon = \begin{cases} +1 & \text{(Riemannian or spacelike normal)}, \\ -1 & \text{(timelike normal in Lorentzian signature)}. \end{cases}$$

Step 2: Shoot out orthogonal geodesics. For each $p \in S$, consider the unique geodesic $\gamma_p(t)$ such that

$$\gamma_p(0) = p, \qquad \dot{\gamma}_p(0) = n^a(p),$$

where t is an affine parameter along γ_p. These geodesics are initially orthogonal to S.

Let us define the coordinate

$$x^0 := t$$

to be the affine parameter along these geodesics. Thus, moving in the x^0 direction corresponds to moving along a normal geodesic, away from or toward S.

Step 3: Extend coordinates from S along the geodesics. Choose coordinates $(x^1, \dots, x^{n-1})$ on the hypersurface S. To define coordinates off of S, we transport these labels along the orthogonal geodesics by holding them fixed:

$$\frac{dx^i}{dt} = 0 \quad \text{along each normal geodesic.}$$

In other words, each geodesic γ_p carries the coordinates of its starting point $p \in S$ along with it.

Altogether, this defines a coordinate system

$$(x^0, x^1, \dots, x^{n-1})$$

in some neighborhood of S.

Metric form in Gaussian normal coordinates. Let us denote the coordinate basis vectors by

$$\left(\frac{\partial}{\partial x^0}\right)^a, \qquad \left(\frac{\partial}{\partial x^i}\right)^a.$$

By construction,

- $\partial/\partial x^0$ is tangent to the normal geodesics,
- $\partial/\partial x^i$ are tangent to the hypersurfaces $x^0 = \text{constant}$,
- at $x^0 = 0$ (on S), the normal vector is $n^a = (\partial/\partial x^0)^a$.

Since the normal geodesics are everywhere orthogonal to the surfaces $x^0 = \text{constant}$, we have

$$g_{0i} = g_{ab}\left(\frac{\partial}{\partial x^0}\right)^a\left(\frac{\partial}{\partial x^i}\right)^b = 0.$$

Moreover, along each normal geodesic, the normalization of n^a is preserved:

$$g_{00} = g_{ab}\left(\frac{\partial}{\partial x^0}\right)^a\left(\frac{\partial}{\partial x^0}\right)^b = g_{ab}n^a n^b = \epsilon = \pm 1.$$

Thus, in Gaussian normal coordinates the metric takes the block form

$$g_{00} = \pm 1, \qquad g_{0i} = 0,$$

and the remaining components $g_{ij}(x^0, x^k)$ encode the intrinsic geometry of the hypersurfaces $x^0 = \text{constant}$ and their extrinsic curvature in the ambient spacetime.

Preservation of orthogonality (idea). One can check (as Wald does) that if a vector field X^a is initially tangent to S (so $n_a X^a = 0$ at $x^0 = 0$), then $n_a X^a = 0$ remains true along the normal geodesics. Equivalently, the coordinate basis vectors $\partial/\partial x^i$ stay orthogonal to n^a for all x^0. This is the geometric reason the cross terms g_{0i} vanish everywhere.

Why Gaussian normal coordinates are useful.
Gaussian normal coordinates are adapted to a hypersurface S and its orthogonal geodesics. They are fundamental for:

- splitting spacetime into "time + space" (e.g. in ADM formalism),
- analyzing the geometry of hypersurfaces (extrinsic curvature),
- setting up initial value problems in general relativity,
- studying gravitational collapse, cosmology, and singularity theorems.

In these coordinates, x^0 measures proper distance or proper time along the normals, while $(x^1, \dots, x^{n-1})$ label the "points on S" that normal geodesics emanate from. The absence of mixed terms g_{0i} reflects the fact that the coordinate grid is built from surfaces that are everywhere orthogonal to the normal geodesics.

3.5.6 Extremizing the Length Functional

Geodesics may also be characterized as curves that *extremize the length functional.* Wald states the resulting Euler–Lagrange equations, but omits the intermediate steps; we supply them here.

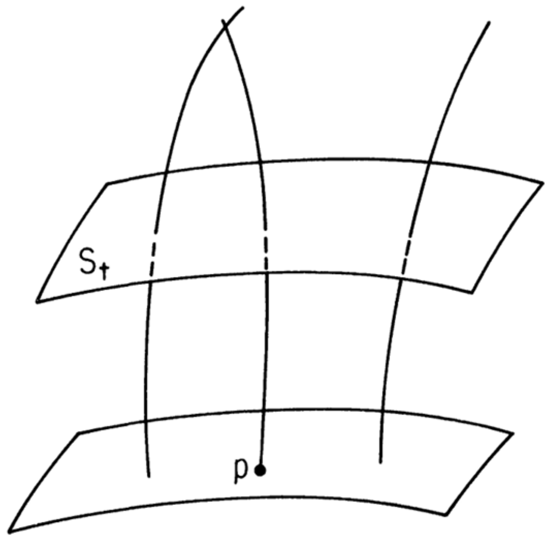

Figure 3.2: Construction of Gaussian normal coordinates from a hypersurface S: geodesics are emitted orthogonally from S with initial tangent n^a, and the hypersurface coordinates x^i are held fixed along each geodesic.

The length functional. For a smooth curve $x^\mu(t)$, define

$$L[x^\mu] = \int \sqrt{g_{\mu\nu}(x)\,\dot{x}^\mu \dot{x}^\nu}\,dt. \tag{3.3.11}$$

The Lagrangian is

$$\mathcal{L}(x,\dot{x}) = \sqrt{g_{\mu\nu}(x)\,\dot{x}^\mu \dot{x}^\nu}.$$

Euler–Lagrange equation. Extremizing L gives

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu}\right) - \frac{\partial \mathcal{L}}{\partial x^\mu} = 0, \tag{3.3.12}$$

which is Wald's equation (3.3.12).

Computing the derivatives (expanded calculation). We find

$$\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = \frac{g_{\mu\nu}\dot{x}^\nu}{\mathcal{L}}, \qquad \frac{\partial \mathcal{L}}{\partial x^\mu} = \frac{1}{2\mathcal{L}}\,\partial_\mu g_{\rho\sigma}\,\dot{x}^\rho \dot{x}^\sigma.$$

Take the total derivative:

$$\frac{d}{dt}\left(\frac{g_{\mu\nu}\dot{x}^\nu}{\mathcal{L}}\right) = \frac{1}{\mathcal{L}}\left(g_{\mu\nu}\ddot{x}^\nu + \partial_\lambda g_{\mu\nu}\dot{x}^\lambda \dot{x}^\nu\right) - \frac{\dot{\mathcal{L}}}{\mathcal{L}^2} g_{\mu\nu}\dot{x}^\nu.$$

Substituting into (3.3.12) and rearranging yields the expanded, *non-affinely parameterized* Euler–Lagrange equation:

$$\ddot{x}^\mu + \Gamma^\mu{}_{\nu\lambda}\dot{x}^\nu \dot{x}^\lambda = \frac{\dot{\mathcal{L}}}{\mathcal{L}}\,\dot{x}^\mu. \tag{3.3.13*}$$

Interpretation. The right-hand side is tangent to the curve. It reflects the freedom to reparameterize the curve: if we change $t \mapsto f(t)$, the left-hand side does not stay invariant, but the added term is always proportional to $\dot{x}^\mu$.

Affine parametrization. A parameter t is affine if $\dot{\mathcal{L}} = 0$, i.e. $\mathcal{L}$ is constant. Under this choice, the tangent term vanishes and we obtain Wald's equation (3.3.13):

$$\ddot{x}^\mu + \Gamma^\mu{}_{\nu\lambda}\dot{x}^\nu\dot{x}^\lambda = 0, \tag{3.3.13}$$

which is Wald's geodesic equation in coordinate form.

Summary. Equation (3.3.13*) is the fully general Euler–Lagrange form. Wald's (3.3.13) is the special case obtained when the parameter is affine.

3.5.7 Geodesic Deviation

The geodesic deviation equation describes how nearby geodesics accelerate toward or away from each other. This is the mathematical expression of tidal forces, and it is the first place where the curvature tensor appears in a dynamical way.

Consider a smooth one–parameter family of geodesics $\gamma_s(t)$, where t is an affine parameter and s labels the different geodesics in the family. Thus each fixed s gives a geodesic $t \mapsto \gamma_s(t)$, and varying s slides us from one geodesic to a nearby one.

Define vector fields

$$T^a := \left(\frac{\partial}{\partial t}\right)^a, \qquad X^a := \left(\frac{\partial}{\partial s}\right)^a.$$

Here,

- T^a is tangent to the geodesics (t–curves),
- X^a points from one geodesic to a nearby geodesic (the deviation vector).

Because T^a and X^a are coordinate vector fields on the (t,s) surface, they *commute*:

$$T^b\nabla_b X^a = X^b\nabla_b T^a. \tag{3.3.16}$$

Geometric Meaning of Commutativity. Both T^a and X^a arise from the (t,s) coordinate chart on the 2–dimensional surface swept out by the family of geodesics. Coordinate vector fields always commute: their Lie bracket vanishes. This expresses the fact that "moving forward in t and then in s" always leads to the same point as "moving in s and then in t".

Because each curve of fixed s is a geodesic,

$$T^b\nabla_b T^a = 0. \tag{3.3.1 revisited}$$

Relative velocity. The rate at which the separation vector X^a changes along the geodesics is the *relative velocity*:

$$v^a := T^b \nabla_b X^a. \tag{3.3.17}$$

Relative acceleration. Differentiating again along T^a gives the *relative acceleration*:

$$a^a := T^c \nabla_c v^a = T^c \nabla_c (T^b \nabla_b X^a).$$

We now compute a^a and relate it to curvature.

Using the commutation relation (3.3.16) and the geodesic equation $T^b \nabla_b T^a = 0$, Wald obtains

$$\begin{aligned} a^a &= T^c \nabla_c (T^b \nabla_b X^a) \\ &= T^c \nabla_c (X^b \nabla_b T^a) \\ &= (T^c \nabla_c X^b)(\nabla_b T^a) + X^b \, T^c \nabla_c \nabla_b T^a \\ &= X^b \, T^c \nabla_c \nabla_b T^a \\ &= -\, R^a{}_{bcd} \, X^b T^c T^d. \end{aligned} \tag{3.3.18}$$

Key Step. The Riemann tensor appears through the identity

$$(\nabla_c \nabla_b - \nabla_b \nabla_c) T^a = R^a{}_{dcb} T^d,$$

which is exactly the commutator formula for covariant derivatives acting on a vector. This is the mathematical core of geodesic deviation: the failure of second covariant derivatives to commute encodes curvature.

Thus we obtain the **geodesic deviation equation** (or Jacobi equation):

$$T^b \nabla_b \left(T^c \nabla_c X^a \right) = -\, R^a{}_{bcd} \, T^b T^d X^c. \tag{3.3.18}$$

This equation states that the relative acceleration of nearby geodesics is determined by the curvature components in the plane spanned by T^a and X^a. In particular, positive curvature tends to make geodesics focus (accelerate toward each other), while negative curvature tends to make them spread apart.

Physical Meaning. In general relativity, the geodesic deviation equation expresses tidal gravity. Two freely falling particles with initial separation X^a do not maintain constant separation: spacetime curvature causes their worldlines to accelerate relative to each other. The curvature tensor $R^a{}_{bcd}$ encodes the tidal forces produced by matter via the Einstein equation.

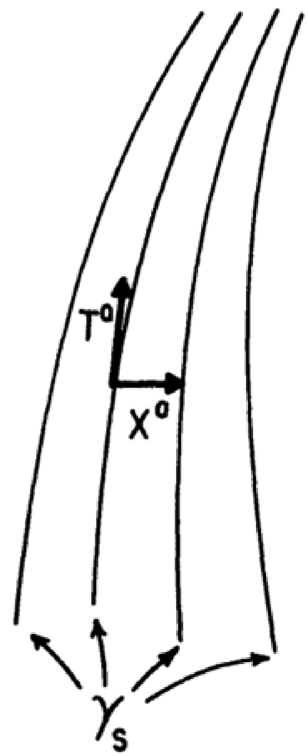

Figure 3.3: A one-parameter family of geodesics $\gamma_s(t)$ with tangent T^a and deviation vector X^a.

3.6 Methods for Computing Curvature

In Section 3.3, we *defined* the Riemann curvature tensor by proving the existence of a tensor field $R_{abc}{}^d$ satisfying the commutator formula

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\,\omega_c = R_{abc}{}^d\,\omega_d$$

for all smooth covector fields ω_a. This existence theorem is conceptually important: it tells us that curvature is the obstruction to commuting covariant derivatives. However, it does *not* tell us how to actually *compute* the components of $R_{abc}{}^d$ given a metric g_{ab}.

In general relativity, being able to compute curvature is essential: the Einstein field equation,

$$G_{ab} = 8\pi T_{ab},$$

involves the Ricci tensor R_{ab}, which is a contraction of the Riemann tensor. Thus, given a metric, one must be able to compute its associated curvature.

The purpose of this section is to develop *practical* methods for calculating curvature. Wald presents two principal approaches:

1. **Coordinate Component Method** (compute Christoffel symbols $\Gamma^a{}_{bc}$ from the metric, then apply the coordinate formula for $R_{abc}{}^d$), and

2. **Orthonormal Basis (Tetrad) Method** (introduce a non-coordinate orthonormal basis; compute curvature using connection 1-forms and the Ricci rotation coefficients).

Both methods have advantages and disadvantages. The coordinate approach is straightforward but algebraically tedious. The tetrad approach often simplifies computations, especially when the metric has symmetries, but it requires additional geometric structure and care with the commutation relations of the basis vectors.

Two Ways to Compute Curvature.

- In a *coordinate basis*, the Christoffel symbols ${\Gamma^a}_{bc}$ are complicated but the basis vectors commute.
- In an *orthonormal basis*, the connection coefficients (Ricci rotation coefficients) are simpler, but the basis vectors need not commute.

Both approaches ultimately compute the same geometric object, the Riemann tensor, but with quite different algebraic machinery.

We now develop each method carefully, filling in the steps omitted in Wald and providing geometric interpretation where helpful.

3.6.1 Coordinate Component Method

The most direct way to compute curvature is to work in an arbitrary coordinate system and express the covariant derivative in terms of the partial derivatives ∂_a and the Christoffel symbols ${\Gamma^c}_{ab}$. This method is conceptually straightforward—everything is written in coordinates—but often computationally heavy.

Let ω_a be a covector field. From the coordinate expression for the covariant derivative of a covector,

$$\nabla_b \omega_c = \partial_b \omega_c - {\Gamma^d}_{bc}\, \omega_d, \tag{3.4.1}$$

we may compute its second covariant derivative (see Eq. 3.1.14):

$$\nabla_a \nabla_b \omega_c = \partial_a (\nabla_b \omega_c) - {\Gamma^e}_{ab} \nabla_e \omega_c - {\Gamma^e}_{ac} \nabla_b \omega_e. \tag{3.4.2}$$

More detail shown in subsection 3.6.1.

Substituting Eq. 3.4.1 into Eq. 3.4.2 and expanding, one obtains a long but entirely mechanical expression. Wald does not write out every term explicitly, but the structure is clear: each ∇ introduces both a partial derivative and a Christoffel symbol.

Now insert this into the defining relation for the curvature tensor,

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)\, \omega_c = {R_{abc}}^d\, \omega_d. \tag{3.2.3 revisited}$$

After canceling terms that appear with opposite signs, and using the commutativity of the coordinate derivatives $\partial_a \partial_b = \partial_b \partial_a$, the remaining expression is

$${R_{abc}}^d\, \omega_d = \left[-2\, \partial_{[a} {\Gamma^d}_{b]c} + 2\, {\Gamma^e}_{c[a}\, {\Gamma^d}_{b]e} \right] \omega_d. \tag{3.4.3}$$

Since this holds for all ω_d, we may drop ω_d and obtain the familiar coordinate formula for the Riemann tensor:

$$R_{\mu\nu\rho}{}^{\sigma} = \partial_\nu \Gamma^{\sigma}{}_{\mu\rho} - \partial_\mu \Gamma^{\sigma}{}_{\nu\rho} + \Gamma^{\sigma}{}_{\mu\lambda}\Gamma^{\lambda}{}_{\nu\rho} - \Gamma^{\sigma}{}_{\nu\lambda}\Gamma^{\lambda}{}_{\mu\rho}. \tag{3.4.4}$$

How to compute curvature in practice. Given a metric $g_{\mu\nu}$:

1. Compute the inverse metric $g^{\mu\nu}$.
2. Compute the Christoffel symbols using

$$\Gamma^{\rho}{}_{\mu\nu} = \tfrac{1}{2} g^{\rho\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}). \tag{3.1.30 revisited}$$

3. Compute the Riemann tensor using Eq. 3.4.4.
4. Contract indices to obtain the Ricci tensor:

$$R_{\mu\rho} = R_{\mu\nu\rho}{}^{\nu}. \tag{3.4.5}$$

Geometric Meaning. The coordinate method is brute-force: it expands the commutator of covariant derivatives directly. All curvature information is encoded in how Christoffel symbols vary from point to point. If the $\Gamma^{\rho}{}_{\mu\nu}$ are constant in a coordinate chart, then all partial derivatives vanish and the curvature reduces to products of Christoffels. If the Christoffels vanish (as in Riemann normal coordinates at p), the curvature is determined entirely by their *first derivatives*.

Index-Free to Index-Full: Second Covariant Derivative

We want to understand clearly how

$$\nabla_a \nabla_b \omega_c = \partial_a(\nabla_b \omega_c) - \Gamma^{e}{}_{ab}\nabla_e \omega_c - \Gamma^{e}{}_{ac}\nabla_b \omega_e \tag{3.4.2}$$

arises from the general rules for covariant differentiation.

Step 0: Type of the objects.

- ω_a is a covector field: a $(0,1)$ tensor.
- $\nabla_b \omega_c$ is a $(0,2)$ tensor (one derivative index b and one covector index c).

The key rule is:

Covariant derivative rule (type $(0,2)$).
If T_{bc} is a $(0,2)$ tensor, then

$$\nabla_a T_{bc} = \partial_a T_{bc} - \Gamma^{e}{}_{ab} T_{ec} - \Gamma^{e}{}_{ac} T_{be}.$$

Every lower index contributes a *minus* Christoffel term with that index replaced by a dummy index e.

Step 1: First covariant derivative (type $(0,1)$). Start with the covector field ω_c. By definition of the covariant derivative of a covector (Wald eq. (3.1.12)):

$$\nabla_b \omega_c = \partial_b \omega_c - \Gamma^{d}{}_{bc}\,\omega_d. \tag{3.4.1}$$

This object $\nabla_b \omega_c$ has two lower indices (b, c), so it is a $(0,2)$ tensor.

Step 2: Recognize $\nabla_b \omega_c$ as a $(0,2)$ tensor. Define

$$T_{bc} := \nabla_b \omega_c.$$

Then T_{bc} is a general $(0,2)$ tensor, so we *must* use the $(0,2)$ covariant-derivative rule when differentiating it.

Step 3: Apply the $(0,2)$ rule. Using the general formula for $\nabla_a T_{bc}$:

$$\nabla_a T_{bc} = \partial_a T_{bc} - \Gamma^e{}_{ab} T_{ec} - \Gamma^e{}_{ac} T_{be},$$

we substitute $T_{bc} = \nabla_b \omega_c$:

$$\nabla_a(\nabla_b \omega_c) = \partial_a(\nabla_b \omega_c) - \Gamma^e{}_{ab}(\nabla_e \omega_c) - \Gamma^e{}_{ac}(\nabla_b \omega_e).$$

This is precisely

$$\nabla_a \nabla_b \omega_c = \partial_a(\nabla_b \omega_c) - \Gamma^e{}_{ab} \nabla_e \omega_c - \Gamma^e{}_{ac} \nabla_b \omega_e,$$

which is Wald's equation (3.4.2).

Step 4: The "one Γ per index" rule. Conceptually, the pattern is:

- For a $(0,1)$ tensor S_c, we have

$$\nabla_a S_c = \partial_a S_c - \Gamma^e{}_{ac} S_e \quad \Rightarrow \quad \text{one lower index} \Rightarrow \text{one } \Gamma \text{ term}.$$

- For a $(0,2)$ tensor T_{bc}, we have

$$\nabla_a T_{bc} = \partial_a T_{bc} - \Gamma^e{}_{ab} T_{ec} - \Gamma^e{}_{ac} T_{be} \quad \Rightarrow \quad \text{two lower indices} \Rightarrow \text{two } \Gamma \text{ terms}.$$

In our case, $\nabla_b \omega_c$ has two lower indices (b, c), so the second covariant derivative $\nabla_a(\nabla_b \omega_c)$ must contain *two* Christoffel correction terms: one for the b index and one for the c index.

Takeaway. Rather than tracking every symbol individually, remember: *"Covariant derivative = partial derivative + one correction term for each index."* Once you recognize the type of $\nabla_b \omega_c$ as $(0,2)$, it is automatic that $\nabla_a(\nabla_b \omega_c)$ will produce exactly two Christoffel terms, one correcting each of the two lower indices.

Determinant identities. Wald takes this opportunity to record useful formulas involving the determinant of the metric,

$$g = \det(g_{\mu\nu}), \tag{3.4.6}$$

since such expressions frequently arise when computing divergences and contracted Christoffel symbols.

From the definition of $\Gamma^\mu{}_{\mu\nu}$, one shows

$$\Gamma^\mu{}_{\mu\nu} = \frac{1}{2g}\,\partial_\nu g = \partial_\nu \ln \sqrt{|g|}. \tag{3.4.9}$$

This identity is extremely useful when simplifying divergences of vector fields:

$$\nabla_a T^a = \frac{1}{\sqrt{|g|}}\,\partial_\mu\left(\sqrt{|g|}\,T^\mu\right). \tag{3.4.10}$$

Why the Coordinate Method Can Be Hard. Although conceptually straightforward, the coordinate component method is notorious for producing long chains of partial derivatives and index contractions. Even in simple metrics (e.g. Schwarzschild, FLRW), dozens of terms may appear before simplifications occur. Errors in signs or indices are common without symbolic software.

3.6.2 Orthonormal Basis (Tetrad) Methods

Coordinate methods work directly with the components $g_{\mu\nu}$ and $\Gamma^\rho{}_{\mu\nu}$ in some chart. Often, however, it is more convenient to work in an *orthonormal basis* of vector fields, especially when the metric has a simple local form (e.g. Minkowski metric) but the coordinate components $g_{\mu\nu}$ are messy. This is the idea behind *tetrad* (or orthonormal frame) methods.

Idea. Instead of expanding all tensors in a coordinate basis $\{\partial_\mu\}$, we choose a *moving orthonormal frame* $\{(e_\mu)^a\}$ so that in this basis the metric is locally $\eta_{\mu\nu} = \mathrm{diag}(-1,\dots,1)$. All geometric information then sits in how the frame *rotates* from point to point, encoded in the *connection 1-forms* (Ricci rotation coefficients).

Orthonormal Frames and Tetrads

A *tetrad* (in n dimensions, more generally an orthonormal frame) is a collection of smooth vector fields

$$\{(e_\mu)^a\}, \qquad \mu = 0, 1, \dots, n-1,$$

such that at every point

$$(e_\mu)^a (e_\nu)_a = \eta_{\mu\nu}, \tag{3.4.11}$$

where $(e_\nu)_a := g_{ab}(e_\nu)^b$ and $\eta_{\mu\nu} = \mathrm{diag}(-1,\dots,1)$ is the flat Minkowski (or Euclidean) metric in this frame.

