

General Relativity

Lecture Notes ¹

Jericho Cain, PhD

Fall 2025

Contents

Preface: From Pure Geometry to Spacetime	6
1 Spacetime as a Differentiable Manifold	7
1.1 Manifolds and Coordinate Systems	7
1.2 Smooth Maps Between Manifolds	15
1.3 Tangent Vectors and Tangent Spaces	19
1.4 The Differential and Pushforward	24
1.5 Covectors and the Pullback	31
1.6 Tensors and Tensor Fields	35
1.7 The Metric Tensor	37
1.8 Orientation and Volume Forms (Optional)	40
1.9 Summary and Outlook	41
2 Curvature	43
2.1 Derivative Operators and Parallel Transport	44
2.2 The Riemann Curvature Tensor	55
2.2.1 Curvature and Parallel Transport Around a Loop	56
2.3 Geodesics	60
2.3.1 Definition via Parallel Transport	60
2.3.2 Coordinate Expression of the Geodesic Equation	61

Contents	3
2.3.3 The Exponential Map	62
2.3.4 Riemannian Normal Coordinates	63
2.3.5 Gaussian Normal Coordinates	65
2.3.6 Extremizing the Length Functional	67
2.3.7 Geodesic Deviation	69
2.4 Methods for Computing Curvature	71
2.4.1 Coordinate Component Method	72
2.4.2 Orthonormal Basis (Tetrad) Methods	73
3 Einstein's Equations	78
3.1 The Geometry of Space in Prerelativity Physics: General and Special Covariance . .	78
3.2 Special Relativity	81
3.2.1 Overview and Inertial Frames	81
3.2.2 Geometry of Minkowski Spacetime	82
3.2.3 Stress–Energy and Matter in Special Relativity	84
3.2.4 Fields in Special Relativity	90
3.3 General Relativity	93
3.3.1 From Special Relativity to a New Theory of Gravity	93
3.3.2 The Equivalence Principle and the Failure of Background Observers	93
3.3.3 The Central Hypothesis: Gravity <i>is</i> Curved Spacetime	94
3.3.4 Why Spacetime Need Not Be $\mathbb{R}^4$: Lorentz Metrics on General Manifolds . . .	95
3.3.5 The Principles Governing Physics in Curved Spacetime	96
3.3.6 Motion, 4–Momentum, and Forces in Curved Spacetime	98
3.3.7 Scalar Fields and Maxwell Fields in Curved Spacetime	99
3.3.8 From Tidal Forces to Einstein's Equation	101
3.3.9 Remarks on the Nature of Einstein's Equation	103

3.4	Linearized Gravity: The Newtonian Limit and Gravitational Radiation	104
3.4.1	The Linearized Einstein Equation	104
3.4.2	The Newtonian Limit	110
3.4.3	Gravitational Radiation (Vacuum)	115
3.4.4	Gravitational Radiation (Sourced)	127
4	The Schwarzschild Solution	136
4.1	Derivation of the Schwarzschild Metric	136
4.1.1	General Static, Spherically Symmetric Ansatz	136
4.1.2	Connection and Curvature Components	138
4.1.3	Solving the Vacuum Equations	141
4.2	Geodesics of Schwarzschild	142
4.2.1	Killing vectors and constants of motion	143
4.2.2	Effective Potentials: Newtonian vs. Schwarzschild	147
4.3	Experimental Tests	150
4.3.1	The Precession of Perihelia	150
4.3.2	Gravitational Redshift	155
4.4	Causal Structure of the Schwarzschild Geometry	157
4.5	The Maximally Extended Schwarzschild Solution	164
4.6	Physical Relalizations: Stars and Black Holes	167
A	Mathematical Methods and Tensor Identities	168
A.1	Derivation Cheat Sheet	168
A.1.1	Covariant Derivatives in Components	168
A.2	Useful Identities and Tensor Manipulations	169
A.2.1	Index Placement and Summation	169
A.2.2	Minkowski Metric and Raising/Lowering	169

A.2.3	Contractions	169
A.2.4	Trace Reversal (Eq. 3.108)	169
A.2.5	Gauge Transformations in Linearized Gravity	170
A.2.6	Metric as a Bilinear Map (Geometric Meaning)	170
A.2.7	Linearized Riemann Tensor (Used in Geodesic Deviation)	170
A.2.8	Plane-Wave Polarizations	171
A.2.9	Stress–Energy Tensor Manipulations	171
A.2.10	Summary Table	172
B	Code Listings	173
B.1	Effective Potentials in Newtonian Gravity	173
B.2	Effective Potentials in General Relativity	174

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 standard general relativity texts. The two texts share the same geometric foundation, but speak in different dialects: Lee emphasizes rigor and abstraction, while the textbook 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.

Chapter 1

Spacetime as a Differentiable Manifold

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 $T_p M$, 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 building blocks 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.

1.1 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 1.1. 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 1.1 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.

the textbook'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 $\varphi : U \rightarrow \mathbb{R}^4$ 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 $\pi^\mu : \mathbb{R}^4 \rightarrow \mathbb{R}$ denotes the projection onto the μ th 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 $U \cap V \neq \emptyset$, the *transition map* $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$ must be smooth.
- A collection of mutually compatible charts forms an *atlas*; its maximal extension defines the smooth structure on M .

Common Terminology.

- **Bijjective.** A map $f : X \rightarrow Y$ is *bijjective* 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 $f : X \rightarrow Y$ 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 $f : M \rightarrow N$ 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. the textbook'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) ,

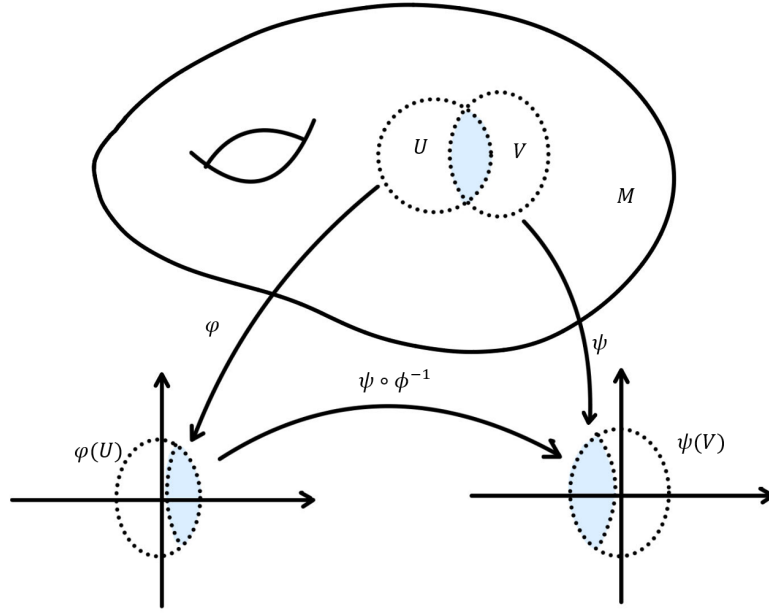


Figure 1.1: Overlapping coordinate charts (U, φ) and (V, ψ) on a manifold M . The transition map $\psi \circ \varphi^{-1}$ relates the coordinates assigned to the same events in the overlap $U \cap V$. The smoothness of these maps is what makes M a smooth manifold.

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. 1.1).

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 1.1: *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 1.2).

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

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

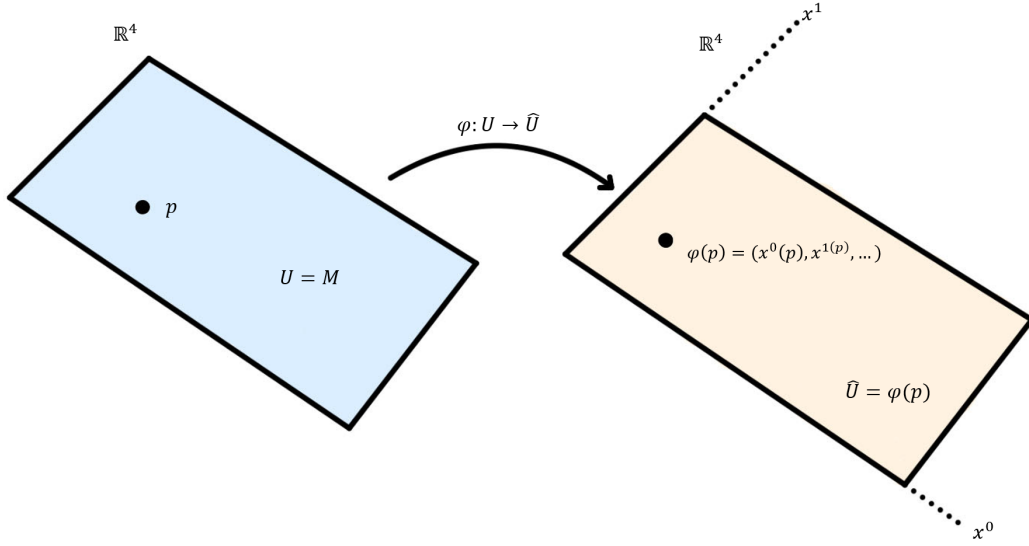


Figure 1.2: A schematic representation of the global chart for Minkowski spacetime. A point $p \in M$ on the manifold is mapped by $\varphi : M \rightarrow \mathbb{R}^4$ to its coordinate image $\varphi(p) = (x^0(p), x^1(p), x^2(p), x^3(p))$. Only two coordinate directions (x^0, x^1) are shown; the remaining 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.

not in the coordinate labels themselves.

* * *

Note 1.1. In the notation $x^\mu(p)$, each $x^\mu : U \rightarrow \mathbb{R}$ 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 $\varphi : U \rightarrow \hat{U}$ in diagrams emphasizes that coordinates are assigned values in $\mathbb{R}^n$, not that the manifold itself lives there.

Example 1.2: The 2-sphere 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

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

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

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

These are the six open hemispheres as shown in Fig. 1.3. On each $U_i^\pm$ we use the two

remaining Cartesian coordinates as chart coordinates:

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

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).
- $f_i^\pm : U_i^\pm \rightarrow \mathbb{R}^2$ 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 $p = (x, y, z) \in U_3^+$ ($z > 0$). 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 $f_1^- : U_1^- \rightarrow \mathbb{R}^2$ 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 $f_i^\pm : U_i^\pm \rightarrow \mathbb{R}^2$ “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 $f_1^+ : U_1^+ \rightarrow \mathbb{R}^2$ assigns the local coordinates

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

- Applying the inverse chart recovers the geometric point:

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

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

Figures 1.4, 1.5 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*, $\sigma_S(x, y, z)$,

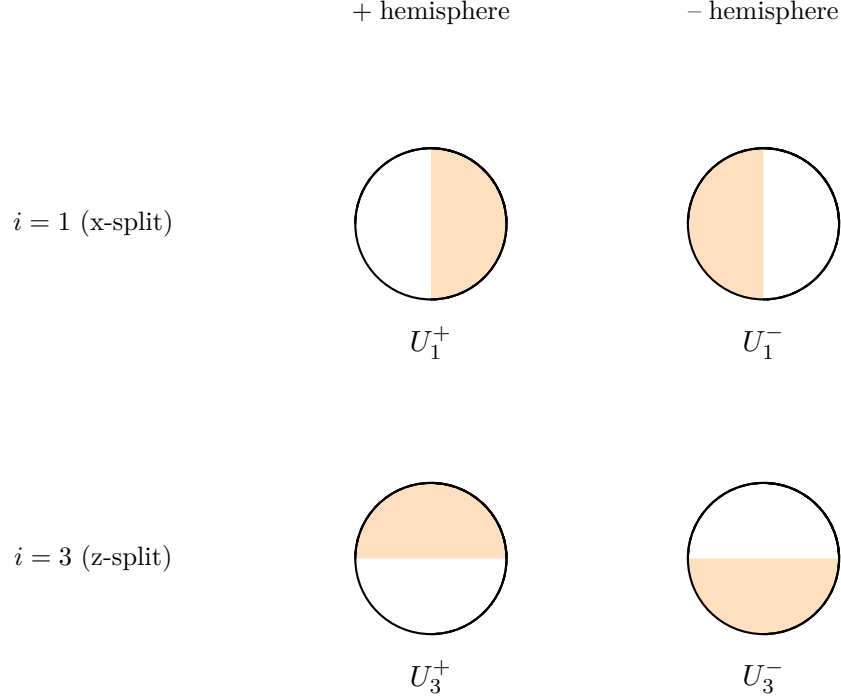


Figure 1.3: The six open hemispheres $U_i^\pm = \{(x, y, z) \in S^2 : \pm x_i > 0\}$ for $i = 1, 2, 3$ (with $x_1 = x$, $x_2 = y$, $x_3 = z$). Each chart $f_i^\pm$ uses the other two Cartesian coordinates as local coordinates on $U_i^\pm$; e.g. on $U_3^\pm$, $f_3^\pm(x, y, z) = (x, y)$. Shading indicates 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.

$$\begin{aligned} \sigma_N : S^2 \setminus \{N\} &\longrightarrow \mathbb{R}^2, & \sigma_N(x, y, z) &= \left(\frac{x}{1-z}, \frac{y}{1-z} \right), \\ \sigma_S : S^2 \setminus \{S\} &\longrightarrow \mathbb{R}^2, & \sigma_S(x, y, z) &= \left(\frac{x}{1+z}, \frac{y}{1+z} \right). \end{aligned}$$

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

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

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

* * *

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

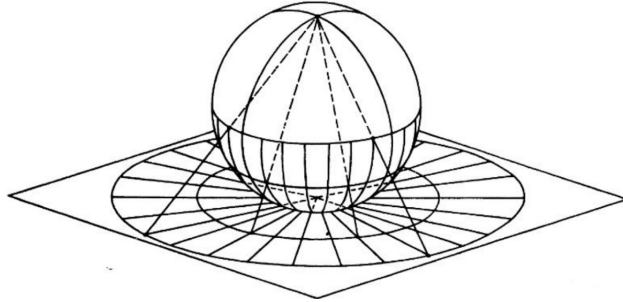


Figure 1.4: Stereographic projection from the north pole $N = (0, 0, 1)$ of the sphere onto the equatorial plane $z = 0$. Each line through N and a point $(x, y, z) \in S^2 \setminus \{N\}$ intersects the plane at a unique point $(u, v) \in \mathbb{R}^2$, defining the stereographic chart $\sigma_N(x, y, z) = (x/(1 - z), y/(1 - z))$. The entire sphere minus N is thus mapped smoothly onto the plane.

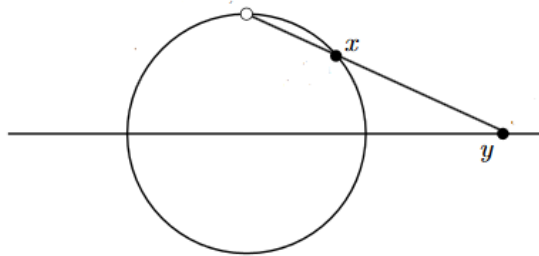


Figure 1.5: Cross-sectional view of stereographic projection. The point x on the circle represents a point on S^2 , and the line through the north pole (open circle) meets the plane at the image point y . This geometric construction yields the coordinate map $\sigma_N : S^2 \setminus \{N\} \rightarrow \mathbb{R}^2$, and a similar projection from the south pole S defines σ_S .

- 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.

1.2 Smooth Maps Between Manifolds

Up to now, we have described how coordinate systems are installed on regions of a manifold M by smooth maps $\varphi : U \rightarrow \mathbb{R}^n$. We now generalize this idea: a *smooth map* between manifolds $F : M \rightarrow N$ 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 $F : M \rightarrow N$ 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

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

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. 1.6.)

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 $f : M \rightarrow \mathbb{R}$ (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.

the textbook's perspective. A smooth function $f : M \rightarrow \mathbb{R}$ assigns a real number (a scalar field) to each event in spacetime. More generally, a smooth map $F : M \rightarrow N$ 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 $F : M \rightarrow N$ and $G : N \rightarrow P$ are smooth, then $G \circ F : M \rightarrow P$ is smooth. In coordinates this is exactly the multivariable chain rule, and it 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 $S^2 \rightarrow \mathbb{R}$, 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

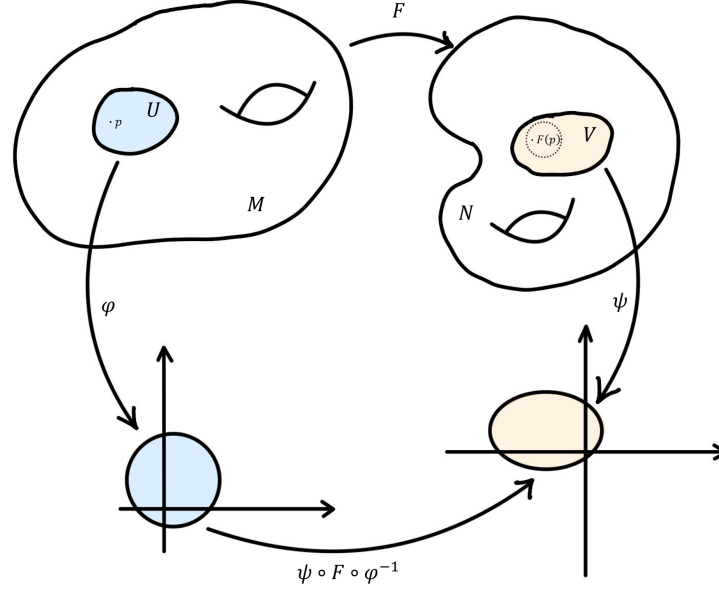


Figure 1.6: A smooth map $F : M \rightarrow N$ between manifolds. In local coordinates, the composition $\psi \circ F \circ \varphi^{-1}$ is an ordinary smooth map between regions of $\mathbb{R}^m$ and $\mathbb{R}^n$.

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 1.3: A scalar field on S^2

Define the *height function*

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

- 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

$$H : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad H(x, y, z) = z.$$

- In local coordinates this smoothness is visible directly. For example, on the chart (U_3^+, f_3^+)

where

$$f_3^+(x, y, z) = (x, y) = (u, v), \quad (f_3^+)^{-1}(u, v) = (u, v, \sqrt{1 - u^2 - v^2}),$$

the coordinate expression of h is the map

$$h \circ (f_3^+)^{-1} : \mathbb{D} \rightarrow \mathbb{R}, \quad (h \circ (f_3^+)^{-1})(u, v) = \sqrt{1 - u^2 - v^2},$$

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 $h : S^2 \rightarrow \mathbb{R}$ is a smooth map between manifolds.

This construction is illustrated in Fig. 1.7, where the smoothness of $h : S^2 \rightarrow \mathbb{R}$ 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 $H : \mathbb{R}^3 \rightarrow \mathbb{R}$ is an ambient extension used to prove smoothness, but the map we study geometrically is its restriction $h : S^2 \rightarrow \mathbb{R}$, which is the arrow shown in the diagram.

Note: Although a smooth function $h : M \rightarrow \mathbb{R}$ resembles a coordinate map, it does not define a coordinate chart unless $\dim M = 1$. Coordinate charts must be locally invertible and have as many component functions as the dimension of the manifold.

* * *

Example 1.4: 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 $f_3^+ : U_3^+ \rightarrow \mathbb{R}^2$ 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

$$f_1^+ \circ (f_3^+)^{-1} : \mathbb{D} \rightarrow \mathbb{R}^2, \quad (u, v) \mapsto (v, \sqrt{1 - u^2 - v^2}),$$

is called a *transition map* (or *change of coordinates*) between the two charts.

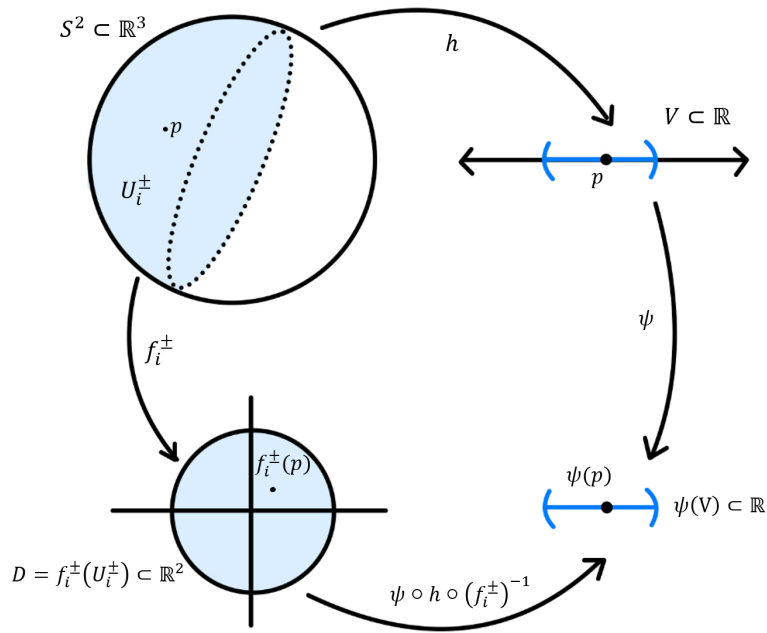


Figure 1.7: Example I: A smooth map $h : S^2 \rightarrow \mathbb{R}$ expressed in local coordinates. An open set $U_i^\pm \subset S^2$ containing p is mapped by the chart $f_i^\pm$ to its coordinate image $D = f_i^\pm(U_i^\pm) \subset \mathbb{R}^2$. An open interval $V \subset \mathbb{R}$ containing $h(p)$ 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 $\psi \circ h \circ (f_i^\pm)^{-1} : D \rightarrow \psi(V) \subset \mathbb{R}$, which is an ordinary smooth function between Euclidean spaces.

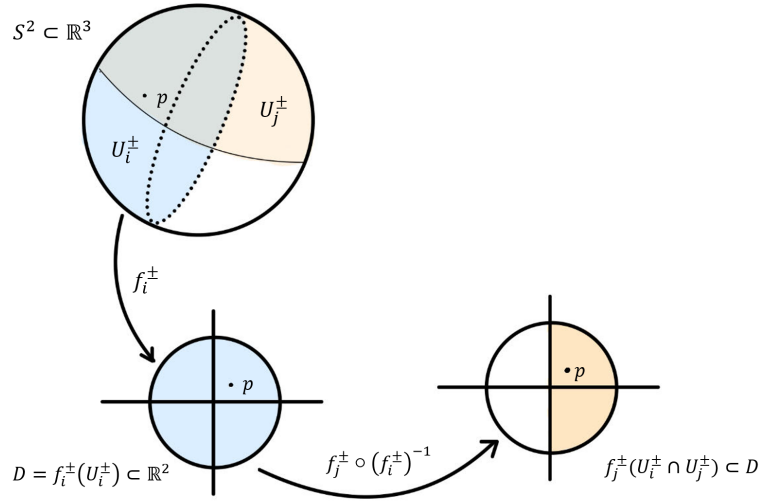


Figure 1.8: A transition map between overlapping coordinate charts on S^2 . Two chart domains $U_i^\pm$ and $U_j^\pm$ overlap on the sphere, and a point p lies in their intersection. The chart $f_i^\pm$ maps $U_i^\pm$ to its coordinate image $D_i = f_i^\pm(U_i^\pm) \subset \mathbb{R}^2$. The transition 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.

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 open subsets of Euclidean space. This relationship between overlapping chart domains and their coordinate images is illustrated in Fig. 1.8, 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.

* * *

1.3 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 defined using local coordinate charts. We now introduce a new type of geometric object that captures how such functions change: *tangent vectors*.

Where a smooth function assigns a number to each point of a manifold, a tangent vector encodes a possible *direction of change* at a point. These directions form the tangent space and provide the local linear approximation to the manifold. Tangent vectors and tangent spaces are the fundamental objects from which all tensorial quantities in general relativity are built.

Geometric picture. Let M be a smooth manifold and let $p \in M$. A tangent vector at p can be understood geometrically as the velocity of a smooth curve passing through p . If

$$\gamma : (-\varepsilon, \varepsilon) \rightarrow M$$

is a smooth curve with $\gamma(0) = p$, then its velocity at p , denoted $\dot{\gamma}(0)$, represents a direction along the manifold at that point. The parameter t is an ordinary real variable, and $(-\varepsilon, \varepsilon)$ denotes a small open time interval around $t = 0$ on which the curve is defined. Differentiation is taken only after composing with a smooth function on M , so the definition of a tangent vector does not rely on coordinates or on an embedding of the manifold.

The collection of all such velocity vectors forms the *tangent space* $T_p M$, which is a real vector space of the same dimension as the manifold.

From curves to derivations. The geometric notion of velocity can be expressed intrinsically, without reference to an embedding, by letting tangent vectors act on smooth functions. Given a curve $\gamma(t)$ with $\gamma(0) = p$, we define a linear operator $v : C^\infty(M) \rightarrow \mathbb{R}$ by

$$v(f) := \left. \frac{d}{dt} [f(\gamma(t))] \right|_{t=0}. \quad (1.1)$$

This operator measures the rate of change of the scalar field f along the curve at the point p .

It is straightforward to verify that v is linear and satisfies the Leibniz rule,

$$v(fg) = f(p)v(g) + g(p)v(f), \quad (1.2)$$

and is therefore a *derivation* at p . This derivation is called the *tangent vector* to the curve γ at p , denoted $\dot{\gamma}(0)$.

Example 1.5: A tangent vector on S^2 from a curve

Let

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

and consider the point

$$p = (0, 0, 1) \in S^2.$$

We construct a tangent vector at p using a smooth curve on the sphere.

Define a curve

$$\gamma : (-\varepsilon, \varepsilon) \rightarrow S^2, \quad \gamma(t) = (\sin t, 0, \cos t).$$

Then $\gamma(0) = p$, and $\gamma(t)$ lies entirely on S^2 for all t . Geometrically, this curve moves along a meridian of the sphere passing through the north pole.

Now let $f : S^2 \rightarrow \mathbb{R}$ be any smooth function. For example, consider the height function

$$f(x, y, z) = z.$$

The composition $f \circ \gamma$ is an ordinary real-valued function of t :

$$(f \circ \gamma)(t) = f(\sin t, 0, \cos t) = \cos t.$$

Differentiating at $t = 0$ gives

$$\left. \frac{d}{dt}(f \circ \gamma)(t) \right|_{t=0} = 0.$$

More generally, for any smooth function $f \in C^\infty(S^2)$, we define

$$v(f) := \left. \frac{d}{dt}[f(\gamma(t))] \right|_{t=0}.$$

This assignment takes a smooth function on S^2 and returns a real number. It measures the rate of change of f along the curve γ at the point p .

The operator $v : C^\infty(S^2) \rightarrow \mathbb{R}$ is linear and satisfies the Leibniz rule,

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

and therefore defines a *derivation at p* . This derivation is called the *tangent vector* to the curve γ at p , denoted $\dot{\gamma}(0)$. Different curves through p define different tangent vectors, and the collection of all such vectors forms the tangent space $T_p S^2$.

* * *

the textbook's perspective: A tangent vector at a point $p \in M$ is defined to be a derivation $v : C^\infty(M) \rightarrow \mathbb{R}$ satisfying linearity and the Leibniz rule. The set of all such derivations forms the tangent space $T_p M$.

Lee's formal framework:

- A *tangent vector at p* is a linear map $v : C^\infty(M) \rightarrow \mathbb{R}$ satisfying

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

- The collection of all tangent vectors at p forms a real vector space, the *tangent space* $T_p M$.
- If $(U, x^1, \dots, x^n)$ is a coordinate chart containing p , then the operators

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

form a basis for $T_p M$, where

$$\left. \frac{\partial}{\partial x^i} \right|_p (f) := \left. \frac{\partial(f \circ x^{-1})}{\partial x^i} \right|_{x(p)}.$$

- Any tangent vector may therefore be written as

$$v = v^i \left. \frac{\partial}{\partial x^i} \right|_p, \quad v^i \in \mathbb{R}.$$

Equivalence of viewpoints. Every velocity vector $\dot{\gamma}(0)$ of a smooth curve defines a derivation, and conversely, every derivation arises as the velocity of some curve through p . Thus, the geometric and algebraic definitions of tangent vectors are equivalent. In practice, we freely switch between thinking of tangent vectors as directions along curves and as directional derivatives acting on scalar fields.