Equation (3.4.11) says that the tetrad vectors form an orthonormal basis at each point, with respect to g_{ab}. From (3.4.11) it follows that

$$\sum_{\mu,\nu} \eta^{\mu\nu} (e_\mu)^a (e_\nu)_b = \delta^a{}_b, \tag{3.4.12}$$

where $\eta^{\mu\nu}$ is the inverse matrix to $\eta_{\mu\nu}$. Thus $\{(e_\mu)^a\}$ is a basis and $(e_\mu)_a$ the dual coframe.

Any tensor may be expanded in this orthonormal basis. For example, a vector V^a has frame components

$$V^\mu := (e^\mu)_a V^a, \qquad V^a = V^\mu (e_\mu)^a,$$

where $(e^\mu)_a$ is the dual coframe satisfying $(e^\mu)_a (e_\nu)^a = \delta^\mu{}_\nu$.

Connection 1-Forms and Ricci Rotation Coefficients

In a coordinate basis the connection is encoded by the Christoffel symbols $\Gamma^c{}_{ab}$. In an orthonormal basis, it is more natural to encode the connection through the *connection 1-forms*

$$\omega_{a\mu\nu} := (e_\mu)^b \nabla_a (e_\nu)_b. \tag{3.4.13}$$

The components of these 1-forms in the tetrad basis,

$$\omega_{\lambda\mu\nu} := (e_\lambda)^a \omega_{a\mu\nu} = (e_\lambda)^a (e_\mu)^b \nabla_a (e_\nu)_b, \tag{3.4.14}$$

are called the *Ricci rotation coefficients*.

Metric compatibility. Using orthonormality (3.4.11) and $\nabla_a g_{bc} = 0$ we get

$$0 = \nabla_a \big[(e_\mu)^b (e_\nu)_b\big] = \omega_{a\mu\nu} + \omega_{a\nu\mu}.$$

Thus the connection 1-forms are antisymmetric in $\mu\nu$:

$$\omega_{a\mu\nu} = -\,\omega_{a\nu\mu}, \tag{3.4.15}$$

and similarly for the Ricci rotation coefficients,

$$\omega_{\lambda\mu\nu} = -\,\omega_{\lambda\nu\mu}. \tag{3.4.16}$$

Geometric Picture. The connection 1-forms tell you how the orthonormal frame "rotates" as you move around the manifold. At each point, the frame looks like a local Minkowski (or Euclidean) basis. Curvature then appears as the failure of these local frames to fit together consistently across finite regions.

Curvature in an Orthonormal Basis

We can express the components of the Riemann tensor in the tetrad basis as

$$R_{\rho\sigma\mu\nu} := R_{abcd} (e_\rho)^a (e_\sigma)^b (e_\mu)^c (e_\nu)^d. \tag{3.4.17}$$

Using the definition of curvature as the commutator of covariant derivatives and the connection 1-forms (3.4.13), Wald shows that

$$R_{\rho\sigma\mu\nu} = (e_\rho)^a (e_\sigma)^b \big(\nabla_a \omega_{b\mu\nu} - \nabla_b \omega_{a\mu\nu}\big) - \sum_{\alpha,\beta} \eta^{\alpha\beta} \big(\omega_{\rho\mu\alpha}\omega_{\sigma\nu\beta} - \omega_{\rho\nu\alpha}\omega_{\sigma\mu\beta}\big). \tag{3.4.20}$$

Equivalently, in a slightly more compact index arrangement,

$$R_{\rho\sigma\mu\nu} = (e_\rho)^a \nabla_a \omega_{\sigma\mu\nu} - (e_\sigma)^a \nabla_a \omega_{\rho\mu\nu} - \sum_{\alpha,\beta} \eta^{\alpha\beta} \big(\omega_{\rho\mu\alpha}\omega_{\sigma\nu\beta} - \omega_{\sigma\mu\alpha}\omega_{\rho\nu\beta}\big). \tag{3.4.21}$$

Ricci tensor in a tetrad. Once $R_{\rho\sigma\mu\nu}$ is known, the Ricci tensor components in the orthonormal frame are simply the contraction

$$R_{\mu\nu} = \sum_{\rho,\sigma} \eta^{\rho\sigma} R_{\rho\mu\sigma\nu}. \tag{3.4.22}$$

Why the Tetrad Method is Useful. In n dimensions there are $n^2(n+1)/2$ independent Christoffel symbols $\Gamma^{\rho}{}_{\mu\nu}$, but only $n(n-1)(n)/2$ independent Ricci rotation coefficients $\omega_{\lambda\mu\nu}$, thanks to the antisymmetry in $\mu\nu$. For example, in 4 dimensions there are 40 independent $\Gamma^{\rho}{}_{\mu\nu}$ but only 24 independent $\omega_{\lambda\mu\nu}$. Moreover, $\eta_{\mu\nu}$ is constant in the orthonormal frame, so the metric itself does not clutter the calculations.

Torsion-Free Condition in a Tetrad

To fully determine the connection 1-forms, we must encode the fact that our derivative operator is torsion-free. Wald gives two equivalent forms of this condition.

(i) Commutation relations of basis vectors. From the general torsion-free condition

$$\nabla_a X_b - \nabla_b X_a = [X, Y]_a$$

(see equation (3.1.2)), applied to the tetrad vectors $(e_\mu)^a$, one obtains

$$(e_\sigma)_a [e_\mu, e_\nu]^a = \omega_{\mu\nu\sigma} - \omega_{\nu\mu\sigma}. \tag{3.4.23}$$

Thus the antisymmetric part of the Ricci rotation coefficients is fixed by the commutators of the tetrad basis vectors.

(ii) Antisymmetrized derivative of the coframe. From the definition of $\omega_{a\mu\nu}$ one also finds

$$\nabla_{[a}(e_\sigma)_{b]} = \sum_{\mu,\nu} \eta^{\mu\nu} (e_\mu)_{[a}\, \omega_{b]\sigma\nu}. \tag{3.4.24}$$

But torsion-freedom implies that the antisymmetrized derivative $\nabla_{[a}(e_\sigma)_{b]}$ is independent of the connection; we may therefore replace ∇ by ∂:

$$\partial_{[a}(e_\sigma)_{b]} = \sum_{\mu,\nu} \eta^{\mu\nu} (e_\mu)_{[a}\, \omega_{b]\sigma\nu}. \tag{3.4.25}$$

Conversely, if (3.4.25) holds for all basis vectors, the connection is torsion-free.

Practical Strategy. In tetrad calculations one typically:

1. Chooses an orthonormal coframe $(e_\mu)_a$ adapted to the symmetries of the problem.
2. Uses (3.4.25) (or the commutators (3.4.23)) to solve for the $\omega_{\lambda\mu\nu}$.
3. Substitutes these into (3.4.21) to obtain $R_{\rho\sigma\mu\nu}$ and hence the Ricci tensor via (3.4.22).

Often the coframe can be chosen so cleverly that step (2) can almost be done by inspection.

Differential-Forms Formulation

Finally, Wald recasts the tetrad method in the language of differential forms, which is especially compact.

Let $\{\mathbf{e}_\mu\}$ denote the coframe 1-forms

$$\mathbf{e}_\mu := (e_\mu)_a \, dx^a,$$

and let $\omega_\mu{}^\nu$ be the connection 1-forms,

$$\omega_\mu{}^\nu := \omega_{a\mu}{}^\nu \, dx^a, \qquad \omega_{\mu\nu} = \eta_{\nu\sigma} \, \omega_\mu{}^\sigma,$$

so that $\omega_{\mu\nu} = -\omega_{\nu\mu}$.

Then the torsion-free condition (3.4.25) can be written compactly as

$$d\mathbf{e}_\sigma = \sum_\mu \mathbf{e}_\mu \wedge \omega^\mu{}_\sigma. \tag{3.4.27}$$

The curvature 2-forms

$$\mathbf{R}_\mu{}^\nu := \tfrac{1}{2} R_\mu{}^\nu{}_{\rho\sigma} \, \mathbf{e}^\rho \wedge \mathbf{e}^\sigma$$

are obtained from the connection 1-forms by

$$\mathbf{R}_\mu{}^\nu = d\omega_\mu{}^\nu + \sum_\alpha \omega_\mu{}^\alpha \wedge \omega_\alpha{}^\nu. \tag{3.4.28}$$

Equations (3.4.27) and (3.4.28) are sometimes called the *Cartan structure equations*.

Summary of Methods. We now have two main computational strategies:

- **Coordinate basis:** start from $g_{\mu\nu}$, compute $\Gamma^\rho{}_{\mu\nu}$ and then $R_{\mu\nu\rho}{}^\sigma$ via Eq 3.4.4.
- **Orthonormal (tetrad) basis:** choose an orthonormal coframe, solve for the connection 1-forms from torsion-free conditions Eq. 3.4.23 or Eq. 3.4.25, then obtain curvature from the Cartan equations Eq. 3.4.27–Eq. 3.4.28.

In highly symmetric spacetimes, a clever tetrad choice often makes the second method far more efficient than brute-force coordinate computation.

Problems

1. Let property (5) (the “torsion free” condition) be dropped from the definition of derivative operator ∇_a in section 3.1.
   (a) Show that there exists a tensor $T^c{}_{ab}$ (called the torsion tensor) such that for all smooth functions f, we have
   $$\nabla_a \nabla_b f - \nabla_b \nabla_a f = -T^c{}_{ab} \nabla_c f.$$
   (Hint: repeat the derivation of eq. (3.1.8), letting ∇_a be a torsion-free derivative operator.)
   (b) Show that for any smooth vector fields X^a, Y^a we have
   $$T^c{}_{ab} X^a Y^b = X^a \nabla_a Y^c - Y^a \nabla_a X^c - [X, Y]^c.$$
   (c) Given a metric g_{ab}, show that there exists a unique derivative operator ∇_a with torsion $T^c{}_{ab}$ such that $\nabla_c g_{ab} = 0$. Derive the analog of equation (3.1.29), expressing this derivative operator in terms of an ordinary derivative ∂_a and $T^c{}_{ab}$.

2. Let M be a manifold with metric g_{ab} and associated derivative operator ∇_a. A solution of the equation $\nabla^a \nabla_a \alpha = 0$ is called a harmonic function. In the case where M is a two-dimensional manifold, let α be harmonic and let ϵ_{ab} be an antisymmetric tensor field satisfying $\epsilon_{ab}\epsilon^{ab} = 2(-1)^s$, where s is the number of minuses occurring in the signature of the metric. Consider the equation

$$\nabla_a \beta = \epsilon_{ab} \nabla^b \alpha.$$

   (a) Show that the integrability conditions (see problem 5 of chapter 2 or appendix B) for this equation are satisfied, and thus, locally, there exists a solution β. Show that β is also harmonic, $\nabla^a \nabla_a \beta = 0$. ($\beta$ is called the harmonic function conjugate to α.)
   (b) By choosing α and β as coordinates, show that the metric takes the form

$$ds^2 = \pm \Omega^2(\alpha, \beta) \left[d\alpha^2 + (-1)^s d\beta^2 \right].$$

3. (a) Show that $R_{abcd} = R_{cdab}$.
   (b) In n dimensions, the Riemann tensor has n^4 components. However, on account of the symmetries (3.2.13), (3.2.14), and (3.2.15), not all of these components are independent. Show that the number of independent components is $n^2(n^2 - 1)/12$.
4. (a) Show that in two dimensions, the Riemann tensor takes the form

$$R_{abcd} = R\, g_{a[c} g_{d]b}.$$

   (Hint: use the result of problem 3(b) to show that $g_{a[c} g_{d]b}$ spans the vector space of tensors having the symmetries of the Riemann tensor.)
   (b) By similar arguments, show that in three dimensions the Weyl tensor vanishes identically; i.e., for $n = 3$, equation (3.2.28) holds with $C_{abcd} = 0$.
5. (a) Show that any curve whose tangent satisfies equation (3.3.2) can be reparameterized so that equation (3.3.1) is satisfied.
   (b) Let t be an affine parameter of a geodesic γ. Show that all other affine parameters of γ take the form $at + b$, where a and b are constants.
6. The metric of Euclidean $\mathbb{R}^3$ in spherical coordinates is

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$

   (see problem 8 of chapter 2).
   (a) Calculate the Christoffel components $\Gamma^{\sigma}{}_{\mu\nu}$ in this coordinate system.
   (b) Write down the components of the geodesic equation in this coordinate system and verify that the solutions correspond to straight lines in Cartesian coordinates.
7. As shown in problem 2, an arbitrary Lorentz metric on a two-dimensional manifold locally always can be put in the form

$$ds^2 = \Omega^2(x, t) \left[-dt^2 + dx^2 \right].$$

   Calculate the Riemann curvature tensor of this metric (a) by the coordinate basis methods of section 3.4a, and (b) by the tetrad methods of section 3.4b.
8. Using the antisymmetry of $\omega_{\lambda\mu\nu}$ in μ and ν, equation (3.4.15), show that

$$\omega_{\lambda\mu\nu} = 3\, \omega_{[\lambda\mu\nu]} - 2\, \omega_{[\nu\mu]\lambda}.$$

   Use this formula together with equation (3.4.23) to solve for $\omega_{\lambda\mu\nu}$ in terms of commutators (or antisymmetrized derivatives) of the orthonormal basis vectors.

Chapter 4

Einstein's Equations

In this chapter we develop a mathematically precise formulation of the ideas behind general relativity. We begin by reviewing the geometry of space in prereleativity physics, then discuss special relativity, and finally introduce spacetime and Einstein's equation. Throughout, we emphasize the tensorial formulation of physical laws and the principles of *general* and *special covariance.*

4.1 The Geometry of Space in Prerelativity Physics: General and Special Covariance

Before relativity, physics assumed that physical space is the manifold $\mathbb{R}^3$. One further assumes that points in space may be labeled by Cartesian coordinates (x^1, x^2, x^3), obtained by constructing a "rigid rectilinear grid" of meter sticks. Although many Cartesian coordinate systems are possible—related by rotations and translations—the *distance* between points is coordinate-invariant.

If two points have coordinate values x^μ and $\bar{x}^\mu$, the Euclidean distance D between them is

$$D^2 = (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2. \tag{4.1.1}$$

For "nearby" points, this becomes

$$(\delta D)^2 = (\delta x^1)^2 + (\delta x^2)^2 + (\delta x^3)^2. \tag{4.1.2}$$

suggesting that the metric of space is

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \tag{4.1.3}$$

or, in index notation,

$$h_{ab} = \sum_{\mu,\nu} h_{\mu\nu} (dx^\mu)_a (dx^\nu)_b, \qquad h_{\mu\nu} = \mathrm{diag}(1, 1, 1). \tag{4.1.4}$$

To get ds^2 from the Eq. 4.1.4, for a small displacement from a point $x^\mu \to x^\mu + dx^\mu$, think of a tangent vector v^a whose components in these coordinates are $v^\mu = dx^\mu$,

$$\begin{aligned} ds^2 &= h_{ab}v^a v^b = h_{\mu\nu}(dx^\mu)_a(dx^\nu)_b v^a v^b \\ &= h_{ab}v^a v^b = h_{\mu\nu}(dx^\mu)_a v^a (dx^\nu)_b v^b \\ &= h_{\mu\nu}v^\mu v^\mu = h_{\mu\nu}dx^\mu dx^\nu \end{aligned}$$

Since the components $h_{\mu\nu}$ are constant in Cartesian coordinates, we have

$$\partial_\alpha h_{\mu\nu} = 0. \tag{4.1.5}$$

Thus the Christoffel symbols in this coordinate system vanish:

$$\Gamma^\mu{}_{\nu\lambda} = \frac{1}{2}h^{\mu\nu}(\partial_\nu h_{\sigma\lambda} + \partial_\lambda h_{\sigma\nu} - \partial_\sigma h_{\nu\lambda}) = \frac{1}{2}h^{\mu\nu}(0+0+0) = 0$$

$$\Gamma^\mu{}_{\nu\lambda} = 0.$$

The associated derivative operator is simply the ordinary partial derivative, covariant derivatives commute, and the curvature tensor is identically zero. Hence h_{ab} is a flat Riemannian metric.

Geometric Meaning. A vanishing Riemann tensor implies that the manifold is flat. Geodesics become exactly straight lines in Cartesian coordinates. Thus the geometric structure described by h_{ab} reproduces standard Euclidean space.

Because geodesics remain straight lines and initially parallel geodesics remain parallel, one may construct rigid grids of meter sticks across space. Consequently, the fundamental assumption of prereleativity physics is equivalent to the statement:

Space is the manifold $\mathbb{R}^3$ equipped with a flat Riemannian metric.

Tensorial Nature of Physical Quantities

Although every physical measurement yields a number, many physical quantities (e.g. electromagnetic fields, stress tensors) cannot be expressed meaningfully as numbers without choosing basis vectors. These quantities are naturally represented as tensor fields.

Why tensors? Any analytic map from vectors and dual vectors to numbers can be expanded as a sum of multilinear maps. Thus the class of tensor fields is sufficiently broad to describe every measurable physical quantity in prereleativity, special relativity, and general relativity.

Therefore the laws of physics must be expressible as *tensor equations*—equalities between tensor fields that hold independently of coordinates.

General Covariance

The central principle is:

The laws of physics must refer only to geometric structures intrinsic to space.

In prereleativity physics this means:

- the only geometric structure associated with space is the metric h_{ab},
- there are no preferred vector fields,
- the laws of physics must hold in all coordinate systems.

Coordinate expressions that fail to include the geometric structure (e.g. writing equations using only partial derivatives rather than covariant derivatives) may *appear* non-tensorial, but this reflects an incomplete formulation rather than a violation of covariance.

A key implication is that a Christoffel symbol ${\Gamma^c}_{ab}$ may *never* appear by itself (i.e. outside of a covariant derivative) in any physical law, since it is not a tensor.

Special Covariance

The metric h_{ab} in Eq. 4.1.4 has a six-parameter isometry group (rotations and translations). The principle of *special covariance* states:

If two observers are related by an isometry of space, they must agree on all physical measurements.

Thus special covariance expresses invariance under the subgroup of coordinate transformations corresponding to isometries, whereas general covariance concerns invariance under arbitrary coordinate transformations.

Special covariance will play a central role in the development of special relativity and in motivating the dynamical laws for fields.

Note: *This is not referring to two observers in relative motion (more on that later). It refers to observers measuring the same space with coordinates that are isometries of each other (think a cartesian coordinate system rotated - the rotated version is an isometry of the non-rotated version).*

4.2 Special Relativity

4.2.1 Overview and Inertial Frames

Special relativity begins with the assumption that spacetime has the structure of $\mathbb{R}^4$. A global coordinate system $\{x^\mu\} = (t, x^1, x^2, x^3)$ is introduced by a family of special observers—the inertial observers—who set up synchronized clocks and a rigid rectilinear grid of meter sticks. Such a coordinate system is called a *global inertial coordinate system*.

Different choices of global inertial coordinates are possible: any two such coordinate systems are related by an element of the 10-parameter Poincaré group (rotations, translations, boosts, and a discrete parity transformation). Thus, the numerical values of (t, x^1, x^2, x^3) at an event have no intrinsic meaning. Nevertheless, these observers agree on one structure: the *spacetime interval* between two events.

Given two events x and $\bar{x}$, the spacetime interval I is defined by

$$I = -(x^0 - \bar{x}^0)^2 + (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2, \tag{4.2.1}$$

where we have chosen units in which $c = 1$. All global inertial observers compute the same value of I; thus I is an intrinsic geometric property of spacetime.

The geometric meaning of the interval. Equation 4.2.1 generalizes the Euclidean distance formula but with a crucial sign difference. The minus sign in front of the time component encodes the causal structure of spacetime: timelike, null, and spacelike separations are distinguished by the sign of I. This structure is preserved under Poincaré transformations.

The spacetime interval suggests introducing the *metric of spacetime* η_{ab}. In the coordinate basis associated with any global inertial coordinate system, we define

$$\eta_{ab} = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu}\,(dx^\mu)_a (dx^\nu)_b, \tag{4.2.2}$$

with components

$$\eta_{\mu\nu} = \mathrm{diag}(-1,\, 1,\, 1,\, 1).$$

The tensor field η_{ab} so defined is independent of the particular choice of inertial coordinates. Furthermore, because the components $\eta_{\mu\nu}$ are constant, the ordinary derivative operator ∂_a satisfies

$$\partial_a \eta_{bc} = 0, \tag{4.2.3}$$

and is therefore the derivative operator compatible with η_{ab}.

Since ordinary derivatives commute, Eq. 4.2.3 implies that the curvature of η_{ab} vanishes. Thus, special relativity models spacetime as a flat Lorentzian manifold $(\mathbb{R}^4, \eta_{ab})$.

Flat Lorentzian geometry. The metric η_{ab} defines all standard geometric notions of special relativity: distances, angles between vectors, classification of curves as timelike, spacelike, or null, and the causal light cone structure. All inertial observers agree on these geometric features.

Consequently, the central geometric assertion of special relativity can be summarized succinctly:

Spacetime is the manifold $\mathbb{R}^4$ equipped with a flat, Lorentz-signature metric η_{ab}.

Conversely, beginning with this assumption alone, one can reconstruct the entire framework of special relativity: the existence of global inertial coordinates, the form Eq. 4.2.1 of the interval, and the interpretation of inertial motion as straight lines (i.e., geodesics) in Minkowski spacetime.

4.2.2 Geometry of Minkowski Spacetime

In special relativity, the flat Lorentzian metric η_{ab} introduced in Eq. 4.2.2 governs all geometric properties of spacetime. Since its components are constant in any global inertial coordinate system, the curvature associated with η_{ab} vanishes. Consequently, the geodesics of η_{ab}—i.e. the curves satisfying

$$T^a \partial_a T^b = 0,$$

where T^a is the tangent—are precisely the straight lines in global inertial coordinates.

Interpretation. A geodesic of η_{ab} is the worldline of a freely falling particle in special relativity. Thus, straight-line motion in (t, x^i) coordinates corresponds to force-free motion, reflecting the homogeneity of Minkowski spacetime.

Since η_{ab} has signature $(-,+,+,+)$, it allows us to classify tangent vectors and curves.

Timelike, null, and spacelike curves. Let T^a be the tangent to a curve γ. We define:

$$\begin{aligned} \eta_{ab} T^a T^b &< 0 \quad \text{timelike}, \\ \eta_{ab} T^a T^b &= 0 \quad \text{null}, \\ \eta_{ab} T^a T^b &> 0 \quad \text{spacelike}. \end{aligned}$$

Timelike curves represent possible worldlines of particles with nonzero rest mass. Null curves represent the propagation of light (photons). Spacelike curves correspond to motions exceeding the speed of light and have no physical realization.

Note: *A metric product with two vectors like, $\eta_{ab} T^a T^b$, gives the inner product between the vectors — the spacetime version of a dot product, but with signature $(-,+,+,+)$. If this is less than 0, it means the motion is mostly through time. If it is greater than 0, the motion is mostly through space. If it is zero, it is right on the edge of the light cone.*

Light cones. At each event, the null vectors of η_{ab} form a double cone: the *future* and *past* light cones. This structure encodes causality. All observers agree on the light cone structure because η_{ab} is the same tensor field for all inertial frames.

Proper Time

Given a timelike worldline with tangent T^a, we define the *proper time* τ by

$$\tau = \int \sqrt{-\eta_{ab} T^a T^b}\, dt, \tag{4.2.4}$$

where t is any parameter that increases toward the future. This quantity represents the time measured by a physical clock carried along the curve.

Different parameterizations generally give different integrals in Eq. 4.2.4, but the resulting value of τ is invariant: it depends only on the geometric curve itself. Thus, proper time is a geometric length for timelike curves.

Proper Time in Special Relativity
Let $t = x^0$ be the coordinate time of an inertial observer, and write

$$x^a(t) = (t, \vec{x}(t)), \qquad T^a = \frac{dx^a}{dt} = (1, \vec{v}).$$

Using the Minkowski metric $\eta_{ab} = \mathrm{diag}(-1, +1, +1, +1)$, the squared norm of T^a is

$$\eta_{ab} T^a T^b = -(T^0)^2 + (T^1)^2 + (T^2)^2 + (T^3)^2 = -1 + |\vec{v}|^2.$$

Therefore the proper time along the worldline is

$$\tau \;=\; \int \sqrt{-\eta_{ab} T^a T^b}\, dt \;=\; \int \sqrt{1 - |\vec{v}|^2}\; dt.$$

Differentiating gives the infinitesimal relation

$$d\tau = \sqrt{1 - |\vec{v}|^2}\; dt,$$

which is the standard **special relativity** time-dilation formula.

4-Velocity and Lorentz Factor in Special Relativity
Let $t = x^0$ be the coordinate time of an inertial observer, and write

$$x^a(t) = (t, \vec{x}(t)), \qquad T^a = \frac{dx^a}{dt} = (1, \vec{v}),$$

where $\vec{v} = d\vec{x}/dt$ is the *ordinary* 3-velocity.
From the proper-time relation,

$$d\tau = dt\sqrt{1 - |\vec{v}|^2}, \qquad \frac{dt}{d\tau} = \frac{1}{\sqrt{1 - |\vec{v}|^2}} \equiv \gamma.$$

Define the 4-velocity using proper time:

$$U^a = \frac{dx^a}{d\tau} = \frac{dx^a}{dt}\frac{dt}{d\tau} = \gamma(1, \vec{v}).$$

Its Minkowski norm is

$$\eta_{ab}U^aU^b = \gamma^2(-1 + |\vec{v}|^2) = -1,$$

so U^a is automatically unit timelike.
Thus the standard special relativity result is

$$\boxed{U^a = \gamma(1, \vec{v}), \qquad \gamma = \frac{1}{\sqrt{1 - |\vec{v}|^2}}}.$$

Note 4.1. **Proper time**, τ, is local and physical. Measured by your clock. It uses the locally flat metric.
Coordinate time, t, is global and mathematical. It is defined by someone else. It can be influenced by curvature and coordinate choices.
$t = \tau$ happens only in flat spacetime (when $\vec{v} = 0$) or in a shared inertial reference frame.

Four-Velocity

Using proper time τ as a parameter along a timelike curve, we define the *four-velocity*:

$$u^a = \frac{dx^a}{d\tau}.$$

By construction,

$$u^a u_a = -1, \tag{4.2.5}$$

reflecting the normalization of the tangent with respect to η_{ab}.

A freely falling particle experiences no force; hence its worldline is a geodesic. Expressed in inertial coordinates, this yields

$$u^a\,\partial_a u^b = 0. \tag{4.2.6}$$

Physical meaning. Equation 4.2.6 states that the components of u^b are constant in global inertial coordinates: free particles move with constant velocity unless acted upon by external forces—the special relativistic form of Newton's first law.

Energy and Momentum

All material particles possess a rest mass m. The *four-momentum* is defined by

$$p^a = mu^a. \tag{4.2.7}$$

For an observer with four-velocity v^a, the *energy* of the particle measured by that observer is

$$E = -p_a v^a. \tag{4.2.8}$$

If the observer is comoving with the particle ($v^a = u^a$), then Eq. 4.2.8 reduces to

$$E = m,$$

the rest energy (in units with $c = 1$), i.e. the relativistic version of mc^2.

Interpretation. Energy is the "time component" of four-momentum relative to the observer's frame. Because η_{ab} is flat and parallel transport is path independent, the concept of energy is globally well-defined for inertial observers.

Energy and Momentum in Special Relativity
For a particle of rest mass m, the four-momentum is

$$p^a = mU^a, \qquad U^a = \gamma(1, \vec{v}), \qquad \gamma = \frac{1}{\sqrt{1 - |\vec{v}|^2}}.$$

Thus

$$p^a = m\gamma(1, \vec{v}) = (\, m\gamma,\ m\gamma\vec{v}\,).$$

For an inertial observer at rest in these coordinates, the energy is the negative inner product with the observer's four-velocity $v^a = (1, 0, 0, 0)$:

$$E = -p_a v^a = p^0 = m\gamma.$$

The spatial components give the relativistic momentum,

$$\vec{p} = m\gamma\, \vec{v}.$$

Therefore the standard **special relativity** relations are

$$\boxed{E = m\gamma, \qquad \vec{p} = m\gamma\vec{v}, \qquad p^a = (E, \vec{p}).}$$

These satisfy the invariant mass-shell condition

$$\boxed{E^2 - |\vec{p}|^2 = m^2,}$$

which follows from $p^a p_a = -m^2$.

4.2.3 Stress–Energy and Matter in Special Relativity

In special relativity, the physical content of matter is encoded in a symmetric rank-2 tensor field T_{ab}, called the *stress–energy–momentum tensor*. This tensor incorporates the energy density, momentum density, and internal stresses of continuous matter distributions. For any observer with four-velocity v^a, the quantity

$$T_{ab} v^a v^b \geq 0 \tag{4.2.9}$$

represents the mass–energy density measured by that observer. For ordinary matter, this quantity is nonnegative; this is the flat-spacetime version of the *energy condition.*

If x^a is a vector orthogonal to v^a, then the component $-T_{ab}v^a x^b$ is interpreted as the momentum density in the x^a direction. Similarly, if y^a is another vector orthogonal to v^a, the quantity $T_{ab}x^a y^b$ represents a spatial stress component.

Interpretation. The components of T_{ab} with one index contracted into v^a give densities and fluxes as measured in the observer's rest frame. The fully spatial components encode internal stresses (pressure, shear). Thus T_{ab} organizes all mechanical properties of matter into one covariant object.

Why the Stress–Energy Tensor Appears in Special Relativity
In special relativity, any matter distribution has three familiar quantities:

- energy density,
- momentum density,
- stresses (pressure and shear).

In different inertial frames these quantities mix under Lorentz transformations: energy can appear as momentum, stresses can appear as energy flux, etc. The only mathematical object that transforms correctly to keep track of all these densities and fluxes in *every* frame is a symmetric rank-2 tensor, the stress–energy tensor T_{ab}.
Contracting T_{ab} with the observer's four-velocity v^a projects out the physical quantities measured in that observer's rest frame: $T_{ab}v^a v^b$ gives energy density, $-T_{ab}v^a x^b$ gives momentum density, and $T_{ab}x^a y^b$ gives spatial stresses.

Perfect Fluids

A *perfect fluid* is an idealized form of matter with no viscosity, no heat conduction, and no anisotropic stresses. In the instantaneous rest frame of a fluid element, the matter must therefore appear isotropic:

- its energy density is a scalar ρ,
- its pressure P acts equally in all spatial directions.

Let u^a be the fluid's unit timelike four-velocity. In the rest frame of the fluid, $u^a = (1, 0, 0, 0)$, and the symmetry assumptions imply that there is

- no momentum flow ($T_{0i} = 0$),
- equal pressure in every spatial direction ($T_{ij} = P\delta_{ij}$).

Thus the stress–energy tensor in the fluid rest frame must be

$$(T_{ab})_{\text{rest}} = \begin{pmatrix} \rho & 0 \\ 0 & P\delta_{ij} \end{pmatrix},$$

a diagonal matrix with energy density in the time–time component and pressure in the spatial components.

To write this in a Lorentz-invariant form valid in *any* frame, we proceed systematically. The only geometric ingredients available for constructing a symmetric rank-2 tensor are the metric η_{ab} and the fluid four-velocity u^a. Isotropy forbids introducing any preferred spatial direction, so the most general form compatible with symmetry is

$$T_{ab} = A\, u_a u_b + B\, \eta_{ab},$$

for some scalars A and B to be determined.

Evaluating this ansatz in the fluid rest frame:

$$T_{00} = A - B = \rho, \qquad T_{ij} = B\, \delta_{ij} = P\, \delta_{ij}.$$

Hence $B = P$ and $A = \rho + P$, giving

$$T_{ab} = (\rho + P) u_a u_b + P\, \eta_{ab}.$$

It is traditional to introduce the *spatial projector*

$$h_{ab} = \eta_{ab} + u_a u_b,$$

which satisfies $h_{ab} u^b = 0$ and reduces in the rest frame to $h_{ij} = \delta_{ij}$, $h_{00} = h_{0i} = 0$. Using h_{ab}, the perfect-fluid tensor takes the compact form

$$T_{ab} = \rho\, u_a u_b + P\, h_{ab} = \rho\, u_a u_b + P(\eta_{ab} + u_a u_b). \tag{4.2.10}$$

Meaning of the terms. The factor $\rho\, u_a u_b$ represents the rest-frame energy density carried along the fluid worldlines. The term $P(\eta_{ab} + u_a u_b)$ projects onto the spatial directions orthogonal to u^a and encodes isotropic pressure acting equally in all directions. The adjective "perfect" indicates the absence of viscosity, heat flow, and anisotropic stresses.

Equations of Motion

In flat spacetime, the conservation of stress–energy takes its simplest form:

$$\partial^a T_{ab} = 0. \tag{4.2.11}$$

This single tensor equation encodes both energy conservation and momentum conservation.

Writing Eq. 4.2.11 in terms of the variables ρ, P, and projecting parallel to u^b,

$$\begin{aligned}
0 = \partial^a T_{ab} &= \partial^a[(\rho + P) u_a u_b + P\eta_{ab}] \\
&= \partial^a[(\rho + P) u_a u_b] + \partial_a[P\eta_{ab}] \\
&= \partial^a(\rho + P) u_a u_b + (\rho + P)\partial^a u_a u_b + \partial^a(P\eta_{ab}) \\
&= \partial^a(\rho + P) u_a u_b + (\rho + P)\partial^a u_a u_b + \eta_{ab}\partial^a P \\
&= \partial^a(\rho + P) u_a u_b + (\rho + P)\partial^a u_a u_b + \partial_b P \\
&= \partial^a(\rho + P) u_a u_b + (\rho + P)[(\partial^a u_a) u_b + u_a(\partial^a u_b)] + \partial_b P
\end{aligned}$$

Now we contract with u^b using $u_b u^b = -1$, go term by term;

- $u_b\partial^a(\rho+P)u_a u_b = \partial^a(\rho+P)u_a u^b u_b = -\partial^a(\rho+P)u_a$
- $u^b(\rho+P)(\partial^a u_a)u_b = (\rho+P)(\partial^a u_a)u^b u_b = -(\rho+P)(\partial^a u_a)$
- $u^b(p+P)u_a(\partial^a u_b) = (p+P)u_a(u^b\partial^a u_b) = 0$
- $u^b\partial_b P$

Put it all together,

$$\begin{aligned} & -\partial^a(\rho+P)u_a - (\rho+P)(\partial^a u_a) + u^b\partial_b P \\ =& -\partial^a(\rho)u^a - \partial^a(P)u^a - \rho(\partial^a u_a) - P(\partial^a u_a) + u^b\partial_b P \\ =& -\partial^a(\rho)u^a - \rho(\partial^a u_a) - P(\partial^a u_a) \\ =& -[u^a\partial^a\rho + (\rho+P)\partial^a u_a] = 0 \end{aligned}$$

which is our first equation,

$$u^a\partial^a\rho + (\rho+P)\partial^a u_a = 0$$

Writing Eq. 4.2.11 in terms of the variables ρ, P, u^a,and u^b, and projecting orthogonal to u^b, and using using the spatial projector,

$$h^c_b = \delta^c_b + u^c u_b \quad \text{(equivalently} \quad h_{ab} = \eta_{ab} + u_a u_b),$$

and that $h^c_b u^b = 0$,

$$\begin{aligned} 0 &= h^c_b[\partial^a(\rho+P)u_a u_b + (\rho+P)[(\partial^a u_a)u_b + u_a(\partial^a u_b)] + \partial_b P] \\ &= h^c_b[\partial^a(\rho+P)u_a u_b + (\rho+P)(\partial^a u_a)u_b) + (\rho+P)u_a(\partial^a u_b) + \partial_b P] \\ &= (\rho+P)h^c_b u_a(\partial^a u_b) + h^c_b\partial_b P \end{aligned}$$

Since h^c_b is just a projector, and $u_a\partial^a u^b$ is already orthogonal to u^b, $h^c_b u_a\partial^a u_b = u_a\partial^a u^c$ and,

$$0 = (\rho+P)u_a\partial^a u^c + h^{cb}\partial_b P$$

Lower the free index $c \to b$ will

$$0 = (\rho+P)u^a\partial_a u_b + h_{ab}\partial^a P$$

Replacing $h_{ab} = \eta_{ab} + u_a u_b$ in that last equation, Wald obtains the two equations:

$$u^a\partial_a\rho + (\rho+P)\,\partial_a u^a = 0, \tag{4.2.12}$$

$$(\rho+P)u^a\partial_a u_b + (\eta_{ab} + u_a u_b)\partial^a P = 0. \tag{4.2.13}$$

Interpretation. Equation 4.2.12 describes the change in energy density along fluid flow (compression increases ρ, expansion decreases it). Equation 4.2.13 gives the relativistic Euler equation: the fluid accelerates in response to pressure gradients.

Nonrelativistic Limit

In the nonrelativistic limit, $P \ll \rho$ and $u^a \approx (1, \vec{v})$, and we also assume that $dP/dt \ll |\vec{\nabla} P|$. Under these approximations, Eqs. 4.2.12 and 4.2.13 reduce to:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0, \tag{4.2.14}$$

$$\rho \left(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla}) \vec{v} \right) = -\vec{\nabla} P. \tag{4.2.15}$$

These are precisely the familiar continuity equation and Euler's equation of fluid dynamics.

Key Point. Special relativity is fully compatible with classical fluid dynamics: Newtonian results emerge from Eq. 4.2.11 once velocities and pressures are sufficiently small.

Mass–Energy Current

For a family of inertial observers with four-velocity v^a, define the mass–energy current vector

$$J_a = -T_{ab} v^b. \tag{4.2.16}$$

Using Eq. 4.2.11, it follows immediately that

$$\partial^a J_a = 0. \tag{4.2.17}$$

Integrating Eq. 4.2.17 over a spacetime region and applying Gauss's law yields

$$\int_S J^a n_a \, dS = 0, \tag{4.2.18}$$

where n^a is the unit normal to the boundary S. This expresses the local conservation of energy as measured by inertial observers.

Geometric view. Equation 4.2.18 states that the net energy flux through any closed 3-surface in Minkowski spacetime is zero. Energy cannot be created or destroyed inside the region.

4.2.4 Fields in Special Relativity

In addition to particles and fluids, special relativity naturally accommodates classical fields. In this subsection we illustrate the tensorial formulation of field theory in flat spacetime by considering two examples: the scalar field and the electromagnetic field.

Scalar Field

Although no classical scalar field is known to exist in nature, scalar fields provide a useful model. A real scalar field ϕ on Minkowski spacetime satisfies the Klein–Gordon equation

$$\partial^a \partial_a \phi - m^2 \phi = 0. \tag{4.2.19}$$

Its stress–energy tensor is

$$T_{ab} = \partial_a \phi \, \partial_b \phi - \frac{1}{2} \eta_{ab} (\partial^c \phi \, \partial_c \phi + m^2 \phi^2). \tag{4.2.20}$$

Interpretation. The first term represents directional derivatives of ϕ (kinetic energy). The second term removes the trace contribution so that T_{ab} represents the local flux of energy and momentum carried by the field. This tensor satisfies the conservation equation $\partial^a T_{ab} = 0$ by virtue of the field equation 4.2.19.

Electromagnetic Field

In prereativity physics, the electric field $\vec{E}$ and magnetic field $\vec{B}$ are separate spatial vectors. Special relativity unifies them into a single antisymmetric tensor field F_{ab}, called the *electromagnetic field tensor*. This tensor has 6 independent components, corresponding to the 3 components of $\vec{E}$ and the 3 components of $\vec{B}$.

For an observer with four-velocity v^a, the electric field measured by that observer is

$$E_a = F_{ab} v^b, \tag{4.2.21}$$

while the magnetic field is given by

$$B_a = -\frac{1}{2} \epsilon_{abcd} F^{cd} v^b. \tag{4.2.22}$$

Here ϵ_{abcd} is the totally antisymmetric tensor with $\epsilon_{0123} = 1$ in a right-handed orthonormal basis, and with $\epsilon_{abcd}\epsilon^{abcd} = -24$.

Geometric meaning. Equations 4.2.21–4.2.22 decompose the antisymmetric tensor F_{ab} into spatial electric and magnetic components relative to an observer. Different observers see different $\vec{E}$ and $\vec{B}$ fields, but the underlying tensor F_{ab} is the true physical object.

Maxwell's Equations

In tensor form, Maxwell's equations take the elegant and compact form

$$\partial^a F_{ab} = -4\pi j_b, \tag{4.2.23}$$

$$\partial_{[a} F_{bc]} = 0, \tag{4.2.24}$$

where j^a is the electric four-current. The antisymmetry of F_{ab} implies

$$0 = \partial^b \partial^a F_{ab} = -4\pi \, \partial^b j_b, \tag{4.2.25}$$

so Maxwell's equations imply charge conservation.

The motion of a particle of charge q and mass m in the electromagnetic field is governed by the Lorentz-force law:

$$u^a \partial_a u_b = \frac{q}{m} F_{bc} u^c, \tag{4.2.26}$$

where u^a is its future-directed unit timelike four-velocity.

Stress–Energy Tensor of the Electromagnetic Field

The electromagnetic field has stress–energy tensor

$$T_{ab} = \frac{1}{4\pi} \left(F_{ac} F_b{}^c - \frac{1}{4} \eta_{ab} F_{cd} F^{cd} \right). \tag{4.2.27}$$

This tensor satisfies the energy condition (4.2.9). Furthermore,

$$\partial^a T_{ab} = 0$$

whenever Maxwell's equations hold and $j^a = 0$.

The Vector Potential

By the Poincaré lemma (Appendix B), the equation $\partial_{[a} F_{bc]} = 0$ implies the existence of a vector field A_a such that

$$F_{ab} = \partial_a A_b - \partial_b A_a. \tag{4.2.28}$$

In terms of A_a, Maxwell's equation 4.2.23 becomes

$$\partial^a (\partial_a A_b - \partial_b A_a) = -4\pi j_b. \tag{4.2.29}$$

Gauge freedom. The potential A_a is not unique: replacing A_a by $A_a + \partial_a \chi$ leaves F_{ab} unchanged. The function χ is called a *gauge function.* This freedom is essential for understanding electromagnetic waves.

To fix the gauge, we may choose χ such that the Lorenz gauge condition holds:

$$\partial^a A_a = 0. \tag{4.2.31}$$

Using the commutativity of derivatives in flat spacetime, Eq. 4.2.29 becomes

$$\partial^a \partial_a A_b = -4\pi j_b. \tag{4.2.32}$$

Electromagnetic Waves

We now seek source-free ($j^a = 0$) solutions of the form

$$A_a = C_a\, e^{iS}, \tag{4.2.33}$$

where C_a is a constant vector field (constant norm and parallel-transported), and S is the *phase*. Substitution into Eq. 4.2.32 yields the eikonal equations

$$\partial^a \partial_a S = 0, \tag{4.2.34}$$
$$\partial_a S\, \partial^a S = 0, \tag{4.2.35}$$
$$C_a \partial^a S = 0. \tag{4.2.36}$$

Null hypersurfaces and light propagation. Equation 4.2.35 states that the gradient $k_a = \partial_a S$ is null: $k_a k^a = 0$. Thus the surfaces of constant phase are null hypersurfaces. Differentiating Eq. 4.2.35 shows that the integral curves of k^a satisfy

$$k^b \partial_b k_a = 0,$$

i.e. they are null geodesics. Hence electromagnetic waves propagate along null geodesics of Minkowski spacetime—the mathematical statement that light travels on the light cone.

The frequency of the wave as measured by an observer with four-velocity v^a is

$$\omega = -v^a \partial_a S = -v^a k_a. \tag{4.2.38}$$

The most important solutions of the form 4.2.33 are plane waves:

$$S = \sum_{\mu=0}^{3} k_\mu x^\mu, \tag{4.2.39}$$

where k_μ are constants, and hence k^a is a constant null vector. All well-behaved electromagnetic solutions at large distances can be expressed as superpositions of such plane waves.

Conclusion. Maxwell's equations predict that light propagates along null geodesics. This justifies the terminology "light cone" and connects directly with the causal structure of spacetime.

4.3 General Relativity

4.3.1 From Special Relativity to a New Theory of Gravity

Maxwell's theory provides a unified and remarkably successful description of electricity, magnetism, and light, and its formulation is naturally compatible with the framework of special relativity. One might therefore expect the next logical step to be the development of a relativistic theory of gravitation analogous to the way Maxwell's theory generalizes Coulomb's law. In such a theory, gravitational effects would be incorporated into special relativity without altering its underlying conception of spacetime.

However, Einstein chose a completely different path. Rather than attempting to modify Newtonian gravitation within the flat spacetime of special relativity, he proposed an entirely new theory: *general relativity*, a theory in which gravitation is inseparable from the geometry of spacetime itself. As already noted in the introduction, Einstein was guided by two major ideas: the equivalence principle and Mach's principle.

Conceptual summary. Maxwell's theory fits neatly into special relativity, but Newtonian gravity does not. Einstein's key insight was that gravitation could not be treated as an ordinary force within the flat geometry of special relativity. Instead, gravity *is geometry*: a manifestation of the curvature of spacetime.

To appreciate the role of the equivalence principle in shaping this viewpoint, consider how one measures electromagnetic fields in special relativity. One first introduces "background observers" who are not subject to electromagnetic forces (for example, observers that are electrically neutral and carry no magnetic moment). These observers, being free of non-gravitational forces, move along geodesics of the flat spacetime metric and therefore define what we call *inertial observers*. A charged test body released in the vicinity of such observers will then deviate from inertial motion according to the electromagnetic field, as described by Eq. 4.2.26. In this way, the electromagnetic field can be experimentally determined.

Key idea. Electromagnetism can be measured relative to inertial observers because such observers can be constructed—they are bodies which feel no electromagnetic forces.

4.3.2 The Equivalence Principle and the Failure of Background Observers

If we attempt to apply the same measurement procedure to gravitation that we used for electromagnetism, we immediately encounter a fundamental obstacle. In electromagnetism, "background observers" are chosen so that they are insulated from electromagnetic forces—electrically neutral, no dipole moment, etc.—and therefore move on inertial (i.e. geodesic) world lines of the flat metric. The deviation of a charged test body from these geodesics reveals the electromagnetic field.

The equivalence principle states that all bodies fall in exactly the same way in a gravitational field: all freely falling bodies follow the same trajectories regardless of their composition or internal structure. Consequently, there is *no physical procedure* by which we can insulate an observer from gravitational forces. Any observer will fall in precisely the same way as a test body.

Thus, there is no natural "background motion" against which gravitational effects can be measured. A freely falling observer near a test body will move in exactly the same way as the test body, so we cannot define a gravitational field by comparing their motions. We lack an analogue of the electromagnetic construction of inertial observers.

Implication of the equivalence principle. Because all bodies fall alike, gravitation cannot be viewed as an ordinary force field that acts differently on different materials. There is no way to separate out the gravitational influence on an observer in order to measure a "gravitational force field" in the same manner as an electromagnetic field.