Coordinate functions and tangent vectors. When we choose a coordinate chart $(U, (x^1, \dots, x^n))$ on a manifold M , we are not merely assigning numerical labels to points. Rather, we are selecting n distinguished smooth real-valued functions $x^i \in C^\infty(M)$ defined on U .

Because a tangent vector $v \in T_p M$ is defined abstractly as a derivation acting on smooth functions, it produces numerical components by acting on these coordinate functions:

$$v^i := v(x^i).$$

These numbers are the components of v in the chosen coordinate system.

The coordinate basis vectors

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

are precisely the derivations characterized by the property

$$\left. \frac{\partial}{\partial x^i} \right|_p (x^j) = \delta_i^j.$$

That is, each $\partial/\partial x^i|_p$ extracts the i -th coordinate component of a tangent vector.

With this identification, any tangent vector admits the coordinate representation

$$v = v^i \left. \frac{\partial}{\partial x^i} \right|_p. \quad (1.3)$$

which reconciles the abstract derivation-based definition with the familiar coordinate expression.

Example 1.6: *Coordinate representation of a tangent vector on S^2*

We continue the previous example on the unit sphere S^2 at the point

$$p = (0, 0, 1).$$

Recall that the curve

$$\gamma(t) = (\sin t, 0, \cos t)$$

defines a tangent vector $v = \dot{\gamma}(0) \in T_p S^2$ via

$$v(f) = \left. \frac{d}{dt} [f(\gamma(t))] \right|_{t=0}, \quad f \in C^\infty(S^2).$$

To express this tangent vector in coordinates, we choose a coordinate chart around p . Since p lies in the upper hemisphere $U_3^+ = \{z > 0\}$, we use the chart

$$f_3^+ : U_3^+ \rightarrow \mathbb{R}^2, \quad f_3^+(x, y, z) = (x, y).$$

The associated coordinate functions on U_3^+ are

$$x(x, y, z) = x, \quad y(x, y, z) = y.$$

The components of the tangent vector v in this coordinate system are obtained by letting v act on the coordinate functions:

$$\begin{aligned} v(x) &= \frac{d}{dt}[x(\gamma(t))]\Big|_{t=0} = \frac{d}{dt}(\sin t)\Big|_{t=0} = 1, \\ v(y) &= \frac{d}{dt}[y(\gamma(t))]\Big|_{t=0} = \frac{d}{dt}(0)\Big|_{t=0} = 0. \end{aligned}$$

Thus, in the coordinate chart (U_3^+, x, y) , the tangent vector v has components

$$(v^x, v^y) = (1, 0),$$

and may be written as

$$v = \frac{\partial}{\partial x}\Big|_p.$$

This expression depends on the chosen coordinates, but the tangent vector itself does not. Different coordinate charts would assign different components to the same abstract vector v , while representing the same geometric direction along the sphere at the point p .

* * *

Vector fields and commutators. A *vector field* assigns a tangent vector to each point of M . Given two smooth vector fields X and Y , their *commutator* (or *Lie bracket*) is defined by

$$[X, Y](f) := X(Yf) - Y(Xf), \quad f \in C^\infty(M). \quad (1.4)$$

This measures the failure of the two directional derivatives to commute. The Leibniz rule ensures that $[X, Y]$ is again a derivation and hence a vector field. The space of smooth vector fields $\mathfrak{X}(M)$ is therefore closed under the Lie bracket and forms a Lie algebra.

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

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

The tangent space $T_p M$ consists of all possible velocity vectors an observer at that event could have. Fields of tangent vectors describe physical quantities such as velocity and momentum fields.

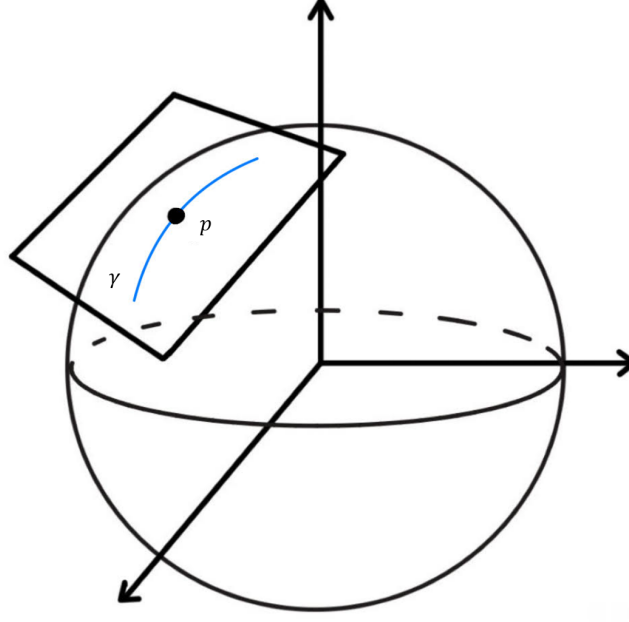


Figure 1.9: Tangent plane $T_p M$ to a two-dimensional manifold M (the sphere) at the point $p = \gamma(0)$. A smooth curve $\gamma(t)$ on M passes through p , and its velocity vector $\dot{\gamma}(0)$ lies in the tangent space. This illustrates how directions along the manifold correspond to vectors in the tangent space.

Summary.

- Tangent vectors represent directions of change at points on a manifold.
- They may be understood as velocities of curves or as derivations acting on smooth functions.
- In coordinates, tangent vectors are linear combinations of partial derivatives.

1.4 The Differential and Pushforward

In the previous section, we defined tangent vectors as abstract objects: derivations acting on smooth functions. We now ask how tangent vectors behave under smooth maps between manifolds.

Given a smooth map $F : M \rightarrow N$ and a point $p \in M$, the map F sends the point p to the point $F(p)$. Since tangent vectors represent directions of motion or change at a point, it is natural to ask: *how does a direction at p map to a direction at $F(p)$?*

The answer is provided by the *differential* of F at p , also called the *pushforward*. It is a linear map

$$dF_p : T_p M \longrightarrow T_{F(p)} N$$

that carries tangent vectors at p to tangent vectors at $F(p)$.

Definition via derivations. The definition of the differential is most naturally expressed using the derivation viewpoint of tangent vectors.

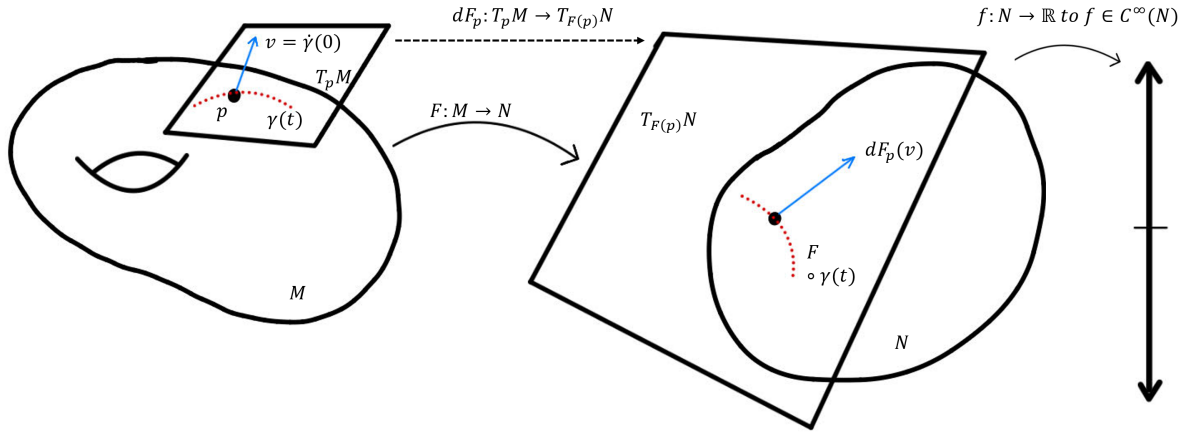


Figure 1.10: Geometric picture of the pushforward dF_p : a tangent vector $v \in T_p M$, realized as the velocity of a curve γ at p , is mapped to the tangent vector $dF_p(v) \in T_{F(p)} N$ of the image curve $F \circ \gamma$ at $F(p)$. Acting on a smooth function $f \in C^\infty(N)$, the pushforward satisfies $(dF_p v)(f) = v(f \circ F)$.

Definition (Lee). Let $F : M \rightarrow N$ be a smooth map and let $p \in M$. For each tangent vector $v_p \in T_p M$, the *differential* of F at p is the tangent vector $dF_p(v_p) \in T_{F(p)} N$ defined by its action on smooth functions $f \in C^\infty(N)$:

$$(dF_p(v_p))(f) := v_p(f \circ F). \quad (1.5)$$

That is, the new tangent vector differentiates a function on the target manifold N by first pulling it back to M along F , then applying the original tangent vector v_p . Linearity of dF_p follows immediately from the linearity of v_p .

Geometric interpretation using curves. This definition has a direct geometric meaning. Recall from Eq. 1.1 that every tangent vector $v_p \in T_p M$ can be realized as the velocity of a curve $\gamma(t)$ with $\gamma(0) = p$:

$$v_p(f) = \frac{d}{dt} [f(\gamma(t))] \Big|_{t=0}.$$

Applying the definition of the pushforward Eq. 1.5,

$$(dF_p(v_p))(f) = \frac{d}{dt} [f(F(\gamma(t)))] \Big|_{t=0}. \quad (1.6)$$

Thus $dF_p(v_p)$ is precisely the velocity at $F(p)$ of the image curve $F \circ \gamma$.

In words: *the differential maps the velocity of a curve at p to the velocity of its image curve at $F(p)$.* This interpretation is illustrated in Fig. 1.10.

Example 1.7: *The differential of the height function on S^2*

Let $M = S^2 \subset \mathbb{R}^3$, and consider the height function

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

Choose the point

$$p = (0, 0, 1) \in S^2,$$

and consider the smooth curve

$$\gamma(t) = (\sin t, 0, \cos t),$$

which lies on S^2 and satisfies $\gamma(0) = p$. The velocity of this curve at $t = 0$ is

$$v_p = \dot{\gamma}(0) = (1, 0, 0),$$

which defines a tangent vector $v_p \in T_p S^2$.

To relate this back to the definition from Lee stated above,

- $M = S^2$
- $N = \mathbb{R}$
- $F = h : S^2 \rightarrow \mathbb{R}$
- $v_p \in T_p S^2$

The pushforward of v_p under h is defined by

$$(dh_p)(v_p) = \left. \frac{d}{dt} [h(\gamma(t))] \right|_{t=0}.$$

Since $h(\gamma(t)) = \cos t$, we obtain

$$(dh_p)(v_p) = \left. \frac{d}{dt} (\cos t) \right|_{t=0} = 0.$$

Thus, although the curve moves along the sphere, its initial direction has no vertical component, as shown in Fig. 1.11. The differential dh_p detects this by assigning the value 0 to the tangent vector v_p . In contrast, a curve with a nonzero vertical component would yield a nonzero value under dh_p .

This example shows explicitly how the differential of a smooth function acts on tangent vectors by measuring rates of change along curves.

In the special case $F : M \rightarrow \mathbb{R}$, the differential dF_p maps tangent vectors at p to real numbers, measuring the rate of change of the function F along those directions (recall in this example $F = h$).

* * *

Coordinate expression. The abstract definition becomes concrete once coordinates are chosen. Let (x^i) be local coordinates on M near p , and let (y^α) be local coordinates on N near $F(p)$. In these coordinates, the map F is written as

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

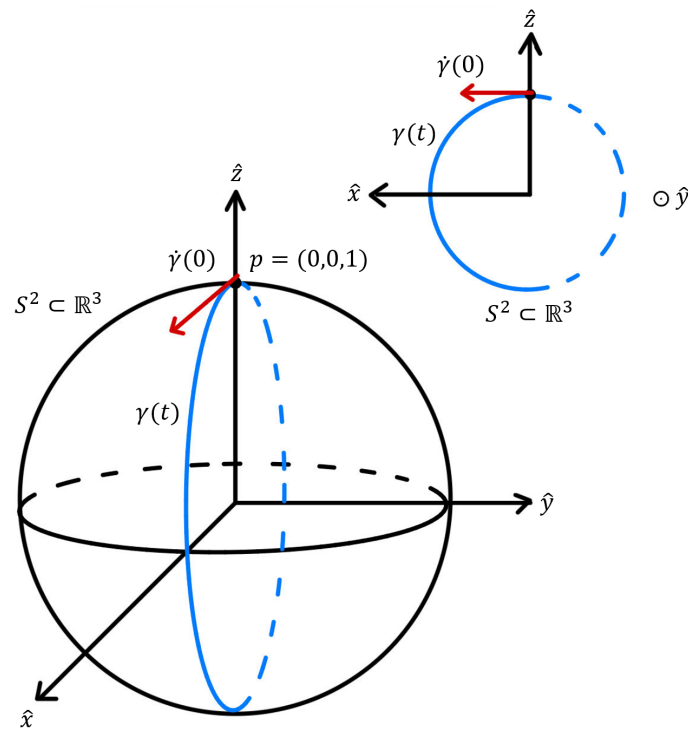


Figure 1.11: A curve $\gamma(t) = (\sin t, 0, \cos t)$ on the unit sphere $S^2 \subset \mathbb{R}^3$ passing through the point $p = (0, 0, 1)$. The velocity vector $\dot{\gamma}(0)$ lies entirely in the horizontal (x) direction and has no vertical (z) component. Since the height function $h(x, y, z) = z$ measures vertical change, the differential dh_p assigns the value 0 to this tangent vector. The inset shows the same curve viewed in the x - z plane, with the y -direction pointing out of the page.

Applying the definition of dF_p to the coordinate basis vectors $\frac{\partial}{\partial x^i}\big|_p$, one finds

$$dF_p\left(\frac{\partial}{\partial x^i}\bigg|_p\right) = \frac{\partial F^\alpha}{\partial x^i}(p) \frac{\partial}{\partial y^\alpha}\bigg|_{F(p)}. \quad (1.7)$$

Thus, in coordinates, the differential is represented by the Jacobian matrix

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

This is the precise sense in which the differential generalizes the Jacobian of multivariable calculus to smooth maps between manifolds. The coordinate relationship is illustrated in Fig. 1.12.

Derivation of the coordinate formula.

Let $v_p = \frac{\partial}{\partial x^i}\big|_p \in T_p M$. By definition of the differential,

$$(dF_p(v_p))(f) = v_p(f \circ F), \quad f \in C^\infty(N).$$

Evaluating the right-hand side in local coordinates and applying the chain rule,

$$\frac{\partial}{\partial x^i}\bigg|_p (f \circ F) = \sum_\alpha \frac{\partial(f \circ y^{-1})}{\partial y^\alpha}\bigg|_{y(F(p))} \frac{\partial F^\alpha}{\partial x^i}(p).$$

Recognizing that

$$\frac{\partial}{\partial y^\alpha}\bigg|_{F(p)}(f) = \frac{\partial(f \circ y^{-1})}{\partial y^\alpha}\bigg|_{y(F(p))},$$

we obtain

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

Example 1.8: *Coordinate expression of the differential on S^2*

We continue with the height function

$$h : S^2 \rightarrow \mathbb{R}, \quad h(x, y, z) = z,$$

and the point $p = (0, 0, 1) \in S^2$.

Near p , use the chart (U_3^+, f_3^+) , where

$$f_3^+(x, y, z) = (u, v) = (x, y), \quad (f_3^+)^{-1}(u, v) = (u, v, \sqrt{1 - u^2 - v^2}).$$

Thus (u, v) are local coordinates on S^2 near p , and the coordinate representation of h is

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

where $\hat{h}$ is the coordinate representation of the smooth map h with respect to the chart (U_3^+, f_3^+) .

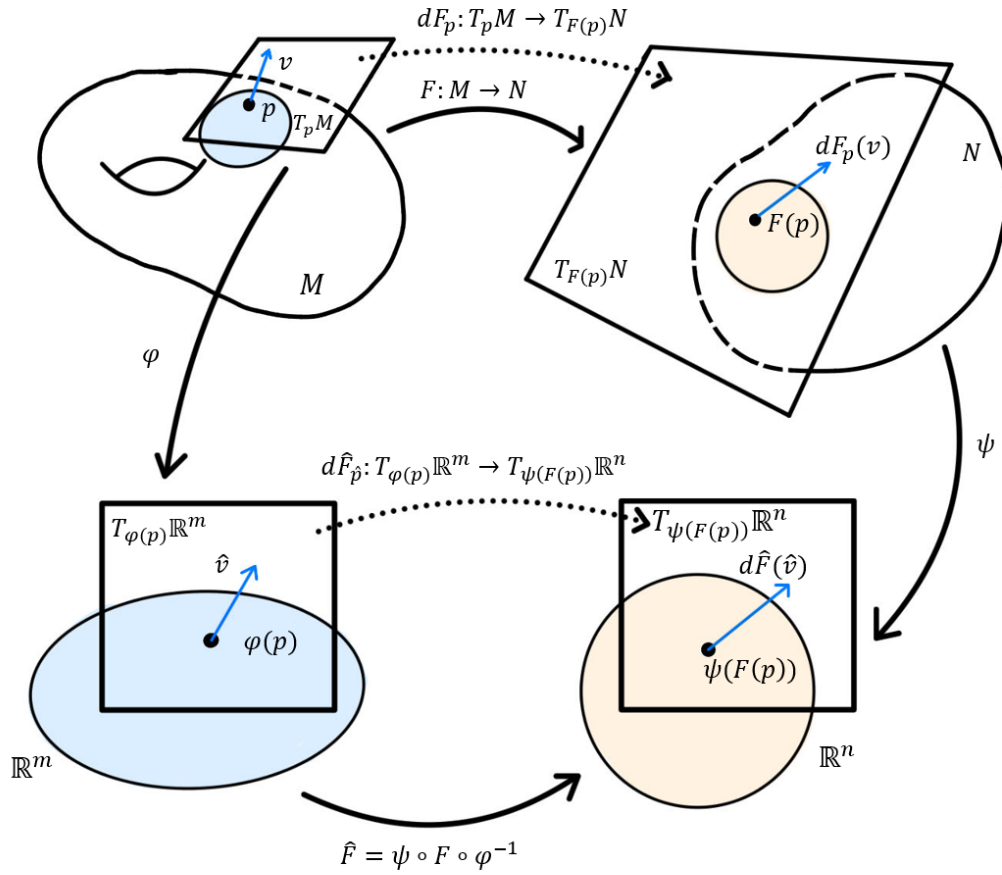


Figure 1.12: The differential dF_p relates tangent spaces of two manifolds M and N under a smooth map $F : M \rightarrow N$. In local coordinates (U, φ) and (V, ψ) , it corresponds to the ordinary Jacobian of the coordinate representation $\hat{F} = \psi \circ F \circ \varphi^{-1}$.

Since the target manifold is $\mathbb{R}$, we use the standard coordinate $y = z$ on $\mathbb{R}$. The Jacobian of $\hat{h}$ is therefore

$$\frac{\partial \hat{h}}{\partial u} = -\frac{u}{\sqrt{1-u^2-v^2}}, \quad \frac{\partial \hat{h}}{\partial v} = -\frac{v}{\sqrt{1-u^2-v^2}}.$$

Evaluating at the coordinate point corresponding to p , namely $(u, v) = (0, 0)$, we find

$$\left. \frac{\partial \hat{h}}{\partial u} \right|_{(0,0)} = 0, \quad \left. \frac{\partial \hat{h}}{\partial v} \right|_{(0,0)} = 0.$$

Thus, in coordinates, the differential at p is

$$dh_p = 0 du + 0 dv,$$

and for any tangent vector

$$v_p = a \left. \frac{\partial}{\partial u} \right|_p + b \left. \frac{\partial}{\partial v} \right|_p \in T_p S^2,$$

we have

$$(dh_p)(v_p) = 0.$$

This agrees with the geometric result obtained earlier: at the north pole, the height function has zero first-order variation in all tangent directions.

Note: Although the expression $h(x, y, z) = z$ resembles a coordinate function, it is not a chart on S^2 ; rather, it is a smooth scalar field obtained by restricting the ambient coordinate function $z : \mathbb{R}^3 \rightarrow \mathbb{R}$ to the submanifold $S^2 \subset \mathbb{R}^3$. The target manifold $\mathbb{R}$ already carries a global coordinate, so no additional chart on the codomain is required.

* * *

Physical interpretation. In physics, particularly in general relativity, the differential is usually called the *pushforward* and written F_* . It appears whenever directions or velocities are transported between spaces:

- Coordinate transformations act on tangent vectors via Jacobian matrices.
- Embeddings of submanifolds push tangent vectors into ambient spacetimes.
- Worldlines $\gamma(\tau)$ generate four-velocity vectors that are naturally interpreted via pushforwards.

Summary. The differential dF_p is the geometric mechanism by which a smooth map carries tangent vectors from one manifold to another. It is defined intrinsically using derivations, admits a clear interpretation in terms of curves, and reduces in coordinates to the familiar Jacobian matrix.

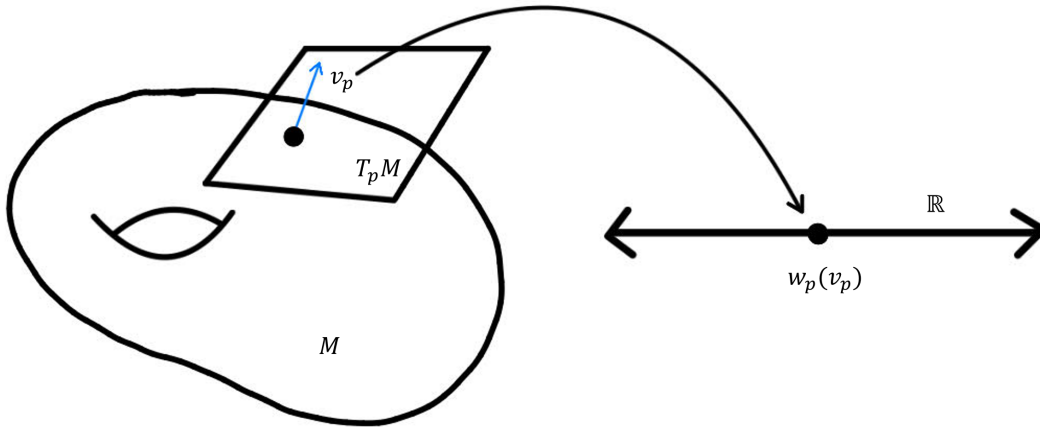


Figure 1.13: A covector $\omega_p \in T_p^*M$ acts as a linear map from the tangent space T_pM to the real numbers. Given a tangent vector $v_p \in T_pM$, the covector assigns a real value $\omega_p(v_p) \in \mathbb{R}$. Thus, while tangent vectors represent directions at a point, covectors measure the response of scalar quantities along those directions.

1.5 Covectors and the Pullback

Tangent vectors describe *directions of motion* on a manifold. Their dual objects, called *covectors* or *one-forms*, are linear maps that assign real numbers to those directions. Where a tangent vector tells us *how to move*, a covector tells us *how a quantity is evaluated along that motion*. In particular, the differential of a scalar field is a covector that measures the rate of change of the field along a given direction. This dual pairing underlies gradients, differentials, and ultimately the tensor calculus used in general relativity.

Cotangent space. At a point $p \in M$, the *cotangent space* to M is defined as

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

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**. Here $\text{Hom}(T_pM, \mathbb{R})$ denotes the set of all linear maps from the vector space T_pM to $\mathbb{R}$; this set itself forms a vector space under pointwise addition and scalar multiplication.

If $v_p \in T_pM$ is a tangent vector and $\omega_p \in T_p^*M$ is a covector, their pairing

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

is a real number, shown in Fig. 1.13. This pairing is bilinear:

$$\omega_p(av_1 + bv_2) = a\omega_p(v_1) + b\omega_p(v_2), \quad a, b \in \mathbb{R}.$$

Coordinate representation. Let $(x^1, \dots, x^n)$ be local coordinates on M near p . The coordinate basis for the tangent space $T_p M$ is

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

and the dual basis for the cotangent space $T_p^* M$ is

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

defined by

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

Any covector at p can therefore be written as

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

and for a tangent vector

$$v_p = v^j \frac{\partial}{\partial x^j} \Big|_p,$$

their pairing is

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

with summation over repeated indices understood.

Differentials of functions. Every smooth scalar field $f : M \rightarrow \mathbb{R}$ naturally defines a covector at each point. Its *differential* at p , denoted df_p , is the covector defined by

$$df_p(v_p) = v_p(f), \quad v_p \in T_p M. \quad (1.8)$$

If $v_p = \dot{\gamma}(0)$ is the velocity of a curve $\gamma(t)$ with $\gamma(0) = p$, then

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

Thus df_p measures the rate of change of f in the direction specified by v_p .

In local coordinates,

$$df = \frac{\partial f}{\partial x^i} dx^i. \quad (1.9)$$

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

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

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

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

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

Pullback of covectors. While tangent vectors are pushed *forward* by a smooth map, covectors are naturally pulled *backward*.

Let $F : M \rightarrow N$ be a smooth map and let $\omega_{F(p)} \in T_{F(p)}^*N$. The *pullback* of ω by F is the covector $F^*\omega_{F(p)} \in T_p^*M$ defined by

$$(F^*\omega_{F(p)})(v_p) = \omega_{F(p)}(dF_p(v_p)), \quad v_p \in T_pM. \quad (1.10)$$

Thus the pullback acts by first pushing the vector v_p forward to $T_{F(p)}N$, and then applying the covector ω .

In coordinates, if

$$\omega = \omega_\alpha dy^\alpha, \quad y^\alpha = F^\alpha(x^i),$$

then

$$F^*\omega = \omega_\alpha \frac{\partial F^\alpha}{\partial x^i} dx^i. \quad (1.11)$$

Example 1.9: Pullback of a covector on S^2

Let $M = S^2 \subset \mathbb{R}^3$, and consider the height function

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

As shown previously, its differential dh is a covector field on S^2 .

Now define a smooth map

$$F : S^2 \rightarrow S^2,$$

given by a rotation about the y -axis through an angle θ :

$$F(x, y, z) = (\cos \theta x + \sin \theta z, y, -\sin \theta x + \cos \theta z).$$

This map preserves the sphere and moves points along meridians, as shown in Fig. 1.14.

The covector $dh_{F(p)} \in T_{F(p)}^*S^2$ measures the rate of change of height at the point $F(p)$.

The pullback $F^*dh_p \in T_p^*S^2$ is defined by

$$(F^*dh_p)(v_p) = dh_{F(p)}(dF_p(v_p)),$$

for any tangent vector $v_p \in T_pS^2$.

Equivalently, since $dh = d(h)$, the pullback satisfies

$$F^*dh = d(h \circ F).$$

Thus the pulled-back covector measures how the *height after rotation* changes along directions on the original sphere.