It is, of course, logically possible that sufficiently complicated and delicate experimental procedures might eventually allow one to construct inertial observers in the flat-spacetime sense and thereby measure a gravitational force field. If special relativity were the correct description of spacetime, one could in principle determine the (flat) spacetime metric by metrological operations—using clocks, metersticks, and the behavior of freely falling bodies—and thereby construct inertial observers whose geodesics are determined by that flat metric.

However, if inertial observers had to carry rocket engines to counteract gravitational effects, then—aside from this external propulsion—the gravitational force field could be measured in a manner analogous to electromagnetism. In this case, the equivalence principle might appear as an odd quirk of the Newtonian gravitational force law.

But this entire picture changes if we take seriously the possibility that special relativity is *not* valid globally, and that spacetime may not possess a flat metric at all.

4.3.3 The Central Hypothesis: Gravity *is* Curved Spacetime

The basic framework of general relativity arises from adopting the opposite viewpoint from the one just discussed. Rather than attempting—even in principle—to construct inertial observers in the special–relativistic sense and then regard deviations from inertial motion as effects of a "gravitational force field," we instead make the following bold hypothesis:

> *The spacetime metric is not flat, as was assumed in special relativity. The world lines of freely falling bodies in a gravitational field are the geodesics of the (curved) spacetime metric.*

In this picture, the freely falling observers *themselves* define the geodesics of the true spacetime geometry. These geodesics now play the role that the "background observers" played in electromagnetism. The motion that special relativity would have interpreted as acceleration due to a gravitational force is now reinterpreted as geodesic motion in a curved spacetime.

Consequently, "absolute gravitational force" ceases to have meaning. However, the *relative* gravitational force—that is, the tidal force between neighboring freely falling bodies—does have meaning. This relative acceleration is governed by the geodesic deviation equation (Eq. 3.3.18).

Key idea. Gravity is not a force field that pushes or pulls bodies; rather, it is a manifestation of spacetime curvature. Freely falling bodies follow geodesics of the metric g_{ab}, and the only physically meaningful gravitational effects are tidal effects encoded in the curvature tensor.

This viewpoint raises an immediate question: how can we reconcile the absence of a well-defined gravitational force with the familiar Newtonian picture of a gravitational acceleration of 980 cm/s^2 at the Earth's surface? Consider an object at rest on the Earth's surface. In Newtonian theory, this object remains stationary because the upward force from the surface balances the downward gravitational force. In general relativity, by contrast, the only force acting on the object is the force from the surface. A freely falling body in the Earth's gravitational field accelerates downward at 980 cm/s^2, but the object on the surface *does not fall* because the surface exerts an upward force which causes its worldline to deviate from a geodesic by exactly this amount.

Thus, "gravitational force" in the Newtonian sense corresponds in general relativity to the deviation of a worldline from a geodesic. If there is a time translation symmetry in the spacetime region—as near the surface of the Earth—one may define a family of preferred observers using this symmetry and thus define a gravitational force field in this restricted sense. But in general, when no such symmetry exists, no preferred background observers can be defined, and the only meaningful remnant of gravity is the tidal acceleration between nearby geodesics.

4.3.4 Why Spacetime Need Not Be $\mathbb{R}^4$: Lorentz Metrics on General Manifolds

In special relativity, spacetime is taken to be the vector space $\mathbb{R}^4$ equipped with the flat Minkowski metric η_{ab}. General relativity, by contrast, removes both of these assumptions:

1. Spacetime need not be a vector space.
2. The metric need not be flat.

Instead, spacetime is a smooth 4–dimensional manifold M, and the geometry is encoded in a Lorentzian metric g_{ab} defined on M. All physical predictions depend only on (M, g_{ab}); no preferred global coordinates are assumed.

Local structure. Although M need not be globally $\mathbb{R}^4$, each point $p \in M$ possesses a tangent space T_pM which *is* a Lorentzian vector space. Thus the "local Minkowski" properties of special relativity remain valid pointwise, even though the global structure of spacetime may be curved or may have nontrivial topology.

Equivalence principle revisited. Because each tangent space carries the metric $g_{ab}(p)$, we can always choose coordinates at p such that

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \qquad \Gamma^{\rho}{}_{\mu\nu}(p) = 0.$$

These are Riemannian normal coordinates. Consequently, the laws of special relativity always hold *at a point*, which is the mathematical content of the equivalence principle.

Freedom in global structure. Nothing in the local Lorentzian structure fixes the global form of M. Spacetime could be:

- topologically $\mathbb{R}^4$,

- or $\mathbb{R}^3 \times S^1$ (closed time direction),
- or $S^3 \times \mathbb{R}$ (closed spatial slices),
- or any smooth 4–manifold that admits a Lorentz metric.

The global topology has physical consequences—e.g. compact spatial slices affect cosmology—but these global questions cannot be answered by the local geometry alone.

Key idea. General relativity is a *local* theory: curvature tells you how nearby geodesics behave, not what the global shape of spacetime is. Einstein's equations constrain g_{ab} locally, but the global structure of spacetime is determined by both the equations and the topology of M.

4.3.5 The Principles Governing Physics in Curved Spacetime

Having introduced the idea that spacetime is a manifold M equipped with a Lorentzian metric g_{ab}, we now examine how the *laws of physics* are to be formulated in this new geometric setting. General relativity does not merely modify Newtonian gravity; it reshapes the very meaning of physical quantities. Wald emphasizes that physics in curved spacetime is guided by two foundational principles.

1. General Covariance. The first principle states:

> *The equations of physics must be expressed entirely in terms of spacetime tensor fields and must hold in all coordinate systems.*

Since there is no flat background metric in general relativity, all geometric and physical objects must be defined directly in terms of g_{ab} and the structures derived from it.

Key idea. General covariance is not merely coordinate–independence. It asserts that the only meaningful quantities in physics are tensor fields on (M, g_{ab}). There is no "hidden" flat metric η_{ab}.

Phrases such as "spacetime vector" or "spacetime gradient" acquire meaning only through g_{ab}, which determines lengths, angles, causal structure, and the connection ∇_a.

2. Reduction to Special Relativity. The second principle:

> *Wherever gravitational effects become negligible (i.e. where $g_{ab} \approx \eta_{ab}$), the laws of physics must reduce to those of special relativity.*

This ensures that the well-tested flat–spacetime physics is recovered in the appropriate limit. Concretely, one follows the heuristic *minimal substitution rule*:

$$\eta_{ab} \rightarrow g_{ab}, \qquad \partial_a \rightarrow \nabla_a,$$

and leaves the form of the physical equations otherwise unchanged.

Warning. This "rule" is only a heuristic. Wald emphasizes that it is not a mathematically precise prescription, and it fails in certain important cases. We will later encounter counterexamples where additional structure is needed.

Implications for Physical Fields. Even though the manifold M need not be $\mathbb{R}^4$, and the metric need not be flat, the *types* of tensorial physical fields remain the same as in special relativity:

- a particle is described by a unit timelike tangent vector u^a;
- a perfect fluid by (ρ, P, u^a);
- electromagnetic fields by an antisymmetric tensor F_{ab};
- stress–energy by a symmetric tensor T_{ab}.

But now all contractions, norms, and derivatives use g_{ab} and its compatible derivative operator ∇_a.

Key idea. General relativity modifies the geometric background of physics, not the *representation type* of physical fields. The same tensor fields appear, but now they live on a curved manifold and are coupled to g_{ab}.

The Geodesic Equation Reappears. The special–relativistic force–free condition

$$u^a \partial_a u^b = 0$$

lifts via minimal substitution to the covariant form

$$u^a \nabla_a u^b = 0. \tag{4.3.1}$$

This is exactly the invariant geodesic equation introduced in Chapter 3. It expresses the central idea of general relativity:

A freely falling particle moves along a geodesic of the spacetime metric.

This marks the beginning of Wald's development of motion, forces, and stress–energy in curved spacetime. In the next subsection, we follow Wald into the discussion of 4–momentum and the generalization of the Lorentz force.

4.3.6 Motion, 4–Momentum, and Forces in Curved Spacetime

Equation (4.3.1),

$$u^a \nabla_a u^b = 0, \tag{4.3.1}$$

states that a freely falling particle moves on a geodesic of the spacetime metric g_{ab}: its 4–velocity u^a is parallel transported along its own worldline.

In the presence of a non-gravitational force, the particle no longer moves geodesically. For a charged particle of (rest) mass m and charge q moving in an electromagnetic field F_{ab}, Wald takes over from special relativity the Lorentz force law, replacing η_{ab} by g_{ab} and ∂_a by ∇_a:

$$u^a \nabla_a u^b = \frac{q}{m} F^b{}_c u^c. \tag{4.3.2}$$

Here ∇_a is the derivative operator compatible with g_{ab}. Indices are raised and lowered using g_{ab} and its inverse g^{ab}, so $F^b{}_c = g^{bd} F_{dc}$.

Orthogonality of the force. Contracting Eq. (4.3.2) with $g_{ab}u^a$ and using the antisymmetry of F_{ab} shows that $u_b F^b{}_c u^c = 0$. Thus the 4–acceleration is orthogonal to the 4–velocity, just as in special relativity: the Lorentz force changes the particle's direction but not its rest mass.

4–momentum and energy. The (rest) mass m of the particle is defined so that

$$p^a \equiv m u^a \tag{4.3.3}$$

is its 4–momentum. Given an observer with 4–velocity v^a at the same event on the worldline, the energy of the particle *as measured by that observer* is

$$E = -p_a v^a. \tag{4.3.4}$$

Here the minus sign reflects the fact that both u^a and v^a are future directed timelike vectors with $g_{ab} v^a v^b = -1$.

No global inertial observers. In curved spacetime there is, in general, no natural way to compare vectors at widely separated points; parallel transport depends on the curve used. Thus there is no "global family" of inertial observers, and no invariant definition of the energy of a distant particle. Energy is always defined *relative to a local observer* at the event where the measurement is made.

Stress–energy of a perfect fluid. Continuous matter distributions are described by a stress–energy tensor T_{ab}. For a perfect fluid, characterized by proper density ρ, pressure P, and 4–velocity u^a, Wald writes

$$T_{ab} = \rho\, u_a u_b + P\big(g_{ab} + u_a u_b\big), \tag{4.3.5}$$

the same form as in special relativity but with g_{ab} in place of η_{ab}. The first term represents the energy density carried with the fluid's motion; the second term encodes the isotropic pressure.

The motion of matter and the exchange of energy and momentum are encoded in the covariant conservation law

$$\nabla^a T_{ab} = 0. \tag{4.3.6}$$

This equation replaces $\partial^a T_{ab} = 0$ from flat spacetime.

To see what Eq. (4.3.6) implies for a perfect fluid, we project along and orthogonal to u^a. Contracting with u^b gives

$$u^b \nabla^a T_{ab} = 0 \implies u^a \nabla_a \rho + (\rho + P)\, \nabla_a u^a = 0, \tag{4.3.7}$$

which is the relativistic energy-balance equation for the fluid.

Next, project orthogonally to u^a with the spatial projector $h^c{}_b \equiv g^c{}_b + u^c u_b$. Using $h^c{}_b u^b = 0$ and Eq. (4.3.5), the spatial projection of $\nabla^a T_{ab} = 0$ yields

$$(\rho + P)\, u^a \nabla_a u_b + (g_{ab} + u_a u_b) \nabla^a P = 0. \tag{4.3.8}$$

This is the relativistic Euler equation: it describes how pressure gradients accelerate the fluid worldlines, with the effective inertial mass density given by $\rho + P$.

Local vs. global conservation. In flat spacetime one can often find a timelike vector field v^a that is covariantly constant, $\nabla_a v^b = 0$, leading to a global conservation law for energy. In curved spacetime such a field typically does not exist. Equation (4.3.6) therefore expresses a *local* conservation of energy–momentum, valid in small regions, rather than a global conservation law defined with respect to a preferred family of inertial observers.

4.3.7 Scalar Fields and Maxwell Fields in Curved Spacetime

We have seen how perfect fluids and point particles generalize naturally from special relativity to curved spacetime. Wald now considers how the same minimal–substitution principles apply to field equations, beginning with the simplest case: a real scalar field.

The Klein–Gordon field. In special relativity, a (real) scalar field ϕ of mass m satisfies

$$\eta^{ab} \partial_a \partial_b \phi - m^2 \phi = 0.$$

Following the prescription $\eta_{ab} \to g_{ab}$ and $\partial_a \to \nabla_a$, the natural curved–spacetime generalization is

$$\nabla^a \nabla_a \phi - m^2 \phi = 0. \tag{4.3.9}$$

This is the Klein–Gordon equation on a curved background.

Stress–energy tensor of a scalar field. The scalar field carries energy–momentum. The stress–energy tensor derived from the minimally coupled Lagrangian is

$$T_{ab} = (\nabla_a \phi)(\nabla_b \phi) - \frac{1}{2} g_{ab} \big[(\nabla_c \phi)(\nabla^c \phi) + m^2 \phi^2 \big]. \tag{4.3.10}$$

It satisfies the covariant conservation law $\nabla^a T_{ab} = 0$ as a direct consequence of Eq. (4.3.9).

Non-uniqueness of minimal substitution. Wald emphasizes that minimal substitution is not the only reasonable generalization. For example, one could add a curvature term:

$$\nabla^a \nabla_a \phi - m^2 \phi - \alpha R \phi = 0, \tag{4.3.11}$$

where α is a constant. This equation still respects general covariance and reduces to the flat-space equation when $R = 0$. The choice $\alpha = 1/6$ is special because it yields conformal invariance for $m = 0$.

Maxwell's equations in curved spacetime. The generalization of Maxwell's equations is more straightforward. The electromagnetic field is still represented by an antisymmetric 2-form F_{ab}, and the equations become

$$\nabla^a F_{ab} = -4\pi j_b, \tag{4.3.12}$$

$$\nabla_{[a} F_{bc]} = 0. \tag{4.3.13}$$

Equation (4.3.13) is unchanged because it expresses the closure of F_{ab}: there exists a vector potential A_a such that $F_{ab} = \nabla_a A_b - \nabla_b A_a$.

The electromagnetic stress–energy tensor is obtained from Eq. (4.2.27) by simply replacing η_{ab} with g_{ab}:

$$T_{ab} = \frac{1}{4\pi} \left(F_{ac} F_b{}^c - \frac{1}{4} g_{ab} F_{cd} F^{cd} \right). \tag{4.3.14}$$

Lorentz gauge and curvature. If we impose the Lorentz gauge $\nabla^a A_a = 0$, then the field equation for A_a becomes

$$\nabla^a \nabla_a A_b - R_b{}^a A_a = -4\pi j_b. \tag{4.3.15}$$

The extra curvature term $R_b{}^a A_a$ arises because covariant derivatives do not commute.

Insight. Had we naïvely replaced $\partial_a \partial^a A_b$ with $\nabla_a \nabla^a A_b$, we would *miss* the curvature term. Equation (4.3.15) demonstrates again that minimal substitution is not a precise algorithm.

Geometrical optics approximation. If the electromagnetic field varies rapidly compared to the scale of the curvature, the vector potential may be written in the WKB form

$$A_a = C_a e^{iS}, \tag{4.3.16}$$

where derivatives of C_a are small. Substituting into the homogeneous Maxwell equation (with $j_b = 0$) and keeping only dominant terms gives

$$\nabla_a S \, \nabla^a S = 0. \tag{4.3.17}$$

Thus $k_a \equiv \nabla_a S$ is a null vector, and the rays of light follow null geodesics:

$$k^a \nabla_a k^b = 0.$$

This matches our earlier geometrical-optics interpretation from flat spacetime, now justified in curved geometry.

Conclusion. Light follows null geodesics of the spacetime metric. Curvature affects not only the path but also the local frequency and amplitude of electromagnetic waves.

4.3.8 From Tidal Forces to Einstein's Equation

We have now described how general relativity treats gravitation in terms of curved spacetime geometry and how the standard laws of physics are modified in this framework. The remaining ingredient is the equation that relates the geometry of spacetime to the matter distribution. Wald obtains this by comparing the description of tidal forces in Newtonian gravity and in general relativity.

Tidal forces in Newtonian theory. In the Newtonian theory, the gravitational field is derived from a potential ϕ. The tidal acceleration between two nearby particles, with separation vector $\vec{x}$, is

$$-(\vec{x}\cdot\vec{\nabla})\vec{\nabla}\phi,$$

i.e. it is governed by second spatial derivatives of ϕ.

Tidal forces in general relativity. In general relativity, the relative acceleration of two nearby geodesics with separation vector x^a and tangent 4–velocity u^a is given by the geodesic deviation equation (Eq. 3.3.18),

$$\frac{D^2x^a}{D\tau^2} = -R^a{}_{bcd}\,u^b x^c u^d.$$

Thus tidal forces are encoded in the curvature tensor R_{abcd}.

Comparing these two descriptions, Wald is led to the correspondence

$$R_{cbd}{}^a\,u^c u^d \;\longleftrightarrow\; \partial_b\partial^a\phi. \tag{4.3.18}$$

Relating curvature to energy density. In Newtonian gravity, the potential satisfies Poisson's equation

$$\nabla^2\phi = 4\pi\rho, \tag{4.3.19}$$

where ρ is the (mass) density of matter.

In special and general relativity, the energy properties of matter are encoded in the stress–energy tensor T_{ab}. For an observer with 4–velocity u^a, the local energy density measured by that observer is

$$\rho_{(\mathrm{obs})} = T_{ab}u^a u^b.$$

Thus Wald is led to the correspondence

$$T_{ab}u^a u^b \;\longleftrightarrow\; \rho, \tag{4.3.20}$$

at least in the regime where Newtonian theory should be valid.

Putting (3.3.18), (4.3.9) and the correspondences (3.3.18), (4.3.9)–(4.3.14) together suggests that an equation of the form

$$R_{ab} \sim T_{ab}$$

should hold, with a proportionality constant chosen so that the Newtonian limit is reproduced. A natural first guess is

$$R_{ab} = 4\pi T_{ab}.$$

Indeed, this equation was originally postulated by Einstein. However, it turns out to be incompatible with the conservation law $\nabla^a T_{ab} = 0$.

Role of the Bianchi identity. The contracted Bianchi identity (Eq. (3.2.31)) states that

$$\nabla^a\left(R_{ab} - \tfrac{1}{2}g_{ab}R\right) = 0.$$

If we required $R_{ab} = 4\pi T_{ab}$, then taking the divergence and using $\nabla^a T_{ab} = 0$ would imply

$$\nabla_b R = 0,$$

so R would have to be constant throughout the universe. This is an unphysical restriction on the matter distribution, and it forces us to reject $R_{ab} = 4\pi T_{ab}$ as the fundamental field equation.

The Einstein tensor. The above difficulty suggests the correct modification. Define the *Einstein tensor*

$$G_{ab} \equiv R_{ab} - \frac{1}{2}Rg_{ab}.$$

By construction, $\nabla^a G_{ab} = 0$ identically, for any metric g_{ab}. We are therefore led to the equation

$$G_{ab} \equiv R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab}. \tag{4.3.21}$$

This is *Einstein's field equation.* It relates the curvature of spacetime directly to the stress–energy tensor of matter and fields.

Taking the trace of (4.3.21), we find

$$R = -8\pi T, \tag{4.3.22}$$

where $T \equiv T^a{}_a$ is the trace of the stress–energy tensor. Substituting back into (4.3.21) yields

$$R_{ab} = 8\pi\left(T_{ab} - \frac{1}{2}g_{ab}T\right). \tag{4.3.23}$$

Newtonian limit check. In situations where Newtonian gravity should be valid, the energy density ρ measured by an observer roughly at rest with respect to the matter dominates over stresses (with $c = 1$). In this case

$$T \approx -\rho \approx -T_{ab}u^a u^b,$$

and Eqs. (3.3.18), (4.3.9), and (4.3.14) lead to

$$R_{ab}u^a u^b \approx 4\pi T_{ab}u^a u^b,$$

which reproduces the Newtonian correspondence between curvature and mass density. Thus (4.3.21) has the correct Newtonian limit.

Summary. Einstein's equation

$$G_{ab} = 8\pi T_{ab}$$

is the central dynamical law of general relativity. It states that the Einstein tensor G_{ab}—a specific combination of curvature components with vanishing divergence—is determined by the stress–energy tensor T_{ab} of matter and fields. Geometry and matter are inseparably linked.

4.3.9 Remarks on the Nature of Einstein's Equation

Wald closes the section with several important comments about the structure and interpretation of Eq. (4.3.21).

Mathematical character. Expressed in a coordinate system, the metric components $g_{\mu\nu}$ satisfy a coupled system of nonlinear second–order partial differential equations. For a Lorentzian metric, these equations are of hyperbolic (wave) type and admit a well–posed initial value formulation (see Chapter 10). Much of the rest of the book is devoted to studying solutions of these equations and their properties.

Analogy with Maxwell's equations. In one sense, Eq. (4.3.21) is analogous to Maxwell's equation $\nabla_a F^{ab} = 4\pi j^b$, with T_{ab} playing the role of the source j^a. However, there is an important difference: for Maxwell's equation one may think of specifying j^a first, and then solving for F_{ab}. In general relativity, the stress–energy tensor of realistic matter fields (fluids, electromagnetic fields, scalar fields, etc.) *contains* the metric g_{ab} explicitly. Thus T_{ab} and g_{ab} must be solved for simultaneously, not separately.

Equations of motion from $\nabla^a T_{ab} = 0$. As we have set things up, the equations of motion of matter fields are postulated first (e.g. the geodesic equation for particles, the Euler equation for fluids, Maxwell's equations for the electromagnetic field). Einstein's equation then relates the resulting stress–energy tensor to curvature.

However, Einstein's equation implies $\nabla^a T_{ab} = 0$ identically, and this relation encodes a great deal of information about the motion of matter. For a perfect fluid with stress–energy tensor

$$T_{ab} = \rho u_a u_b + P(g_{ab} + u_a u_b),$$

the condition $\nabla^a T_{ab} = 0$ reproduces the fluid equations of motion, including the geodesic motion of dust when $P = 0$. More generally, it can be shown that any sufficiently small body whose self–gravity is weak must move on a geodesic of the spacetime metric. Thus Einstein's equation is consistent with—and in appropriate limits actually implies—the geodesic hypothesis.

Final summary of the section. The content of general relativity can be encapsulated as follows:

- Spacetime is a manifold M equipped with a Lorentzian metric g_{ab}.
- Matter and non-gravitational fields are described by tensor fields on (M, g_{ab}) with stress–energy tensor T_{ab}.
- Test bodies move on geodesics of g_{ab}; tidal effects are encoded in the curvature tensor R_{abcd}.
- The metric is not fixed but dynamical, determined by Einstein's equation $G_{ab} = 8\pi T_{ab}$.

4.4 Linearized Gravity: The Newtonian Limit and Gravitational Radiation

4.4.1 The Linearized Einstein Equation

The full Einstein equation is highly nonlinear, which makes exact solutions difficult to obtain. However, in many physically important situations—such as weak gravitational fields, small perturbations of flat spacetime, and the generation and propagation of gravitational waves—the deviations of the spacetime metric from the flat Minkowski metric are very small.

- The **goal** of this section is to derive the *linearized Einstein equation*, valid when the gravitational field is weak.
- This linearization allows us to treat gravity as a small perturbation γ_{ab} of flat spacetime and to obtain a wave equation for these perturbations.
- This approximation is foundational for understanding gravitational waves, the Newtonian limit of GR, and the interpretation of gravity as a massless spin-2 field.

Throughout this section we work in a global inertial coordinate system and retain only terms *linear* in the perturbation γ_{ab}.

4.4.1.1 Perturbing the Metric

We assume that the physical metric g_{ab} differs only slightly from the flat Minkowski metric η_{ab}, and we write

$$g_{ab} = \eta_{ab} + \gamma_{ab}, \tag{4.4.1}$$

where $|\gamma_{ab}| \ll 1$ in some global inertial coordinate system. (Since there is no positive definite norm on tensors in relativity, "smallness" means simply that the components $\gamma_{\mu\nu}$ are $\ll 1$ in these coordinates.)

All index raising and lowering in linearized gravity will be performed with η_{ab} and its inverse η^{ab}. This is because using g_{ab} to raise/lower indices would introduce terms of order γ_{ab}^2, which are discarded in the linear approximation.

A notable exception is that g^{ab} itself is defined as the inverse of g_{ab}. To linear order, we solve

$$(\eta^{ac} - \gamma^{ac})(\eta_{cb} + \gamma_{cb}) = \delta^a{}_b$$

and obtain

$$g^{ab} = \eta^{ab} - \gamma^{ab}. \tag{4.4.2}$$

4.4.1.2 Linearized Christoffel Symbols

We denote by ∂_a the flat derivative compatible with η_{ab}. Substituting (4.4.1) into the definition of the Christoffel symbols,

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\nu} - \partial_{\sigma\nu} g_{\mu\nu}\right)$$
$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}(\eta^{\rho\sigma} - \gamma^{\rho\sigma}) \left(\partial_\mu(\eta_{\sigma\nu} + \gamma_{\sigma\nu}) + \partial_\nu(\eta_{\sigma\nu} + \gamma_{\sigma\nu}) - \partial_{\sigma\nu}(\eta_{\mu\nu} + \gamma_{\mu\nu})\right)$$

and keeping only terms linear in γ_{ab} yields

$$\Gamma^c{}_{ab} = \frac{1}{2}\, \eta^{cd}(\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab}). \tag{4.4.3}$$

after recognizing that $\partial_a \eta_{bc} = 0$ and throwing away $\gamma^{\rho\sigma}\partial_\mu \gamma_{\sigma\mu}$ terms because they are second order in γ.

4.4.1.3 Linearized Ricci Tensor

The Ricci tensor is

$$R_{ab} = \partial_c \Gamma^c{}_{ab} - \partial_a \Gamma^c{}_{cb} + \Gamma^c_{ab}\Gamma^d_{cd} - \Gamma^c_{ad}\Gamma^d_{cb}$$

The linearized Ricci Tensor is,

$$R^{(1)}_{ab} = \partial^c \partial_{(a}\gamma_{b)c} - \frac{1}{2}\,\partial^c\partial_c \gamma_{ab} - \frac{1}{2}\,\partial_a \partial_b \gamma \tag{4.4.4}$$

where $\gamma \equiv \gamma^c{}_c$ is the trace and the symmetrization $\partial^c \partial_{(a}\gamma_{b)c} = \partial^c \partial_a \gamma_{bc} + \gamma^c \partial_b \gamma_{ac}$.

Ricci Tensor (Details of the Linearization) The full derivation of the linearized Ricci Tensor follows.Recall that to linear order in γ_{ab} the Christoffel symbols are

$$\Gamma^c{}_{ab} = \frac{1}{2}\, \eta^{cd}(\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab}). \tag{4.4.3}$$

The Ricci tensor is defined by

$$R_{ab} = \partial_c \Gamma^c{}_{ab} - \partial_a \Gamma^c{}_{cb} + \Gamma^c{}_{ab}\Gamma^d{}_{cd} - \Gamma^c{}_{ad}\Gamma^d{}_{cb}.$$

In the linear approximation we discard all terms quadratic in γ_{ab}, so the $\Gamma\Gamma$ terms are dropped. Hence

$$R^{(1)}_{ab} = \partial_c \Gamma^c{}_{ab} - \partial_a \Gamma^c{}_{cb}. \tag{*}$$

Step 1: Compute $\partial_c \Gamma^c{}_{ab}$. Using (4.4.3),

$$\begin{aligned}\partial_c \Gamma^c{}_{ab} &= \partial_c \left[\frac{1}{2}\,\eta^{cd}(\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab})\right] \\ &= \frac{1}{2}\,\eta^{cd}(\partial_c\partial_a \gamma_{bd} + \partial_c\partial_b \gamma_{ad} - \partial_c\partial_d \gamma_{ab}),\end{aligned}$$

since $\partial_c \eta^{cd} = 0$ in inertial coordinates.

Now raise the index c with η^{cd}:

$$\eta^{cd}\partial_c = \partial^d, \qquad \eta^{cd}\partial_c\partial_d = \partial^c\partial_c.$$

So

$$\partial_c \Gamma^c{}_{ab} = \frac{1}{2}\Big(\partial^c\partial_a \gamma_{bc} + \partial^c\partial_b \gamma_{ac} - \partial^c\partial_c \gamma_{ab}\Big). \tag{1}$$

Step 2: Compute $\partial_a \Gamma^c{}_{cb}$. First contract $\Gamma^c{}_{cb}$ using (4.4.3):

$$\Gamma^c{}_{cb} = \frac{1}{2}\,\eta^{cd}(\partial_c \gamma_{bd} + \partial_b \gamma_{cd} - \partial_d \gamma_{cb}).$$

Use symmetry $\gamma_{cd} = \gamma_{dc}$ to simplify the last two terms:

$$\eta^{cd}\partial_b \gamma_{cd} = \partial_b(\eta^{cd}\gamma_{cd}) = \partial_b \gamma,$$

where $\gamma \equiv \gamma^c{}_c$ is the trace.

Also,

$$\eta^{cd}\partial_d \gamma_{cb} = \partial^c \gamma_{cb}.$$

Thus,

$$\Gamma^c{}_{cb} = \frac{1}{2}\Big(\partial^c \gamma_{bc} + \partial_b \gamma - \partial^c \gamma_{cb}\Big) = \frac{1}{2}\,\partial_b \gamma,$$

because the first and third terms are identical.

Therefore,

$$\partial_a \Gamma^c{}_{cb} = \frac{1}{2}\,\partial_a\partial_b \gamma. \tag{2}$$

Step 3: Subtract. Insert (1) and (2) into (*):

$$R^{(1)}_{ab} = \frac{1}{2}\Big(\partial^c\partial_a\gamma_{bc} + \partial^c\partial_b\gamma_{ac} - \partial^c\partial_c\gamma_{ab}\Big) - \frac{1}{2}\,\partial_a\partial_b\gamma.$$

Combine the first two terms using symmetrization:

$$\partial^c\partial_{(a}\gamma_{b)c} \equiv \frac{1}{2}\left(\partial^c\partial_a\gamma_{bc} + \partial^c\partial_b\gamma_{ac}\right).$$

Hence the linearized Ricci tensor is

$$\boxed{R^{(1)}_{ab} = \partial^c\partial_{(a}\gamma_{b)c} - \frac{1}{2}\,\partial^c\partial_c\gamma_{ab} - \frac{1}{2}\,\partial_a\partial_b\gamma}. \tag{4.4.4}$$

Contraction of the Linearized Ricci Tensor Starting from the linearized Ricci tensor,

$$R^{(1)}_{ab} = \partial^c\partial_{(a}\gamma_{b)c} - \frac{1}{2}\,\partial^c\partial_c\gamma_{ab} - \frac{1}{2}\,\partial_a\partial_b\gamma, \qquad \gamma \equiv \gamma^c{}_c, \tag{4.4.4}$$

we now contract with η^{ab} to obtain the linearized Ricci scalar.

Step 1: Contract the first term. Using

$$\partial_{(a}\gamma_{b)c} = \frac{1}{2}(\partial_a\gamma_{bc} + \partial_b\gamma_{ac}),$$

we obtain

$$\eta^{ab}\partial_{(a}\gamma_{b)c} = \partial^a\gamma_{ac}.$$

Thus

$$\eta^{ab}\partial^c\partial_{(a}\gamma_{b)c} = \partial^c\partial^a\gamma_{ac}.$$

Step 2: Contract the second term. Since $\eta^{ab}\gamma_{ab} = \gamma$,

$$\eta^{ab}\partial^c\partial_c\gamma_{ab} = \partial^c\partial_c\,\gamma.$$

Hence

$$-\frac{1}{2}\,\eta^{ab}\partial^c\partial_c\gamma_{ab} = -\frac{1}{2}\,\partial^c\partial_c\gamma.$$

Step 3: Contract the third term.

$$\eta^{ab}\partial_a\partial_b\gamma = \partial^a\partial_a\gamma = \partial^c\partial_c\gamma,$$

so

$$-\frac{1}{2}\,\eta^{ab}\partial_a\partial_b\gamma = -\frac{1}{2}\,\partial^c\partial_c\gamma.$$

Result: Linearized Ricci Scalar. Combining all three pieces,

$$R^{(1)} = \partial^a \partial^b \gamma_{ab} - \partial^c \partial_c \gamma. \tag{4.4.5a}$$

4.4.1.4 Linearized Einstein Tensor

The linearized Einstein tensor is

$$G^{(1)}_{ab} = R^{(1)}_{ab} - \frac{1}{2}\,\eta_{ab} R^{(1)}.$$

Substituting equations (4.4.4) and (4.4.5a), we obtain the expression for the linearized Einstein tensor,

$$G^{(1)}_{ab} = \partial^c \partial_{(a}\gamma_{b)c} - \frac{1}{2}\,\partial^c\partial_c \gamma_{ab} - \frac{1}{2}\,\partial_a\partial_b\gamma - \frac{1}{2}\,\eta_{ab}\Big(\partial^c\partial^d\gamma_{cd} - \partial^e\partial_e\gamma\Big), \tag{4.4.5}$$

Eq. 4.4.5 is algebraically complicated. It contains both γ_{ab} and its trace $\gamma = \gamma^c{}_c$ in several places, which obscures the underlying structure of the equations.

Why introduce a trace-reversed field? In analogy with electromagnetism, where working with the potential A_a becomes simpler in Lorenz gauge, it is useful to redefine the metric perturbation so that the Einstein tensor takes a simpler, more symmetric form. In particular, many trace terms combine naturally under a trace-reversal operation.

Definition (trace reversal). Define the trace-reversed perturbation

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2}\,\eta_{ab}\gamma, \qquad \gamma = \gamma^c{}_c. \tag{4.4.6}$$

Taking the trace of this equation shows that

$$\bar{\gamma} \equiv \bar{\gamma}^c{}_c = -\gamma, \qquad \text{so} \qquad \gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2}\,\eta_{ab}\bar{\gamma}.$$

Substituting into $G^{(1)}_{ab}$. Inserting the expression

$$\gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2}\,\eta_{ab}\bar{\gamma}$$

into equation (4.4.5), and rearranging terms, one finds after a short calculation that

$$G^{(1)}_{ab} = -\frac{1}{2}\left(\partial^c\partial_c\,\bar{\gamma}_{ab} - \partial^c\partial_a\bar{\gamma}_{bc} - \partial^c\partial_b\bar{\gamma}_{ac} + \eta_{ab}\partial^c\partial^d\bar{\gamma}_{cd}\right). \tag{4.4.7a}$$

4.4.1.5 Lorenz Gauge

The perturbation γ_{ab} is defined only up to an infinitesimal coordinate transformation (a gauge transformation). Under $x^a \to x^a + \xi^a$, the trace-reversed perturbation transforms as

$$\bar{\gamma}_{ab} \to \bar{\gamma}_{ab} + \partial_a\xi_b + \partial_b\xi_a - \eta_{ab}\,\partial^c\xi_c.$$

Using this freedom, one may impose the *Lorenz gauge*:

$$\partial^a \bar{\gamma}_{ab} = 0. \tag{4.4.11}$$

This is the exact gravitational analogue of the Lorenz gauge condition $\partial^a A_a = 0$ in electromagnetism.

Simplification in Lorenz gauge. With $\partial^a \bar{\gamma}_{ab} = 0$, the last three terms of (4.4.7a) vanish, leaving the remarkably simple equation

$$G^{(1)}_{ab} = -\frac{1}{2}\, \partial^c \partial_c\, \bar{\gamma}_{ab}. \tag{4.4.7b}$$

Lorenz Gauge in Linearized Gravity: What It Really Means

In electromagnetism, gauge freedom reflects the fact that the potential A_a is not unique: adding a gradient $A_a \to A_a + \partial_a \Lambda$ does not change the physical field F_{ab}. The Lorenz gauge $\partial^a A_a = 0$ is a convenient condition that simplifies Maxwell's equations while leaving all observables unchanged.

In linearized gravity, the analogous freedom

$$\gamma_{ab} \;\to\; \gamma_{ab} - \partial_a \xi_b - \partial_b \xi_a$$

arises from infinitesimal coordinate transformations $x^a \to x^a + \xi^a$. Different choices of coordinates change the components of the metric perturbation γ_{ab} but *do not* change physical tidal forces or curvature. Thus this "gauge" freedom reflects *coordinate redundancy*, not freedom in a physical field.

Introducing the trace-reversed field $\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma$ allows many trace terms in the linearized Einstein tensor to combine naturally. Because the transformation of $\bar{\gamma}_{ab}$ retains the same structure, one may use the coordinate freedom to impose the Lorenz gauge

$$\partial^a \bar{\gamma}_{ab} = 0.$$

In this gauge, the linearized Einstein tensor collapses to the simple wave equation

$$G^{(1)}_{ab} = -\tfrac{1}{2}\,\Box\,\bar{\gamma}_{ab},$$

so that

$$\Box\,\bar{\gamma}_{ab} = -16\pi T_{ab}.$$

Thus, the Lorenz gauge is a *choice of coordinates* that isolates the true radiative degrees of freedom of the gravitational field and places the linearized Einstein equation in a clean, wave-like form.

The d'Alembertian (Wave Operator)

The symbol $\Box$ denotes the flat-spacetime wave operator (d'Alembertian),

$$\Box \equiv \partial^c \partial_c = \eta^{ab} \partial_a \partial_b.$$

In inertial coordinates,

$$\Box = -\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Thus,

$$-\tfrac{1}{2}\,\Box\,\bar{\gamma}_{ab} = -\frac{1}{2}\,\eta^{cd}\partial_c\partial_d\,\bar{\gamma}_{ab}$$

is simply the Minkowski wave operator acting on $\bar{\gamma}_{ab}$.

4.4.1.6 Linearized Einstein Equation

Einstein's equation to linear order is

$$G^{(1)}_{ab} = 8\pi T_{ab}.$$

Using (4.4.7b), the linearized Einstein equation becomes

$$\partial^c\partial_c\,\bar{\gamma}_{ab} = -16\pi T_{ab}, \tag{4.4.12}$$

which has exactly the form of the wave equation with source.

Thus the trace-reversed metric perturbation $\bar{\gamma}_{ab}$ behaves like a massless spin-2 field propagating on the flat background spacetime (M, η_{ab}).

4.4.2 The Newtonian Limit

The first major test of general relativity is that its predictions must reduce to Newtonian gravity in the regime where Newton's theory is already known to be accurate. This corresponds to the case in which:

- gravitational fields are *weak*,
- sources move slowly compared to the speed of light,
- material stresses are very small compared to mass density,
- a global inertial coordinate system of the background Minkowski metric η_{ab} is available.

Our goal is to show that, under these approximations, the linearized Einstein equation implies:

$$\vec{a} = -\vec{\nabla}\phi,$$

i.e. the Newtonian gravitational force law, where ϕ is the Newtonian gravitational potential satisfying Poisson's equation $\nabla^2\phi = 4\pi\rho$.

Matter in the Newtonian Limit

To connect the linearized Einstein equation with Newtonian gravity, we must understand the form of the stress–energy tensor in the Newtonian regime. The assumptions of this limit are:

- matter moves slowly, so $|\vec{v}| \ll 1$,
- pressure is small relative to rest–mass energy density, $P \ll \rho$,
- gravitational fields are weak, allowing the use of global inertial coordinates,

- proper time τ differs negligibly from coordinate time t.

We begin with the exact stress–energy tensor of a perfect fluid,

$$T_{ab} = (\rho + P)\, u_a u_b + P\, \eta_{ab},$$

and expand each term under the above assumptions.

Four–velocity in the slow–motion limit. In inertial coordinates (t, x^i),

$$u^a = \gamma(1, v^i), \qquad \gamma = (1 - |\vec{v}|^2)^{-1/2}.$$

For $|\vec{v}| \ll 1$, we have $\gamma \approx 1$, so

$$u^a \approx (1, v^i), \qquad u_a = \eta_{ab} u^b \approx (-1, v_i).$$

Dominant energy component. The time–time component becomes

$$T_{00} = (\rho + P) u_0 u_0 + P \eta_{00} \approx \rho(-1)(-1) = \rho,$$

since $P \ll \rho$. This is the dominant term.

Momentum densities. The mixed components satisfy

$$T_{0i} = (\rho + P) u_0 u_i + P \eta_{0i} \approx \rho(-1) v_i = -\rho v_i,$$

which are small because $|v_i| \ll 1$.

Spatial stresses. The purely spatial components are

$$T_{ij} = (\rho + P) u_i u_j + P \eta_{ij} \approx \rho\, v_i v_j + P\, \delta_{ij}.$$

Both terms are negligible compared to ρ, since $v_i v_j \ll 1$ and $P \ll \rho$.

Resulting form. Therefore,

$$T_{00} \approx \rho, \qquad T_{0i} \ll \rho, \qquad T_{ij} \ll \rho.$$

This means that, to excellent approximation, the stress–energy tensor has only one significant component: the mass density measured in an inertial rest frame.

Let $t^a = (\partial/\partial x^0)^a$ be the unit future-directed time vector of that frame. Then

$$t_a = \eta_{ab} t^b = (-1, 0, 0, 0),$$

and the stress–energy tensor takes the Newtonian-limit form

$$T_{ab} \approx \rho\, t_a t_b. \tag{4.4.13}$$

This expresses the fact that slowly moving matter with negligible pressure carries almost exclusively rest–mass energy, with insignificant momentum densities or internal stresses.

Under these assumptions, the linearized Einstein equation

$$\partial^c \partial_c \bar{\gamma}_{ab} = -16\pi T_{ab}$$

simplifies drastically. Since the sources vary slowly, we neglect time derivatives of γ_{ab}. The field equations reduce to:

$$\nabla^2 \bar{\gamma}_{\mu\nu} = 0, \qquad \mu, \nu \neq 0, \tag{4.4.14}$$

$$\nabla^2 \bar{\gamma}_{00} = -16\pi\rho. \tag{4.4.15}$$

The unique solution that vanishes at spatial infinity is $\bar{\gamma}_{\mu\nu} = 0$ for $\mu, \nu \neq 0$.

Recovering the Newtonian Potential

Define the Newtonian potential by

$$\phi \equiv -\frac{1}{4}\bar{\gamma}_{00},$$

so that (4.4.15) becomes Poisson's equation:

$$\nabla^2 \phi = 4\pi\rho. \tag{4.4.17}$$

Using the definition of the trace-reversed perturbation,

$$\gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2}\eta_{ab}\bar{\gamma},$$

one finds that, in the Newtonian limit,

$$\gamma_{ab} = -(4t_a t_b + 2\eta_{ab})\phi. \tag{4.4.16}$$

Geodesic Motion and Newton's Second Law

The motion of a freely falling test particle is governed by the geodesic equation,

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\rho\sigma} \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau} = 0. \tag{4.4.18}$$

To recover Newtonian gravity, we impose the assumptions of the Newtonian regime:

- **Velocities are small:** In the Newtonian limit, the spatial velocity satisfies $|\vec{v}| \ll 1$, meaning that the motion of matter is slow compared with the speed of light (in units where $c = 1$).

- **Weak gravitational fields:** The spacetime metric is only slightly perturbed from flat Minkowski space,
$$g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu}, \qquad |\gamma_{\mu\nu}| \ll 1,$$
so one may treat $\gamma_{\mu\nu}$ as first-order small quantities.
- **Proper time is nearly coordinate time:** For a timelike worldline,
$$d\tau^2 = -g_{\mu\nu} dx^\mu dx^\nu.$$
When $|\vec{v}| \ll 1$ and $g_{00} \approx -1$, this becomes
$$d\tau = dt\sqrt{1 - |\vec{v}|^2 + \mathcal{O}(\gamma)} \approx dt,$$
so τ differs negligibly from t.

With these assumptions, the 4–velocity is
$$u^\mu = \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau} \approx \frac{dx^\mu}{dt},$$
since $dt/d\tau \approx 1$.

Thus
$$u^\mu = \frac{dx^\mu}{d\tau} \approx \left(\frac{dt}{dt}, \frac{dx^i}{dt}\right) = (1,\, v^i),$$
where
$$v^i = \frac{dx^i}{dt}, \qquad |v^i| \ll 1.$$

Lowering the index with the nearly flat metric gives
$$u_\mu = g_{\mu\nu} u^\nu \approx \eta_{\mu\nu} u^\nu = (-1,\, v_i),$$
again up to corrections small in $|\gamma_{\mu\nu}|$ and $|\vec{v}|$.

Step 1: Keeping only leading-order terms. In the geodesic equation, the term
$$\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau}$$
contains products $v^\rho v^\sigma$, which are negligible unless both indices are $\rho = \sigma = 0$. Thus, to leading order,
$$\Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} \approx \Gamma^\mu_{00}\left(\frac{dx^0}{d\tau}\right)^2 \approx \Gamma^\mu_{00}.$$

Therefore the spatial components of (4.4.18) reduce to:
$$\frac{d^2x^i}{d\tau^2} + \Gamma^i_{00} = 0, \qquad (i = 1, 2, 3). \tag{4.1}$$

Using $\tau \approx t$, this becomes
$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{00}. \tag{4.4.19}$$

Step 2: Compute the Christoffel symbol Γ^i_{00}. In the linearized theory,

$$g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu},$$

and the Christoffel symbols are

$$\Gamma^\mu_{\rho\sigma} = \frac{1}{2}\eta^{\mu\lambda}\left(\partial_\rho \gamma_{\sigma\lambda} + \partial_\sigma \gamma_{\rho\lambda} - \partial_\lambda \gamma_{\rho\sigma}\right).$$

Setting $\rho = \sigma = 0$ and taking $\mu = i$ (a spatial index), we obtain

$$\Gamma^i_{00} = \frac{1}{2}\eta^{i\lambda}\left(2\,\partial_0 \gamma_{0\lambda} - \partial_\lambda \gamma_{00}\right).$$

In the Newtonian limit:

- the gravitational field is static $\Rightarrow \partial_0 \gamma_{\mu\nu} = 0$,
- $\eta^{i\lambda}$ picks out $\lambda = i$ with $\eta^{ii} = +1$,

so

$$\Gamma^i_{00} = -\frac{1}{2}\partial_i \gamma_{00}. \tag{*}$$

Step 3: Insert the Newtonian form of the metric. Earlier we showed that in the weak-field limit the metric takes the form

$$\gamma_{00} = -2\phi, \qquad \text{so} \qquad g_{00} = -1 - 2\phi,$$

with ϕ the Newtonian potential.

Substituting $\gamma_{00} = -2\phi$ into (*) gives

$$\Gamma^i_{00} = -\frac{1}{2}\partial_i(-2\phi) = \partial_i \phi. \tag{4.4.20}$$

Step 4: Recover Newton's second law. Plugging (4.4.20) into (4.4.19), we obtain

$$\frac{d^2 x^i}{dt^2} = -\partial_i \phi.$$

In vector notation:

$$\vec{a} = \frac{d^2 \vec{x}}{dt^2} = -\vec{\nabla}\phi. \tag{4.4.21}$$

This is exactly Newton's equation for the gravitational acceleration in a potential ϕ, showing that geodesic motion reproduces Newton's second law in the appropriate limit.