Explicitly,

$$(h \circ F)(x, y, z) = -\sin \theta x + \cos \theta z,$$

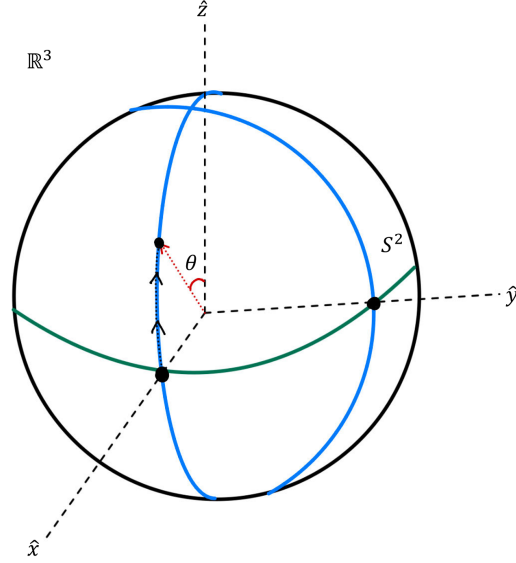


Figure 1.14: An ambient rotation in $\mathbb{R}^3$ about the y -axis. The map $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is defined by rotating points through an angle θ about the y -axis. Since this rotation preserves the sphere $S^2 \subset \mathbb{R}^3$, it restricts to a smooth map $F : S^2 \rightarrow S^2$. The motion of a point along a meridian illustrates the action of F on S^2 .

so

$$F^*dh = d(-\sin \theta x + \cos \theta z) = -\sin \theta dx + \cos \theta dz.$$

This covector assigns to each tangent vector at p the rate of change of the rotated height function along that direction.

We now evaluate the pulled-back covector on a specific tangent vector to obtain a concrete numerical result.

Consider the point

$$p = (0, 0, 1) \in S^2,$$

the north pole. The tangent space at p is

$$T_p S^2 = \{(a, b, c) \in \mathbb{R}^3 : (0, 0, 1) \cdot (a, b, c) = 0\} = \{(a, b, 0)\}.$$

Choose the tangent vector

$$v_p = (1, 0, 0) \in T_p S^2,$$

which points along the meridian through p in the positive x -direction.

Since

$$F^*dh = -\sin \theta dx + \cos \theta dz,$$

we evaluate this covector on v_p by applying it to the coordinate functions:

$$dx(v_p) = 1, \quad dz(v_p) = 0.$$

Therefore,

$$(F^*dh)_p(v_p) = -\sin \theta \, dx(v_p) + \cos \theta \, dz(v_p) = -\sin \theta.$$

This number represents the rate of change of the *rotated height function* $h \circ F$ at the point p in the direction v_p . In particular, if $\theta = \pi/6$, then

$$(F^*dh)_p(v_p) = -\frac{1}{2}.$$

Covectors are pulled back because they must act on tangent vectors at the original point of the map.

This example illustrates the essential feature of the pullback: covectors are transported *backward* along a smooth map so that their action on vectors reproduces the same physical or geometric measurement. While tangent vectors are pushed forward by dF_p , covectors must be pulled back by F^* to preserve scalar quantities.

* * *

Physical interpretation. Covectors represent gradients, momenta, and field differentials.

- The differential df of a scalar field gives the local gradient as a covector.
- Under a coordinate transformation $x^\mu \mapsto x^{\mu'}(x)$, covector components transform with the inverse Jacobian:

$$\omega_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \omega_\mu.$$

- The pullback F^* explains how scalar gradients, potentials, and fluxes are transported between manifolds.

1.6 Tensors and Tensor Fields

Having introduced tangent vectors and cotangent vectors, we now consider objects built from them. A *tensor* at a point $p \in M$ is a multilinear map that takes a fixed number of tangent and cotangent vectors as input and returns a real number. Tensors generalize vectors and covectors and provide the basic language for expressing physical laws in a coordinate-independent way.

Tensors as multilinear maps. Let M be a smooth manifold and $p \in M$. A tensor of type (r, s) at p is a multilinear map

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

Each input slot is linear in its own argument, and the order and type of the slots matter: covector slots and vector slots are not interchangeable. The collection of all tensors of type (r, s) at p forms a vector space, denoted $T_s^r(T_pM)$.

the textbook'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. In coordinates, its 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. This transformation behavior ensures that tensorial equations have the same form in all coordinate systems.

Lee's formal framework:

- For any vector space V , its dual is $V^* = \{\text{linear maps } V \rightarrow \mathbb{R}\}$.
- The tensor product $V^{\otimes r} \otimes (V^*)^{\otimes s}$ may be identified with the space of multilinear maps $(V^*)^r \times V^s \rightarrow \mathbb{R}$, called tensors of type (r, s) .
- On a smooth manifold M , this construction applied pointwise defines the tensor bundle

$$T_s^r(M) = \bigcup_{p \in M} T_s^r(T_p M),$$

and smooth sections of this bundle are called tensor fields.

Coordinate representation. Once a coordinate chart $(U, x^1, \dots, x^n)$ is chosen around p , a tensor $T \in T_s^r(T_p M)$ can be expressed in components. If $\{\partial/\partial x^i|_p\}$ is the associated basis of $T_p M$ and $\{dx^i|_p\}$ the dual basis of $T_p^* M$, then any tensor may be written as

$$T = T^{a_1 \dots a_r}_{b_1 \dots b_s} \frac{\partial}{\partial x^{a_1}} \Big|_p \otimes \dots \otimes \frac{\partial}{\partial x^{a_r}} \Big|_p \otimes dx^{b_1} \Big|_p \otimes \dots \otimes dx^{b_s} \Big|_p.$$

Evaluating the tensor on covectors ω_i and vectors v_j corresponds to contracting these components with the coordinate components of the inputs.

Bridging the viewpoints. The abstract multilinear definition and the component-based index notation describe the same geometric object. At each point $p \in M$, a tensor of type (r, s) acts on r covectors and s vectors to produce a real number:

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

The position of the indices reflects which slots are filled by vectors and which by covectors.

Tensor fields. A *tensor field* assigns to each point $p \in M$ a tensor of fixed type (r, s) in a smooth way. Smoothness means that the components of the tensor vary smoothly in any coordinate chart. Tensor fields allow local geometric data to be compared from point to point across the manifold.

Physical interpretation. Tensors provide the natural language of field theory:

- The *metric tensor* g_{ab} assigns lengths and angles to tangent vectors.
- The *electromagnetic field* F_{ab} is an antisymmetric tensor field encoding electric and magnetic fields.

- The *stress-energy tensor* T_{ab} encodes the local distribution of energy, momentum, and pressure.

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

Looking ahead. Among all tensor fields, the metric tensor plays a distinguished role. It provides a natural way to convert vectors into covectors and to measure lengths, angles, and spacetime intervals. We turn to this object next.

Summary.

- A tensor of type (r, s) is a multilinear map acting on r covectors and s vectors.
- Tensors admit coordinate representations, but their definition is independent of any particular coordinate system.
- A tensor field assigns such an object smoothly to each point of a manifold.
- Tensorial transformation laws ensure coordinate-independent physical and geometric statements.

1.7 The Metric Tensor

With tangent vectors, covectors, and general tensors in hand, we now introduce a distinguished tensor field: the *metric tensor*. The metric provides a systematic way to compare tangent vectors at a point. It measures lengths and angles and, in relativity, encodes the causal structure of spacetime.

The metric as a tensor field. Let M be a smooth manifold. A metric assigns to each point $p \in M$ a bilinear map

$$g_p : T_p M \times T_p M \rightarrow \mathbb{R},$$

smoothly varying with p . Thus, the metric is a tensor field of type $(0, 2)$.

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

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

interpreted as their inner product. The signature $(-+++)$ distinguishes timelike, null, and spacelike directions.

Lee's formal framework: A *Riemannian metric* on a smooth manifold M is a smooth assignment $p \mapsto g_p$ such that, for each $p \in M$,

- g_p is bilinear in each argument;
- 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 $v = 0$.
 If $g_p(v, v) > 0$ for all nonzero v , the metric is *Riemannian*. If the metric has indefinite signature, it is *pseudo-Riemannian*; the Lorentzian case is used in general relativity.

Geometric meaning. The metric equips each tangent space $T_p M$ with an inner product. In particular, it allows us to define:

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

In Lorentzian geometry, the value of $g_p(v, v)$ may be positive, negative, or zero. These cases correspond to spacelike, timelike, and null directions respectively. Thus the metric determines not only distances, but the causal structure of spacetime.

Lowering and raising indices. Because the metric is non-degenerate, it defines a natural identification between vectors and covectors. For each point p , define the *lowering map*

$$\flat : T_p M \rightarrow T_p^* M, \quad v_p \mapsto v_p^\flat,$$

by

$$v_p^\flat(w_p) := g_p(v_p, w_p).$$

This associates to every vector a covector.

The inverse map,

$$\sharp : T_p^* M \rightarrow T_p M, \quad \omega_p \mapsto \omega_p^\sharp,$$

is defined implicitly by

$$g_p(\omega_p^\sharp, w_p) = \omega_p(w_p) \quad \text{for all } w_p \in T_p M.$$

In coordinates, these operations correspond to

$$v_a = g_{ab} v^b, \quad v^a = g^{ab} v_b,$$

where g^{ab} is the inverse matrix of g_{ab} .

Physical interpretation. The metric provides the geometric foundation of spacetime physics:

- It defines proper time and proper distance via $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$.
- It determines light cones and causal relationships.
- It enables the conversion between vectors and covectors, allowing physical laws to be written in covariant form.

Example 1.10: Minkowski metric

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

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

For tangent vectors v^μ and 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 leads to the classification

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

* * *

Why the metric fits the tensor definition. By definition, a tensor at a point p is a multilinear map

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

The metric at a point,

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

is therefore a tensor of type $(0, 2)$.

When we write components $g_{\mu\nu}$, we are referring not to a single bilinear form, but to the entire *metric tensor field*: a smooth assignment $p \mapsto g_p$. In Minkowski spacetime the components are constant, so the same matrix represents g_p at every point. In curved spacetime, the components $g_{\mu\nu}(x)$ vary with position, and this distinction becomes essential.

Example 1.11: Recovering $\hat{x}$ from the manifold definition

Let $M = \mathbb{R}^3$ with global coordinates (x, y, z) , viewed as a smooth manifold. At any point $p \in \mathbb{R}^3$, the coordinate derivations

$$\left. \frac{\partial}{\partial x} \right|_p, \quad \left. \frac{\partial}{\partial y} \right|_p, \quad \left. \frac{\partial}{\partial z} \right|_p$$

form a basis of the tangent space $T_p\mathbb{R}^3$.

Introduce the Euclidean metric

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

Then

$$g\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial x}\right) = 1, \quad g\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = 0,$$

and similarly for the other basis vectors. Thus the coordinate basis is orthonormal.

With this identification, the derivation

$$\frac{\partial}{\partial x}$$

corresponds exactly to the familiar unit vector

$$\hat{x} = (1, 0, 0).$$

Therefore, the standard basis vectors of Euclidean space arise as metric-normalized coordinate basis vectors.

* * *

Summary. The metric tensor is a symmetric, non-degenerate $(0, 2)$ -tensor field that assigns an inner product to each tangent space. It enables:

- measurement of lengths and angles;
- identification of vectors and covectors;
- and, in relativity, the classification of causal directions.

The metric is the geometric heart of both Riemannian geometry and general relativity.

1.8 Orientation and Volume Forms (Optional)

Orientation and volume forms provide the final geometric ingredients needed to define integration on manifolds in a coordinate-independent way. Orientation supplies a consistent notion of “handedness,” while a volume form specifies how to measure signed volume. These structures play a central role in flux integrals, conservation laws, and action principles in general relativity, but are not strictly required for the development that follows. For this reason, the present section may be read as optional.

the textbook’s perspective: Spacetime admits a smooth, nowhere-vanishing volume element ϵ_{abcd} , a completely antisymmetric tensor field of type $(0, 4)$. In any positively oriented coordinate system,

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

and under a change of coordinates it transforms with the sign of the Jacobian determinant. This object defines the invariant spacetime volume element used in integration and in expressing physical conservation laws.

Lee’s formal framework: A smooth manifold M is *orientable* if it admits an atlas whose transition maps all have positive Jacobian determinant. An *orientation* is a choice of such an equivalence class of atlases.

On an oriented n -dimensional manifold, a *volume form* is a smooth, nowhere-vanishing differential form of top degree,

$$\omega \in \Omega^n(M).$$

At each point p , the value $\omega_p(v_1, \dots, v_n)$ gives the signed volume spanned by the tangent vectors v_i .

If the manifold carries a Riemannian or Lorentzian metric g , the metric determines a canonical volume form,

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

in any oriented coordinate chart.

Geometric and physical meaning. The mathematical and physical descriptions are two views of the same object. The tensor ϵ_{abcd} familiar from physics is the component representation of the volume form ω with respect to a positively oriented coordinate basis:

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

Because ω transforms covariantly under coordinate changes, integrals defined using it are invariant:

$$\int_M f \omega = \int f \sqrt{|\det g|} d^n x.$$

In spacetime physics, orientation distinguishes future from past and fixes sign conventions, while the volume form provides the invariant measure needed to define conserved quantities. It enters directly into flux integrals, the definition of the Hodge dual, and the Einstein–Hilbert action

$$S = \int R \sqrt{|\det g|} d^4 x.$$

Orientation and volume forms complete the geometric framework required to formulate physical laws in a coordinate-independent way. With these tools in place, we now possess the full kinematical structure of spacetime geometry. What remains is to introduce dynamics: connections and curvature, which describe how geometry itself evolves and interacts with matter.

1.9 Summary and Outlook

Summary. In this chapter we have built the geometric foundation required for general relativity using the language of differential geometry:

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

Physical picture. At this stage we possess a complete mathematical language for describing spacetime as a differentiable manifold: points correspond to events, tangent vectors to directions of motion, and tensor fields to 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 spacetime itself evolves or curves. In the next chapter we introduce connections and curvature, which supply the dynamical content of the theory.

Interdisciplinary Insight: Geometry Beyond Physics

The following perspective is not required for the remainder of the course, but illustrates the broader relevance of the geometric ideas developed in this chapter.

The mathematical structures introduced here – manifolds, tangent spaces, covectors, and tensors – form not only the language of general relativity, but also an increasingly important foundation for modern data science and machine learning.

Many real-world data sets do not occupy the full Euclidean space in which they are embedded. Instead, they lie on smooth *manifolds* of much lower intrinsic dimension. Analyzing and comparing data on such curved spaces requires precisely the tools of differential geometry.

- **Manifold Learning:** Algorithms such as ISOMAP, LLE, and T-SNE are designed to uncover the low-dimensional manifold structure underlying high-dimensional data.
- **Geometric Deep Learning:** Extends neural networks to non-Euclidean domains—such as graphs, surfaces, and manifolds—using ideas of symmetry and invariance.
- **Riemannian Optimization:** Many optimization problems involve curved parameter spaces (for example, spheres, orthogonal matrices, or covariance manifolds), where gradients and geodesics are defined using a metric.
- **Information Geometry:** Interprets families of probability distributions as points on a manifold endowed with the Fisher information metric, leading to the concept of a “natural gradient.”
- **Topological Data Analysis:** Uses tools from topology, such as persistent homology, to capture the global shape and connectivity of data sets in a way that is robust to noise.

Across these fields, the central questions echo those of spacetime geometry:

What is the shape of the space we are working in, and how do objects change and relate within it?

Differential geometry provides a unified language for addressing this question, whether the manifold under study is spacetime itself or a high-dimensional space of data.

Chapter 2

Curvature

In this chapter, we give precise mathematical meaning to what it means for spacetime to be curved.

We begin with a basic but subtle question: how can curvature be detected using only information that lives within the space itself? For familiar surfaces like a sphere, curvature can be seen by how the surface bends when embedded in a higher-dimensional space. Spacetime, however, is not assumed to sit inside anything larger. Any notion of curvature must therefore be intrinsic, defined entirely in terms of the geometry of the manifold.

To make such a definition possible, we must first address a more primitive issue. At each point of a manifold there is a tangent space, but there is no natural way to compare vectors living at different points. Without a rule for comparison, statements such as “a vector has changed” or “two directions are the same” are meaningless. We therefore introduce a geometric structure that allows vectors to be transported along curves while remaining “as unchanged as possible” according to the geometry. This notion of parallel transport provides the foundation for everything that follows.

Once vectors can be transported, curvature reveals itself naturally. If a vector is carried around a small closed loop and returns to its starting point unchanged, the geometry is locally flat. If it returns rotated or distorted, the geometry is curved. Curvature is thus identified with the failure of parallel transport to be path-independent.

This failure has a precise mathematical expression: transporting a vector in two different infinitesimal directions and comparing the results does not, in general, give the same answer. The object that measures this discrepancy is the Riemann curvature tensor, which encodes how the geometry twists and turns independently of any coordinate system.

Curvature is not merely an abstract property of vectors. It has direct consequences for the motion of particles and light. In curved spacetime, nearby geodesics that begin parallel can accelerate toward or away from one another. This relative acceleration is governed by the geodesic deviation equation and provides a direct link between curvature and physical tidal effects.

The chapter concludes by showing how these geometric ideas are implemented in practice. Given a metric, we outline systematic methods for computing the associated curvature quantities, preparing

the ground for the Einstein field equations developed in the next chapter.

2.1 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 $T_p M$ and $T_q M$ 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 on abstract indices

The symbols $a_1, \dots, a_k$ and $b_1, \dots, b_l$ are *abstract indices*: they label the tensor’s type, not its components. Upper indices correspond to contravariant (vector) slots, and lower indices to covariant (covector) slots. They tell us how the tensor contracts and transforms, but do not take numerical values until a coordinate basis is chosen.

For example, a rank $(3, 2)$ tensor field $T^{a_1 a_2 a_3}_{b_1 b_2} \in \mathcal{T}(3, 2)$ assigns to each point $p \in M$ a tensor $T^{a_1 a_2 a_3}_{b_1 b_2}(p) \in T_p^{(3,2)} M$, which is a multilinear map taking three vectors and two covectors at p and returning a real number.

Definition of a derivative operator

Definition 2.1. Derivative Operator A derivative operator ∇ is a map that takes a smooth tensor field

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

to another tensor field

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

such that the following five properties hold:

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

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

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

$$\begin{aligned} \nabla_e \left(A^{a_1 \dots a_k}_{b_1 \dots b_l} B^{c_1 \dots c_{k'}}_{d_1 \dots d_{l'}} \right) &= (\nabla_e A^{a_1 \dots a_k}_{b_1 \dots b_l}) B^{c_1 \dots c_{k'}}_{d_1 \dots d_{l'}} \\ &\quad + A^{a_1 \dots a_k}_{b_1 \dots b_l} (\nabla_e B^{c_1 \dots c_{k'}}_{d_1 \dots d_{l'}}). \end{aligned}$$

3. **Commutativity with contraction.** Derivative operators commute with index contraction. For all $A \in \mathcal{T}(k, l)$,

$$\nabla_d(A^{a_1 \dots c \dots a_k}_{b_1 \dots c \dots b_l}) = \nabla_d A^{a_1 \dots c \dots a_k}_{b_1 \dots c \dots b_l}.$$

Here the repeated index c , appearing once contravariantly and once covariantly, denotes index contraction. In abstract index notation, contraction is encoded by the index pattern itself rather than written as a separate operation. The equality above therefore asserts that contracting a tensor field and then applying ∇_d gives the same result as first applying ∇_d to the uncontracted tensor field and then contracting the same pair of indices. In this sense, covariant differentiation does not interfere with purely algebraic index operations.

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

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

5. **Torsion-free condition (scalar form).** For all smooth functions f ,

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

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

Geometric intuition

Linearity and the Leibniz rule ensure that ∇ behaves like a genuine differentiation operation compatible with the tensor algebra. Commutation with contraction guarantees that purely algebraic index operations are respected by differentiation. Equation 4 anchors ∇ to the familiar directional derivative on scalar fields. Finally, the torsion-free condition 5 excludes an additional geometric freedom (torsion) that is not part of standard GR; later we will express torsion as an antisymmetric tensor built from ∇ acting on vector fields.

Commutator of Vector Fields

Using properties (4) and (5), together 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 ,

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

Applying the Leibniz rule to each term gives

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

Subtracting these expressions, we obtain

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

The second term vanishes because the torsion-free condition implies that $\nabla_a \nabla_b f = \nabla_b \nabla_a f$, so the second derivative of a scalar is symmetric in the indices a and b . Contracting this symmetric expression with the antisymmetric combination $v^a w^b - w^a v^b$ therefore gives zero.

Thus,

$$[v, w](f) = \left(v^a \nabla_a w^b - w^a \nabla_a v^b \right) \nabla_b f.$$

Since this holds for all smooth functions f , we obtain

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

Thus, when the derivative operator is torsion-free, the Lie bracket of vector fields can be written entirely in terms of covariant derivatives.

If the vector fields v and w commuted, their Lie bracket would vanish. In general they do not, and the expression above shows that, for a torsion-free connection, this failure of commutativity is fully captured by the covariant derivatives of the vector fields themselves. Vector fields fail to commute because of their intrinsic behavior as generators of motion. The requirements that the derivative operator be linear, consistent with directional derivatives, and torsion-free ensure that this failure of commutativity is expressed entirely in terms of the vector fields themselves, without any additional contribution from the connection.

The Lie bracket tells us how vector fields fail to commute as generators of motion. Parallel transport, however, introduces a new kind of comparison: it allows us to compare vectors at nearby points.

This leads to a natural question: if a vector is transported first in one direction and then in another, does the final result depend on the order in which the transports are performed?

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.

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 $T^{a_1 \dots a_k}_{b_1 \dots b_l} \in \mathcal{T}(k, l)$, we define an operator ∂_a by taking the partial derivatives of its components in this coordinate basis:

$$(\partial_c T)^{a_1 \dots a_k}_{b_1 \dots b_l} = \frac{\partial T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l}}{\partial x^\sigma},$$

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.

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 $T^{a_1 \dots a_k}_{b_1 \dots b_l}$ will have different component functions $T^{\mu'_1 \dots \mu'_k}_{\nu'_1 \dots \nu'_l}$. The corresponding partial derivatives $\partial/\partial x^{\sigma'}$ of these components will, in general, not be related to the original ones by the tensor transformation law.

Equivalently, although $\partial_a T^{a_1 \dots a_k}_{b_1 \dots b_l}$ is defined abstractly by taking partial derivatives of components in a chosen coordinate chart, the resulting components in different charts are not related tensorially. Thus the operator ∂_a is not a geometric object intrinsic to the manifold; it depends on the coordinate system used to define it.

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 $\nabla_a T^{b_1 \dots b_k}_{c_1 \dots c_l}$ *does* transform tensorially.

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?

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

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

for any smooth function $f \in \mathcal{F}(M)$. 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.

To investigate the difference between two derivative operators, it is useful to begin with covector fields. Let ω_b be a smooth covector field and consider

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

The reason for starting with covectors is that their contraction with scalar fields allows us to isolate pointwise behavior.

Multiply ω_b by an arbitrary scalar field f and apply both derivative operators. Using the Leibniz rule, we find

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

Since $\tilde{\nabla}_a f = \nabla_a f$ for all scalar fields f , this reduces to

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

This shows that the difference $(\tilde{\nabla}_a - \nabla_a)\omega_b$ depends only on the value of ω_b at a point, not on how it varies in a neighborhood.

More precisely, if two covector fields ω_b and ω'_b agree at a point p , then

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

Thus, the difference operator is *linear and pointwise*.

Any operation that is linear and depends only on the value of its argument at a single point defines a tensor. Accordingly, there exists a smooth tensor field C^a_{bc} of type $(1, 2)$ such that, for every vector field v^a ,

$$(\tilde{\nabla}_b - \nabla_b)v^a = C^a_{bc}v^c. \tag{2.1}$$

This tensor C^a_{bc} completely characterizes how one derivative operator differs from another.

Several important consequences follow immediately. First, there is not a unique derivative operator on a manifold: infinitely many exist, differing by the choice of a tensor C^a_{bc} . Second, the space of all derivative operators forms an *affine space*: given any one operator ∇_a , adding a tensor C^a_{bc} produces a new one,

$$\tilde{\nabla}_a = \nabla_a + C^b_{ac}.$$

Finally, when a coordinate system is chosen, the components of C^a_{bc} appear as the familiar *connection coefficients*, or *Christoffel symbols*. When a metric is introduced, additional conditions will single out a preferred choice among these possibilities.

Difference of Derivative Operators on Dual Vector Fields

The tensor C^a_{bc} was defined by the relation

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

for all vector fields v^a . Once this definition is fixed, the action of the derivative operators on covector fields is no longer independent. It is determined by the requirement that derivative operators commute with contraction.

Let ω_b be a smooth covector field and v^b an arbitrary vector field. Since $\omega_b v^b$ is a scalar field, all derivative operators must agree on its derivative:

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

Using the Leibniz rule for each operator, we expand this expression:

$$(\tilde{\nabla}_a \omega_b - \nabla_a \omega_b)v^b + \omega_b(\tilde{\nabla}_a v^b - \nabla_a v^b) = 0.$$

Substituting the defining relation for C^a_{bc} , we obtain

$$(\tilde{\nabla}_a \omega_b - \nabla_a \omega_b)v^b + \omega_b C^b_{ac}v^c = 0.$$

Because this equation must hold for all vector fields v^c , it follows that

$$\tilde{\nabla}_a \omega_b - \nabla_a \omega_b = -C^c_{ab} \omega_c.$$

Equivalently,

$$\nabla_a \omega_b = \tilde{\nabla}_a \omega_b - C^c_{ab} \omega_c. \quad (2.2)$$

The minus sign is therefore not a convention but a direct consequence of requiring consistency with contraction.

We now use the torsion-free condition to further constrain the tensor C^a_{bc} . Recall that torsion-freeness implies that covariant derivatives commute on scalar fields:

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

for all smooth functions f . Since all derivative operators agree on scalars, the gradient $\nabla_b f = \tilde{\nabla}_b f$ defines a covector field that is independent of the choice of derivative operator.

Setting $\omega_b = \nabla_b f$ in 2.2, we find

$$\nabla_a \nabla_b f = \tilde{\nabla}_a \tilde{\nabla}_b f - C^c_{ab} \nabla_c f.$$

Both second derivatives on the left-hand side are symmetric in the indices a and b . Since $\nabla_c f$ is arbitrary, the correction term must also be symmetric, and therefore

$$C^c_{ab} = C^c_{ba}.$$

Thus, for torsion-free derivative operators, the difference tensor C^a_{bc} is symmetric in its lower indices.

This symmetry has a clear geometric interpretation. In coordinates, it corresponds to the symmetry of the Christoffel symbols $\Gamma^a_{bc} = \Gamma^a_{cb}$. Geometrically, torsion-freeness means that parallel transport around an infinitesimal loop closes independently of the order in which the two directions are traversed. In a curved but torsion-free spacetime, vectors may rotate under transport, but they do not acquire an independent twist.

Difference of Derivative Operators: From Vectors to General Tensor Fields

We have shown that the difference between two derivative operators ∇_a and $\tilde{\nabla}_a$ acting on covector fields is given by

$$(\tilde{\nabla}_a - \nabla_a)\omega_b = C^c_{ab}\omega_c. \quad (2.3)$$

We now determine how the two operators differ when acting on vector fields.

Let t^b be a smooth vector field and ω_b an arbitrary covector field. The contraction $\omega_b t^b$ is a scalar field, and since all derivative operators agree on scalars, we have

$$(\tilde{\nabla}_a - \nabla_a)(\omega_b t^b) = 0. \quad (2.4)$$

Using the Leibniz rule, we expand this expression:

$$(\tilde{\nabla}_a - \nabla_a)(\omega_b t^b) = (\tilde{\nabla}_a - \nabla_a)\omega_b t^b + \omega_b(\tilde{\nabla}_a - \nabla_a)t^b. \quad (2.5)$$

Substituting Eq. 2.3 into the first term gives

$$C^c_{ab}\omega_c t^b + \omega_b(\tilde{\nabla}_a - \nabla_a)t^b = 0. \quad (2.6)$$

Relabeling dummy indices and factoring out ω_b , this may be written as

$$\omega_b \left(C^b_{ac} t^c + (\tilde{\nabla}_a - \nabla_a)t^b \right) = 0. \quad (2.7)$$

Because this equation must hold for *all* covector fields ω_b , the expression in parentheses must vanish. Hence,

$$(\tilde{\nabla}_a - \nabla_a)t^b = -C^b_{ac} t^c. \quad (2.8)$$

Rearranging, we obtain the action of the two derivative operators on vector fields:

$$\nabla_a t^b = \tilde{\nabla}_a t^b + C^b_{ac} t^c. \quad (2.9)$$

Thus, once the difference tensor C^a_{bc} is fixed by the action of the operators on covector fields, its action on vector fields is uniquely determined by compatibility with contraction.