Interpretation

Equations (4.4.17) and (4.4.21) reproduce the foundational equations of Newtonian gravity:

$$\nabla^2\phi = 4\pi\rho, \qquad \vec{a} = -\nabla\phi.$$

Thus, general relativity reduces to Newtonian gravity in the weak-field, slow-motion limit. The physical interpretation, however, is different:

- In Newtonian theory, the Sun exerts a force on the Earth.
- In general relativity, the Sun curves spacetime; the Earth follows a geodesic in this curved geometry.

The predictions agree, but the underlying mechanisms differ fundamentally.

4.4.3 Gravitational Radiation (Vacuum)

One of the most important conceptual changes that occurs when one passes *from Coulomb's electrostatics to Maxwell's electrodynamics* is that the electromagnetic field becomes a genuinely dynamical system: electromagnetic disturbances can propagate freely through spacetime in the form of *electromagnetic radiation.*

A completely analogous phenomenon occurs when one passes from Newtonian gravity to general relativity. In Newtonian theory, gravity is instantaneous: there is no concept of propagating gravitational disturbances. In general relativity, however, gravitational perturbations can propagate as *gravitational radiation*, i.e. ripples in the curvature of spacetime.

In the context of linearized gravity, this dynamical behavior follows directly from the linearized Einstein equation in Lorenz gauge. In vacuum ($T_{ab} = 0$), we have from equations (4.4.11) and (4.4.12)

$$\partial^a\bar{\gamma}_{ab} = 0, \tag{4.4.25}$$

and

$$\partial^c\partial_c\bar{\gamma}_{ab} = 0. \tag{4.4.26}$$

Thus each component of the trace-reversed metric perturbation $\bar{\gamma}_{ab}$ satisfies the *source-free wave equation* on the flat background spacetime (M, η_{ab}).

Consequently, in the linear approximation, gravitational disturbances propagate at the speed of light and behave mathematically like a massless spin-2 field. This is the gravitational analogue of the propagation of electromagnetic waves in Maxwell's theory.

4.4.3.1 Gauge Freedom in Linearized Gravity

In linearized gravity we write the metric as

$$g_{ab} = \eta_{ab} + \gamma_{ab}, \qquad |\gamma_{ab}| \ll 1,$$

and consider infinitesimal coordinate transformations

$$x^a \to x'^a = x^a + \xi^a(x),$$

with ξ^a small. To first order in ξ^a and γ_{ab}, the metric perturbation transforms as

$$\gamma'_{ab} = \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a. \tag{4.4.G1}$$

This is the *gauge freedom* of linearized gravity. Different choices of ξ^a correspond to different coordinate descriptions of the same physical geometry. Consequently, many of the components of γ_{ab} have no physical meaning; they can be changed arbitrarily through suitable choices of ξ^a.

It is therefore essential to impose gauge conditions that eliminate non-physical components. We now develop a systematic procedure for doing so.

4.4.3.2 Lorenz Gauge: First Stage of Simplification

Define the trace-reversed perturbation

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma, \qquad \gamma = \gamma^c{}_c.$$

The *Lorenz gauge* condition is

$$\partial^a \bar{\gamma}_{ab} = 0. \tag{4.4.25}$$

This is directly analogous to the electromagnetic Lorenz gauge $\partial^a A_a = 0$. In this gauge, the linearized Einstein equation reduces to the simple wave equation

$$\Box \bar{\gamma}_{ab} = -16\pi T_{ab}, \qquad \Box \equiv \partial^c \partial_c.$$

Gauge freedom remaining in Lorenz gauge. Applying a gauge transformation (4.4.G1),

$$\begin{aligned}
\overline{\gamma}'_{ab} &= \gamma'_{ab} - \frac{1}{2}\eta_{ab}\gamma' \\
&= (\overline{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2}\eta_{ab}\gamma'^c_c \\
&= (\overline{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2}\eta_{ab}\eta^{cd}\gamma'_{cd} \\
&= (\overline{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2}\eta_{ab}\eta^{cd}(\gamma_{cd} + \partial_c \xi_d + \partial_d \xi_c) \\
&= (\overline{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2}\eta_{ab}(\gamma + \partial^c \xi_c + \partial^d \xi_d) \\
&= (\overline{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2}\eta_{ab}(\gamma + 2\partial^c \xi_c) \\
&= (\gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma) + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab}\partial^c \xi_c
\end{aligned}$$

so the trace-reversed field transforms as,

$$\bar{\gamma}'_{ab} = \bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab}\, \partial^c \xi_c. \tag{4.4.G2}$$

Imposing the Lorenz condition on the transformed field,

$$\begin{aligned}
\partial^\alpha \bar{\gamma}'_{ab} &= \partial^a \bar{\gamma}_{ab} + \partial^a \partial_a \xi_b + \partial^a \partial_b \xi_a - \partial^a \eta_{ab} \partial^c \xi_c \\
&= \partial^a \bar{\gamma}_{ab} + \Box \xi_b + \partial_b \partial^a \xi_a - \eta_{ab} \partial^a \partial^c \xi_c \\
&= \partial^a \bar{\gamma}_{ab} + \Box \xi_b + \partial_b \partial^a \xi_a - \partial_b \partial^c \xi_c
\end{aligned}$$

and recognizing that $\partial_b \partial^a \xi_a = \partial_b \partial^c \xi_c$ gives,

$$\partial^a \bar{\gamma}'_{ab} = \partial^a \bar{\gamma}_{ab} + \Box \xi_b.$$

Thus the Lorenz condition is preserved if and only if

$$\Box \xi_b = 0. \tag{4.4.27}$$

Interpretation. Lorenz gauge eliminates four combinations of γ_{ab}, but it leaves intact all gauge transformations generated by vector fields ξ^a satisfying the homogeneous wave equation (4.4.27). These constitute the *residual gauge freedom*.

This freedom will now be used to impose further gauge conditions.

4.4.3.3 The Radiation Gauge

In a vacuum region ($T_{ab} = 0$), the physically relevant part of the field is the radiative portion—the transverse, traceless components that propagate at the speed of light. To isolate these components, it is advantageous to impose a gauge analogous to the Coulomb (radiation) gauge of electromagnetism, where one sets

$$A_0 = 0, \qquad \nabla \cdot \vec{A} = 0.$$

In gravity, the useful analogue is to choose coordinates so that

$$\gamma = 0, \qquad \gamma_{0\mu} = 0. \tag{4.4.RG1}$$

The first condition removes the trace; the second removes the time–space components. Combined with Lorenz gauge, these reduce the 10 components of γ_{ab} to the 2 physical polarizations of gravitational waves.

However, imposing (4.4.RG1) requires solving explicitly for the gauge vector ξ^a. We now derive the necessary equations.

Deriving the Radiation Gauge Conditions

Under a gauge transformation,

$$\gamma'_{ab} = \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a,$$

the trace becomes

$$\gamma' = \gamma + 2(\partial^a \xi_a) = \gamma + 2(-\partial_t \xi_0 + \vec{\nabla} \cdot \vec{\xi}).$$

Requiring $\gamma' = 0$ yields

$$2\Big(-\partial_t \xi_0 + \vec{\nabla} \cdot \vec{\xi}\Big) = -\gamma. \tag{4.4.34a}$$

Similarly, for the time–space components,

$$\gamma'_{0\mu} = \gamma_{0\mu} + \partial_0 \xi_\mu + \partial_\mu \xi_0,$$

and requiring $\gamma'_{0\mu} = 0$ gives

$$\partial_t \xi_\mu + \partial_\mu \xi_0 = -\gamma_{0\mu}, \qquad \mu = 1, 2, 3. \tag{4.4.34c}$$

Differentiating (4.4.34a) and (4.4.34c) yields second-order elliptic equations for ξ_0 and ξ_μ:

$$2\Big[-\nabla^2 \xi_0 + \vec{\nabla} \cdot (\partial_t \vec{\xi})\Big] = -\partial_t \gamma, \tag{4.4.34b}$$

$$\nabla^2 \xi_\mu + \partial_\mu (\partial_t \xi_0) = -\partial_t \gamma_{0\mu}. \tag{4.4.34d}$$

Equations (4.4.34a)–(4.4.34d) determine the values of ξ^a and its time derivatives on an initial hypersurface $t = t_0$.

Extension into spacetime. To extend ξ^a off the initial slice, we use the fact that residual gauge freedom requires

$$\Box \xi_a = 0,$$

so ξ^a with the initial data determined by (4.4.34a)–(4.4.34d) is evolved throughout the vacuum region by solving the wave equation.

Final result. With this choice of ξ^a, the gauge-transformed field satisfies:

$$\partial^a \bar{\gamma}'_{ab} = 0, \qquad \gamma' = 0, \qquad \gamma'_{0\mu} = 0.$$

This is the *radiation gauge.* In vacuum, the remaining components γ'_{ij} contain only the two transverse, traceless gravitational-wave polarizations.

What a Gauge Condition Really Means
In linearized gravity, the metric perturbation γ_{ab} contains far more mathematical components than there are physical degrees of freedom. A *gauge condition* eliminates this redundancy by selecting one convenient representative from a whole family of equivalent descriptions. This idea has simple analogues in elementary mathematics:

1. Coordinate Redundancy (polar vs. Cartesian). A point in the plane can be written as (r, θ), but (r, θ) and $(r, \theta + 2\pi)$ describe the *same* point. Choosing $0 \leq \theta < 2\pi$ removes this redundancy. Likewise, many tensors γ_{ab} describe the same physical spacetime; a gauge condition chooses a standard representative.

2. Electromagnetic Potentials. The electric and magnetic fields are unchanged when $A_a \to A_a + \partial_a \chi$. The potentials contain "pure gauge" information that does not affect physics. One imposes the Lorenz or Coulomb gauge to remove this ambiguity. In linearized gravity, $\gamma_{ab} \to \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a$ plays the same role.

3. Fixing an Equivalence Class. Consider all functions differing by $ax+b$: $f(x) \sim f(x)+ax+b$. These represent the same physical object. Imposing $f(0) = 0$ and $f'(0) = 0$ selects one representative. In gravity, gauge conditions such as $\partial^a \bar{\gamma}_{ab} = 0$ (Lorenz gauge) or $\gamma = 0$, $\gamma_{0\mu} = 0$ (radiation gauge) perform the same task for γ_{ab}.

Summary. A gauge condition removes the non-physical freedom in γ_{ab}, just as choosing an angular range in polar coordinates or fixing an electromagnetic potential removes representational ambiguity. Different gauge choices are simply different coordinate conventions for expressing the *same* physical gravitational field.

Wave Equation in Radiation Gauge

Starting from the linearized Einstein equation in Lorenz gauge,

$$\Box \bar{\gamma}_{ab} = -16\pi T_{ab}, \tag{4.4.12}$$

and restricting to a vacuum region where $T_{ab} = 0$, we obtain the source-free wave equation

$$\Box \bar{\gamma}_{ab} = 0. \tag{4.4.12*}$$

In the radiation gauge we impose (cf. the construction leading to equations (4.4.34a)–(4.4.34d))

$$\gamma = 0, \qquad \gamma_{0\mu} = 0.$$

When the trace vanishes, the trace-reversed perturbation reduces to the original one. Using Wald's trace-reversal definition (4.4.6),

$$\bar{\gamma}_{ab} = \gamma_{ab} - \tfrac{1}{2}\eta_{ab}\gamma, \tag{4.4.6}$$

we have, in radiation gauge,

$$\bar{\gamma}_{ab} = \gamma_{ab}. \tag{4.4.6*}$$

Thus in this gauge the vacuum equation (4.4.12*) becomes

$$\Box \gamma_{ab} = 0. \tag{4.4.12**}$$

Because the radiation gauge sets all components $\gamma_{0\mu}$ to zero and preserves this condition under evolution, the only nonvanishing dynamical components are the spatial ones. Therefore, for $\mu, \nu = 1, 2, 3$, we obtain the simple wave equation

$$\partial^c \partial_c \, \gamma_{\mu\nu} = 0, \qquad (\mu, \nu = 1, 2, 3), \tag{4.4.37}$$

which governs the propagation of the physical gravitational-wave degrees of freedom.

Why Certain Components Become Nonphysical in the Linearized, Vacuum Regime
In a region far from sources, the linearized Einstein equation reduces to

$$\Box \bar{\gamma}_{ab} = 0,$$

together with the Lorenz gauge constraint

$$\partial^a \bar{\gamma}_{ab} = 0.$$

These two relations dramatically restrict which components of the metric perturbation can represent genuine curvature. In this regime, the mixed space–time components $\gamma_{0\mu}$ (and the trace γ) solve the *homogeneous* wave equation but do not contribute to the transverse, propagating degrees of freedom allowed in vacuum. Their solutions describe non-radiative, longitudinal distortions of the metric which carry *no physical curvature.*
Because these components have no dynamical or gauge-invariant influence on the tidal field measured by freely falling observers, they become "pure gauge": they can be eliminated by an infinitesimal coordinate transformation without changing any physical observable. Consequently, in this setting one can impose the radiation gauge,

$$\gamma = 0, \qquad \gamma_{0\mu} = 0,$$

leaving only the transverse, traceless spatial components $\gamma_{ij}^{\mathrm{TT}}$. These two remaining degrees of freedom correspond to the physical $+$ and $\times$ polarizations of a massless spin-2 field.
Summary. The combined effect of (i) being in vacuum, (ii) linearizing the equations, and (iii) exploiting residual gauge freedom causes the non-radiative parts of the metric perturbation to drop out entirely. What remains are only the gauge-invariant, propagating components that constitute true gravitational radiation.

Plane-Wave Solutions in Radiation Gauge

In the radiation gauge ($\gamma = 0$, $\gamma_{0\mu} = 0$) and in vacuum ($T_{ab} = 0$), the linearized Einstein equations reduce to the simple wave equation

$$\partial^c \partial_c \, \gamma_{\mu\nu} = 0, \qquad (\mu, \nu = 1, 2, 3). \tag{4.4.37}$$

We look for plane-wave solutions of the form

$$\gamma_{\mu\nu}(x) = \Re\Big\{ A_{\mu\nu} \, e^{i k_a x^a} \Big\}, \tag{4.2}$$

where $A_{\mu\nu}$ is a constant polarization tensor and k_a is a constant wavevector. It is sufficient to work with the complex field

$$\tilde{\gamma}_{\mu\nu}(x) = A_{\mu\nu}e^{ik_a x^a},$$

since the real part may be taken at the end.

First derivative.

$$\partial_c\tilde{\gamma}_{\mu\nu} = \partial_c\Big(A_{\mu\nu}e^{ik_a x^a}\Big) = A_{\mu\nu}\,\partial_c e^{ik_a x^a} = ik_c\,A_{\mu\nu}e^{ik_a x^a}. \tag{4.3}$$

Second derivative. Applying $\partial^c = \eta^{cd}\partial_d$ gives

$$\partial^c\partial_c\tilde{\gamma}_{\mu\nu} = \eta^{cd}\partial_d\left(ik_c A_{\mu\nu}e^{ik_a x^a}\right) = i\,\eta^{cd}k_c A_{\mu\nu}\partial_d e^{ik_a x^a} \tag{4.4}$$

$$= i\,\eta^{cd}k_c A_{\mu\nu}\left(ik_d e^{ik_a x^a}\right) \tag{4.5}$$

$$= -\,\eta^{cd}k_c k_d\,A_{\mu\nu}e^{ik_a x^a} \tag{4.6}$$

$$= -\,k_a k^a\,A_{\mu\nu}e^{ik_b x^b}. \tag{4.7}$$

Substituting into the wave equation. Equation (4.4.37) requires

$$\partial^c\partial_c\tilde{\gamma}_{\mu\nu} = 0,$$

so the above expression implies

$$-\,k_a k^a A_{\mu\nu}e^{ik_b x^b} = 0.$$

For a nontrivial solution ($A_{\mu\nu} \neq 0$), this must hold for all x, and therefore

$$k_a k^a = 0. \tag{4.4.38}$$

Thus the wavevector is null, showing that linearized gravitational perturbations in vacuum propagate at the speed of light and behave like a massless spin-2 field.

Gauge conditions. In radiation gauge we impose

$$\gamma = 0, \qquad \gamma_{0\mu} = 0,$$

and we now determine what these conditions imply for the polarization tensor of a plane-wave solution.

We seek plane-wave solutions of the form

$$\gamma_{\mu\nu}(x) = A_{\mu\nu}\,e^{ik_a x^a},$$

where $A_{\mu\nu}$ is a constant symmetric tensor and k_a is a constant wavevector. Substituting this ansatz into the gauge conditions produces purely algebraic constraints on $A_{\mu\nu}$.

(1) Tracelessness. The condition $\gamma = \gamma^{\mu}{}_{\mu} = 0$ becomes

$$\gamma^{\mu}{}_{\mu} = A^{\mu}{}_{\mu}\, e^{ik\cdot x} = 0.$$

Since the exponential never vanishes, we obtain

$$A^{\mu}{}_{\mu} = 0. \tag{4.4.37a}$$

(2) Vanishing time–space components. The radiation gauge requires

$$\gamma_{0\mu} = 0.$$

Substituting the plane-wave form gives

$$\gamma_{0\mu} = A_{0\mu}\, e^{ik\cdot x} = 0,$$

and hence

$$A_{0\mu} = 0. \tag{4.4.37b}$$

Thus the polarization tensor has no components with a time index.

(3) Lorenz gauge transversality. The Lorenz gauge condition,

$$\partial^{a}\bar{\gamma}_{ab} = 0,$$

reduces for a plane wave to the algebraic condition

$$k^{a}A_{ab} = 0. \tag{4.4.37c}$$

We now apply this to determine which *spatial* components of $A_{\mu\nu}$ are allowed.

Choose coordinates so that the wave propagates in the $+z$ direction. Then a null wavevector has the form

$$k^{\mu} = (\omega,\, 0,\, 0,\, \omega)$$

with $k_a k^a = 0$. Applying the transversality condition gives

$$k^{\mu}A_{\mu\nu} = \omega\,(A_{0\nu} + A_{3\nu}) = 0.$$

Using (4.4.37b), $A_{0\nu} = 0$, this reduces to

$$A_{3\nu} = 0. \tag{4.4.37d}$$

Resulting form of the polarization tensor. The conditions (4.4.37a)–(4.4.37d) imply that

$$A_{0\mu} = 0, \qquad A_{3\mu} = 0,$$

so the only nonzero components of $A_{\mu\nu}$ lie in the x–y plane:

$$A_{11}, \quad A_{22}, \quad A_{12} = A_{21}.$$

The tracelessness condition $A^1{}_1 + A^2{}_2 = 0$ then leaves exactly two independent components. These correspond to the two physical polarization states of a gravitational wave.

In addition, the residual Lorenz gauge condition,

$$\partial^a \bar{\gamma}_{ab} = 0, \tag{4.4.25}$$

reduces in vacuum to the transversality condition

$$k^\mu A_{\mu\nu} = 0. \tag{4.4.40}$$

Counting independent components. The symmetric 3×3 tensor $A_{\mu\nu}$ initially has six independent components. The transversality condition (4.4.40) imposes three independent constraints. The traceless condition (4.4.39) removes one further degree of freedom. Hence

$$6 \;-\; 3 \;-\; 1 \;=\; 2$$

independent degrees of freedom remain.

These two degrees of freedom correspond to the two physical polarization states of the gravitational field.

Explicit polarization tensors. Consider a wave propagating in the $+z$-direction, so that

$$k^a = \omega(1, 0, 0, 1).$$

The transversality condition $k^\mu A_{\mu\nu} = 0$ forces all components with a z index to vanish. The only nonzero components are the spatial "transverse" ones in the x and y directions.

A convenient basis of polarization tensors is

$$\varepsilon^{(+)}_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \varepsilon^{(\times)}_{\mu\nu} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{4.4.41}$$

with indices restricted to $\mu, \nu = x, y$.

These are the familiar "plus" and "cross" polarization states of a gravitational wave.

From the polarization tensor to the TT metric perturbation. The analysis above shows that, for a plane-wave solution of the linearized Einstein equation in radiation gauge, the metric perturbation takes the form

$$\gamma_{\mu\nu}(x) = A_{\mu\nu}\, e^{ik\cdot x},$$

with the polarization tensor $A_{\mu\nu}$ satisfying

$$A_{0\mu} = 0, \qquad A_{3\mu} = 0, \qquad A^\mu{}_\mu = 0.$$

These are precisely the *transverse* and *traceless* conditions defining the TT gauge. Consequently, the spatial components of the plane-wave perturbation are

$$\gamma_{ij}(t, z) = A_{ij}\, e^{i\omega(t-z)}, \qquad i, j = 1, 2,$$

and we may identify

$$\gamma_{ij}^{\mathrm{TT}}(t,z) \equiv \gamma_{ij}(t,z) = A_{ij}\, e^{i\omega(t-z)}. \qquad \text{(TT-identification)}$$

Thus the algebraic object A_{ij} obtained from the gauge conditions is *precisely* the physical TT metric perturbation that enters observable quantities such as the Riemann tensor and the geodesic deviation equation. Only after this identification does the curvature tensor acquire nonzero, measurable components, leading to the tidal accelerations characteristic of gravitational waves.

Transverse–Traceless (TT) Gauge.
In the plane-wave analysis, a gravitational wave solution takes the form

$$\gamma_{ij}(x) = A_{ij}\, e^{ik\cdot x},$$

where A_{ij} is a constant symmetric polarization tensor. After imposing the full set of radiation–gauge conditions

$$\gamma_{0\mu} = 0, \qquad \gamma = 0, \qquad k^i A_{ij} = 0,$$

the tensor A_{ij} satisfies the same transverse and traceless constraints. Thus the *physical* metric perturbation is encoded entirely in the *transverse–traceless* spatial tensor

$$\gamma_{ij}^{\mathrm{TT}}(x) = A_{ij}\, e^{ik\cdot x}.$$

In practical applications—such as computing tidal forces or interferometer strain—the spatial dependence is irrelevant and the oscillatory factor is absorbed into the time-dependent amplitudes $h_+(t)$ and $h_\times(t)$. The remaining tensor A_{ij} represents the *polarization pattern* of the wave. For a wave propagating in the $+z$–direction, the TT tensor takes the standard form

$$\gamma_{ij}^{\mathrm{TT}}(t) = \begin{pmatrix} h_+(t) & h_\times(t) & 0 \\ h_\times(t) & -h_+(t) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $h_+(t)$ and $h_\times(t)$ encode the observable oscillations of the two polarization states of a massless spin-2 field.
In TT gauge the entire physical content of a gravitational wave—its tidal stretching and squeezing of freely falling test masses—is contained in this 3×3 spatial tensor.

4.4.3.4 Effect on Free Test Particles: Geodesic Deviation

To see the physical effect of gravitational waves, we study the relative motion of nearby freely falling test particles. Their separation is governed by the geodesic deviation equation,

$$\frac{D^2 X^a}{d\tau^2} = -R^a{}_{bcd}\, u^b X^c u^d,$$

where u^a is the four-velocity of the reference geodesic and X^a is the separation vector to a neighboring geodesic.

In the weak-field, slow-motion limit appropriate for gravitational-wave detectors, we may choose coordinates such that, in the absence of the wave, the test particles are at rest:

$$u^a \approx t^a = (1,0,0,0), \qquad X^a = (0, X^i),$$

and proper time τ can be identified with coordinate time t.