Applying the same reasoning repeatedly, one finds that for any tensor field $T^{b_1 \dots b_k}_{c_1 \dots c_l}$, the difference between the two derivative operators takes the general form

$$\nabla_a T^{b_1 \dots b_k}_{c_1 \dots c_l} = \tilde{\nabla}_a T^{b_1 \dots b_k}_{c_1 \dots c_l} + \sum_{i=1}^k C^{b_i}_{ad} T^{b_1 \dots d \dots b_k}_{c_1 \dots c_l} - \sum_{j=1}^l C^d_{ac_j} T^{b_1 \dots b_k}_{c_1 \dots d \dots c_l}. \quad (2.10)$$

Each contravariant index contributes a positive correction term, while each covariant index contributes a negative one, reflecting the opposite transformation behavior of vectors and covectors.

Equation 2.10 shows that the tensor C^a_{bc} completely characterizes the difference between any two derivative operators. Given $\tilde{\nabla}_a$ and C^a_{bc} , one can reconstruct ∇_a ; conversely, any smooth symmetric tensor field C^a_{bc} defines a valid derivative operator when added in this way. In coordinates, the components of C^a_{bc} appear as the familiar connection coefficients, or Christoffel symbols.

Freedom in Choosing a Derivative Operator and Christoffel Symbols

Equation 2.10 shows that the difference between any two derivative operators ∇_a and $\tilde{\nabla}_a$ is completely determined by a smooth tensor field C^a_{bc} of type $(1, 2)$, symmetric in its lower indices. Conversely, given any such tensor field C^a_{bc} , equation 2.10 defines a new derivative operator ∇_a in terms of $\tilde{\nabla}_a$.

Thus, the smooth manifold structure alone does not single out a unique derivative operator. There is substantial freedom in how tensors at different points are compared, unless additional geometric structure is introduced.

On an n -dimensional manifold, a tensor C^a_{bc} symmetric in its lower indices has

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

independent components at each point. Each independent choice of these components defines a distinct derivative operator. Hence, a manifold admits infinitely many derivative operators, none of which is preferred a priori.

A particularly important special case arises when we take $\tilde{\nabla}_a$ to be the *ordinary derivative operator* ∂_a associated with a chosen coordinate system. In this case, the difference tensor C^a_{bc} is conventionally denoted by

$$\Gamma^a_{bc} \equiv C^a_{bc},$$

and is called a *Christoffel symbol*.

Substituting $\tilde{\nabla}_a = \partial_a$ into equation 2.9, the covariant derivative of a vector field t^a takes the form

$$\nabla_a t^b = \partial_a t^b + \Gamma^b_{ac} t^c.$$

Here the partial derivative $\partial_a t^b$ captures the coordinate variation of the components of t^b , while the Christoffel symbols Γ^b_{ac} encode how the coordinate basis vectors themselves vary from point to point.

Because the coordinate basis $\{\partial/\partial x^\mu\}$ changes under coordinate transformations, the associated ordinary derivative operator ∂_a also changes. As a result, the Christoffel symbols Γ^a_{bc} do not transform according to the tensor transformation law. Instead, they acquire additional inhomogeneous terms that precisely compensate for the non-tensorial behavior of partial derivatives.

This shows that the Christoffel symbols are *not* the components of a tensor field. Rather, they encode the coordinate-dependent correction needed to convert ordinary partial derivatives into a geometric, coordinate-independent covariant derivative.

Metric Compatibility and the Levi-Civita Condition

So far, we have seen that a smooth manifold admits infinitely many possible derivative operators, differing by the choice of a tensor C^a_{bc} or, in coordinates, by a choice of Christoffel symbols Γ^a_{bc} .

The manifold structure alone does not single out a preferred way to compare vectors at different points.

If, however, the manifold is equipped with a metric g_{ab} , the metric provides a natural criterion for selecting a distinguished derivative operator.

Suppose we parallel transport two vector fields v^a and w^a along a curve with tangent vector t^a . It is natural to require that their inner product, as measured by the metric, remain unchanged along the curve. That is,

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

This condition expresses the idea that parallel transport should preserve lengths and angles.

Using the Leibniz rule and the defining property of parallel transport,

$$t^a \nabla_a v^b = 0, \quad t^a \nabla_a w^b = 0,$$

the above condition reduces to

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

Since this must hold for all curves and for all vectors v^a and w^a that are parallel transported along them, we are led to impose the condition

$$\nabla_a g_{bc} = 0. \tag{2.11}$$

Equation 2.11 is called the *metric compatibility condition*. It ensures that the covariant derivative preserves the metric under parallel transport. Equivalently,

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

Geometrically, metric compatibility means that the connection respects the geometry defined by g_{ab} : the inner product of any two vectors that are parallel transported along a curve remains invariant.

Together with the torsion-free condition,

$$C^a_{bc} = C^a_{cb},$$

metric compatibility uniquely determines the derivative operator. This distinguished connection is known as the *Levi-Civita connection*.

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

We have seen that a smooth manifold admits infinitely many derivative operators, differing by a tensor C^a_{bc} and, in coordinates, by a choice of Christoffel symbols. No particular derivative operator is preferred by the manifold alone.

If, however, the manifold is equipped with a metric g_{ab} , the metric provides additional geometric structure. In this section we show that the requirements of *metric compatibility* and *torsion-freeness*

together select a unique derivative operator. This distinguished connection is called the *Levi-Civita connection*.

Theorem 2.1 (Levi-Civita). Let (M, g) be a smooth manifold with metric g_{ab} . There exists a unique derivative operator ∇_a such that:

1. $\nabla_a g_{bc} = 0$ (metric compatibility),
2. $\nabla_a \nabla_b f = \nabla_b \nabla_a f$ for all smooth f (torsion-free).

Geometric role of the assumptions. Metric compatibility ensures that parallel transport preserves the notions of length and angle defined by g_{ab} . Torsion-freeness removes any intrinsic “twist” in how tangent spaces are related, requiring that infinitesimal parallelograms close. Together, these conditions eliminate all ambiguity in the choice of connection.

Uniqueness of the Levi-Civita Connection

Suppose ∇_a and $\tilde{\nabla}_a$ are two derivative operators on (M, g) that are both torsion-free and metric-compatible. We show that they must coincide.

From earlier results, the difference between two derivative operators is described by a tensor C^a_{bc} of type $(1, 2)$, defined by

$$(\tilde{\nabla}_a - \nabla_a)v^c = C^c_{ab}v^b \quad (2.12)$$

for every vector field v^b .

Because both connections are torsion-free, the tensor C^c_{ab} is symmetric in its lower indices:

$$C^c_{ab} = C^c_{ba}. \quad (2.13)$$

Now consider metric compatibility. Since both ∇_a and $\tilde{\nabla}_a$ annihilate the metric,

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

To evaluate this difference, we apply the general transformation rule for covariant derivatives of a $(0, 2)$ tensor, Eq. 2.10, where $T_{bc} = g_{bc}$,

- $k = 0$ (no contravariant indices) $\Rightarrow + \sum_{i=1}^k$ term vanishes.
- $l = 2$ (two covariant indices), so the $-\sum_{j=1}^2$ term contributes two corrections.

$$\nabla_a g_{bc} = \tilde{\nabla}_a g_{bc} - \sum_{j=1}^2 C^d_{ac_j} g_{c_1 \dots d \dots c_2}$$

Here $(c_1, c_2) = (b, c)$ so we write the two terms explicitly. Expand the $j = 1$ and $j = 2$ terms,

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

and the difference becomes

$$0 = -C_{ab}^d g_{dc} - C_{ac}^d g_{bd}. \quad (2.14)$$

Contracting with arbitrary vectors and using the symmetry 2.13, the only solution of 2.14 is

$$C_{ab}^c = 0.$$

Therefore,

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

and hence

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

This proves the uniqueness of the torsion-free, metric-compatible connection.

Existence of the Levi–Civita Connection

We now show that such a connection exists.

Let $\tilde{\nabla}_a$ be any torsion-free derivative operator. (For example, one may take the ordinary partial derivative operator in a coordinate chart.) We construct a new derivative operator ∇_a by adding a correction term:

$$\nabla_a v^c := \tilde{\nabla}_a v^c + C_{ab}^c v^b, \quad (2.15)$$

where C_{ab}^c is a tensor symmetric in a, b , to be determined.

Because C_{ab}^c is a tensor, the operator ∇_a defined by 2.15 satisfies all defining properties of a derivative operator. By construction, it is torsion-free. We now choose C_{ab}^c so that ∇_a is metric-compatible.

Applying ∇_a to the metric yields

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

Metric compatibility requires $\nabla_a g_{bc} = 0$, hence

$$0 = \tilde{\nabla}_a g_{bc} - C_{ab}^d g_{dc} - C_{ac}^d g_{bd}. \quad (2.16)$$

To solve for C_{ab}^c , write 2.16 together with its permutations under $a \leftrightarrow b$ and $a \leftrightarrow c$, then add the first two and subtract the third. Using the symmetry of C_{ab}^c and of the metric, the result reduces to

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

Raising the index with g^{cd} gives

$$C^c_{ab} = \frac{1}{2} g^{cd} (\tilde{\nabla}_a g_{bd} + \tilde{\nabla}_b g_{ad} - \tilde{\nabla}_d g_{ab}). \quad (2.17)$$

This uniquely determines C^c_{ab} and hence the derivative operator ∇_a . By construction, ∇_a is torsion-free and metric-compatible. Therefore, the Levi-Civita connection exists.

Coordinate expression. If $\tilde{\nabla}_a$ is chosen to be the ordinary derivative operator ∂_a associated with a coordinate system, then 2.17 reduces to the familiar formula for the Christoffel symbols:

$$\Gamma^c_{ab} = \frac{1}{2} g^{cd} (\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}). \quad (2.18)$$

This completes the construction of the Levi-Civita connection.

2.2 The Riemann Curvature Tensor

In the previous section we developed the notion of a derivative operator ∇_a (also called a *connection*) and showed how it determines 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 version 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 covector field. For a scalar field f , we compute

$$\begin{aligned} \nabla_a \nabla_b (f \omega_c) &= \nabla_a (\omega_c \nabla_b f + f \nabla_b \omega_c) \\ &= (\nabla_a \nabla_b f) \omega_c + (\nabla_b f) \nabla_a \omega_c + (\nabla_a f) \nabla_b \omega_c + f \nabla_a \nabla_b \omega_c. \end{aligned}$$

Subtracting the same expression with a and b exchanged, the first three terms cancel pairwise, yielding

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)(f \omega_c) = f (\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c. \quad (2.19)$$

As in our earlier analysis of derivative operators in Section 2.1, 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 how ω_c varies in a neighborhood of p . In particular, for any scalar field f one has

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)(f \omega_c) = f (\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c,$$

which is precisely the criterion used earlier to identify tensorial operators.

Consequently, at each point $p \in M$ the assignment

$$\omega_c(p) \mapsto [(\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c](p)$$

defines a linear map from T_p^*M to itself.

The indices a and b label the two covariant derivatives and hence correspond to two covariant slots, while the index c labels the input covector. Since the output of the commutator is itself a covector, the associated tensor must also carry one contravariant index to represent this output. Therefore the object defined by this commutator has one contravariant and three covariant indices.

Definition 2.2 (Riemann curvature tensor). There exists a unique tensor field $R_{abc}{}^d$ such that for all covector fields ω_c ,

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c = R_{abc}{}^d \omega_d. \quad (2.20)$$

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

2.2.1 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 small closed loop. Let $p \in M$ and choose a two-dimensional surface S through p . Introduce local coordinates (t, s) on S with p corresponding to $(0, 0)$.

Consider the small rectangular loop shown in Fig. 2.1:

$$(0, 0) \rightarrow (\Delta t, 0) \rightarrow (\Delta t, \Delta s) \rightarrow (0, \Delta s) \rightarrow (0, 0).$$

Let v^a be a vector at p (not assumed tangent to S), and 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 covector field ω_a .

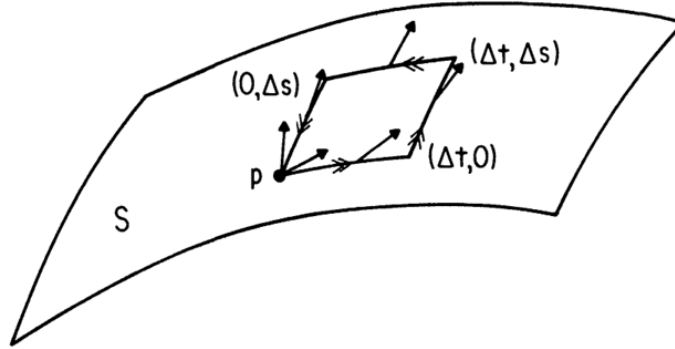


Figure 2.1: Parallel transport of a vector v^a around a small rectangular loop in the (t, s) surface through p . The leading nontrivial change in v^a is governed by the Riemann tensor.

First-order variation along one edge. Along the first leg we move from $(0, 0)$ to $(\Delta t, 0)$, i.e. along $\gamma(t) = (t, 0)$, $0 \leq t \leq \Delta t$. To first order in Δt ,

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

The tangent to γ is $T^b = (\partial/\partial t)^b$. Since $v^a \omega_a$ is a scalar, its derivative along γ is the directional derivative:

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

(using that all derivative operators agree with directional derivatives on scalars). Applying the Leibniz rule,

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

Here T^a denotes the tangent vector to the curve, so $T^b \nabla_b$ is the directional covariant derivative along that curve. Recall that a vector field v^a is said to be parallel transported along a curve with tangent T^a if its covariant derivative along the curve vanishes:

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

as introduced in Section 2.1. Hence

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

and therefore

$$\delta_1 = \Delta t v^a T^b \nabla_b \omega_a \big|_{(\Delta t/2, 0)}. \quad (2.21)$$

The first-order calculation serves an important conceptual role. Along a single edge of the loop, the change in the scalar $v^a \omega_a$ arises entirely from the variation of the covector field ω_a ; the transported vector v^a itself contributes no change. This is precisely what parallel transport is designed to accomplish: a vector is carried along a curve in such a way that it remains as constant as possible, so that any first-order variation comes solely from the object against which it is being tested. At this order, the result depends only on the directional derivative of ω_a along the curve and contains no intrinsic geometric information about the manifold. In particular, this behavior is identical in flat and curved spaces. The first-order analysis therefore confirms that parallel transport along a single direction is locally trivial and path-independent to leading order. Curvature cannot appear at this level; it can only arise when transport is compared along two independent directions, which necessarily introduces second-order (area) effects.

Second-order contribution from the full loop. Comparing opposite sides of the rectangle and expanding to first order in Δs produces a net second-order change. One finds

$$\delta_1 + \delta_3 = -\Delta t \Delta s v^a S^c \nabla_c (T^b \nabla_b \omega_a), \quad (2.22)$$

where $S^c = (\partial/\partial s)^c$. Repeating the calculation for $\delta_2 + \delta_4$ and adding both contributions yields

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

Because coordinate vector fields commute, $[T, S] = 0$, we may rewrite this as

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

Using the definition 2.20,

$$\delta(v^a \omega_a) = \Delta t \Delta s v^a T^c S^b R_{cba}{}^d \omega_d.$$

Since this holds for all ω_d , the change in v^a itself is

$$\delta v^a = \Delta t \Delta s v^d T^b S^c R_{cbd}{}^a. \quad (2.23)$$

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

The appearance of curvature at second order has a clear geometric origin. At first order in the displacement, parallel transport is path-independent: transporting a vector a short distance along one direction produces no intrinsic information about curvature. This is why all first-order contributions cancel once the vector itself is parallel transported.

Curvature emerges only when we compare *two distinct directions of transport*. The second-order term measures the failure of successive parallel transports along the t and s directions to commute. Equivalently, it captures the discrepancy between transporting first along T^a and then along S^a , versus reversing the order. This noncommutativity is precisely what the Riemann tensor encodes.

Thus the Riemann tensor isolates the genuinely geometric content of parallel transport: it vanishes if and only if transport around every infinitesimal loop returns vectors to their original values. Flatness corresponds to the absence of any second-order loop effect, while nonzero curvature is detected exactly by this failure at order $\Delta t \Delta s$.

Curvature Acting on Vectors and General Tensors

We now derive the action of the curvature on vector fields. Let t^c be a vector field and let ω_c be an arbitrary covector field. Since $t^c \omega_c$ is a scalar, the commutator vanishes on it:

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)(t^c \omega_c) = 0.$$

Expanding with the Leibniz rule and using 2.20,

$$0 = \omega_c(\nabla_a \nabla_b - \nabla_b \nabla_a)t^c + t^c(\nabla_a \nabla_b - \nabla_b \nabla_a)\omega_c = \omega_c(\nabla_a \nabla_b - \nabla_b \nabla_a)t^c + t^c \omega_d R_{abc}{}^d.$$

Since ω_c is arbitrary, we conclude

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)t^c = -R_{abd}{}^c t^d. \quad (2.24)$$

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

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)T^{c_1 \dots c_k}{}_{d_1 \dots d_l} = -\sum_{i=1}^k R_{abe}{}^{c_i} T^{c_1 \dots e \dots c_k}{}_{d_1 \dots d_l} + \sum_{j=1}^l R_{abd_j}{}^e T^{c_1 \dots c_k}{}_{d_1 \dots e \dots d_l}. \quad (2.25)$$

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

Algebraic Properties and Bianchi Identities

The curvature tensor satisfies several fundamental identities.

Antisymmetry in the commutator indices. From the definition 2.20 we immediately obtain

$$R_{abc}{}^d = -R_{bac}{}^d. \quad (2.26)$$

First (algebraic) Bianchi identity. For a torsion-free connection, curvature also obeys the cyclic identity

$$R_{[abc]}{}^d = 0. \quad (2.27)$$

(Geometrically, this encodes the fact that the leading-order net effect of curvature on an infinitesimal parallelepiped is independent of how it is decomposed into loops; a standard derivation uses torsion-freeness together with the Jacobi identity for iterated commutators of covariant derivatives.)

Additional symmetries for the Levi-Civita connection. When ∇_a is metric-compatible, we define $R_{abcd} := g_{de} R_{abc}{}^e$. Applying 2.25 to the metric and using $\nabla_a g_{bc} = 0$ gives

$$0 = (\nabla_a \nabla_b - \nabla_b \nabla_a) g_{cd} = R_{abc}{}^e g_{ed} + R_{abd}{}^e g_{ce} = R_{abcd} + R_{abdc},$$

so

$$R_{abcd} = -R_{abdc}. \quad (2.28)$$

Together with 2.26 and 2.27 one can show the pair-exchange symmetry

$$R_{abcd} = R_{cdab}. \quad (2.29)$$

Differential Bianchi identity. The curvature also satisfies the (second) Bianchi identity

$$\nabla_{[a} R_{bc]d}{}^e = 0. \quad (2.30)$$

(A standard derivation applies the Jacobi identity for operator commutators to $\nabla_a, \nabla_b, \nabla_c$ acting on a covector field and then substitutes 2.20.)

Ricci Tensor, Scalar Curvature, and Weyl Tensor

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

$$R_{ac} \equiv R_{abc}{}^b. \quad (2.31)$$

For the Levi-Civita connection, R_{ab} is symmetric.

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

$$R \equiv R_a{}^a = g^{ab} R_{ab}. \quad (2.32)$$

It is often useful to decompose R_{abcd} into a “trace part” and a “trace-free part.” The trace-free part is the *Weyl tensor* C_{abcd} . For $n \geq 3$,

$$R_{abcd} = C_{abcd} + \frac{2}{n-2} (g_{a[c} R_{d]b} - g_{b[c} R_{d]a}) - \frac{2}{(n-1)(n-2)} R g_{a[c} g_{d]b}. \quad (2.33)$$

Geometrically, C_{abcd} encodes the “shape-changing” (conformal) part of curvature, while the Ricci tensor encodes the part related to volume change.

Contracted Bianchi Identity and the Einstein Tensor

Contracting the differential Bianchi identity 2.30 yields an identity for the Ricci tensor, and a further contraction gives the twice-contracted Bianchi identity:

$$\nabla^a G_{ab} = 0, \quad (2.34)$$

where the *Einstein tensor* is defined by

$$G_{ab} \equiv R_{ab} - \frac{1}{2}R g_{ab}. \quad (2.35)$$

The identity 2.34 will play a central role in general relativity: in Einstein's field equations $G_{ab} = 8\pi T_{ab}$, it guarantees consistency with local conservation of stress-energy, $\nabla^a T_{ab} = 0$.

2.3 Geodesics

2.3.1 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 :

$$T^a \nabla_a T^b = 0. \quad (2.36)$$

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

$$T^a \nabla_a T^b = \alpha T^b, \quad (2.37)$$

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 2.37 vanishes, yielding the affinely parametrized form 2.36. A parameter for which 2.36 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.”

2.3.2 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

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

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

Along the curve $\gamma(t)$ we have coordinate functions $x^\mu(t)$ and

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

In a coordinate chart, the operator $T^a \partial_a$ acts on any scalar field $F(x)$ by the chain rule:

$$(T^a \partial_a)F = \frac{dx^\nu}{dt} \frac{\partial F}{\partial x^\nu} = \frac{d}{dt}(F(x(t))).$$

Apply this with the scalar field $F(x) := T^\mu(x)$ (the μ th component of T viewed as a function of the coordinates). Then

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

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

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

$$\frac{d^2 x^\mu}{dt^2} + \Gamma^\mu_{\nu\lambda} \frac{dx^\nu}{dt} \frac{dx^\lambda}{dt} = 0. \quad (2.38)$$

Equation 2.38 is the textbook'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 2.38. 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 2.38 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.

2.3.3 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 $T_p M$ be its tangent space. Given any vector $T^a \in T_p M$, consider the geodesic equation introduced in Eq. 2.36,

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

with initial conditions

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

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

$$\gamma_T : I \rightarrow M,$$

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 the textbook and define a map from the tangent space into the manifold by evaluating each geodesic at unit affine parameter:

$$\exp_p(T^a) := \gamma_T(1). \quad (2.39)$$

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

$$(d \exp_p)_{0^a} = \text{Id}_{T_p M}.$$

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 $0^a \in T_p M$ 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 $T_p M$,
- 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:

$$\exp_p : T_p M \rightarrow M$$

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

2.3.4 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:

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \quad \Gamma^\rho_{\mu\nu}(p) = 0.$$

These coordinates arise directly from the exponential map.

Step 1: Use the exponential map to label nearby points. For each vector $T^a \in T_p M$, 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 $e_{(\mu)}^a$ of $T_p M$. Every tangent vector can be uniquely expanded as

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

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 $e_{(\mu)}^a$.

Geometric meaning of coordinate axes. In RNC: - The coordinate lines through p are geodesics. - The coordinate basis vectors at p equal the original basis $e_{(\mu)}^a$. - 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 $t = x^\mu$:

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

Along such a curve,

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

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 $\{e_{(\mu)}^a\}$ orthonormally:

$$g_{\mu\nu}(p) = g_{ab} e_{(\mu)}^a e_{(\nu)}^b = \eta_{\mu\nu}.$$

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. the textbook briefly states, and one can show, that:

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

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:

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \quad \partial_\lambda g_{\mu\nu}(p) = 0, \quad \Gamma^\rho_{\mu\nu}(p) = 0.$$

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.*

2.3.5 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

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

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

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

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

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

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

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

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 the textbook 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.

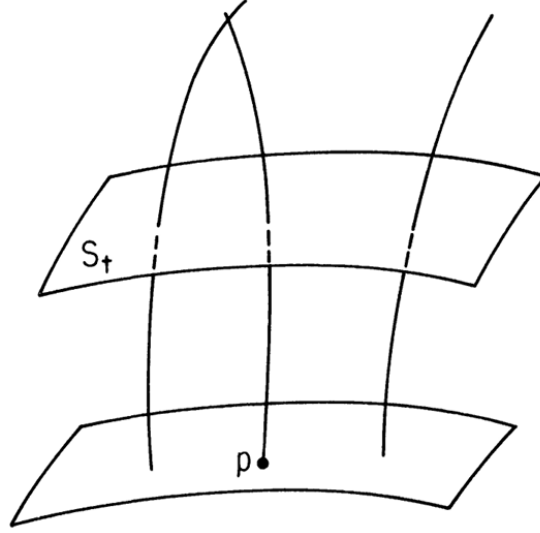


Figure 2.2: Construction of Gaussian normal coordinates from a hypersurface S : geodesics are emitted orthogonally from S with initial tangent n^a , and the hypersurface coordinates x^i are held fixed along each geodesic.

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.

2.3.6 Extremizing the Length Functional

Geodesics may also be characterized as curves that *extremize the length functional*. the textbook states the resulting Euler–Lagrange equations, but omits the intermediate steps; we supply them here.

The length functional. For a smooth curve $x^\mu(t)$, define

$$L[x^\mu] = \int \sqrt{g_{\mu\nu}(x) \dot{x}^\mu \dot{x}^\nu} dt. \quad (2.40)$$

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

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} \right) - \frac{\partial \mathcal{L}}{\partial x^\mu} = 0. \quad (2.41)$$

Computing the derivatives (expanded calculation). We find

$$\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = \frac{g_{\mu\nu} \dot{x}^\nu}{\mathcal{L}}, \quad \frac{\partial \mathcal{L}}{\partial x^\mu} = \frac{1}{2\mathcal{L}} \partial_\mu g_{\rho\sigma} \dot{x}^\rho \dot{x}^\sigma.$$

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 2.41 and rearranging yields the expanded, *non-affinely parameterized* Euler–Lagrange equation:

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\lambda} \dot{x}^\nu \dot{x}^\lambda = \frac{\dot{\mathcal{L}}}{\mathcal{L}} \dot{x}^\mu. \quad (2.42)$$

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 the textbook’s equation (3.3.13):

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\lambda} \dot{x}^\nu \dot{x}^\lambda = 0, \quad (2.43)$$

which is the textbook’s geodesic equation in coordinate form.

Summary. Equation 2.42 is the fully general Euler–Lagrange form. Equation 2.43 is the special case obtained when the parameter is affine.

Relation to the textbook’s derivation. the textbook derives the geodesic equation by varying the length functional and working entirely in coordinates, keeping all metric derivatives explicit until the final step. In contrast, we have already introduced the connection coefficients $\Gamma^\mu_{\nu\lambda}$ and rewrite the Euler–Lagrange equations directly in terms of them. Expanding $\Gamma^\mu_{\nu\lambda}$ in terms of derivatives of the metric shows that our equation is algebraically identical to the textbook’s coordinate expression. The difference is purely organizational: we recognize the Christoffel symbols earlier, avoiding unnecessary index expansions and making the geometric content more transparent.