With these choices, the spatial components of geodesic deviation reduce to

$$\frac{d^2X^i}{dt^2} = -R^i{}_{0j0}\, X^j. \tag{4.4.42}$$

Riemann tensor in the radiation gauge. In linearized gravity, the Riemann tensor to first order in γ_{ab} is

$$R^{(1)}_{abcd} = \frac{1}{2}(\partial_c\partial_b\gamma_{ad} + \partial_d\partial_a\gamma_{bc} - \partial_c\partial_a\gamma_{bd} - \partial_d\partial_b\gamma_{ac}).$$

In the radiation gauge for a wave in vacuum we have

$$\gamma_{00} = 0, \qquad \gamma_{0i} = 0, \qquad \gamma_{ij} = \gamma^{\rm TT}_{ij}(t,z),$$

with $\gamma^{\rm TT}_{ij}$ transverse and traceless.

Taking $a=0$, $b=i$, $c=0$, $d=j$, we obtain

$$\begin{aligned} R^{(1)}_{0i0j} &= \frac{1}{2}(\partial_0\partial_i\gamma_{0j} + \partial_j\partial_0\gamma_{0i} - \partial_0\partial_0\gamma_{ij} - \partial_j\partial_i\gamma_{00}) \\ &= -\frac{1}{2}\,\partial_0^2\gamma_{ij}, \end{aligned}$$

since $\gamma_{00} = \gamma_{0i} = 0$ in radiation gauge. Thus

$$R^{(1)}_{0i0j} = -\frac{1}{2}\,\ddot{\gamma}_{ij}, \tag{4.4.43}$$

where dots denote derivatives with respect to t.

Raising the first index with η^{ik}, we have

$$R^i{}_{0j0} = \eta^{ik} R_{k0j0} = -\frac{1}{2}\,\ddot{\gamma}^i{}_j.$$

Relative acceleration of test particles. Substituting this into the geodesic deviation equation (4.4.42) gives

$$\frac{d^2X^i}{dt^2} = \frac{1}{2}\,\ddot{\gamma}^i{}_j\, X^j. \tag{4.4.44}$$

Thus a passing gravitational wave produces a time-dependent tidal acceleration on the separation vector X^i between neighboring freely falling particles. In particular, for a wave propagating in the $+z$-direction and a ring of test particles lying in the x–y plane, the "plus" and "cross" polarization states $\gamma^{\rm TT}_{ij}$ produce the characteristic alternating stretching and squeezing patterns associated with gravitational waves.

4.4.3.5 Response of an Interferometric Detector

To connect the geodesic deviation result with real gravitational-wave experiments, consider a simple model of an interferometric detector. Two test masses are placed at fixed coordinate positions along orthogonal spatial arms (e.g. the x-arm and the y-arm). In the absence of gravitational waves, the proper lengths of these arms remain constant. A passing gravitational wave influences the proper distance between freely falling test masses.

Assume a plane gravitational wave traveling in the $+z$-direction, written in TT gauge as

$$\gamma_{ij}^{\mathrm{TT}}(t-z) = h_+(t-z)\,\varepsilon_{ij}^{(+)} + h_\times(t-z)\,\varepsilon_{ij}^{(\times)},$$

with polarization tensors

$$\varepsilon_{ij}^{(+)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \varepsilon_{ij}^{(\times)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Because the TT gauge is constructed so that freely falling detectors remain at constant spatial coordinates, the coordinate positions of the end masses do not change. Instead, the *proper distance* between them changes.

For a mass separated by a vector X^i from the beam splitter, the geodesic deviation equation gives

$$\frac{d^2X^i}{dt^2} = \frac{1}{2}\,\ddot{\gamma}^{\mathrm{TT}\,i}{}_j\,X^j.$$

Integrating twice (and assuming the wave amplitude is small and the initial velocity is zero) yields

$$\delta X^i(t) = \frac{1}{2}\,\gamma^{\mathrm{TT}\,i}{}_j(t)\,X_0^j,$$

where X_0^j is the unperturbed separation.

Effect on detector arms. Take an interferometer with arms of coordinate lengths

$$L_x = L, \qquad L_y = L,$$

lying along the x- and y-axes. The proper length of the x-arm becomes

$$L_x(t) = L + \delta L_x(t) = L + \tfrac{1}{2}\,\gamma_{xx}^{\mathrm{TT}}(t)\,L,$$

and similarly for the y-arm.

For a $+$-polarized wave,

$$\gamma_{xx}^{\mathrm{TT}} = h_+, \qquad \gamma_{yy}^{\mathrm{TT}} = -h_+,$$

so the arm lengths change in opposite directions:

$$\frac{\delta L_x}{L} = \tfrac{1}{2}\,h_+, \qquad \frac{\delta L_y}{L} = -\tfrac{1}{2}\,h_+.$$

Detected strain. Interferometers measure the *difference* in fractional arm-length changes:

$$h(t) = \frac{\delta L_x}{L} - \frac{\delta L_y}{L} = h_+(t).$$

Thus the measurable strain of a detector is directly the amplitude of the transverse–traceless metric perturbation:

$$h(t) = \gamma_{xx}^{\rm TT}(t) - \gamma_{yy}^{\rm TT}(t).$$

This linear relationship between detector response and wave amplitude is the foundation for gravitational-wave interferometry, as implemented in LIGO, VIRGO, KAGRA, and similar observatories.

4.4.4 Gravitational Radiation (Sourced)

In the previous sections we analyzed vacuum solutions of the linearized Einstein equation, $\Box\bar{\gamma}_{ab} = 0$, and showed that they describe freely propagating gravitational waves in transverse–traceless form. We now turn to the more general case in which matter is present. In linearized gravity, the sourced field equation

$$\Box\bar{\gamma}_{ab} = -16\pi T_{ab}$$

is a set of inhomogeneous wave equations, entirely analogous to Maxwell's equation for the electromagnetic four-potential. The physically relevant solution is the *retarded* one, which expresses the metric perturbation at a point in terms of the past behavior of the stress–energy tensor. This retarded integral provides the bridge between matter dynamics in the near zone and gravitational radiation observed in the far zone.

Retarded Green's function. The wave equation for a scalar field ψ,

$$\partial^c\partial_c\psi = -4\pi f,$$

admits the retarded solution

$$\psi(t,\vec{x}) = \int d^3x' \, \frac{f\big(t_{\rm ret},\vec{x}'\big)}{|\vec{x}-\vec{x}'|}, \qquad t_{\rm ret} = t - |\vec{x}-\vec{x}'|.$$

This follows from the retarded Green's function in flat spacetime,

$$G_{\rm ret}(x,x') = \frac{\delta\big(t' - t + |\vec{x}-\vec{x}'|\big)}{|\vec{x}-\vec{x}'|}.$$

Since each component $\bar{\gamma}_{ab}$ satisfies the same wave equation (4.4.12) with source $-16\pi T_{ab}$, we may write the solution by direct analogy:

$$\bar{\gamma}_{ab}(t,\vec{x}) = 4\int d^3x' \, \frac{T_{ab}\big(t_{\rm ret},\vec{x}'\big)}{|\vec{x}-\vec{x}'|}, \qquad t_{\rm ret} = t - |\vec{x}-\vec{x}'|. \tag{4.4.45}$$

This is the *retarded solution* of the linearized Einstein equation: the metric perturbation at $(t,\vec{x})$ is determined by the past behavior of the stress–energy tensor evaluated on the past light cone of that point.

Fourier transform. To obtain Wald's Eq. 4.4.44, we Fourier transform the retarded solution Eq. 4.4.45 in time only. Insert Eq. 4.4.45 into the Fourier transform:

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega,\vec{x}) = \frac{4}{\sqrt{2\pi}} \int d^3x' \frac{1}{|\vec{x}-\vec{x}'|} \int_{-\infty}^{\infty} dt\, T_{\mu\nu}(t-|\vec{x}-\vec{x}'|,\vec{x}')\, e^{i\omega t}.$$

Make the substitution $u = t - |\vec{x}-\vec{x}'|$, so $t = u + |\vec{x}-\vec{x}'|$ and $dt = du$. Then

$$\int dt\, T_{\mu\nu}(t-|\vec{x}-\vec{x}'|,\vec{x}')\, e^{i\omega t} = e^{\,i\omega|\vec{x}-\vec{x}'|} \int_{-\infty}^{\infty} T_{\mu\nu}(u,\vec{x}')\, e^{i\omega u} du.$$

Using the Fourier transform of the stress–energy tensor,

$$\hat{T}_{\mu\nu}(\omega,\vec{x}') = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} T_{\mu\nu}(u,\vec{x}')\, e^{i\omega u} du,$$

the inner integral becomes

$$\int T_{\mu\nu}(u,\vec{x}') e^{i\omega u} du = \sqrt{2\pi}\, \hat{T}_{\mu\nu}(\omega,\vec{x}').$$

Substituting back and cancelling the factors of $\sqrt{2\pi}$, we obtain Wald's eq. (4.4.44):

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega,\vec{x}) = 4 \int d^3x'\, \frac{\hat{T}_{\mu\nu}(\omega,\vec{x}')}{|\vec{x}-\vec{x}'|} e^{\,i\omega|\vec{x}-\vec{x}'|}. \tag{4.4.44}$$

Gauge relation for the time-space components. Wald uses the linearized conservation law $\partial^a T_{ab} = 0$ together with the Lorenz gauge condition to deduce his key relation:

$$-i\omega\, \hat{\bar{\gamma}}_{0\mu} = \sum_{\nu=1}^{3} \frac{\partial}{\partial x^\nu} \hat{\bar{\gamma}}_{\nu\mu}. \tag{4.4.45}$$

Thus the spatial derivatives of the $\bar{\gamma}_{ij}$ determine $\bar{\gamma}_{0i}$, and only the *spatial–spatial* components must be solved explicitly.

Slow-Motion Sources and the Quadrupole Approximation

We now evaluate the retarded solution (4.4.42) in the regime of *slow-moving, spatially compact* sources,

$$|\vec{v}| \ll 1, \qquad \text{source size } L \ll \text{radiation distance } R.$$

Far-zone expansion of the spatial components. In the radiation zone ($R \gg L$) the phase factor $e^{i\omega|\vec{x}-\vec{x}'|}$ can be approximated by

$$e^{i\omega|\vec{x}-\vec{x}'|} \approx e^{i\omega R},$$

and pulled out of the integral.

We now fill in the steps leading to Wald's relation

$$\int \hat{T}^{\mu\nu} d^3x = -\frac{\omega^2}{2} \int \hat{T}^{00}\, x^\mu x^\nu\, d^3x, \tag{4.4.46}$$

showing how it follows from stress–energy conservation and integration by parts.

Step 1: Start from local conservation. The local conservation law is

$$\partial_a T^{ab} = 0 \quad \Rightarrow \quad \partial_t T^{0b} + \partial_k T^{kb} = 0,$$

where Latin spatial indices $k = 1, 2, 3$ and $b = 0, 1, 2, 3$.

Step 2: First integration by parts (one time derivative).

Take $b = 0$ and multiply by $x^\mu x^\nu$, then integrate over space:

$$\int d^3x\, x^\mu x^\nu \left(\partial_t T^{00} + \partial_k T^{k0}\right) = 0.$$

The first term gives

$$\int d^3x\, x^\mu x^\nu\, \partial_t T^{00} = \frac{d}{dt}\left(\int d^3x\, x^\mu x^\nu T^{00}\right),$$

since $x^\mu x^\nu$ are time-independent.

For the second term, integrate by parts in space:

$$\int d^3x\, x^\mu x^\nu \partial_k T^{k0} = \int d^3x\, \partial_k\left(x^\mu x^\nu T^{k0}\right) - \int d^3x\, \left(\partial_k x^\mu x^\nu\right) T^{k0}.$$

The divergence term is a surface integral at infinity and vanishes if T^{k0} has compact support or decays fast enough. Using $\partial_k(x^\mu x^\nu) = \delta^\mu_k x^\nu + \delta^\nu_k x^\mu$, we obtain

$$\int d^3x\, x^\mu x^\nu \partial_k T^{k0} = -\int d^3x\, \left(x^\nu T^{\mu 0} + x^\mu T^{\nu 0}\right).$$

The conservation law therefore gives

$$\frac{d}{dt}\left(\int d^3x\, x^\mu x^\nu T^{00}\right) = \int d^3x\, \left(x^\nu T^{\mu 0} + x^\mu T^{\nu 0}\right). \tag{4.4.46*}$$

Step 3: Second integration by parts (two time derivatives).

Differentiate Eq. 4.4.46* with respect to time:

$$\frac{d^2}{dt^2}\left(\int d^3x\, x^\mu x^\nu T^{00}\right) = \int d^3x\, \left(x^\nu \partial_t T^{\mu 0} + x^\mu \partial_t T^{\nu 0}\right).$$

Now use conservation again, this time with $b = \mu$ and $b = \nu$:

$$\partial_t T^{\mu 0} = -\partial_k T^{k\mu}, \qquad \partial_t T^{\nu 0} = -\partial_k T^{k\nu}.$$

Thus

$$\frac{d^2}{dt^2}\left(\int d^3x\, x^\mu x^\nu T^{00}\right) = -\int d^3x\,(x^\nu \partial_k T^{k\mu} + x^\mu \partial_k T^{k\nu}).$$

Integrate by parts once more in space. For the first term,

$$\int d^3x\, x^\nu \partial_k T^{k\mu} = \int d^3x\, \partial_k(x^\nu T^{k\mu}) - \int d^3x\,(\partial_k x^\nu) T^{k\mu} = -\int d^3x\, T^{\nu\mu},$$

again dropping the surface term. Similarly,

$$\int d^3x\, x^\mu \partial_k T^{k\nu} = -\int d^3x\, T^{\mu\nu}.$$

Therefore

$$\frac{d^2}{dt^2}\left(\int d^3x\, x^\mu x^\nu T^{00}\right) = 2\int d^3x\, T^{\mu\nu}.$$

Equivalently,

$$\int d^3x\, T^{\mu\nu}(t,\vec{x}) = \frac{1}{2}\frac{d^2}{dt^2}\left(\int d^3x\, x^\mu x^\nu T^{00}(t,\vec{x})\right). \qquad (4.4.46^\dagger)$$

Step 4: Go to Fourier space.

Define the temporal Fourier transform (Wald's eq. (4.4.43))

$$\hat{T}^{ab}(\omega,\vec{x}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} T^{ab}(t,\vec{x})\, e^{i\omega t}\, dt.$$

Fourier transforming Eq. $4.4.46^\dagger$ with respect to t, and using

$$\frac{d^2}{dt^2} \longrightarrow -\omega^2 \quad \text{in Fourier space,}$$

we obtain

$$\int d^3x\, \hat{T}^{\mu\nu}(\omega,\vec{x}) = -\frac{\omega^2}{2}\int d^3x\, x^\mu x^\nu \hat{T}^{00}(\omega,\vec{x}),$$

which is precisely Wald's eq. (4.4.46):

$$\int \hat{T}^{\mu\nu} d^3x = -\frac{\omega^2}{2}\int \hat{T}^{00}\, x^\mu x^\nu\, d^3x. \qquad (4.4.46)$$

Thus (4.4.46) is just the Fourier-space version of the time-domain statement that the *spatial components* of the stress–energy tensor are determined by the *second time derivative of the mass quadrupole moment* $\int T^{00} x^\mu x^\nu d^3x$.

Far-zone result for the Fourier-space field. Substituting (4.4.46) into (4.4.44), Wald arrives at the far-zone field

$$\hat{\bar{\gamma}}_{\mu\nu}(\omega, \vec{x}) = -\frac{2\omega^2}{3R}\, e^{i\omega R}\, \hat{q}_{\mu\nu}(\omega), \qquad (\mu, \nu = 1, 2, 3), \tag{4.4.47}$$

where $\hat{q}_{\mu\nu}$ is the Fourier transform of the mass quadrupole moment tensor.

The quadrupole moment. Wald defines (eq. 4.4.48)

$$q_{\mu\nu}(t) = 3 \int T^{00}(t, \vec{x})\, x_\mu x_\nu \, d^3x. \tag{4.4.48}$$

Inverse Fourier transform: the time-domain waveform. Taking the inverse transform of (4.4.47), Wald obtains the quadrupole formula for the spatial components of the metric perturbation:

$$\bar{\gamma}_{\mu\nu}(t, \vec{x}) = \frac{2}{3R} \left.\frac{d^2 q_{\mu\nu}}{dt^2}\right|_{t'=t-R}, \qquad (\mu, \nu = 1, 2, 3). \tag{4.4.49}$$

This is Wald's final result for slow-motion gravitational radiation: *the metric perturbation is proportional to the second time derivative of the mass quadrupole moment, evaluated at the retarded time.*

4.4.4.1 Energy Flux and Radiated Power

Although gravitational waves do not possess a true local stress–energy tensor, one may define an *effective* stress–energy tensor that correctly describes the averaged energy and momentum flux carried by weak, rapidly varying gravitational radiation. Following the standard linearized-gravity treatment, we work exclusively with the TT part of the metric perturbation, $\gamma_{ij}^{\mathrm{TT}}$.

Averaging procedure. Define the effective stress–energy tensor by

$$t_{\mu\nu}^{\mathrm{GW}} = \frac{1}{32\pi} \left\langle \partial_\mu \gamma_{ij}^{\mathrm{TT}}\, \partial_\nu \gamma_{ij}^{\mathrm{TT}} \right\rangle, \tag{GW.1}$$

where $\langle \cdots \rangle$ denotes an average over several wavelengths. This is a standard post-Wald extension and introduces no new equations in Wald's numbering.

Energy flux. For outward-propagating radiation in direction $\hat{n}$,

$$\frac{dE}{dA\, dt} = t_{0i}^{\mathrm{GW}} n^i.$$

In the far zone, the TT field has the form

$$\gamma_{ij}^{\mathrm{TT}}(t, r, \hat{n}) = \frac{1}{r}\, A_{ij}(t - r),$$

implying

$$\partial_t \gamma_{ij}^{\mathrm{TT}} = -\partial_r \gamma_{ij}^{\mathrm{TT}} = \frac{1}{r}\, \dot{A}_{ij}(t-r).$$

Thus the radial flux becomes

$$t_{0r}^{\mathrm{GW}} = \frac{1}{32\pi r^2} \left\langle \dot{A}_{ij}(t-r)\, \dot{A}_{ij}(t-r) \right\rangle. \qquad \text{(GW.2)}$$

Total radiated power. Integrating over a sphere of radius r gives

$$\frac{dE}{dt} = \frac{1}{32\pi} \int d\Omega \left\langle \dot{A}_{ij}(t-r)\, \dot{A}_{ij}(t-r) \right\rangle. \qquad \text{(GW.3)}$$

Relation to the quadrupole moment. Using Wald's waveform (4.4.55),

$$\gamma_{ij}(t, \vec{x}) = \frac{2}{R}\, \ddot{Q}_{ij}(t-R),$$

the TT part is

$$\gamma_{ij}^{\mathrm{TT}} = \frac{2}{R}\, \ddot{Q}_{ij}^{\mathrm{TT}}(t-R),$$

so the amplitude is

$$A_{ij}(t) = 2\, \ddot{Q}_{ij}^{\mathrm{TT}}(t), \qquad \dot{A}_{ij}(t) = 2\, \dddot{Q}_{ij}^{\mathrm{TT}}(t).$$

Substituting into (GW.3) yields

$$\frac{dE}{dt} = \frac{1}{8\pi} \int d\Omega \left\langle \dddot{Q}_{ij}^{\mathrm{TT}}\, \dddot{Q}_{ij}^{\mathrm{TT}} \right\rangle. \qquad \text{(GW.4)}$$

Using the standard angular identity

$$\int d\Omega\, \dddot{Q}_{ij}^{\mathrm{TT}}\, \dddot{Q}_{ij}^{\mathrm{TT}} = \frac{4\pi}{5}\, \dddot{Q}_{ij} \dddot{Q}_{ij},$$

we obtain the standard quadrupole power formula:

$$\frac{dE}{dt} = -\frac{1}{5} \left\langle \dddot{Q}_{ij}\, \dddot{Q}_{ij} \right\rangle. \qquad \text{(GW.5)}$$

This expression is *not* in Wald; it is a standard modern extension consistent with Wald's notation and conventions.

Orders of Magnitude and Astrophysical Sources

The quadrupole power formula,

$$\frac{dE}{dt} = -\frac{1}{5} \left\langle \dddot{Q}_{ij}\, \dddot{Q}_{ij} \right\rangle, \qquad \text{(GW.5)}$$

and the radiative waveform,

$$\gamma_{ij}^{\mathrm{TT}}(t, \vec{x}) = \frac{2}{R}\, \ddot{Q}_{ij}^{\mathrm{TT}}(t-R), \qquad (4.4.57)$$

describe how gravitational waves depend on the dynamics of the source. We now ask how large these effects are in realistic astrophysical systems.

Dimensional estimate of the strain. Let a source have total mass M, characteristic size L, and internal velocity scale $v \ll 1$ (in units where $c = 1$). The magnitude of the mass quadrupole moment is

$$Q \sim ML^2,$$

and if the motion varies on timescale $T \sim L/v$, then

$$\ddot{Q} \sim \frac{Q}{T^2} \sim ML^2 \left(\frac{v}{L}\right)^2 \sim Mv^2.$$

At distance R in the radiation zone the strain amplitude is

$$h \sim \frac{\ddot{Q}}{R} \sim \frac{Mv^2}{R}. \tag{GW.6}$$

For a Keplerian binary with separation L,

$$v^2 \sim \frac{M}{L},$$

which gives

$$h \sim \frac{M^2}{LR}. \tag{GW.7}$$

This estimate shows that strong gravitational radiation requires both large masses and small separations — i.e., compact objects.

Scaling of the radiated power. A periodic source with characteristic frequency $\omega \sim v/L$ has

$$\dddot{Q} \sim \omega^3 Q \sim \frac{v^3}{L^3} ML^2 \sim M \frac{v^3}{L}.$$

Hence

$$\dddot{Q}_{ij} \dddot{Q}_{ij} \sim M^2 \frac{v^6}{L^2},$$

and the power radiated is

$$\frac{dE}{dt} \sim -M^2 \frac{v^6}{L^2}. \tag{GW.8}$$

Using again $v^2 \sim M/L$,

$$\frac{dE}{dt} \sim -\frac{M^5}{L^5}.$$

Restoring factors of G and c, one finds

$$\frac{dE}{dt} \sim -\frac{G}{c^5} M^2 L^4 \omega^6 \sim -\frac{G^4}{c^5} \frac{M^5}{L^5}.$$

The factor G/c^5 suppresses radiation from all laboratory-sized sources.

Laboratory sources are negligible. For a system with

$$M \sim 10^3 \text{ kg}, \qquad L \sim 1 \text{ m}, \qquad v \ll 1,$$

the strain at astrophysical distances is

$$h \sim \frac{GMv^2}{c^4 R} \ll 10^{-30},$$

far below detectable levels.

Astrophysical sources. Consider instead a compact binary of two neutron stars or black holes with

$$M \sim M_\odot, \qquad L \sim \frac{GM}{c^2}.$$

Then

$$v^2 \sim \frac{GM}{Lc^2} \sim 1,$$

and at distance $R \sim 100-500$ Mpc,

$$h \sim \frac{GM}{c^2 R} \sim 10^{-21},$$

the level at which modern interferometers such as LIGO and Virgo operate.

Summary. These scaling arguments show:

- gravitational waves from laboratory systems are utterly negligible,
- compact binaries are the dominant astrophysical sources,
- linearized gravity is an excellent approximation in the far zone, even for strong-field compact objects.

Observational Confirmation: The Hulse–Taylor Binary Pulsar

The quadrupole power formula,

$$\frac{dE}{dt} = -\frac{1}{5}\left\langle \dddot{Q}_{ij}\, \dddot{Q}_{ij} \right\rangle, \tag{GW.5}$$

predicts that a binary system steadily loses orbital energy to gravitational radiation, causing the orbit to shrink.