2.3.7 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

$$T^a := \left(\frac{\partial}{\partial t} \right)^a, \quad X^a := \left(\frac{\partial}{\partial s} \right)^a.$$

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*:

$$T^b \nabla_b X^a = X^b \nabla_b T^a. \quad (2.44)$$

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,

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

Relative velocity. The rate at which the separation vector X^a changes along the geodesics is the *relative velocity*:

$$v^a := T^b \nabla_b X^a.$$

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 2.44 and the geodesic equation $T^b \nabla_b T^a = 0$, the textbook obtains

$$\begin{aligned} a^a &= T^c \nabla_c (T^b \nabla_b X^a) \\ &= T^c \nabla_c (X^b \nabla_b T^a) \\ &= (T^c \nabla_c X^b) (\nabla_b T^a) + X^b T^c \nabla_c \nabla_b T^a. \end{aligned}$$

Now use 2.44 again to rewrite $T^c \nabla_c X^b = X^c \nabla_c T^b$:

$$(T^c \nabla_c X^b)(\nabla_b T^a) = (X^c \nabla_c T^b)(\nabla_b T^a) = X^b (\nabla_b T^c)(\nabla_c T^a),$$

(where we relabeled dummy indices in the last step). Hence

$$a^a = X^b (\nabla_b T^c)(\nabla_c T^a) + X^b T^c \nabla_c \nabla_b T^a.$$

Next commute the derivatives acting on T^a :

$$(\nabla_c \nabla_b - \nabla_b \nabla_c)T^a = R^a_{\text{ } dcb} T^d,$$

so

$$T^c \nabla_c \nabla_b T^a = T^c \nabla_b \nabla_c T^a + T^c R^a_{\text{ } dcb} T^d.$$

Apply the Leibniz rule to the first term:

$$T^c \nabla_b \nabla_c T^a = \nabla_b (T^c \nabla_c T^a) - (\nabla_b T^c)(\nabla_c T^a).$$

Since $T^c \nabla_c T^a = 0$ along each geodesic, the first term vanishes, and we get

$$T^c \nabla_c \nabla_b T^a = -(\nabla_b T^c)(\nabla_c T^a) + T^c R^a_{\text{ } dcb} T^d.$$

Substituting into a^a produces the crucial cancellation:

$$\begin{aligned} a^a &= X^b (\nabla_b T^c)(\nabla_c T^a) + X^b \left[-(\nabla_b T^c)(\nabla_c T^a) + T^c R^a_{\text{ } dcb} T^d \right] \\ &= X^b T^c T^d R^a_{\text{ } dcb}. \end{aligned}$$

Relabeling dummy indices to match conventions yields the geodesic deviation equation

$$a^a = T^b \nabla_b (T^c \nabla_c X^a) = -R^a_{\text{ } bcd} T^b T^d X^c.$$

Key Step. The Riemann tensor appears through the identity

$$(\nabla_c \nabla_b - \nabla_b \nabla_c)T^a = R^a_{\text{ } dcb} T^d,$$

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):

$$T^b \nabla_b (T^c \nabla_c X^a) = -R^a_{\text{ } bcd} T^b T^d X^c. \quad (2.45)$$

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.

Conceptual role of geodesic deviation. A single freely falling observer cannot detect curvature through their own motion: along any one geodesic, the tangent vector parallel transports itself, and local physics appears inertial. Curvature becomes observable only when we compare *nearby* geodesics. The geodesic deviation equation shows that curvature manifests itself as a relative acceleration between neighboring free-fall trajectories. Thus, curvature is not a force acting on individual particles, but a property of spacetime that governs how families of geodesics converge or diverge. This is the precise mathematical expression of tidal gravity.

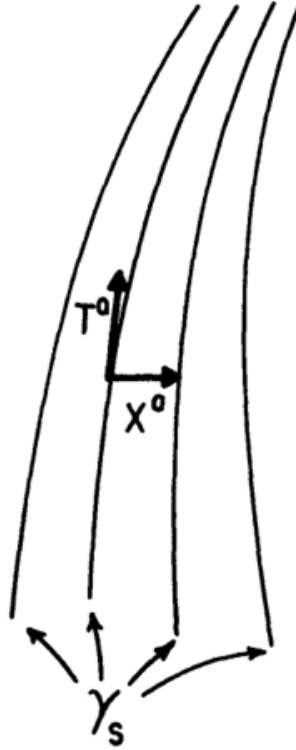


Figure 2.3: A one-parameter family of geodesics $\gamma_s(t)$ with tangent T^a and deviation vector X^a .

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.

2.4 Methods for Computing Curvature

In Section 2.2 we defined the Riemann curvature tensor as the unique tensor field $R_{abc}{}^d$ satisfying the commutator identity

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c = R_{abc}{}^d \omega_d, \quad (2.46)$$

for every smooth covector field ω_c . This characterization is conceptually central: curvature is the obstruction to commuting covariant derivatives.

In practice, however, we are usually given a metric g_{ab} and need an explicit recipe for computing $R_{abc}{}^d$ and its contractions. This matters in general relativity because Einstein's equation,

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

involves the Einstein tensor

$$G_{ab} = R_{ab} - \frac{1}{2}R g_{ab},$$

built from the Ricci tensor R_{ab} and scalar curvature R .

Accordingly, this section is computational: starting from g_{ab} , we build the associated connection and then compute curvature and its contractions,

$$g_{ab} \longrightarrow \Gamma^a{}_{bc} \longrightarrow R_{abc}{}^d \longrightarrow R_{ab}, R, G_{ab}.$$

the textbook emphasizes two standard pipelines:

1. **Coordinate component method.** Work in a coordinate basis $\{\partial/\partial x^\mu\}$, compute the Christoffel symbols $\Gamma^\rho{}_{\mu\nu}$ from the metric, and then evaluate the coordinate formula for $R^\rho{}_{\sigma\mu\nu}$.
2. **Orthonormal frame (tetrad) method.** Choose an orthonormal frame $\{e_{(\alpha)}^a\}$, compute the connection coefficients (Ricci rotation coefficients / connection 1-forms), and then compute curvature using the corresponding structure equations. This often reduces algebra when the spacetime has symmetries, at the cost of tracking noncommuting frame vectors.

Two computational pipelines.

- In a *coordinate basis*, the basis vectors commute, but the connection coefficients $\Gamma^\rho{}_{\mu\nu}$ can be algebraically complicated.
- In an *orthonormal frame*, the connection coefficients are often simpler, but the frame vectors need not commute, and their commutators enter the curvature calculation.

Both approaches compute the same geometric tensor $R_{abc}{}^d$; they simply organize the algebra differently.

We now develop each method, filling in intermediate steps when useful and emphasizing a workflow you can apply directly to metrics of interest.

2.4.1 Coordinate Component Method

The most direct way to compute curvature is to work in a coordinate basis and express the covariant derivative in terms of partial derivatives and Christoffel symbols $\Gamma^c{}_{ab}$. This method is conceptually straightforward but often algebraically heavy.

Let ω_a be a covector field. Using the coordinate expression for the covariant derivative,

$$\nabla_b \omega_c = \partial_b \omega_c - \Gamma^d{}_{bc} \omega_d, \quad (2.47)$$

we compute the second covariant derivative (using the general tensor rule:

$$\nabla_a \nabla_b \omega_c = \partial_a (\nabla_b \omega_c) - \Gamma^e{}_{ab} \nabla_e \omega_c - \Gamma^e{}_{ac} \nabla_b \omega_e. \quad (2.48)$$

(See Appendix A.1.1 for a detailed derivation of Eq. 2.48.)

Substituting 2.47 into 2.48 and expanding yields a lengthy but mechanical expression. Inserting the result into the defining commutator relation

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

and using the commutativity of coordinate derivatives $\partial_a \partial_b = \partial_b \partial_a$, one finds

$$R_{abc}{}^d = -2 \partial_{[a} \Gamma^d{}_{b]c} + 2 \Gamma^e{}_{c[a} \Gamma^d{}_{b]e}. \quad (2.49)$$

In coordinate components this becomes the familiar formula

$$R_{\mu\nu\rho}{}^\sigma = \partial_\nu \Gamma^\sigma{}_{\mu\rho} - \partial_\mu \Gamma^\sigma{}_{\nu\rho} + \Gamma^\sigma{}_{\mu\lambda} \Gamma^\lambda{}_{\nu\rho} - \Gamma^\sigma{}_{\nu\lambda} \Gamma^\lambda{}_{\mu\rho}. \quad (2.50)$$

2.4.2 Orthonormal Basis (Tetrad) Methods

The coordinate method computes curvature directly from the metric components $g_{\mu\nu}$ and the associated Christoffel symbols $\Gamma^\rho{}_{\mu\nu}$ in some chart. Although completely systematic, this approach can become algebraically unwieldy.

An alternative is to work in an *orthonormal frame* (or tetrad). Here we keep the same Levi-Civita connection ∇_a , but change the basis in which tensors are expanded.

Instead of the coordinate basis $\{\partial_\mu\}$, we introduce a *moving orthonormal frame* of vector fields

$$\{(e_\mu)^a\}, \quad \mu = 0, 1, \dots, n-1,$$

chosen so that at each point

$$g_{ab}(e_\mu)^a(e_\nu)^b = \eta_{\mu\nu},$$

where $\eta_{\mu\nu} = \text{diag}(-1, 1, \dots, 1)$.

Geometric idea. In a coordinate basis, the metric carries all the geometric information. In an orthonormal frame, the metric is locally flattened to $\eta_{\mu\nu}$, and the geometry instead appears in how the frame itself changes from point to point.

That change is encoded in the *connection 1-forms* (or Ricci rotation coefficients), which measure the “rotation” of the frame relative to parallel transport.

Orthonormal Frames and Tetrads

A *tetrad* (in n dimensions, more generally an orthonormal frame) is a collection of smooth vector fields

$$\{(e_\mu)^a\}, \quad \mu = 0, 1, \dots, n-1,$$

such that at every point

$$g_{ab}(e_\mu)^a(e_\nu)^b = \eta_{\mu\nu}, \quad (2.51)$$

where $\eta_{\mu\nu} = \text{diag}(-1, 1, \dots, 1)$.

Here:

- Latin indices $a, b, \dots$ refer to abstract tensor indices on the manifold,
- Greek indices $\mu, \nu, \dots$ label components in the orthonormal frame.

Equation 2.51 states that the tetrad vectors form an orthonormal basis at each point with respect to g_{ab} .

Lowering an index with the metric gives the dual covectors

$$(e_\mu)_a := g_{ab}(e_\mu)^b.$$

The inverse matrix $\eta^{\mu\nu}$ satisfies $\eta^{\mu\alpha}\eta_{\alpha\nu} = \delta^\mu_\nu$, and one obtains the completeness relation

$$\eta^{\mu\nu}(e_\mu)^a(e_\nu)_b = \delta^a_b. \quad (2.52)$$

Thus $\{(e_\mu)^a\}$ forms a basis of $T_p M$ and $\{(e^\mu)_a\}$ (defined by $(e^\mu)_a(e_\nu)^a = \delta^\mu_\nu$) forms the dual coframe.

Any tensor may be expanded in this orthonormal frame. For example, a vector V^a has frame components

$$V^\mu := (e^\mu)_a V^a, \quad V^a = V^\mu (e_\mu)^a.$$

Connection 1-Forms and Ricci Rotation Coefficients

In a coordinate basis the connection is encoded by the Christoffel symbols Γ^c_{ab} . In an orthonormal basis, it is more natural to encode the connection through the *connection 1-forms*

$$\omega_{a\mu\nu} := (e_\mu)^b \nabla_a (e_\nu)_b. \quad (2.53)$$

The components of these 1-forms in the tetrad basis,

$$\omega_{\lambda\mu\nu} := (e_\lambda)^a \omega_{a\mu\nu} = (e_\lambda)^a (e_\mu)^b \nabla_a (e_\nu)_b, \quad (2.54)$$

are called the *Ricci rotation coefficients*.

Metric compatibility. Using orthonormality 2.51 and $\nabla_a g_{bc} = 0$ we get

$$0 = \nabla_a [(e_\mu)^b (e_\nu)_b] = \omega_{a\mu\nu} + \omega_{a\nu\mu}.$$

Thus the connection 1-forms are antisymmetric in $\mu\nu$:

$$\omega_{a\mu\nu} = -\omega_{a\nu\mu}, \quad (2.55)$$

and similarly for the Ricci rotation coefficients,

$$\omega_{\lambda\mu\nu} = -\omega_{\lambda\nu\mu}. \quad (2.56)$$

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

Given a tetrad $\{(e_\mu)^a\}$, we define the components of the Riemann tensor in the orthonormal frame by projection:

$$R_{\rho\sigma\mu\nu} := R_{abcd}(e_\rho)^a(e_\sigma)^b(e_\mu)^c(e_\nu)^d. \quad (2.57)$$

Recall that curvature is defined by the commutator of covariant derivatives. In a tetrad basis, the covariant derivative is encoded by the connection 1-forms $\omega_{a\mu\nu}$ defined in Eq. 2.53. Substituting this representation into the commutator definition of curvature yields

$$R_{\rho\sigma\mu\nu} = (e_\rho)^a(e_\sigma)^b(\nabla_a\omega_{b\mu\nu} - \nabla_b\omega_{a\mu\nu}) - \eta^{\alpha\beta}(\omega_{\rho\mu\alpha}\omega_{\sigma\nu\beta} - \omega_{\rho\nu\alpha}\omega_{\sigma\mu\beta}). \quad (2.58)$$

Equivalently, projecting the derivative along the tetrad directions,

$$R_{\rho\sigma\mu\nu} = (e_\rho)^a\nabla_a\omega_{\sigma\mu\nu} - (e_\sigma)^a\nabla_a\omega_{\rho\mu\nu} - \eta^{\alpha\beta}(\omega_{\rho\mu\alpha}\omega_{\sigma\nu\beta} - \omega_{\sigma\mu\alpha}\omega_{\rho\nu\beta}). \quad (2.59)$$

Ricci tensor in a tetrad. Once $R_{\rho\sigma\mu\nu}$ is known, the Ricci tensor components in the orthonormal frame are obtained by contraction:

$$R_{\mu\nu} = \eta^{\rho\sigma}R_{\rho\mu\sigma\nu}. \quad (2.60)$$

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 $n^2(n-1)/2$ independent Ricci rotation coefficients $\omega_{\lambda\mu\nu}$, thanks to the antisymmetry $\omega_{\lambda\mu\nu} = -\omega_{\lambda\nu\mu}$. In four dimensions this reduces the number of independent connection coefficients from 40 to 24. Moreover, in the orthonormal frame the metric components are constant, $\eta_{\mu\nu}$, so metric derivatives never appear explicitly. All geometric information is carried by how the frame rotates.

Torsion-Free Condition in a Tetrad

To determine the connection 1-forms $\omega_{\lambda\mu\nu}$ completely, we must impose that the derivative operator is torsion-free. Recall from Section 2.1 that torsion-freeness is equivalent to

$$\nabla_X Y - \nabla_Y X = [X, Y] \quad (2.61)$$

for all vector fields X^a, Y^a .

(i) **Commutation relations of basis vectors.** Apply 2.61 to the tetrad vectors $(e_\mu)^a$ and $(e_\nu)^a$:

$$\nabla_{e_\mu} e_\nu - \nabla_{e_\nu} e_\mu = [e_\mu, e_\nu].$$

Projecting onto the frame using $(e_\sigma)_a$ and using the definition of the Ricci rotation coefficients 2.54, we obtain

$$(e_\sigma)_a [e_\mu, e_\nu]^a = \omega_{\mu\nu\sigma} - \omega_{\nu\mu\sigma}. \quad (2.62)$$

Thus the antisymmetric part of the Ricci rotation coefficients is determined by the commutators of the tetrad basis vectors. If the tetrad happens to be coordinate-induced (so $[e_\mu, e_\nu] = 0$), then $\omega_{\mu\nu\sigma}$ is antisymmetric in $\mu\nu$ automatically.

(ii) **Antisymmetrized derivative of the coframe.** An equivalent formulation arises from differentiating the coframe. From the definition 2.53, one finds

$$\nabla_{[a}(e_{\sigma)}_{b]} = \eta^{\mu\nu} (e_\mu)_{[a} \omega_{b]\sigma\nu}. \quad (2.63)$$

For a torsion-free connection, the antisymmetrized derivative of any covector field depends only on partial derivatives, so we may replace ∇ by ∂ :

$$\partial_{[a}(e_{\sigma)}_{b]} = \eta^{\mu\nu} (e_\mu)_{[a} \omega_{b]\sigma\nu}. \quad (2.64)$$

Conversely, if 2.64 holds for all tetrad 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. Computes the exterior derivatives $\partial_{[a}(e_{\sigma)}_{b]}$ and solves 2.64 for $\omega_{\lambda\mu\nu}$.
3. Substitutes these into 2.59 to obtain $R_{\rho\sigma\mu\nu}$ and then contracts via 2.60.

A well-chosen coframe often makes step (2) almost mechanical.

Differential-Forms Formulation

the textbook further recasts the tetrad method in the language of differential forms, which packages the entire computation into two remarkably compact equations.

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,

$$\omega_\mu{}^\nu := \omega_{a\mu}{}^\nu dx^a, \quad \omega_{\mu\nu} = \eta_{\nu\sigma} \omega_\mu{}^\sigma,$$

so that, by metric compatibility 2.56,

$$\omega_{\mu\nu} = -\omega_{\nu\mu}.$$

First Cartan structure equation (torsion-free). The torsion-free condition 2.64 can be written in differential-form language as

$$d\mathbf{e}_\sigma = \mathbf{e}_\mu \wedge \omega^\mu{}_\sigma. \quad (2.65)$$

This is the *first Cartan structure equation*. It encodes how the orthonormal frame twists as one moves through the manifold.

Second Cartan structure equation (curvature). Define the curvature 2-forms

$$\mathbf{R}_\mu{}^\nu := \frac{1}{2} R_\mu{}^\nu{}_{\rho\sigma} \mathbf{e}^\rho \wedge \mathbf{e}^\sigma.$$

These are obtained directly from the connection 1-forms by

$$\mathbf{R}_\mu{}^\nu = d\omega_\mu{}^\nu + \omega_\mu{}^\alpha \wedge \omega_\alpha{}^\nu. \quad (2.66)$$

This is the *second Cartan structure equation*.

Together, 2.65 and 2.66 replace the lengthy coordinate formulas for Christoffel symbols and curvature components with two compact geometric identities.

Computational Interpretation. The differential-forms method reduces curvature computation to:

1. Choose an orthonormal coframe $\mathbf{e}_\mu$.
2. Compute $d\mathbf{e}_\mu$.
3. Solve the first structure equation 2.65 for $\omega_\mu{}^\nu$.
4. Compute $d\omega_\mu{}^\nu$.
5. Insert into 2.66 to obtain $\mathbf{R}_\mu{}^\nu$ and hence $R_{\mu\nu\rho\sigma}$.

For metrics with symmetry (e.g. Schwarzschild or FLRW), this procedure often collapses what would be pages of index algebra into a short sequence of exterior derivatives and wedge products.

Chapter 3

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 prerelativity 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*.

3.1 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

$$D^2 = (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2, \quad (3.1)$$

where $\mu = 1, 2, 3$ labels spatial coordinates.

For “nearby” points, this becomes

$$(\delta D)^2 = (\delta x^1)^2 + (\delta x^2)^2 + (\delta x^3)^2, \quad (3.2)$$

suggesting that the metric of space is

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2. \quad (3.3)$$

Abstractly, the metric is a smooth, symmetric tensor field h_{ab} . In a Cartesian coordinate system $\{x^\mu\}$, its components are

$$h_{\mu\nu} = \delta_{\mu\nu}, \quad \mu, \nu = 1, 2, 3, \quad (3.4)$$

and the abstract tensor is related to its coordinate components by

$$h_{ab} = h_{\mu\nu}(dx^\mu)_a(dx^\nu)_b. \quad (3.5)$$

To obtain ds^2 from Eq. 3.5, consider a small displacement $x^\mu \rightarrow x^\mu + dx^\mu$ and a tangent vector v^a whose coordinate components are $v^\mu = dx^\mu$. Then

$$\begin{aligned} ds^2 &= h_{ab}v^av^b \\ &= h_{\mu\nu}(dx^\mu)_a(dx^\nu)_bv^av^b \\ &= h_{\mu\nu}v^\mu v^\nu \\ &= h_{\mu\nu}dx^\mu dx^\nu. \end{aligned}$$

Since the components $h_{\mu\nu}$ are constant in Cartesian coordinates, we have

$$\partial_\alpha h_{\mu\nu} = 0. \quad (3.6)$$

Therefore the Christoffel symbols in this coordinate system vanish:

$$\Gamma^\mu{}_{\nu\lambda} = \frac{1}{2}h^{\mu\sigma}(\partial_\nu h_{\sigma\lambda} + \partial_\lambda h_{\sigma\nu} - \partial_\sigma h_{\nu\lambda}) = 0. \quad (3.7)$$

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 locally isometric to Euclidean space. 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 prerelativity 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 ultimately yields a number, many physical quantities (e.g. electromagnetic fields, stress tensors) cannot be specified meaningfully without reference to directions in space. Such quantities assign numbers to vectors and covectors in a manner that is multilinear, and are therefore naturally represented as tensor fields.

Why tensors? Any smooth map that assigns a number to a collection of vectors and dual vectors and is linear in each argument is, by definition, a tensor. Moreover, sufficiently regular physical quantities may be decomposed into sums of such multilinear maps. Thus the class of

tensor fields is broad enough to represent all measurable physical quantities in prerelativity, special relativity, and general relativity.

Consequently, the laws of physics must be expressible as *tensor equations*—equalities between tensor fields that hold independently of any particular coordinate system.

General Covariance

The central principle is:

The laws of physics must refer only to geometric structures intrinsic to space.

In prerelativity 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.

Thus physical laws must be formulated entirely in terms of tensor fields constructed from the metric and its associated derivative operator. A change of coordinates may alter the components of these tensors, but it cannot alter the tensorial relations themselves.

Coordinate expressions that fail to incorporate the geometric structure (for example, writing equations using only partial derivatives rather than the covariant derivative compatible with h_{ab}) may *appear* non-tensorial. This reflects an incomplete formulation rather than a failure of covariance.

A key implication is that the Christoffel symbols Γ^c_{ab} , which are not tensorial objects, may not appear by themselves in any physical law. They may enter only through combinations that form genuine tensors, such as the covariant derivative or the curvature tensor.

Special Covariance

The metric h_{ab} in Eq. 3.4 admits a six-parameter group of isometries (three rotations and three translations). An isometry is a diffeomorphism that preserves the metric, i.e. one under which h_{ab} is invariant.

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 that preserve the metric structure. In contrast, general covariance requires that the laws of physics be

expressible in a tensorial form valid in *all* coordinate systems, whether or not they correspond to symmetries of the metric.

Special covariance therefore reflects a genuine physical symmetry of space, whereas general covariance is a requirement on the mathematical formulation of physical laws.

Note: *This does not refer to observers in relative motion (to be discussed in special relativity). Rather, it concerns observers who describe the same spatial geometry using coordinate systems related by isometries—for example, one Cartesian coordinate system rotated relative to another.*

3.2 Special Relativity

3.2.1 Overview and Inertial Frames

Special relativity begins with the assumption that spacetime has the structure of $\mathbb{R}^4$. A global coordinate system $\{x^\mu\} = (x^0, x^1, x^2, x^3)$ 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 discrete parity and time-reversal transformations). Thus, the numerical values of (x^0, x^1, x^2, x^3) at an event have no intrinsic meaning. Nevertheless, all inertial 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

$$I = -(x^0 - \bar{x}^0)^2 + (x^1 - \bar{x}^1)^2 + (x^2 - \bar{x}^2)^2 + (x^3 - \bar{x}^3)^2, \quad (3.8)$$

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 3.8 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, its components are

$$\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1), \quad \mu, \nu = 0, 1, 2, 3, \quad (3.9)$$

and the abstract tensor is related to its coordinate components by

$$\eta_{ab} = \eta_{\mu\nu} (dx^\mu)_a (dx^\nu)_b. \quad (3.10)$$

The tensor field η_{ab} is independent of the particular choice of inertial coordinates. Because the components $\eta_{\mu\nu}$ are constant in any global inertial coordinate system, we have

$$\partial_\alpha \eta_{\mu\nu} = 0. \quad (3.11)$$

Thus the ordinary derivative operator ∂_a is compatible with η_{ab} and serves as its Levi-Civita connection.

Since ordinary derivatives commute, the associated curvature tensor vanishes. Therefore 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. 3.8 of the interval, and the interpretation of inertial motion as straight lines (i.e., geodesics) in Minkowski spacetime.

3.2.2 Geometry of Minkowski Spacetime

In special relativity, the flat Lorentzian metric η_{ab} introduced in Eq. 3.10 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

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

are precisely the straight lines in global inertial coordinates. In such coordinates, where the connection coefficients vanish, this reduces to

$$T^a \partial_a T^b = 0.$$

Interpretation. A geodesic of η_{ab} is the worldline of a freely falling particle in special relativity. Thus, straight-line motion in (x^0, 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:

$$\begin{aligned} \eta_{ab} T^a T^b &< 0 && \text{timelike,} \\ \eta_{ab} T^a T^b &= 0 && \text{null,} \\ \eta_{ab} T^a T^b &> 0 && \text{spacelike.} \end{aligned}$$

Timelike curves represent possible worldlines of particles with nonzero rest mass. Null curves represent the propagation of light. Spacelike curves cannot represent worldlines of physical particles.

Note: The quantity $\eta_{ab}T^aT^b$ is the spacetime inner product of T^a with itself. With signature $(-, +, +, +)$, a negative value indicates a vector lying inside the light cone (timelike), a positive value lies outside the light cone (spacelike), and zero lies on the light cone (null).

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 inertial observers agree on the light-cone structure because η_{ab} is the same tensor field in every inertial frame.

Proper Time

Given a timelike worldline parameterized by λ , the proper time is defined by

$$\tau = \int \sqrt{-\eta_{ab} \frac{dx^a}{d\lambda} \frac{dx^b}{d\lambda}} d\lambda. \quad (3.12)$$

This quantity depends only on the geometric curve and is independent of the choice of parameter λ . Thus proper time is the Lorentzian arc-length of a timelike curve.

Choosing $\lambda = t = x^0$ in an inertial coordinate system, and writing

$$x^a(t) = (t, \vec{x}(t)), \quad \vec{v} = \frac{d\vec{x}}{dt},$$

we obtain

$$d\tau = \sqrt{1 - |\vec{v}|^2} dt,$$

which is the familiar time-dilation relation of special relativity.

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,

$$u^a u_a = -1. \quad (3.13)$$

A freely falling particle experiences no force; hence its worldline is a geodesic:

$$u^a \nabla_a u^b = 0. \quad (3.14)$$

In global inertial coordinates this reduces to

$$u^a \partial_a u^b = 0,$$

so the components of u^b are constant: free particles move with constant velocity unless acted upon by external forces.

Energy and Momentum

All material particles possess a rest mass m . The *four-momentum* is defined by

$$p^a = mu^a. \quad (3.15)$$

For an observer with four-velocity v^a , the *energy* of the particle measured by that observer is

$$E = -p_a v^a. \quad (3.16)$$

For an inertial observer at rest in these coordinates, $v^a = (1, 0, 0, 0)$, this reduces to

$$E = p^0.$$

Writing $u^a = \gamma(1, \vec{v})$ with

$$\gamma = \frac{1}{\sqrt{1 - |\vec{v}|^2}},$$

we obtain

$$p^a = m\gamma(1, \vec{v}) = (m\gamma, m\gamma\vec{v}).$$

Thus

$$E = m\gamma, \quad \vec{p} = m\gamma\vec{v},$$

and these satisfy the invariant mass-shell condition

$$p^a p_a = -m^2, \quad \text{equivalently} \quad E^2 - |\vec{p}|^2 = m^2.$$

Interpretation. Energy is the time component of four-momentum relative to a chosen observer. Because η_{ab} is flat and parallel transport is path independent, energy and momentum are globally well-defined for inertial observers.

Note 3.1. Proper time, τ , is local and physical: it is measured by a clock carried along a worldline.

Coordinate time, t , is global and observer-dependent.

They coincide only when the particle is at rest in the chosen inertial frame.

3.2.3 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

$$T_{ab} v^a v^b \geq 0 \quad (3.17)$$

represents the mass–energy density measured by that observer. For ordinary matter this quantity is nonnegative; this is the flat-spacetime version of the *weak energy condition*.

If x^a is any vector orthogonal to v^a (i.e. $v_a x^a = 0$), then

$$-T_{ab}v^a x^b$$

is interpreted as the momentum density in the x^a direction as measured in the observer’s rest frame. Similarly, if x^a and y^a are both orthogonal to v^a , the quantity

$$T_{ab}x^a y^b$$

represents a spatial stress component (pressure or shear).

Interpretation. Contracting T_{ab} with the observer’s four-velocity v^a projects out the physical densities and fluxes measured in that observer’s rest frame. The scalar $T_{ab}v^a v^b$ gives energy density. The mixed contractions $T_{ab}v^a x^b$ encode momentum density and energy flux. The fully spatial components encode internal stresses. Thus T_{ab} organizes all mechanical properties of matter into a single 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, and so on. The only mathematical object that transforms correctly under Lorentz transformations and consistently organizes these quantities in every inertial frame is a symmetric rank-2 tensor, the stress–energy tensor T_{ab} . Projecting T_{ab} with the observer’s four-velocity v^a extracts 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 ($u^a u_a = -1$). In the rest frame of the fluid, $u^a = (1, 0, 0, 0)$, and the symmetry assumptions imply:

- 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 has the form

$$(T_{ab})_{\text{rest}} = \begin{pmatrix} \rho & 0 \\ 0 & P\delta_{ij} \end{pmatrix},$$

where we are using the Minkowski signature $(-, +, +, +)$.

To express this in a Lorentz-invariant form valid in *any* frame, we proceed geometrically. The only tensorial ingredients available for constructing a symmetric rank-2 tensor from the fluid variables are the metric η_{ab} and the four-velocity u^a . Isotropy forbids introducing any preferred spatial direction, so the most general symmetric tensor consistent with these requirements is

$$T_{ab} = A u_a u_b + B \eta_{ab},$$

for scalars A and B to be determined.

Evaluating this ansatz in the fluid rest frame,

$$T_{00} = A - B = \rho, \quad T_{ij} = B \delta_{ij} = P \delta_{ij}.$$

Hence $B = P$ and $A = \rho + P$, yielding

$$T_{ab} = (\rho + P) u_a u_b + P \eta_{ab}.$$

It is useful 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 $h_{ij} = \delta_{ij}$ and $h_{00} = h_{0i} = 0$. Using h_{ab} , the perfect-fluid tensor may be written compactly as

$$T_{ab} = \rho u_a u_b + P h_{ab} = \rho u_a u_b + P(\eta_{ab} + u_a u_b). \quad (3.18)$$

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, conservation of stress–energy takes the form

$$\partial^a T_{ab} = 0. \quad (3.19)$$

This single tensor equation encodes both energy and momentum conservation.

Using the perfect-fluid form

$$T_{ab} = (\rho + P)u_a u_b + P \eta_{ab},$$

we compute

$$\begin{aligned} 0 &= \partial^a T_{ab} \\ &= \partial^a [(\rho + P)u_a u_b + P \eta_{ab}] \\ &= \partial^a [(\rho + P)u_a u_b] + \partial^a [P \eta_{ab}] \\ &= \partial^a (\rho + P) u_a u_b + (\rho + P) \partial^a (u_a u_b) + \partial^a (P \eta_{ab}) \\ &= \partial^a (\rho + P) u_a u_b + (\rho + P) \partial^a (u_a u_b) + \eta_{ab} \partial^a P \\ &= \partial^a (\rho + P) u_a u_b + (\rho + P) \partial^a (u_a u_b) + \partial_b P. \end{aligned}$$

Now expand the derivative of the product:

$$\partial^a (u_a u_b) = (\partial^a u_a) u_b + u_a (\partial^a u_b).$$

Substituting,

$$\begin{aligned} 0 &= \partial^a (\rho + P) u_a u_b \\ &\quad + (\rho + P) [(\partial^a u_a) u_b + u_a (\partial^a u_b)] + \partial_b P. \end{aligned}$$

Projection along u^b . Contract with u^b , using $u^b u_b = -1$:

$$\begin{aligned} 0 &= u^b \partial^a (\rho + P) u_a u_b \\ &\quad + u^b (\rho + P) (\partial^a u_a) u_b \\ &\quad + u^b (\rho + P) u_a (\partial^a u_b) \\ &\quad + u^b \partial_b P. \end{aligned}$$

Now evaluate term by term:

- $u^b u_b = -1$, so

$$u^b \partial^a (\rho + P) u_a u_b = -\partial^a (\rho + P) u_a.$$

- Similarly,

$$u^b (\rho + P) (\partial^a u_a) u_b = -(\rho + P) (\partial^a u_a).$$

- For the third term, use

$$u^b \partial^a u_b = \frac{1}{2} \partial^a (u^b u_b) = \frac{1}{2} \partial^a (-1) = 0,$$

so this term vanishes.

- The final term is simply

$$u^b \partial_b P = u^a \partial_a P.$$

Putting everything together,

$$\begin{aligned}
 0 &= -\partial^a(\rho + P)u_a - (\rho + P)(\partial^a u_a) + u^a \partial_a P \\
 &= -u^a \partial_a \rho - u^a \partial_a P - (\rho + P)\partial^a u_a + u^a \partial_a P \\
 &= -u^a \partial_a \rho - (\rho + P)\partial^a u_a.
 \end{aligned}$$

Hence,

$$u^a \partial_a \rho + (\rho + P)\partial^a u_a = 0. \quad (3.20)$$

Projection orthogonal to u^b . Introduce the spatial projector

$$h^c_b = \delta^c_b + u^c u_b, \quad h_{ab} = \eta_{ab} + u_a u_b,$$

which satisfies $h^c_b u^b = 0$.

Projecting orthogonally,

$$\begin{aligned}
 0 &= h^c_b \partial^a T_{ab} \\
 &= h^c_b \left[\partial^a(\rho + P)u_a u_b \right. \\
 &\quad \left. + (\rho + P)(\partial^a u_a)u_b \right. \\
 &\quad \left. + (\rho + P)u_a(\partial^a u_b) + \partial_b P \right].
 \end{aligned}$$

All terms proportional to u_b vanish under projection, leaving

$$(\rho + P)h^c_b u_a(\partial^a u_b) + h^{cb} \partial_b P = 0.$$

Since $u_a \partial^a u^b$ is orthogonal to u^b , the projector does not alter it:

$$h^c_b u_a \partial^a u^b = u_a \partial^a u^c.$$

Thus,

$$(\rho + P)u^a \partial_a u^c + h^{cb} \partial_b P = 0.$$

Lowering the free index,

$$(\rho + P)u^a \partial_a u_b + (\eta_{ab} + u_a u_b)\partial^a P = 0. \quad (3.21)$$

Interpretation. Equation 3.20 describes the change in energy density along the fluid flow (compression increases ρ , expansion decreases it). Equation 3.21 is the relativistic Euler equation: the fluid accelerates in response to pressure gradients.

Nonrelativistic Limit

In the nonrelativistic limit we assume

- $|\vec{v}| \ll 1$,
- $P \ll \rho$,
- time derivatives of P are small compared with spatial gradients.

To leading order in $|\vec{v}|$, the four-velocity becomes

$$u^a \approx (1, \vec{v}),$$

and the Lorentz factor satisfies $\gamma \approx 1$.

Under these approximations, the energy equation 3.20 reduces to

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0, \quad (3.22)$$

which is the classical continuity equation.

Similarly, the spatial components of Eq. 3.21 reduce to

$$\rho \left(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla}) \vec{v} \right) = -\vec{\nabla} P, \quad (3.23)$$

which is the Euler equation of classical fluid dynamics.

Key Point. Classical fluid dynamics emerges as the low-velocity, low-pressure limit of the relativistic conservation law $\partial^a T_{ab} = 0$.

Mass–Energy Current

For a family of inertial observers with constant four-velocity v^a , define the mass–energy current vector

$$J^a = -T^{ab} v_b. \quad (3.24)$$

Since $\partial_a T^{ab} = 0$ and v^a is constant in inertial coordinates, it follows that

$$\partial_a J^a = 0. \quad (3.25)$$

Integrating Eq. 3.25 over a spacetime region $\mathcal{V}$ and applying the divergence theorem gives

$$\int_{\partial \mathcal{V}} J^a n_a dS = 0, \quad (3.26)$$

where n^a is the outward-directed unit normal to the boundary $\partial \mathcal{V}$ and dS is the induced 3-volume element.

This expresses local conservation of energy as measured by the inertial observers with four-velocity v^a .

Geometric view. Equation 3.26 states that the total mass–energy flux through any closed 3-surface in Minkowski spacetime vanishes. Equivalently, the total energy contained within a spatial region can change only by energy flowing across its boundary.

3.2.4 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 fundamental classical scalar field is known to exist, scalar fields provide a useful model. A real scalar field ϕ on Minkowski spacetime satisfies the Klein–Gordon equation

$$\partial^a \partial_a \phi - m^2 \phi = 0. \quad (3.27)$$

Its stress–energy tensor is

$$T_{ab} = \partial_a \phi \partial_b \phi - \frac{1}{2} \eta_{ab} \left(\partial^c \phi \partial_c \phi + m^2 \phi^2 \right). \quad (3.28)$$

Interpretation. The first term represents kinetic contributions from directional derivatives of ϕ . The second term subtracts the appropriate trace contribution so that T_{ab} represents the local flux of energy and momentum carried by the field. Using the field equation 3.27, one verifies directly that $\partial^a T_{ab} = 0$.

Electromagnetic Field

In prerelativity 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 six independent components, corresponding to the three components of $\vec{E}$ and the three components of $\vec{B}$.

For an observer with four-velocity v^a , the electric field measured by that observer is

$$E_a = F_{ab} v^b, \quad (3.29)$$

while the magnetic field is

$$B_a = -\frac{1}{2} \epsilon_{abcd} F^{cd} v^b. \quad (3.30)$$

Here ϵ_{abcd} is the totally antisymmetric tensor with $\epsilon_{0123} = 1$ in a right-handed orthonormal basis.

Geometric meaning. Equations 3.29–3.30 decompose the antisymmetric tensor F_{ab} into spatial electric and magnetic components relative to an observer. Different observers measure different electric and magnetic fields, but the underlying tensor F_{ab} is invariantly defined.

Maxwell's Equations

In tensor form, Maxwell's equations are

$$\partial^a F_{ab} = -4\pi j_b, \quad (3.31)$$

$$\partial_{[a} F_{bc]} = 0, \quad (3.32)$$

where j^a is the electric four-current.

Taking the divergence of Eq. 3.31 and using the antisymmetry of F_{ab} gives

$$\partial^b j_b = 0, \quad (3.33)$$

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:

$$u^a \partial_a u_b = \frac{q}{m} F_{bc} u^c, \quad (3.34)$$

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

$$T_{ab} = \frac{1}{4\pi} \left(F_{ac} F_b{}^c - \frac{1}{4} \eta_{ab} F_{cd} F^{cd} \right). \quad (3.35)$$

Using Maxwell's equations, one finds

$$\partial^a T_{ab} = -F_{bc} j^c,$$

which vanishes in vacuum ($j^a = 0$).

The Vector Potential

The equation $\partial_{[a} F_{bc]} = 0$ implies (by the Poincaré lemma) the existence of a vector field A_a such that

$$F_{ab} = \partial_a A_b - \partial_b A_a. \quad (3.36)$$

In terms of A_a , Eq. 3.31 becomes

$$\partial^a (\partial_a A_b - \partial_b A_a) = -4\pi j_b. \quad (3.37)$$

Gauge freedom. The potential A_a is not unique: $A_a \rightarrow A_a + \partial_a \chi$ leaves F_{ab} unchanged. This gauge freedom reflects redundancy in the description of the physical field.

Choosing the Lorenz gauge,

$$\partial^a A_a = 0, \quad (3.38)$$

Eq. 3.37 reduces to

$$\partial^a \partial_a A_b = -4\pi j_b. \quad (3.39)$$

Electromagnetic Waves

For source-free fields ($j^a = 0$), seek solutions of the form

$$A_a = C_a e^{iS}, \quad (3.40)$$

where C_a is constant and S is the phase.

Substituting into Eq. 3.39 yields

$$\partial^a \partial_a S = 0, \quad (3.41)$$

$$\partial_a S \partial^a S = 0, \quad (3.42)$$

$$C_a \partial^a S = 0. \quad (3.43)$$

Null hypersurfaces and light propagation. Let $k_a = \partial_a S$. Equation 3.42 states that k_a is null. Differentiating $k_a k^a = 0$ gives

$$k^b \partial_b k_a = 0,$$

so the integral curves of k^a are null geodesics. Thus electromagnetic waves propagate along null geodesics of Minkowski spacetime.

The frequency measured by an observer with four-velocity v^a is

$$\omega = -v^a k_a. \quad (3.44)$$

Plane-wave solutions correspond to

$$S = k_\mu x^\mu, \quad (3.45)$$

where k_μ are constants and k^a is a constant null vector.

Conclusion. Maxwell's equations predict that light propagates along null geodesics. This connects directly with the causal structure defined by the Minkowski metric.

3.3 General Relativity

3.3.1 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 naturally into special relativity because it acts differently on different bodies; Newtonian gravity does not share this feature.. 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). In the absence of non-gravitational forces, such observers follow geodesics of the flat spacetime metric and therefore define inertial motion. 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. 3.34. 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.

3.3.2 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 (weak) equivalence principle states that all freely falling bodies follow the same trajectories in a gravitational field, independent of their internal structure.: 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 remains valid locally, but spacetime need not possess a globally flat metric.

3.3.3 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. 2.45).

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, what Newtonian theory calls gravitational force corresponds in general relativity to non-geodesic motion. 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.

3.3.4 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 $T_p M$ 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

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \quad \Gamma^\rho_{\mu\nu}(p) = 0.$$

These are Riemann normal coordinates. Although the Christoffel symbols vanish at p , their derivatives generally do not; curvature cannot be transformed away. 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. Einstein's equations are local differential equations: they determine how curvature behaves pointwise but do not fix the global topology of spacetime. The global structure of (M, g_{ab}) depends on both the field equations and the underlying manifold M .

3.3.5 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. the textbook 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*:

$$\eta_{ab} \rightarrow g_{ab}, \quad \partial_a \rightarrow \nabla_a,$$

and leaves the form of the physical equations otherwise unchanged.

Warning. This “rule” is only a heuristic. the textbook 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

$$u^a \nabla_a u^b = 0. \tag{3.46}$$

This is exactly the invariant geodesic equation introduced in Eq. 2.36. 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 the textbook’s development of motion, forces, and stress-energy in curved spacetime. In the next subsection, we follow the textbook into the discussion of 4-momentum and the generalization of the Lorentz force.

3.3.6 Motion, 4-Momentum, and Forces in Curved Spacetime

Equation 3.46,

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

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} , the textbook takes over from special relativity the Lorentz force law, replacing η_{ab} by g_{ab} and ∂_a by ∇_a :

$$u^a \nabla_a u^b = \frac{q}{m} F^b{}_c u^c. \quad (3.47)$$

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. 3.47 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

$$p^a \equiv m u^a \quad (3.48)$$

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

$$E = -p_a v^a. \quad (3.49)$$

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 , the textbook writes

$$T_{ab} = \rho u_a u_b + P(g_{ab} + u_a u_b), \quad (3.50)$$

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

$$\nabla^a T_{ab} = 0. \quad (3.51)$$

This equation replaces $\partial^a T_{ab} = 0$ from flat spacetime.

To see what Eq. 3.51 implies for a perfect fluid, we project along and orthogonal to u^a . Contracting with u^b gives

$$u^b \nabla^a T_{ab} = 0 \implies u^a \nabla_a \rho + (\rho + P) \nabla_a u^a = 0, \quad (3.52)$$

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. 3.50, the spatial projection of $\nabla^a T_{ab} = 0$ yields

$$(\rho + P) u^a \nabla_a u_b + (g_{ab} + u_a u_b) \nabla^a P = 0. \quad (3.53)$$

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 3.51 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.

3.3.7 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. the textbook 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 $\eta_{ab} \rightarrow g_{ab}$ and $\partial_a \rightarrow \nabla_a$, the natural curved-spacetime generalization is

$$\nabla^a \nabla_a \phi - m^2 \phi = 0. \quad (3.54)$$

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

$$T_{ab} = (\nabla_a \phi)(\nabla_b \phi) - \frac{1}{2} g_{ab} [(\nabla_c \phi)(\nabla^c \phi) + m^2 \phi^2]. \quad (3.55)$$

It satisfies the covariant conservation law $\nabla^a T_{ab} = 0$ as a direct consequence of Eq. 3.54.

Non-uniqueness of minimal substitution. the textbook emphasizes that minimal substitution is not the only reasonable generalization. For example, one could add a curvature term:

$$\nabla^a \nabla_a \phi - m^2 \phi - \alpha R \phi = 0,$$

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

$$\nabla^a F_{ab} = -4\pi j_b, \quad (3.56)$$

$$\nabla_{[a} F_{bc]} = 0. \quad (3.57)$$

Equation 3.57 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. 3.35 by simply replacing η_{ab} with g_{ab} :

$$T_{ab} = \frac{1}{4\pi} \left(F_{ac} F_b{}^c - \frac{1}{4} g_{ab} F_{cd} F^{cd} \right). \quad (3.58)$$

Lorentz gauge and curvature. If we impose the Lorentz gauge $\nabla^a A_a = 0$, then the field equation for A_a becomes

$$\nabla^a \nabla_a A_b - R_b{}^a A_a = -4\pi j_b. \quad (3.59)$$

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 3.59 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

$$A_a = C_a e^{iS},$$

where derivatives of C_a are small. Substituting into the homogeneous Maxwell equation (with $j_b = 0$) and keeping only dominant terms gives

$$\nabla_a S \nabla^a S = 0. \quad (3.60)$$

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.

3.3.8 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. We will obtain this by comparing the description of tidal forces in Newtonian gravity with the corresponding description 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. 2.45),

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

Thus tidal forces are encoded in the curvature tensor R_{abcd} .

Comparing these two descriptions, the textbook is led to the correspondence

$$R_{cbd}{}^a u^c u^d \longleftrightarrow \partial_b \partial^a \phi. \quad (3.61)$$

Relating curvature to energy density. In Newtonian gravity, the potential satisfies Poisson's equation

$$\nabla^2 \phi = 4\pi\rho, \quad (3.62)$$

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

$$\rho_{(\text{obs})} = T_{ab} u^a u^b.$$

Thus the textbook is led to the correspondence

$$T_{ab} u^a u^b \longleftrightarrow \rho, \quad (3.63)$$

at least in the regime where Newtonian theory should be valid.

Combining the geodesic-deviation picture 2.45 with the identification of matter content through the stress–energy tensor (see 3.54–3.58), the natural expectation is an equation of the schematic 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. 2.34 states that

$$\nabla^a (R_{ab} - \tfrac{1}{2} g_{ab} R) = 0.$$

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*

$$G_{ab} \equiv R_{ab} - \tfrac{1}{2} R g_{ab}.$$

By construction, $\nabla^a G_{ab} = 0$ identically, for any metric g_{ab} . We are therefore led to the equation

$$G_{ab} \equiv R_{ab} - \tfrac{1}{2} R g_{ab} = 8\pi T_{ab}. \quad (3.64)$$

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 3.64, we find

$$R = -8\pi T, \quad (3.65)$$

where $T \equiv T^a_a$ is the trace of the stress–energy tensor. Substituting back into 3.64 yields

$$R_{ab} = 8\pi \left(T_{ab} - \tfrac{1}{2} g_{ab} T \right). \quad (3.66)$$

Newtonian limit check. To compare with Newtonian gravity, we evaluate Einstein's equation along the 4-velocity u^a of observers comoving with the matter. This contraction isolates the energy density seen by those observers and is the relativistic analogue of the Newtonian source term ρ in Poisson's equation.

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. 2.45, 3.54, and 3.58 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 3.64 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.

3.3.9 Remarks on the Nature of Einstein's Equation

We close this section with several important comments about the structure and interpretation of Eq. 3.64.

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. This hyperbolic character is what allows initial data on a suitable hypersurface to determine the subsequent spacetime evolution in a causal way. 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. 3.64 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 $G_{ab} = 8\pi T_{ab}$.

3.4 Linearized Gravity: The Newtonian Limit and Gravitational Radiation

3.4.1 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} .

3.4.1.1 Perturbing the Metric

We assume that the physical metric g_{ab} differs only slightly from the flat Minkowski metric η_{ab} , and we write

$$g_{ab} = \eta_{ab} + \gamma_{ab}, \tag{3.67}$$

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

$$g^{ab} = \eta^{ab} - \gamma^{ab}. \quad (3.68)$$

This is consistent with first-order truncation: using η_{ab} to raise/lower indices keeps all expressions closed at $O(\gamma)$, while any additional correction from g_{ab} would enter at $O(\gamma^2)$.

3.4.1.2 Linearized Christoffel Symbols

We denote by ∂_a the flat derivative compatible with η_{ab} . Substituting Eq. 3.67 into the definition of the Christoffel symbols,

$$\begin{aligned} \Gamma_{\mu\nu}^\rho &= \frac{1}{2} g^{\rho\sigma} (\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \\ \Gamma_{\mu\nu}^\rho &= \frac{1}{2} (\eta^{\rho\sigma} - \gamma^{\rho\sigma}) (\partial_\mu (\eta_{\sigma\nu} + \gamma_{\sigma\nu}) + \partial_\nu (\eta_{\sigma\mu} + \gamma_{\sigma\mu}) - \partial_\sigma (\eta_{\mu\nu} + \gamma_{\mu\nu})) \end{aligned}$$

and keeping only terms linear in γ_{ab} yields

$$\Gamma_{ab}^c = \frac{1}{2} \eta^{cd} (\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab}). \quad (3.69)$$

After recognizing that $\partial_a \eta_{bc} = 0$, we discard all terms of the form $\gamma \partial \gamma$, since they are second order in γ .

3.4.1.3 Linearized Ricci Tensor

The Ricci tensor is

$$R_{ab} = \partial_c \Gamma_{ab}^c - \partial_a \Gamma_{cb}^c + \Gamma_{ab}^c \Gamma_{cd}^d - \Gamma_{ad}^c \Gamma_{cb}^d$$

The linearized Ricci tensor is

$$R_{ab}^{(1)} = \partial^c \partial_{(a} \gamma_{b)c} - \frac{1}{2} \partial^c \partial_c \gamma_{ab} - \frac{1}{2} \partial_a \partial_b \gamma \quad (3.70)$$

and we now verify this expression explicitly. where $\gamma \equiv \gamma^c_c$ is the trace and the symmetrization

$$\partial^c \partial_{(a} \gamma_{b)c} = \frac{1}{2} (\partial^c \partial_a \gamma_{bc} + \partial^c \partial_b \gamma_{ac}).$$

Ricci Tensor (Details of the Linearization) The full derivation of the linearized Ricci tensor is as follows. Recall that to linear order in γ_{ab} the Christoffel symbols are

$$\Gamma^c_{ab} = \frac{1}{2} \eta^{cd} (\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab}).$$

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

$$R^{(1)}_{ab} = \partial_c \Gamma^c_{ab} - \partial_a \Gamma^c_{cb}. \quad (3.71)$$

Using Eq. 3.69, we first evaluate $\partial_c \Gamma^c_{ab}$:

$$\begin{aligned} \partial_c \Gamma^c_{ab} &= \partial_c \left[\frac{1}{2} \eta^{cd} (\partial_a \gamma_{bd} + \partial_b \gamma_{ad} - \partial_d \gamma_{ab}) \right] \\ &= \frac{1}{2} \eta^{cd} (\partial_c \partial_a \gamma_{bd} + \partial_c \partial_b \gamma_{ad} - \partial_c \partial_d \gamma_{ab}), \end{aligned}$$

since $\partial_c \eta^{cd} = 0$ in inertial coordinates.

Now raise the index c with η^{cd} :

$$\eta^{cd} \partial_c = \partial^d, \quad \eta^{cd} \partial_c \partial_d = \partial^c \partial_c.$$

So

$$\partial_c \Gamma^c_{ab} = \frac{1}{2} (\partial^c \partial_a \gamma_{bc} + \partial^c \partial_b \gamma_{ac} - \partial^c \partial_c \gamma_{ab}). \quad (3.72)$$

Next, we evaluate $\partial_a \Gamma^c_{cb}$. First contract Γ^c_{cb} using Eq. 3.69:

$$\Gamma^c_{cb} = \frac{1}{2} \eta^{cd} (\partial_c \gamma_{bd} + \partial_b \gamma_{cd} - \partial_d \gamma_{cb}).$$

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,

$$\Gamma^c_{cb} = \frac{1}{2} (\partial^c \gamma_{bc} + \partial_b \gamma - \partial^c \gamma_{cb}) = \frac{1}{2} \partial_b \gamma,$$

because the first and third terms are identical.

Therefore,

$$\partial_a \Gamma^c{}_{cb} = \frac{1}{2} \partial_a \partial_b \gamma. \quad (3.73)$$

Substituting Eqs. 3.72 and 3.73 into Eq. 3.71, we obtain

$$R_{ab}^{(1)} = \frac{1}{2} \left(\partial^c \partial_a \gamma_{bc} + \partial^c \partial_b \gamma_{ac} - \partial^c \partial_c \gamma_{ab} \right) - \frac{1}{2} \partial_a \partial_b \gamma.$$

Combine the first two terms using symmetrization:

$$\partial^c \partial_{(a} \gamma_{b)c} \equiv \frac{1}{2} (\partial^c \partial_a \gamma_{bc} + \partial^c \partial_b \gamma_{ac}).$$

Hence the linearized Ricci tensor is

$$R_{ab}^{(1)} = \partial^c \partial_{(a} \gamma_{b)c} - \frac{1}{2} \partial^c \partial_c \gamma_{ab} - \frac{1}{2} \partial_a \partial_b \gamma.$$

Contraction of the Linearized Ricci Tensor Starting from the linearized Ricci tensor,

$$R_{ab}^{(1)} = \partial^c \partial_{(a} \gamma_{b)c} - \frac{1}{2} \partial^c \partial_c \gamma_{ab} - \frac{1}{2} \partial_a \partial_b \gamma, \quad \gamma \equiv \gamma^c{}_c, \quad (3.74)$$

we now contract with η^{ab} to obtain the linearized Ricci scalar.