A remarkable confirmation of this prediction was provided by the binary pulsar PSR B1913+16, discovered by R. A. Hulse and J. H. Taylor.

The system. PSR B1913+16 consists of two neutron stars in an eccentric, relativistic orbit with period

$$P \approx 8 \text{ hr}.$$

One star is a radio pulsar whose precisely timed pulses allow the orbital parameters to be monitored with extraordinary accuracy.

Predicted orbital decay. General relativity predicts that gravitational-wave emission causes the orbital period to decrease at the rate

$$\dot{P}_{\rm GR} = -\frac{192\pi}{5}\left(\frac{2\pi G}{c^3}\right)^{5/3}\frac{(m_1 m_2)\,(m_1+m_2)^{-1/3}}{a^{5/3}\,(1-e^2)^{7/2}}\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right), \tag{GW.9}$$

a direct consequence of the quadrupole formula applied to an eccentric binary.

Observations. Long-term timing observations yield the measured decay,

$$\dot{P}_{\rm obs},$$

and the result agrees with the general-relativistic prediction to within better than 0.5%:

$$\dot{P}_{\rm obs} \approx \dot{P}_{\rm GR}.$$

Significance. This agreement is the first experimental confirmation that gravitational waves carry energy exactly as predicted by the linearized theory. The cumulative orbital decay — although far too small to detect as a strain at Earth — provides a precise indirect detection of gravitational radiation.

The discovery earned Hulse and Taylor the 1993 Nobel Prize in Physics.

Chapter 5

Homogeneous, Isotropic, Cosmology

In Chapter 4 we developed the description of spacetime as a four-dimensional manifold equipped with a Lorentzian metric, g_{ab}, whose dynamics are determined by Einstein's equation,

$$G_{ab} = 8\pi T_{ab}.$$

A natural question now arises: *which* solutions of Einstein's equation describe the spacetime we actually observe? In this chapter we examine the large-scale structure of our universe as implied by general relativity under the empirical assumption that, on sufficiently large scales, the universe is both homogeneous and isotropic.

5.1 Homogeneity and Isotropy

Note: Throughout this section we work with an idealized cosmological spacetime that is exactly homogeneous and isotropic. These symmetry assumptions are not meant to describe local inhomogeneities such as stars or galaxies, but rather the large-scale structure of the universe. The existence of isometries mapping arbitrary points on Σ_t therefore applies only within this idealized model.

We shall assume that the universe is isotropic, meaning that there exist no preferred directions in space, and that sufficiently large-scale observations yield statistics independent of the direction in which we look. Isotropy about every point then implies homogeneity: no spatial point is distinguished from any other.

These assumptions are strongly supported by modern observations:

- On the largest observable scales, the spatial distribution of galaxies is consistent with homogeneity and isotropy.
- Counts of radio sources and the observed isotropy of the X-ray and γ-ray backgrounds provide additional evidence for large-scale uniformity.
- The Cosmic Microwave Background (CMB), a nearly perfect 3 K blackbody radiation field permeating the universe, is isotropic to extremely high precision.

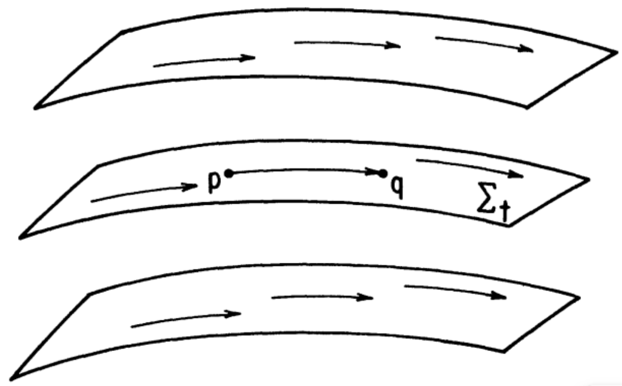

Figure 5.1: The hypersurfaces of spatial homogeneity in spacetime. By definition of homogeneity, for each t and each $p, q \in \Sigma_t$ there exists and isometry of the spacetime which takes p into q.

Even if the universe is not exactly homogeneous and isotropic, these symmetries provide an excellent approximation on sufficiently large scales. Accordingly, throughout this chapter we shall adopt homogeneity and isotropy as our working assumptions and explore their profound implications for the form of the spacetime metric and the dynamics of cosmic evolution.

Homogeneity

Before introducing the formal definition of homogeneity, it is useful to recall the notion of a *hypersurface* in spacetime. A hypersurface Σ is a smooth three-dimensional submanifold of the four-dimensional spacetime manifold. Intuitively, one may think of a hypersurface as representing an "instant of time" or a "slice" of the spacetime, although no preferred foliation is assumed apriori. A foliation is a partition of a manifold into submanifolds. A family of hypersurfaces $\{\Sigma_t\}$, smoothly labeled by a real parameter t, provides a foliation of spacetime in which each event lies on exactly one such slice. When the hypersurfaces are spacelike, each Σ_t may be viewed as the "space" of the universe at parameter time t.

In a homogeneous cosmological model, it is precisely these spacelike hypersurfaces that exhibit the symmetry: for any two points p and q lying on the same hypersurface Σ_t, there exists an isometry of the induced three-metric on Σ_t that maps p to q. Or in other words, there exists a smooth diffeomorphism, $\phi : \Sigma_t \to \Sigma_t$, such that $\phi(p) = q$ where $\phi^* h_{ab} = h_{ab}$ (geometry preservation). To break it down further, take a point $p \in \Sigma_t$ – around p, the metric h_{ab} defines distances, angles, curvature, volumes, and any other local geometric facts. Homogeneity says, there exists another point $q \in \Sigma_t$ such that *everything geometrically true in a neighborhood of p* is also true in a neighborhood of q. The isometry is simply the formal object that encodes this equivalence.

Fig. 5.1 illustrates a one-parameter family of such slices, together with the motion of points under the action of spatial isometries. An *isometry* is a symmetry of the spacetime metric: a smooth map that moves points around without changing any physical distances or the form of the metric. . Thus the concept of a hypersurface provides the geometric setting in which homogeneity is defined: each "spatial" slice of the universe looks the same at every point.

Isotropy

Isotropy in cosmology means that at each point, space looks the same in every direction. But we are in spacetime, not Euclidean space, so "directions" must be defined carefully.

To describe this, we first need a congruence of observers. But what is meant by congruence? Imagine you want to assign an observer to every event in spacetime – one who represents "the cosmic rest frame" (the frame in which the universe looks isotropic). You cannot use just one observer for this, because a single observer's worldline passes through only one event at each moment of their own time.

So instead, you need a whole family of observers, arranged so that:

- Their worldlines never intersect,
- Every event, p, in spacetime lies on exactly one of them,
- Together, they fill the entire spacetime manifold.

This family is called a ***congruence***. Equivalently, it is the integral curves of a smooth, everywhere-defined timelike vector field $u^a(x)$.

Once we have such a congruence, the tangent vector $u^a(p)$ to the worldline passing through an event p defines the "time direction" of the isotropic observer at p. Here, "time direction" refers to the direction in the tangent space along which the observer's proper time increases. All other possible directions a curve may take through p form the tangent space T_pM. Among these, the directions that are *spatial* for the observer are those orthogonal to u^a, that is,

$$V_p = \{\, s^a \in T_pM \mid g_{ab}u^a s^b = 0 \,\}.$$

This 3-dimensional subspace V_p contains all the possible spatial directions the observer at p can point toward. Here, the term "point" refers to selecting a spatial direction in the observer's instantaneous rest space, not to moving in that direction; spatial directions in V_p describe how the observer may orient measurements or aim instruments at a fixed proper time, rather than possible trajectories of motion.

With these notions in place, we can now give a precise, coordinate-free statement of isotropy. Consider Fig. 5.2. A spacetime is said to be *spatially isotropic at* p if, for any two unit vectors $s_1^a, s_2^a \in V_p$, there exists an isometry of the spacetime metric g_{ab} that satisfies the following three conditions:

1. It leaves the event p fixed,
2. It leaves the observer's velocity u^a at p fixed, and
3. It rotates s_1^a into s_2^a.

The first two conditions ensure that we are considering a *purely spatial* symmetry at the event p. The third condition expresses the essential content of isotropy: no spatial direction at p is geometrically distinguished from any other. If such rotations exist for all pairs of spatial directions in V_p, then the observer at p sees no preferred direction in the universe.

Thus, isotropy at a point means that the local spatial geometry around the cosmic observer admits

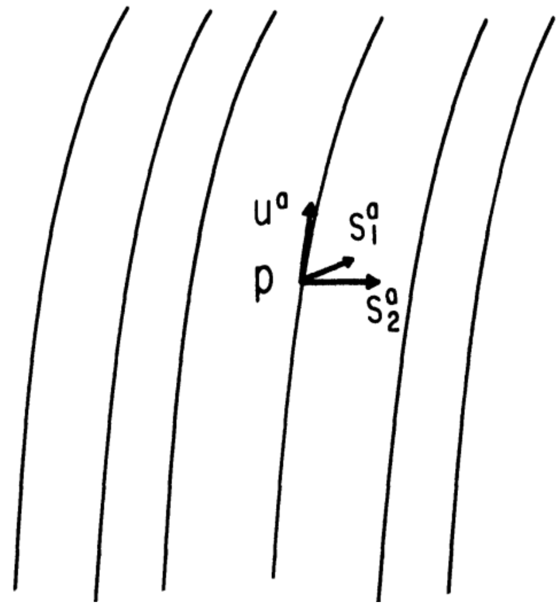

Figure 5.2: The world lines of isotropic observers in spacetime. By definition of isotropy, for any two vectors s_1^a, s_2^a at p which are orthogonal to u^a, there exists an isometry of the spacetime which leaves p fixed and rotates s_1^a into s_2^a.

the full rotation group as a symmetry. In particular, one cannot construct any geometrically preferred spatial vector orthogonal to u^a. This is the precise general-relativistic formulation of the intuitive idea that "space looks the same in every direction."

Note: Rotations are singled out because they are the only isometries that fix the event p and the time direction $u^a(p)$ while mapping one spatial direction in V_p to another; all other isometries either shift the point p (as in translations) or alter the time direction (as in boosts). Thus rotational symmetry of the spatial subspace V_p is precisely the condition for isotropy at p.

5.1.0.1 Spatial Curvature Implied by Homogeneity and Isotropy

Wald claims

> It is not difficult to see that in the case of homogeneous and isotropic spacetime, the surfaces Σ_t of homogeneity must be orthogonal to the tangents, u^a, to the world lines of the isotropic observers. If not, then assuming that the isotropic observers and the family of homogeneous surfaces Σ_t are unique, the failure of the tangent subspace orthogonal to u^a to coincide with the tangent space of Σ_t would enable us to construct a geometrically preferred spatial vector, in violation of isotropy.

Oh yeah. Totally not difficult. I for one saw it instantly - lets go through it anyway.

Given that the hypersurfaces Σ_t of homogeneity are orthogonal to the timelike congruence u^a, we define the induced spatial metric by

$$h_{ab}(t) = g_{ab} + u_a u_b.$$

Here $u_a u_b$ is the tensor product of the covector u_a with itself; it encodes the distinguished timelike direction defined by the comoving observers. The combination $h_{ab} = g_{ab} + u_a u_b$ therefore removes the timelike component of the spacetime metric and yields a positive-definite metric on vectors orthogonal to u^a.

This tensor projects all vectors into the subspace orthogonal to u^a and therefore measures spatial distances within Σ_t, i.e. the geometry seen by comoving observers. Homogeneity and isotropy imply that $h_{ab}(t)$ cannot depend on spatial position, and that each Σ_t must be a maximally symmetric three–manifold. We now derive this result explicitly by examining the intrinsic curvature of (Σ_t, h_{ab}).

Isotropy at a Point. The goal of this section is to show that isotropy at a single point forces the spatial curvature to be identical in all directions. Equivalently, the Riemann tensor at a point cannot distinguish one spatial direction from another.

We consider the Riemann tensor ${}^{(3)}R_{abcd}$ associated with the induced spatial metric h_{ab}. By "associated" we mean that this is not the spacetime Riemann tensor, but rather the intrinsic Riemann tensor of the spatial slice Σ_t, encoding how space itself is curved at fixed cosmic time. Thus, from this point onward we are doing purely three-dimensional Riemannian geometry. This restriction is possible because h_{ab} is the projection metric onto directions orthogonal to the timelike congruence u^a, and hence contains no timelike components (as indicated by the superscript (3)).

The Riemann tensor has four indices. By raising two of them with h^{ab}, we may reinterpret it as a map

$${}^{(3)}R_{ab}{}^{cd} : (\text{antisymmetric index pair } ab) \longrightarrow (\text{antisymmetric index pair } cd).$$

In other words, the Riemann tensor acts naturally on two-forms, i.e. rank–$(0,2)$ antisymmetric tensors. This reflects the geometric fact that curvature is associated with two-dimensional planes rather than individual directions: sectional curvature measures the curvature of a two-plane.

Rather than introducing the abstract operator notation $L : W \to W$, we will write Wald's argument explicitly in index notation, which makes transparent how isotropy constrains the intrinsic curvature of the spatial slices.

Let F_{ab} denote an antisymmetric rank–$(0,2)$ tensor at a point $p \in \Sigma_t$, satisfying

$$F_{ab} = -F_{ba}.$$

In three dimensions, the space of such tensors has dimension three. Geometrically, each two-form corresponds to an oriented two-dimensional plane through the point p.

The action of the Riemann tensor on a two-form is given explicitly by

$$\tilde{F}_{ab} = {}^{(3)}R_{ab}{}^{cd} F_{cd}.$$

Thus, given any antisymmetric tensor F_{cd}, contraction with the Riemann tensor produces another antisymmetric tensor $\tilde{F}_{ab}$. This explicit index expression is all that is meant by viewing the Riemann tensor as a linear map on the space of two-forms; no additional structure is being introduced.

Using the spatial metric h_{ab}, there is a natural inner product on two-forms, and the algebraic symmetries of the Riemann tensor,

$$^{(3)}R_{abcd} = {}^{(3)}R_{cdab}, \qquad {}^{(3)}R_{abcd} = -{}^{(3)}R_{bacd},$$

imply that this action is self-adjoint with respect to that inner product. Consequently, the curvature operator admits an orthonormal basis of eigen–two–forms at each point.

Aside: Self-adjointness of the curvature operator and its consequences.

At a fixed point $p \in \Sigma_t$, consider the vector space of two-forms $W = \Lambda^2(T_p^*\Sigma_t)$ equipped with the inner product induced by the spatial metric h_{ab},

$$\langle F, G\rangle = \frac{1}{2} F_{ab} G_{cd}\, h^{ac} h^{bd}.$$

Define the action of the spatial Riemann tensor on two-forms by

$$(\mathcal{R}F)_{ab} = {}^{(3)}R_{ab}{}^{cd} F_{cd}.$$

The algebraic symmetries of the Riemann tensor imply that for all two-forms F_{ab} and G_{ab},

$$\langle F, \mathcal{R}G\rangle = \langle \mathcal{R}F, G\rangle.$$

This equality is precisely the statement that $\mathcal{R}$ is self-adjoint with respect to the inner product on W. In this context, the adjoint of an operator is defined by the requirement that

$$\langle F, \mathcal{R}G\rangle = \langle \mathcal{R}^\dagger F, G\rangle \quad \text{for all } F, G \in W,$$

and self-adjointness means $\mathcal{R}^\dagger = \mathcal{R}$.

Since W is a finite-dimensional real inner-product space, standard linear algebra guarantees that $\mathcal{R}$ admits an orthonormal basis of eigen–two–forms. That is, there exist two-forms $F_{ab}^{(i)}$ and real eigenvalues λ_i such that

$$^{(3)}R_{ab}{}^{cd} F_{cd}^{(i)} = \lambda_i F_{ab}^{(i)},$$

with the $F_{ab}^{(i)}$ orthonormal under $\langle \cdot, \cdot \rangle$.

Each such eigen–two–form corresponds geometrically to an oriented spatial plane through p, and the associated eigenvalue λ_i gives the sectional curvature of that plane. If the eigenvalues were not all equal, the corresponding eigen–two–forms would single out geometrically preferred planes and hence preferred spatial directions, violating isotropy.

As shown above, isotropy at p therefore requires that the curvature act identically on all two-forms, so that

$$^{(3)}R_{ab}{}^{cd} F_{cd} = K\, F_{ab} \quad \text{for all antisymmetric } F_{ab},$$

for some scalar K. Since this must hold for all two-forms, the Riemann tensor must take the form

$$^{(3)}R_{ab}{}^{cd} = K\,\delta^c{}_{[a}\delta^d{}_{b]}.$$

Lowering indices with h_{ab} yields the constant-curvature expression

$$^{(3)}R_{abcd} = K\,h_{c[a}h_{b]d}. \tag{5.1}$$

Constancy of the Curvature. Homogeneity further requires that K be the same at all points of Σ_t. Substituting (5.1) into the three-dimensional Bianchi identity gives

$$0 = D_{[e}{}^{(3)}R_{ab]cd} = (D_{[e}K)\,h_{|c|a}h_{b]d},$$

which can hold only if $D_eK = 0$. Thus K is constant on each spatial slice.

Conclusion. Each hypersurface Σ_t is therefore a three-dimensional manifold of constant curvature K, i.e. a maximally symmetric Riemannian space. The only allowed spatial geometries are

$$K > 0: \text{ a 3-sphere}, \qquad K = 0: \ \mathbb{R}^3, \qquad K < 0: \text{ hyperbolic 3-space.}$$

Any time dependence of the spatial metric can appear only as an overall scale factor, so the most general induced metric is

$$h_{ab}(t) = a^2(t)\,\tilde{h}_{ab},$$

where $\tilde{h}_{ab}$ is the metric of a fixed maximally symmetric three-manifold.

A maximally symmetric three-manifold is a Riemannian space whose metric admits the maximum number (six) of independent Killing vector fields, and is equivalently characterized by homogeneity, isotropy, and constant curvature.

Note 5.1. Up to this point, no use has been made of Einstein's field equations. The results obtained follow solely from differential geometry together with the assumptions of homogeneity and isotropy. These symmetry requirements fix the spatial geometry of Σ_t and restrict the spacetime metric to the FRW form, leaving only the scale factor $a(t)$ undetermined. The dynamical evolution of $a(t)$ enters only upon imposing Einstein's equations.

Classification of Constant-Curvature Spatial Geometries Having shown that homogeneity and isotropy force each spatial hypersurface Σ_t to have constant curvature, we now classify the possible spatial geometries by enumerating the three-dimensional spaces that realize each value of the curvature parameter K.

Appendix: Einstein Index Notation Cheat Sheet

Einstein Index Notation Cheat Sheet

(Optimized for Linearized Gravity and Wald Ch. 4.4)

1. Index Placement and Summation

- Contravariant (upper index): V^a
- Covariant (lower index): W_b
- Einstein summation: a repeated index (one up, one down) is summed:

$$V^a W_a = \sum_{a=0}^{3} V^a W_a.$$

- Unrepeated indices are never summed.

2. Minkowski Metric and Raising/Lowering

Wald uses signature $(-+++)$:

$$\eta_{ab} = \mathrm{diag}(-1, 1, 1, 1), \qquad \eta^{ab} = \mathrm{diag}(-1, 1, 1, 1).$$

Lowering a vector:

$$V_a = \eta_{ab} V^b.$$

Raising a covector:

$$W^a = \eta^{ab} W_b.$$

Special cases:

$$V_0 = -V^0, \qquad V_i = V^i \quad (i = 1, 2, 3).$$

3. Contractions

General contractions:

$$T^a{}_a = \eta^{ab} T_{ba}, \qquad T_{ab} V^a W^b, \qquad R^a{}_{0a0}.$$

Contraction always means: **raise then sum** (or sum directly if mixed).

4. Trace Reversal (Wald Eq. 4.4.6)

Trace:

$$\gamma = \gamma^a{}_a = \eta^{ab} \gamma_{ab}.$$

Trace-reversed metric perturbation:

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2} \eta_{ab} \gamma.$$

Useful identities:

$$\bar{\gamma} = -\gamma, \qquad (\gamma = 0) \;\Rightarrow\; \bar{\gamma}_{ab} = \gamma_{ab}.$$

5. Gauge Transformations in Linearized Gravity

Basic transformation:

$$\gamma'_{ab} = \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a.$$

Trace-reversed form:

$$\bar{\gamma}'_{ab} = \bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab} \partial^c \xi_c.$$

Lorenz gauge:

$$\partial^a \bar{\gamma}_{ab} = 0.$$

Under a gauge transformation:

$$\partial^a \bar{\gamma}'_{ab} = \partial^a \bar{\gamma}_{ab} + \Box \xi_b.$$

Residual gauge freedom:

$$\Box \xi_b = 0.$$

6. Metric as a Bilinear Map (Geometric Meaning)

Lower a vector:

$$V_b = g_{ba} V^a \qquad \text{('flat' operator)}.$$

Raise a covector:

$$\alpha^a = g^{ab} \alpha_b \qquad \text{('sharp' operator)}.$$

Used in:

- switching between γ_{ab}, γ^{ab}, $\bar{\gamma}_a{}^b$
- manipulating stress–energy components
- Riemann tensor computations

7. Linearized Riemann Tensor (Used in Geodesic Deviation)

$$R^{(1)}_{abcd} = \frac{1}{2}(\partial_c \partial_b \gamma_{ad} + \partial_d \partial_a \gamma_{bc} - \partial_c \partial_a \gamma_{bd} - \partial_d \partial_b \gamma_{ac}).$$

Raise first index:

$$R^a{}_{bcd} = \eta^{ae} R_{ebcd}.$$

Used to show:

$$R^i{}_{0j0} = -\frac{1}{2}\,\ddot{\gamma}^{\mathrm{TT}\,i}{}_j.$$

8. Plane-Wave Polarizations

Plane wave:

$$\gamma_{\mu\nu} = A_{\mu\nu} e^{ik_a x^a}.$$

Null condition:

$$k_a k^a = 0.$$

Lorenz transversality:

$$k^a A_{ab} = 0.$$

Radiation gauge:

$$A_{0\mu} = 0, \qquad A_{3\mu} = 0.$$

Tracelessness:

$$A^\mu{}_\mu = 0.$$

Remaining components:

$$A_{11},\ A_{22},\ A_{12} = A_{21}, \qquad A_{11} + A_{22} = 0.$$

These form the $+$ and $\times$ polarization tensors.

9. Stress–Energy Tensor Manipulations

Lower two indices:

$$T_{ab} = g_{ac} g_{bd} T^{cd}.$$

Mixed tensor:

$$T^a{}_b = g^{ac} T_{cb}.$$

Conservation:

$$\partial_a T^{ab} = 0 \quad \Rightarrow \quad \partial_t T^{0b} = -\partial_i T^{ib}.$$

Used in deriving Wald's equation (4.4.46).

10. Summary Table

Operation	Formula	Used In
Trace	$\gamma = \eta^{ab}\gamma_{ab}$	Radiation gauge
Trace reverse	$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma$	Lorenz gauge form
Raise index	$V^a = \eta^{ab}V_b$	Riemann, TT waves
Lower index	$W_a = \eta_{ab}W^b$	Stress–energy
Divergence	$\partial_a T^{ab} = 0$	Quadrupole derivation
TT conditions	$A^\mu{}_\mu = 0, \; k^\mu A_{\mu\nu} = 0$	Polarization
Geodesic deviation	$R^i{}_{0j0} = -\frac{1}{2}\ddot{\gamma}^i{}_j$	Detector physics

This appendix summarizes the conventions and index operations used throughout the gravitational radiation analysis in linearized gravity.