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

$$\eta^{ab} \partial^c \partial_{(a} \gamma_{b)c} = \partial^c \partial^a \gamma_{ac}.$$

For the second term, since $\eta^{ab} \gamma_{ab} = \gamma$,

$$\eta^{ab} \partial^c \partial_c \gamma_{ab} = \partial^c \partial_c \gamma.$$

Hence

$$-\frac{1}{2} \eta^{ab} \partial^c \partial_c \gamma_{ab} = -\frac{1}{2} \partial^c \partial_c \gamma.$$

For the third term,

$$\eta^{ab} \partial_a \partial_b \gamma = \partial^a \partial_a \gamma = \partial^c \partial_c \gamma,$$

so

$$-\frac{1}{2} \eta^{ab} \partial_a \partial_b \gamma = -\frac{1}{2} \partial^c \partial_c \gamma.$$

Combining all three pieces,

$$R^{(1)} = \partial^a \partial^b \gamma_{ab} - \partial^c \partial_c \gamma. \quad (3.75)$$

3.4.1.4 Linearized Einstein Tensor

The linearized Einstein tensor is

$$G_{ab}^{(1)} = R_{ab}^{(1)} - \frac{1}{2} \eta_{ab} R^{(1)}.$$

Substituting equations 3.70 and 3.75, we obtain the expression for the linearized Einstein tensor,

$$G_{ab}^{(1)} = \partial^c \partial_{(a} \gamma_{b)c} - \frac{1}{2} \partial^c \partial_c \gamma_{ab} - \frac{1}{2} \partial_a \partial_b \gamma - \frac{1}{2} \eta_{ab} \left(\partial^c \partial^d \gamma_{cd} - \partial^e \partial_e \gamma \right), \quad (3.76)$$

Eq. 3.76 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

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2} \eta_{ab} \gamma, \quad \gamma = \gamma^c_c. \quad (3.77)$$

Taking the trace of this equation shows that

$$\bar{\gamma} \equiv \bar{\gamma}^c_c = -\gamma, \quad \text{so} \quad \gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2} \eta_{ab} \bar{\gamma}.$$

Substituting into $G_{ab}^{(1)}$. Inserting the expression

$$\gamma_{ab} = \bar{\gamma}_{ab} - \frac{1}{2} \eta_{ab} \bar{\gamma}$$

into Eq. 3.76, and rearranging terms, one finds after a short calculation that

$$G_{ab}^{(1)} = -\frac{1}{2} \left(\partial^c \partial_c \bar{\gamma}_{ab} - \partial^c \partial_a \bar{\gamma}_{bc} - \partial^c \partial_b \bar{\gamma}_{ac} + \eta_{ab} \partial^c \partial^d \bar{\gamma}_{cd} \right). \quad (3.78)$$

3.4.1.5 Lorenz Gauge

The perturbation γ_{ab} is defined only up to an infinitesimal coordinate transformation (a gauge transformation). Under $x^a \rightarrow x^a + \xi^a$, the trace-reversed perturbation transforms as

$$\bar{\gamma}_{ab} \rightarrow \bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab} \partial^c \xi_c.$$

Using this freedom, one may impose the *Lorenz gauge*:

$$\partial^a \bar{\gamma}_{ab} = 0. \quad (3.79)$$

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 Eq. 3.78 vanish, leaving the remarkably simple equation

$$G_{ab}^{(1)} = -\frac{1}{2} \partial^c \partial_c \bar{\gamma}_{ab}. \quad (3.80)$$

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 $A_a \rightarrow A_a + \partial_a \Lambda$ 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

$$\gamma_{ab} \rightarrow \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a$$

arises from infinitesimal coordinate transformations $x^a \rightarrow x^a + \xi^a$. 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

$$G_{ab}^{(1)} = -\frac{1}{2} \square \bar{\gamma}_{ab},$$

so that

$$\square \bar{\gamma}_{ab} = -16\pi T_{ab}.$$

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 $\square$ denotes the flat-spacetime wave operator (d'Alembertian),

$$\square \equiv \partial^c \partial_c = \eta^{ab} \partial_a \partial_b.$$

In inertial coordinates,

$$\square = -\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Thus,

$$-\frac{1}{2} \square \bar{\gamma}_{ab} = -\frac{1}{2} \eta^{cd} \partial_c \partial_d \bar{\gamma}_{ab}$$

is simply the Minkowski wave operator acting on $\bar{\gamma}_{ab}$.

3.4.1.6 Linearized Einstein Equation

Einstein's equation to linear order is

$$G_{ab}^{(1)} = 8\pi T_{ab}.$$

Using Eq. 3.80, the linearized Einstein equation becomes

$$\partial^c \partial_c \bar{\gamma}_{ab} = -16\pi T_{ab}, \quad (3.81)$$

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}) .

3.4.2 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) ,

$$u^a = \gamma(1, v^i), \quad \gamma = (1 - |\vec{v}|^2)^{-1/2}.$$

For $|\vec{v}| \ll 1$, we have $\gamma \approx 1$, so

$$u^a \approx (1, v^i), \quad u_a = \eta_{ab}u^b \approx (-1, v_i).$$

Dominant energy component. The time-time component becomes

$$T_{00} = (\rho + P)u_0u_0 + P\eta_{00} \approx \rho(-1)(-1) = \rho,$$

since $P \ll \rho$. This is the dominant term.

Momentum densities. The mixed components satisfy

$$T_{0i} = (\rho + P)u_0u_i + P\eta_{0i} \approx \rho(-1)v_i = -\rho v_i,$$

which are small because $|v_i| \ll 1$.

Spatial stresses. The purely spatial components are

$$T_{ij} = (\rho + P)u_iu_j + P\eta_{ij} \approx \rho v_iv_j + P\delta_{ij}.$$

Both terms are negligible compared to ρ , since $v_iv_j \ll 1$ and $P \ll \rho$.

Resulting form. Therefore,

$$T_{00} \approx \rho, \quad T_{0i} \ll \rho, \quad T_{ij} \ll \rho.$$

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

$$T_{ab} \approx \rho t_a t_b. \tag{3.82}$$

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:

$$\nabla^2 \bar{\gamma}_{\mu\nu} = 0, \quad \mu, \nu \neq 0, \quad (3.83)$$

$$\nabla^2 \bar{\gamma}_{00} = -16\pi\rho. \quad (3.84)$$

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 Eq. 3.84 becomes Poisson's equation:

$$\nabla^2 \phi = 4\pi\rho. \quad (3.85)$$

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,

$$\gamma_{ab} = -(4t_a t_b + 2\eta_{ab})\phi. \quad (3.86)$$

Geodesic Motion and Newton's Second Law

The motion of a freely falling test particle is governed by the geodesic equation, which in general affine form is

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0. \quad (3.87)$$

Here λ is an *affine parameter* along the curve. A parameter is affine if linear reparameterizations $\lambda \rightarrow a\lambda + b$ preserve the form of the equation; nonlinear reparameterizations introduce an additional term proportional to $dx^\mu/d\lambda$.

Earlier, when we considered geodesics in Euclidean space in Eq. 2.38, we chose $\lambda = t$. In that context, t was simply a convenient affine parameter labeling points along the curve; it did not represent a physical time. An affine parameter need not correspond to any physical time coordinate at all. For example, in a Euclidean plane one may choose arc length as the affine parameter.

For timelike curves in spacetime, however, there is a distinguished affine parameter: the proper time τ . It is therefore natural to parameterize the motion of a massive test particle by τ , yielding

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau} = 0. \quad (3.88)$$

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,

$$g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu}, \quad |\gamma_{\mu\nu}| \ll 1,$$

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

$$v^i = \frac{dx^i}{dt}, \quad |v^i| \ll 1.$$

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 $|\vec{v}|$.

In the geodesic equation, the term

$$\Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau}$$

contains products $v^\rho v^\sigma$, which are negligible unless both indices are $\rho = \sigma = 0$. Thus, to leading order,

$$\Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau} \approx \Gamma_{00}^\mu \left(\frac{dx^0}{d\tau} \right)^2 \approx \Gamma_{00}^\mu.$$

Therefore the spatial components of Eq. 3.88 reduce to:

$$\frac{d^2 x^i}{d\tau^2} + \Gamma_{00}^i = 0, \quad (i = 1, 2, 3). \quad (3.89)$$

Using $\tau \approx t$, this becomes

$$\frac{d^2 x^i}{dt^2} = -\Gamma_{00}^i. \quad (3.90)$$

In the linearized theory,

$$g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu},$$

and the Christoffel symbols are

$$\Gamma_{\rho\sigma}^\mu = \frac{1}{2}\eta^{\mu\lambda}(\partial_\rho\gamma_{\sigma\lambda} + \partial_\sigma\gamma_{\rho\lambda} - \partial_\lambda\gamma_{\rho\sigma}).$$

Setting $\rho = \sigma = 0$ and taking $\mu = i$ (a spatial index), we obtain

$$\Gamma_{00}^i = \frac{1}{2}\eta^{i\lambda}(2\partial_0\gamma_{0\lambda} - \partial_\lambda\gamma_{00}).$$

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

$$\Gamma_{00}^i = -\frac{1}{2}\partial_i\gamma_{00}. \quad (3.91)$$

Earlier we showed that in the weak-field limit the metric takes the form

$$\gamma_{00} = -2\phi,$$

so,

$$g_{00} = -1 - 2\phi, \quad (3.92)$$

with ϕ the Newtonian potential.

Substituting $\gamma_{00} = -2\phi$ into Eq. 3.91 gives

$$\Gamma_{00}^i = -\frac{1}{2}\partial_i(-2\phi) = \partial_i\phi. \quad (3.93)$$

Plugging Eq. 3.93 into Eq. 3.90, we obtain

$$\frac{d^2 x^i}{dt^2} = -\partial_i\phi.$$

In vector notation:

$$\vec{a} = \frac{d^2 \vec{x}}{dt^2} = -\vec{\nabla}\phi. \quad (3.94)$$

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 3.85 and 3.94 reproduce the foundational equations of Newtonian gravity:

$$\nabla^2\phi = 4\pi\rho, \quad \vec{a} = -\nabla\phi.$$

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.

3.4.3 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 Eqs. 3.79 and 3.81,

$$\partial^a \bar{\gamma}_{ab} = 0, \tag{3.95}$$

and

$$\partial^c \partial_c \bar{\gamma}_{ab} = 0. \tag{3.96}$$

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.

3.4.3.1 Gauge Freedom in Linearized Gravity

In linearized gravity we write the metric as

$$g_{ab} = \eta_{ab} + \gamma_{ab}, \quad |\gamma_{ab}| \ll 1,$$

and consider infinitesimal coordinate transformations

$$x^a \rightarrow x'^a = x^a + \xi^a(x),$$

with ξ^a small. To first order in ξ^a and γ_{ab} , the metric perturbation transforms as

$$\gamma'_{ab} = \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a. \quad (3.97)$$

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.

3.4.3.2 Lorenz Gauge: First Stage of Simplification

Define the trace-reversed perturbation

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2} \eta_{ab} \gamma, \quad \gamma = \gamma^c_c.$$

The *Lorenz gauge* condition is

$$\partial^a \bar{\gamma}_{ab} = 0. \quad (3.98)$$

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

$$\square \bar{\gamma}_{ab} = -16\pi T_{ab}, \quad \square \equiv \partial^c \partial_c.$$

Gauge freedom remaining in Lorenz gauge. Applying the gauge transformation Eq. 3.97,

$$\begin{aligned} \bar{\gamma}'_{ab} &= \gamma'_{ab} - \frac{1}{2} \eta_{ab} \gamma' \\ &= (\bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2} \eta_{ab} \gamma'^c_c \\ &= (\bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2} \eta_{ab} \eta^{cd} \gamma'_{cd} \\ &= (\bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2} \eta_{ab} \eta^{cd} (\gamma_{cd} + \partial_c \xi_d + \partial_d \xi_c) \\ &= (\bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2} \eta_{ab} (\gamma + \partial^c \xi_c + \partial^d \xi_d) \\ &= (\bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a) - \frac{1}{2} \eta_{ab} (\gamma + 2\partial^c \xi_c) \\ &= (\gamma_{ab} - \frac{1}{2} \eta_{ab} \gamma) + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab} \partial^c \xi_c \end{aligned}$$

so the trace-reversed field transforms as,

$$\bar{\gamma}'_{ab} = \bar{\gamma}_{ab} + \partial_a \xi_b + \partial_b \xi_a - \eta_{ab} \partial^c \xi_c. \quad (3.99)$$

Imposing the Lorenz condition on the transformed field,

$$\begin{aligned} \partial^a \bar{\gamma}'_{ab} &= \partial^a \bar{\gamma}_{ab} + \partial^a \partial_a \xi_b + \partial^a \partial_b \xi_a - \partial^a \eta_{ab} \partial^c \xi_c \\ &= \partial^a \bar{\gamma}_{ab} + \square \xi_b + \partial_b \partial^a \xi_a - \eta_{ab} \partial^a \partial^c \xi_c \\ &= \partial^a \bar{\gamma}_{ab} + \square \xi_b + \partial_b \partial^a \xi_a - \partial_b \partial^c \xi_c \end{aligned}$$

and recognizing that $\partial_b \partial^a \xi_a = \partial_b \partial^c \xi_c$ gives,

$$\partial^a \bar{\gamma}'_{ab} = \partial^a \bar{\gamma}_{ab} + \square \xi_b.$$

Thus the Lorenz condition is preserved if and only if

$$\square \xi_b = 0. \quad (3.100)$$

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 Eq. 3.100. These constitute the *residual gauge freedom*.

This freedom will now be used to impose further gauge conditions.

3.4.3.3 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

$$A_0 = 0, \quad \nabla \cdot \vec{A} = 0.$$

In gravity, the useful analogue is to choose coordinates so that

$$\gamma = 0, \quad \gamma_{0\mu} = 0. \quad (3.101)$$

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 3.101 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

$$2(-\partial_t \xi_0 + \vec{\nabla} \cdot \vec{\xi}) = -\gamma. \quad (3.102)$$

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

$$\partial_t \xi_\mu + \partial_\mu \xi_0 = -\gamma_{0\mu}, \quad \mu = 1, 2, 3. \quad (3.103)$$

Differentiating 3.102 and 3.103 yields second-order elliptic equations for ξ_0 and ξ_μ :

$$2[-\nabla^2 \xi_0 + \vec{\nabla} \cdot (\partial_t \vec{\xi})] = -\partial_t \gamma, \quad (3.104)$$

$$\nabla^2 \xi_\mu + \partial_\mu (\partial_t \xi_0) = -\partial_t \gamma_{0\mu}. \quad (3.105)$$

Equations 3.102–3.105 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

$$\square \xi_a = 0,$$

so ξ^a with the initial data determined by 3.102–3.105 is evolved throughout the vacuum region by solving the wave equation.

Final result. With this choice of ξ^a , the gauge-transformed field satisfies:

$$\partial^a \bar{\gamma}'_{ab} = 0, \quad \gamma' = 0, \quad \gamma'_{0\mu} = 0.$$

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 $A_a \rightarrow A_a + \partial_a \chi$. 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, $\gamma_{ab} \rightarrow \gamma_{ab} + \partial_a \xi_b + \partial_b \xi_a$ 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,

$$\square \bar{\gamma}_{ab} = -16\pi T_{ab}, \quad (3.106)$$

and restricting to a vacuum region where $T_{ab} = 0$, we obtain the source-free wave equation

$$\square \bar{\gamma}_{ab} = 0. \quad (3.107)$$

In the radiation gauge we impose (cf. the construction leading to Eqs. 3.102–3.105)

$$\gamma = 0, \quad \gamma_{0\mu} = 0.$$

When the trace vanishes, the trace-reversed perturbation reduces to the original one. Using the trace-reversal definition,

$$\bar{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma, \quad (3.108)$$

we have, in radiation gauge,

$$\bar{\gamma}_{ab} = \gamma_{ab}. \quad (3.109)$$

Thus in this gauge the vacuum equation 3.107 becomes

$$\square \gamma_{ab} = 0. \quad (3.110)$$

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

$$\partial^c \partial_c \gamma_{\mu\nu} = 0, \quad (\mu, \nu = 1, 2, 3), \quad (3.111)$$

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

$$\square \bar{\gamma}_{ab} = 0,$$

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,

$$\gamma = 0, \quad \gamma_{0\mu} = 0,$$

leaving only the transverse, traceless spatial components γ_{ij}^{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

$$\partial^c \partial_c \gamma_{\mu\nu} = 0, \quad (\mu, \nu = 1, 2, 3). \quad (3.112)$$

We look for plane-wave solutions of the form

$$\gamma_{\mu\nu}(x) = \Re \left\{ A_{\mu\nu} e^{ik_a x^a} \right\}, \quad (3.113)$$

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.

$$\partial_c \tilde{\gamma}_{\mu\nu} = \partial_c (A_{\mu\nu} e^{ik_a x^a}) = A_{\mu\nu} \partial_c e^{ik_a x^a} = ik_c A_{\mu\nu} e^{ik_a x^a}. \quad (3.114)$$

Second derivative. Applying $\partial^c = \eta^{cd} \partial_d$ gives

$$\partial^c \partial_c \tilde{\gamma}_{\mu\nu} = \eta^{cd} \partial_d (ik_c A_{\mu\nu} e^{ik_a x^a}) = i \eta^{cd} k_c A_{\mu\nu} \partial_d e^{ik_a x^a} \quad (3.115)$$

$$= i \eta^{cd} k_c A_{\mu\nu} (ik_d e^{ik_a x^a}) \quad (3.116)$$

$$= -\eta^{cd} k_c k_d A_{\mu\nu} e^{ik_a x^a} \quad (3.117)$$

$$= -k_a k^a A_{\mu\nu} e^{ik_b x^b}. \quad (3.118)$$

Substituting into the wave equation. Eq. 3.112 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

$$k_a k^a = 0. \quad (3.119)$$

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

$$\gamma = 0, \quad \gamma_{0\mu} = 0,$$

and we now determine what these conditions imply for the polarization tensor of a plane-wave solution.

Using the same plane-wave ansatz above,

$$\gamma_{\mu\nu}(x) = A_{\mu\nu} e^{ik_a x^a},$$

with constant symmetric $A_{\mu\nu}$ and constant wavevector k_a , substituting 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

$$A^\mu{}_\mu = 0. \quad (3.120)$$

(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

$$A_{0\mu} = 0. \quad (3.121)$$

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

$$k^a A_{ab} = 0. \quad (3.122)$$

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

$$k^a = (\omega, 0, 0, \omega)$$

with $k_a k^a = 0$ from Eq. 3.119. Applying the transversality condition gives

$$k^a A_{a\nu} = \omega (A_{0\nu} + A_{3\nu}) = 0.$$

Using Eq. 3.121, $A_{0\nu} = 0$, this reduces to

$$A_{3\nu} = 0. \quad (3.123)$$

Resulting form of the polarization tensor. The conditions 3.120–3.123 imply that

$$A_{0\mu} = 0, \quad A_{3\mu} = 0,$$

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.

Counting independent components. The symmetric 3×3 tensor $A_{\mu\nu}$ initially has six independent components. The transversality condition Eq. 3.122 imposes three independent constraints. The traceless condition Eq. 3.120 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 $k^a A_{a\nu} = 0$ 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

$$\varepsilon_{\mu\nu}^{(+)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \varepsilon_{\mu\nu}^{(\times)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

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

$$A_{0\mu} = 0, \quad A_{3\mu} = 0, \quad A^\mu{}_\mu = 0.$$

These are precisely the *transverse* and *traceless* conditions defining the TT gauge. Consequently, the spatial components of the plane-wave perturbation are

$$\gamma_{ij}(t, z) = A_{ij} e^{i\omega(t-z)}, \quad i, j = 1, 2,$$

and we may identify

$$\gamma_{ij}^{\text{TT}}(t, z) \equiv \gamma_{ij}(t, z) = A_{ij} e^{i\omega(t-z)}.$$

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

$$\gamma_{0\mu} = 0, \quad \gamma = 0, \quad k^i A_{ij} = 0,$$

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

$$\gamma_{ij}^{\text{TT}}(x) = A_{ij} e^{ik \cdot x}.$$

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

$$\gamma_{ij}^{\text{TT}}(t) = \begin{pmatrix} h_+(t) & h_\times(t) & 0 \\ h_\times(t) & -h_+(t) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

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.

3.4.3.4 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:

$$u^a \approx t^a = (1, 0, 0, 0), \quad X^a = (0, X^i),$$

and proper time τ can be identified with coordinate time t .

With these choices, the spatial components of geodesic deviation reduce to

$$\frac{d^2 X^i}{dt^2} = -R^i{}_{0j0} X^j. \quad (3.124)$$

Riemann tensor in the radiation gauge. In linearized gravity, the Riemann tensor to first order in γ_{ab} is

$$R_{abcd}^{(1)} = \frac{1}{2}(\partial_c \partial_b \gamma_{ad} + \partial_d \partial_a \gamma_{bc} - \partial_c \partial_a \gamma_{bd} - \partial_d \partial_b \gamma_{ac}).$$

In the radiation gauge for a wave in vacuum we have

$$\gamma_{00} = 0, \quad \gamma_{0i} = 0, \quad \gamma_{ij} = \gamma_{ij}^{\text{TT}}(t, z),$$

with γ_{ij}^{TT} transverse and traceless.

Taking $a = 0$, $b = i$, $c = 0$, $d = j$, we obtain

$$\begin{aligned} R_{0i0j}^{(1)} &= \frac{1}{2}(\partial_0 \partial_i \gamma_{0j} + \partial_j \partial_0 \gamma_{0i} - \partial_0 \partial_0 \gamma_{ij} - \partial_j \partial_i \gamma_{00}) \\ &= -\frac{1}{2} \partial_0^2 \gamma_{ij}, \end{aligned}$$

since $\gamma_{00} = \gamma_{0i} = 0$ in radiation gauge. Thus

$$R_{0i0j}^{(1)} = -\frac{1}{2} \ddot{\gamma}_{ij},$$

where dots denote derivatives with respect to t at fixed detector position (approximately fixed z in the local detector frame).

Raising the first index with η^{ik} , we have

$$R^{(1)i}{}_{0j0} = \eta^{ik} R_{k0j0}^{(1)} = -\frac{1}{2} \ddot{\gamma}^i{}_j.$$

Relative acceleration of test particles. Substituting this into the geodesic deviation Eq. 3.124 gives

$$\frac{d^2 X^i}{dt^2} = \frac{1}{2} \ddot{\gamma}^i{}_j X^j. \quad (3.125)$$

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 γ_{ij}^{TT} produce the characteristic alternating stretching and squeezing patterns associated with gravitational waves.

3.4.3.5 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

$$\gamma_{ij}^{\text{TT}}(t-z) = h_+(t-z) \varepsilon_{ij}^{(+)} + h_\times(t-z) \varepsilon_{ij}^{(\times)},$$

with polarization tensors

$$\varepsilon_{ij}^{(+)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \varepsilon_{ij}^{(\times)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Because TT gauge is constructed so that idealized free test masses 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 Eq. 3.125 gives

$$\frac{d^2 X^i}{dt^2} = \frac{1}{2} \ddot{\gamma}^{\text{TT} i}_{\phantom{\text{TT}} j} X^j.$$

Integrating twice (and assuming the wave amplitude is small and the initial velocity is zero) yields

$$\delta X^i(t) = \frac{1}{2} \gamma^{\text{TT} i}_{\phantom{\text{TT}} j}(t) X_0^j,$$

where X_0^j is the unperturbed separation. This uses initial conditions chosen so that the integration constants vanish (e.g. at $t = t_0$, $\delta X^i(t_0) = 0$, $\dot{\delta X}^i(t_0) = 0$, and $\gamma_{ij}^{\text{TT}}(t_0) = 0$).

Effect on detector arms. Take an interferometer with arms of coordinate lengths

$$L_x = L, \quad L_y = L,$$

lying along the x - and y -axes. The proper length of the x -arm becomes

$$L_x(t) = L + \delta L_x(t) = L + \frac{1}{2} \gamma_{xx}^{\text{TT}}(t) L,$$

and similarly for the y -arm.

For a $+$ -polarized wave,

$$\gamma_{xx}^{\text{TT}} = h_+, \quad \gamma_{yy}^{\text{TT}} = -h_+,$$

so the arm lengths change in opposite directions:

$$\frac{\delta L_x}{L} = \frac{1}{2} h_+, \quad \frac{\delta L_y}{L} = -\frac{1}{2} h_+.$$

Detected strain. Interferometers measure the *difference* in fractional arm-length changes:

$$h(t) = \frac{\delta L_x}{L} - \frac{\delta L_y}{L} = h_+(t).$$

for normal incidence and for detector arms aligned with the $+$ polarization axes.

Thus the measurable strain of a detector is directly the amplitude of the transverse–traceless metric perturbation:

$$h(t) = \gamma_{xx}^{\text{TT}}(t) - \gamma_{yy}^{\text{TT}}(t).$$

More generally, for arbitrary sky location and polarization angle, detector response takes the form

$$h(t) = F_+ h_+(t) + F_\times h_\times(t),$$

with antenna factors F_+ and $F_\times$ (for this idealized orientation, $F_+ = 1$ and $F_\times = 0$).

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.

3.4.4 Gravitational Radiation (Sourced)

In the previous sections we analyzed vacuum solutions of the linearized Einstein equation, $\square \bar{\gamma}_{ab} = 0$, 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

$$\square \bar{\gamma}_{ab} = -16\pi T_{ab}$$

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

$$\psi(t, \vec{x}) = \int d^3x' \frac{f(t_{\text{ret}}, \vec{x}')}{|\vec{x} - \vec{x}'|}, \quad t_{\text{ret}} = t - |\vec{x} - \vec{x}'|.$$

This follows from the retarded Green’s function in flat spacetime,

$$G_{\text{ret}}(x, x') = \frac{\delta(t' - t + |\vec{x} - \vec{x}'|)}{|\vec{x} - \vec{x}'|}.$$

Since each component $\bar{\gamma}_{ab}$ satisfies the same wave equation Eq. 3.106 with source $-16\pi T_{ab}$, we may write the solution by direct analogy:

$$\bar{\gamma}_{ab}(t, \vec{x}) = 4 \int d^3x' \frac{T_{ab}(t_{\text{ret}}, \vec{x}')}{|\vec{x} - \vec{x}'|}, \quad t_{\text{ret}} = t - |\vec{x} - \vec{x}'|. \quad (3.126)$$

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 the textbook's Eq. (4.4.44), we Fourier transform the retarded solution Eq. 3.126 in time only. Insert Eq. 3.126 into the Fourier transform:

$$\hat{\gamma}_{\mu\nu}(\omega, \vec{x}) = \frac{4}{\sqrt{2\pi}} \int d^3x' \frac{1}{|\vec{x} - \vec{x}'|} \int_{-\infty}^{\infty} dt T_{\mu\nu}(t - |\vec{x} - \vec{x}'|, \vec{x}') e^{i\omega t}.$$

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 the textbook's eq. (4.4.44):

$$\hat{\gamma}_{\mu\nu}(\omega, \vec{x}) = 4 \int d^3x' \frac{\hat{T}_{\mu\nu}(\omega, \vec{x}')}{|\vec{x} - \vec{x}'|} e^{i\omega|\vec{x} - \vec{x}'|}. \quad (3.127)$$

Gauge relation for the time-space components. Using the linearized conservation law $\partial^a T_{ab} = 0$ together with the Lorenz gauge condition, we obtain

$$-i\omega \hat{\gamma}_{0\mu} = \sum_{\nu=1}^3 \frac{\partial}{\partial x^\nu} \hat{\gamma}_{\nu\mu}. \quad (3.128)$$

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 Eq. 3.126 in the regime of *slow-moving, spatially compact* sources,

$$|\vec{v}| \ll 1, \quad \text{source size } L \ll \text{radiation distance } R.$$

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 the textbook's relation

$$\int \hat{T}^{\mu\nu} d^3x = -\frac{\omega^2}{2} \int \hat{T}^{00} x^\mu x^\nu d^3x,$$

showing how it follows from stress–energy conservation and integration by parts.

We begin from local conservation:

$$\partial_a T^{ab} = 0 \Rightarrow \partial_t T^{0b} + \partial_k T^{kb} = 0,$$

where Latin spatial indices $k = 1, 2, 3$ and $b = 0, 1, 2, 3$.

Set $b = 0$, multiply by $x^\mu x^\nu$, and integrate over space:

$$\int d^3x x^\mu x^\nu (\partial_t T^{00} + \partial_k T^{k0}) = 0.$$

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:

$$\int d^3x x^\mu x^\nu \partial_k T^{k0} = \int d^3x \partial_k (x^\mu x^\nu T^{k0}) - \int d^3x (\partial_k x^\mu x^\nu) T^{k0}.$$

The divergence term is a surface integral at infinity and vanishes if T^{k0} has compact support or decays fast enough. Using $\partial_k (x^\mu x^\nu) = \delta_k^\mu x^\nu + \delta_k^\nu x^\mu$, we obtain

$$\int d^3x x^\mu x^\nu \partial_k T^{k0} = - \int d^3x (x^\nu T^{\mu 0} + x^\mu T^{\nu 0}).$$

The conservation law therefore gives

$$\frac{d}{dt} \left(\int d^3x x^\mu x^\nu T^{00} \right) = \int d^3x (x^\nu T^{\mu 0} + x^\mu T^{\nu 0}).$$

Differentiate Eq. 3.4.4 with respect to time:

$$\frac{d^2}{dt^2} \left(\int d^3x x^\mu x^\nu T^{00} \right) = \int d^3x (x^\nu \partial_t T^{\mu 0} + x^\mu \partial_t T^{\nu 0}).$$

Now use conservation again, this time with $b = \mu$ and $b = \nu$:

$$\partial_t T^{\mu 0} = -\partial_k T^{k\mu}, \quad \partial_t T^{\nu 0} = -\partial_k T^{k\nu}.$$

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,

$$\int d^3x T^{\mu\nu}(t, \vec{x}) = \frac{1}{2} \frac{d^2}{dt^2} \left(\int d^3x x^\mu x^\nu T^{00}(t, \vec{x}) \right). \quad (3.129)$$

Now pass to Fourier space. Define the temporal Fourier transform

$$\hat{T}^{ab}(\omega, \vec{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} T^{ab}(t, \vec{x}) e^{i\omega t} dt.$$

Fourier transforming Eq. 3.129 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 the textbook's eq. (4.4.46):

$$\int \hat{T}^{\mu\nu} d^3x = -\frac{\omega^2}{2} \int \hat{T}^{00} x^\mu x^\nu d^3x. \quad (3.130)$$

Thus Eq. 3.130 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 Eq. 3.130 into Eq. 3.127, the textbook arrives at the far-zone field

$$\hat{\gamma}_{\mu\nu}(\omega, \vec{x}) = -\frac{2\omega^2}{3R} e^{i\omega R} \hat{q}_{\mu\nu}(\omega), \quad (\mu, \nu = 1, 2, 3), \quad (3.131)$$

where $\hat{q}_{\mu\nu}$ is the Fourier transform of the mass quadrupole moment tensor.

The quadrupole moment. Define the quadrupole moment by

$$q_{\mu\nu}(t) = 3 \int T^{00}(t, \vec{x}) x_\mu x_\nu d^3x. \quad (3.132)$$

Inverse Fourier transform: the time-domain waveform. Taking the inverse transform of 3.131, the textbook obtains the quadrupole formula for the spatial components of the metric perturbation:

$$\bar{\gamma}_{\mu\nu}(t, \vec{x}) = \frac{2}{3R} \left. \frac{d^2 q_{\mu\nu}}{dt^2} \right|_{t'=t-R}, \quad (\mu, \nu = 1, 2, 3). \quad (3.133)$$

This is the textbook'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.*

3.4.4.1 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, γ_{ij}^{TT} .

Averaging procedure. Define the effective stress–energy tensor by

$$t_{\mu\nu}^{\text{GW}} = \frac{1}{32\pi} \left\langle \partial_\mu \gamma_{ij}^{\text{TT}} \partial_\nu \gamma_{ij}^{\text{TT}} \right\rangle, \quad (3.134)$$

where $\langle \dots \rangle$ denotes an average over several wavelengths. This is a standard post-textbook extension and introduces no new equations in the textbook's numbering.

Energy flux. For outward-propagating radiation in direction $\hat{n}$,

$$\frac{dE}{dA dt} = t_{0i}^{\text{GW}} n^i.$$

In the far zone, the TT field has the form

$$\gamma_{ij}^{\text{TT}}(t, r, \hat{n}) = \frac{1}{r} A_{ij}(t - r),$$

implying

$$\partial_t \gamma_{ij}^{\text{TT}} = -\partial_r \gamma_{ij}^{\text{TT}} = \frac{1}{r} \dot{A}_{ij}(t - r).$$

Thus the radial flux becomes

$$t_{0r}^{\text{GW}} = \frac{1}{32\pi r^2} \left\langle \dot{A}_{ij}(t - r) \dot{A}_{ij}(t - r) \right\rangle. \quad (3.135)$$

Total radiated power. Integrating over a sphere of radius r gives

$$P_{\text{GW}} \equiv \frac{dE_{\text{GW}}}{dt} = \frac{1}{32\pi} \int d\Omega \left\langle \dot{A}_{ij}(t-r) \dot{A}_{ij}(t-r) \right\rangle. \quad (3.136)$$

Relation to the quadrupole moment. Using the textbook's waveform (4.4.55),

$$\gamma_{ij}(t, \vec{x}) = \frac{2}{R} \ddot{Q}_{ij}(t-R),$$

the TT part is

$$\gamma_{ij}^{\text{TT}} = \frac{2}{R} \ddot{Q}_{ij}^{\text{TT}}(t-R),$$

so the amplitude is

$$A_{ij}(t) = 2 \ddot{Q}_{ij}^{\text{TT}}(t), \quad \dot{A}_{ij}(t) = 2 \dddot{Q}_{ij}^{\text{TT}}(t).$$

Substituting into Eq. 3.136 yields

$$P_{\text{GW}} = \frac{1}{8\pi} \int d\Omega \left\langle \ddot{Q}_{ij}^{\text{TT}} \ddot{Q}_{ij}^{\text{TT}} \right\rangle. \quad (3.137)$$

Using the standard angular identity

$$\int d\Omega \ddot{Q}_{ij}^{\text{TT}} \ddot{Q}_{ij}^{\text{TT}} = \frac{4\pi}{5} \ddot{Q}_{ij} \ddot{Q}_{ij},$$

we obtain the standard quadrupole power formula:

$$\frac{dE_{\text{GW}}}{dt} = \frac{1}{5} \left\langle \ddot{Q}_{ij} \ddot{Q}_{ij} \right\rangle. \quad (3.138)$$

Accordingly, the source loses energy at the rate

$$\frac{dE_{\text{source}}}{dt} = -\frac{dE_{\text{GW}}}{dt} = -\frac{1}{5} \left\langle \ddot{Q}_{ij} \ddot{Q}_{ij} \right\rangle.$$

This expression is *not* in the textbook; it is a standard modern extension consistent with the textbook's notation and conventions.

Orders of Magnitude and Astrophysical Sources

The quadrupole power formula,

$$\frac{dE_{\text{source}}}{dt} = -\frac{1}{5} \left\langle \ddot{Q}_{ij} \ddot{Q}_{ij} \right\rangle,$$

and the radiative waveform,

$$\gamma_{ij}^{\text{TT}}(t, \vec{x}) = \frac{2}{R} \ddot{Q}_{ij}^{\text{TT}}(t-R),$$

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

$$h \sim \frac{\ddot{Q}}{R} \sim \frac{Mv^2}{R}.$$

For a Keplerian binary with separation L ,

$$v^2 \sim \frac{M}{L},$$

which gives

$$h \sim \frac{M^2}{LR}.$$

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

$$\ddot{Q} \sim \omega^3 Q \sim \frac{v^3}{L^3} ML^2 \sim M \frac{v^3}{L}.$$

Hence

$$\ddot{Q}_{ij} \ddot{Q}_{ij} \sim M^2 \frac{v^6}{L^2},$$

and the power radiated is

$$\frac{dE_{\text{source}}}{dt} \sim -M^2 \frac{v^6}{L^2}.$$

Using again $v^2 \sim M/L$,

$$\frac{dE_{\text{source}}}{dt} \sim -\frac{M^5}{L^5}.$$

Restoring factors of G and c , one finds

$$\frac{dE_{\text{source}}}{dt} \sim -\frac{G}{c^5} M^2 L^4 \omega^6 \sim -\frac{G^4}{c^5} \frac{M^5}{L^5}.$$

Equivalently, the emitted gravitational-wave power is $P_{\text{GW}} = -dE_{\text{source}}/dt > 0$.

The factor G/c^5 suppresses radiation from all laboratory-sized sources.

Laboratory sources are negligible. For a system with

$$M \sim 10^3 \text{ kg}, \quad L \sim 1 \text{ m}, \quad v \ll 1,$$

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

$$M \sim M_\odot, \quad L \sim \frac{GM}{c^2}.$$

This corresponds to an order-of-magnitude compact, near-relativistic estimate rather than a universal scaling for all binaries. Then

$$v^2 \sim \frac{GM}{Lc^2} \sim 1,$$

and at distance $R \sim 100\text{--}500$ 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.

These scaling estimates are tested directly by precision timing of compact binary pulsars.

Observational Confirmation: The Hulse–Taylor Binary Pulsar

The quadrupole power formula,

$$\frac{dE}{dt} = -\frac{1}{5} \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle,$$

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

$$\dot{P}_{\text{GR}} = -\frac{192\pi}{5} \left(\frac{2\pi G}{c^3} \right)^{5/3} \frac{(m_1 m_2) (m_1 + m_2)^{-1/3}}{a^{5/3} (1 - e^2)^{7/2}} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right), \quad (3.139)$$

a direct consequence of the quadrupole formula applied to an eccentric binary.

Observations. Long-term timing observations yield the measured decay,

$$\dot{P}_{\text{obs}},$$

and the result agrees with the general-relativistic prediction to within better than 0.5%:

$$\dot{P}_{\text{obs}} \approx \dot{P}_{\text{GR}}.$$

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 4

The Schwarzschild Solution

WORK IN PROGRESS

Here we follow Carroll.

4.1 Derivation of the Schwarzschild Metric

We seek the exterior gravitational field of a static, spherically symmetric, isolated body. In the vacuum region outside matter, Einstein's equation reduces to

$$R_{\mu\nu} = 0. \quad (4.1)$$

Here by static, we mean that our source is not evolving. For now, we will interpret static to mean that all metric components are independent of the time coordinate, and that there are no time-space cross terms in the metric.

4.1.1 General Static, Spherically Symmetric Ansatz

Lets start by considering the metric of Minkowski space in polar coordinates $x^a = (t, r, \theta, \phi)$,

$$ds_{Minkowski}^2 = -dt^2 + dr^2 + r^2 d\Omega^2 \quad (4.2)$$

where $d\Omega^2$ is the metric on a unit two-sphere,

$$d\Omega^2 = d\theta^2 + \sin^2 \theta d\varphi^2 \quad (4.3)$$

To preserve spherical symmetry, Eq. 4.3 must be preserved, but otherwise we can multiply all of the other terms by separate coefficients as long as they are only functions of r – basically, you can

make the sphere bigger, but it still needs to be perfectly round. The most general static, spherically symmetric metric can therefore be written as

$$ds^2 = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + e^{2\gamma(r)}r^2d\Omega^2. \quad (4.4)$$

Here r is the areal radius, so $4\pi r^2$ is the area of the symmetry two-spheres.

Here we chose exponential functions of r so that the signature of the metric doesn't change. At this stage it is useful to note that Wald begins with a slightly more general static, spherically symmetric line element, allowing more coefficient functions than we have kept in our ansatz. This does not lead to a different exterior geometry; it mainly postpones coordinate choices. As in earlier chapters, different-looking metric components can represent the same spacetime when they are related by a smooth, invertible relabeling of coordinates. So Wald's route is more general in presentation, while our route uses symmetry and coordinate freedom earlier to reduce algebra. Both approaches solve the same vacuum equations $R_{\mu\nu} = 0$ and therefore land on the same Schwarzschild family (up to coordinate conventions and parameter naming).

We can make our lives easier by changing coordinates before we apply Einstein's equation. The nice thing about general relativity, is that we simultaneously define coordinates and the metric as a function of those coordinates. So let's choose a coordinate change that will cancel out with the coefficient on the last term in the metric 4.4,

$$\bar{r} = e^{\gamma(r)}r, \quad (4.5)$$

The relabeling 4.5 is a coordinate redefinition, not a new physical assumption. From earlier chapters: coordinates are labels, and tensor equations are invariant under smooth invertible changes of labels (diffeomorphisms). Here we use that freedom to put the angular sector into the standard form $\bar{r}^2 d\Omega^2$, so that $\bar{r}$ is the areal radius (spheres at constant $\bar{r}$ have area $4\pi\bar{r}^2$). This is the same idea used previously when choosing convenient charts: we simplify the metric's appearance without changing the underlying geometry.

When we write basis one-forms such as $dt, dr, d\theta, d\phi$, we are using the dual basis introduced in Chapter 1. A one-form is a linear map that eats a vector and returns a number; e.g. $dt(\partial_t) = 1$, $dt(\partial_r) = 0$, etc. The metric can then be read as a bilinear form built from these basis one-forms, with coefficients that encode geometry. So writing ds^2 in this basis is just the covector-language version of the same tensor object we have been using throughout. The relabeling 4.5 has an associated basis one-form,

$$d\bar{r} = e^{\gamma}dr + e^{\gamma}r d\gamma = \left(1 + r \frac{d\gamma}{dr}\right) e^{\gamma}dr \quad (4.6)$$

or

$$dr^2 = \left(1 + r \frac{d\gamma}{dr}\right)^{-2} e^{-2\gamma} d\bar{r}^2. \quad (4.7)$$

Now the metric 4.4 becomes,

$$ds^2 = -e^{2\alpha(r)} dt^2 + \left(1 + r \frac{d\gamma}{dr}\right)^{-2} e^{2\beta(r)-2\gamma(r)} d\bar{r}^2 + \bar{r}^2 d\Omega^2. \quad (4.8)$$

Lets relabel:

$$\bar{r} \rightarrow r \quad (4.9)$$

$$\left(1 + r \frac{d\gamma}{dr}\right)^{-2} e^{2\beta(r)-2\gamma(r)} \rightarrow e^{2\beta}. \quad (4.10)$$

What sorcery is this? How can we justify doing this? We will repeatedly relabel coordinates and absorb functions into new symbols. This is the same bookkeeping strategy used earlier: fix symmetry first, then spend coordinate freedom to remove redundant functions, then solve the field equations for the genuinely invariant content. The relabeling does not change observables; it only chooses a cleaner chart. In that sense, our “shortcut” is not a shortcut in physics, only in algebra. Our metric 4.8 becomes,

$$ds^2 = -e^{2\alpha(r)} dt^2 + e^{2\beta(r)} dr^2 + r^2 d\Omega^2. \quad (4.11)$$

4.1.2 Connection and Curvature Components

Christoffel Symbols

We begin by evaluating Christoffel symbols. Recall Eq. 2.18 for Christoffel symbols,

$$\Gamma_{\mu\nu}^{\sigma} = \frac{1}{2} g^{\sigma\rho} (\partial_{\mu} g_{\nu\rho} + \partial_{\nu} g_{\rho\mu} - \partial_{\rho} g_{\mu\nu})$$

We use labels, (t, r, θ, ϕ) for $(0, 1, 2, 3)$ so the only non-zero metric coefficients are,

$$\begin{aligned} g_{tt} &= -e^{2\alpha(r)} \\ g_{rr} &= e^{2\beta(r)} \\ g_{\theta\theta} &= r^2 \\ g_{\phi\phi} &= r^2 \sin^2 \theta \end{aligned}$$

which means that our Christoffel symbols start at,

$$\begin{aligned}
\Gamma_{\mu\nu}^t &= \frac{1}{2} g^{tt} (\partial_\mu g_{\nu t} + \partial_\nu g_{t\mu} - \partial_t g_{\mu\nu}) \\
\Gamma_{\mu\nu}^r &= \frac{1}{2} g^{rr} (\partial_\mu g_{\nu r} + \partial_\nu g_{r\mu} - \partial_r g_{\mu\nu}) \\
\Gamma_{\mu\nu}^\theta &= \frac{1}{2} g^{\theta\theta} (\partial_\mu g_{\nu\theta} + \partial_\nu g_{\theta\mu} - \partial_\theta g_{\mu\nu}) \\
\Gamma_{\mu\nu}^\phi &= \frac{1}{2} g^{\phi\phi} (\partial_\mu g_{\nu\phi} + \partial_\nu g_{\phi\mu} - \partial_\phi g_{\mu\nu})
\end{aligned}$$

We shall do a few just to demonstrate and leave the rest up to the reader. Take the first one, $\Gamma_{\mu\nu}^t$. The last term, $\partial_t g_{\mu\nu}$ will be zero as none of the metric coefficients are functions of t so we can just consider it to be zero. The only non-zero metric coefficient involving t is g_{tt} so if we choose $\mu = t$ then ν has to either be r, θ , or ϕ , all of which when combined with $g_{t\mu}$ create metric coefficients that are zero, therefore

$$\Gamma_{t\nu}^t = \frac{1}{2} g^{tt} \partial_\nu g_{tt}$$

$\nu = r$ is the only coordinate whose derivative would survive so,

$$\Gamma_{tr}^t = \frac{1}{2} g^{tt} \partial_r g_{tt} = \frac{1}{2} e^{-2\alpha(r)} \partial_r (e^{2\alpha(r)}) = \frac{1}{2} e^{-2\alpha(r)} (2e^{2\alpha(r)} \partial_r \alpha) = \partial_r \alpha$$

The second Christoffel symbol shown above is $\Gamma_{\mu\nu}^r$:

$$\Gamma_{\mu\nu}^r = \frac{1}{2} g^{rr} (\partial_\mu g_{\nu r} + \partial_\nu g_{r\mu} - \partial_r g_{\mu\nu})$$

The only non-zero Christoffel symbols for $\Gamma_{\mu\nu}^r$ that are non-zero for metric 4.11 are:

$$\Gamma_{rr}^r, \Gamma_{tt}^r, \Gamma_{\theta\theta}^r, \text{ and } \Gamma_{\phi\phi}^r.$$

Let's do a couple of them,

$$\Gamma_{rr}^r = \frac{1}{2} g^{rr} (\partial_r g_{rr} + \partial_r g_{rr} - \partial_r g_{rr}) = \frac{1}{2} g^{rr} \partial_r g_{rr} = \frac{1}{2} e^{-2\beta} \partial_r e^{2\beta} = \partial_r \beta$$

$$\Gamma_{\phi\phi}^r = \frac{1}{2} g^{rr} (\partial_\phi g_{\phi r} + \partial_\phi g_{r\phi} - \partial_r g_{\phi\phi}) = -\frac{1}{2} g^{rr} \partial_r g_{\phi\phi} = -\frac{1}{2} e^{-2\beta} \partial_r (r^2 \sin^2 \theta) = -r e^{-2\beta} \sin^2 \theta$$

The rest of the Christoffel symbols should be computed by the reader. In total:

$$\begin{aligned}
\Gamma_{tt}^r &= e^{2(\alpha-\beta)} \partial_r \alpha & \Gamma_{\theta\theta}^r &= -r e^{-2\beta} & \Gamma_{r\phi}^\phi &= \frac{1}{r} \\
\Gamma_{tr}^t &= \partial_r \alpha & \Gamma_{\phi\phi}^r &= -r \sin^2 \theta e^{-2\beta} & \Gamma_{\theta\phi}^\phi &= \cot \theta \\
\Gamma_{rr}^r &= \partial_r \beta & \Gamma_{r\theta}^\theta &= \frac{1}{r} & \Gamma_{\phi\phi}^\theta &= -\sin \theta \cos \theta
\end{aligned} \tag{4.12}$$

Riemann Tensor Components

Recall the Riemann tensor in coordinate form from Eq. 2.50,

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

The using the Christoffel symbols above, we can construct the Riemann tensor components. The only nonvanishing components of the Riemann tensor:

$$\begin{aligned}
R^t{}_{rtr} &= \partial_r \alpha \partial_r \beta - \partial_r^2 \alpha - (\partial_r \alpha)^2 \\
R^t{}_{\theta t \theta} &= -r e^{-2\beta} \partial_r \alpha \\
R^t{}_{\phi t \phi} &= -r e^{2\beta} \sin^2 \theta \partial_r \alpha \\
R^r{}_{\theta r \theta} &= r e^{-2\beta} \partial_r \beta \\
R^r{}_{\phi r \phi} &= r e^{-2\beta} \sin^2 \theta \partial_r \beta \\
R^\theta{}_{\phi \theta \phi} &= (1 - e^{-2\beta}) \sin^2 \theta
\end{aligned}$$

Ricci Tensor Components

The Ricci tensor was given in Eq. 2.31 as the contraction of the Riemann tensor:

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

Consider $\mu = \nu = t$,

$$R_{tt} = R^\lambda{}_{t\lambda t} = R^r{}_{trt} + R^\theta{}_{t\theta t} + R^\phi{}_{t\phi t}$$

where $R^t{}_{ttt} = 0$ by antisymmetry. We can get a non-vanishing Riemann tensor component using raising and lowering,

$$\begin{aligned}
R^r{}_{trt} &= g^{rr} R_{rtrt} = g^{rr} R_{trtr} = g^{rr} g_{tt} R^t{}_{rtr} = -e^{2(\alpha-\beta)} R^t{}_{rtr} \\
&= -e^{2(\alpha-\beta)} [\partial_r \alpha \partial_r \beta - \partial_r^2 \alpha - (\partial_r \alpha)^2].
\end{aligned}$$

Likewise,

$$R^\theta{}_{t\theta t} = g^{\theta\theta} R_{\theta t \theta t} = g^{\theta\theta} R_{t\theta t \theta} = g^{\theta\theta} g_{tt} R^t{}_{\theta t \theta} = -\frac{1}{r^2} e^{2\alpha} R^t{}_{\theta t \theta} = \frac{1}{r} e^{2(\alpha-\beta)} \partial_r \alpha,$$

and

$$R^\phi{}_{t\phi t} = g^{\phi\phi} R_{\phi t \phi t} = g^{\phi\phi} R_{t\phi t \phi} = g^{\phi\phi} g_{tt} R^t{}_{\phi t \phi} = -\frac{1}{r^2 \sin^2 \theta} e^{2\alpha} R^t{}_{\phi t \phi} = \frac{1}{r} e^{2(\alpha-\beta)} \partial_r \alpha.$$

So,

$$R_{tt} = e^{2(\alpha-\beta)}[\partial_r^2 \alpha + (\partial_r \alpha)^2 - \partial_r \alpha \partial_r \beta + \frac{2}{r} \partial_r \alpha].$$

The remaining non-vanishing Ricci tensors are,

$$\begin{aligned} R_{rr} &= -\partial_r^2 \alpha - (\partial_r \alpha)^2 + \partial_r \alpha \partial_r \beta + \frac{2}{r} \partial_r \beta \\ R_{\theta\theta} &= e^{-2\beta}[r(\partial_r \beta - \partial_r \alpha) - 1] + 1 \\ R_{\phi\phi} &= \sin^2 \theta R_{\theta\theta} \end{aligned}$$

The scalar curvature, R , is given in Eq. 2.32 as,

$$R = R^\mu{}_\mu = g^{\mu\nu} R_{\mu\nu}. \quad (4.13)$$

so using our non-vanishing Ricci tensors,

$$R = -2e^{-2\beta}[\partial_r^2 \alpha + (\partial_r \alpha)^2 - \partial_r \alpha \partial_r \beta + \frac{2}{r}(\partial_r \alpha - \partial_r \beta) + \frac{1}{r^2}(1 - e^{2\beta})] \quad (4.14)$$

4.1.3 Solving the Vacuum Equations

Now we are ready to impose Eq. 4.1 on the Ricci tensor – so we set it to zero. However, notice that R_{tt} and R_{rr} do not have any cross-terms? This means that they vanish independently,

$$\begin{aligned} R_{tt} + R_{rr} &= e^{2(\alpha-\beta)}[\partial_r^2 \alpha + (\partial_r \alpha)^2 - \partial_r \alpha \partial_r \beta + \frac{2}{r} \partial_r \alpha] - \partial_r^2 \alpha - (\partial_r \alpha)^2 + \partial_r \alpha \partial_r \beta + \frac{2}{r} \partial_r \beta \\ &= \frac{2}{r}(\partial_r \alpha + \partial_r \beta) = 0 \end{aligned}$$

For $r \neq 0$,

$$\partial_r \alpha + \partial_r \beta = \partial_r(\alpha + \beta) = 0 \rightarrow \alpha + \beta = c \rightarrow \alpha = -\beta + c,$$

where c is some constant. We can set $c = 0$ by rescaling our time coordinate by $t \rightarrow e^{-c}t$. To see this, if $\alpha = -\beta + c$ then $g_{tt} = -e^{2\alpha} = -e^{2c}e^{-2\beta}$, so the time part of the metric 4.11 is $-e^{2c}e^{-2\beta}dt^2$. If we rescale our time coordinate by $t \rightarrow e^{-c}t$, then $dt \rightarrow e^{-c}dt$ and the time part of the metric rescales as $-e^{2c}e^{-2\beta}dt^2 \rightarrow -e^{2c}e^{-2\beta}e^{-2c}dt^2 = -e^{-2\beta}dt^2$. This allows us to write,

$$\alpha = -\beta.$$

Now let us consider $R_{\theta\theta} = 0$ but substituting $\alpha = -\beta$,

$$R_{\theta\theta} = e^{2\alpha}(2r\partial_r \alpha + 1) = 1,$$

which is equivalent to,

$$\partial_r(re^{2\alpha}) = 1.$$

Integrating with respect to r ,

$$re^{2\alpha} = r + C \rightarrow e^{2\alpha} = 1 + \frac{C}{r},$$

and renaming the integration constant as $C = -R_s$ yields,

$$e^{2\alpha} = 1 - \frac{R_s}{r},$$

where R_s is some constant. Now we can write metric 4.11 as,

$$ds^2 = -\left(1 - \frac{R_s}{r}\right) dt^2 + \left(1 - \frac{R_s}{r}\right)^{-1} dr^2 + r^2 d\Omega.$$

In Eq. 3.92 we showed that in the weak-field limit,

$$g_{tt} = -\left(1 - \frac{2GM}{r}\right)$$

where $\phi = -GM/r$ is the weak-field gravitational potential (Newton's gravitational potential).

Substituting $R_s = 2GM$ yields the Schwarzschild metric in spherical coordinates,

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega \quad (4.15)$$

where, as a reminder,

$$d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$$

Note as $M \rightarrow 0$ we recover the Minkowski metric in spherical coordinates. Also as $r \rightarrow \infty$ we also recover the Minkowski metric – this is known as **asymptotic flatness**.

Notice also that if either $r = 0$ or $r = 2GM$, the metric coefficients become infinite.

4.2 Geodesics of Schwarzschild

Recall the geodesic equation given by Eq. 3.87

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0.$$

We have the non-vanishing Christoffel symbols given in Eq. 4.12, and we know that,

$$e^{2\alpha} = 1 - \frac{2GM}{r} \quad \text{and} \quad e^{2\beta} = \left(1 - \frac{2GM}{r}\right)^{-1}$$

and so we can write the Christoffel symbols as,

$$\begin{aligned} \Gamma_{tt}^r &= \frac{GM}{r^3}(r - 2GM) & \Gamma_{\theta\theta}^r &= -(r - 2GM) & \Gamma_{r\phi}^\phi &= \frac{1}{r} \\ \Gamma_{tr}^t &= \frac{GM}{r(r - 2GM)} & \Gamma_{\phi\phi}^r &= -(r - 2GM) \sin^2 \theta & \Gamma_{\theta\phi}^\phi &= \frac{\cos \theta}{\sin \theta} \\ \Gamma_{rr}^r &= \frac{-GM}{r(r - 2GM)} & \Gamma_{r\theta}^\theta &= \frac{1}{r} & \Gamma_{\phi\phi}^\theta &= -\sin \theta \cos \theta \end{aligned} \quad (4.16)$$

Using the geodesic equation and the above Christoffel symbols we we can now write,

$$\begin{aligned} \frac{d^2 t}{d\lambda^2} + \frac{2GM}{r(r - 2GM)} \frac{dr}{d\lambda} \frac{dt}{d\lambda} &= 0 \\ \frac{d^2 r}{d\lambda^2} + \frac{GM}{r^3}(r - 2GM) \left(\frac{dt}{d\lambda}\right)^2 - \frac{GM}{r(r - 2GM)} \left(\frac{dr}{d\lambda}\right)^2 \\ - (r - 2GM) \left[\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2 \theta \left(\frac{d\phi}{d\lambda}\right)^2 \right] &= 0 \\ \frac{d^2 \theta}{d\lambda^2} + \frac{2}{r} \frac{d\theta}{d\lambda} \frac{dr}{d\lambda} - \sin \theta \cos \theta \left(\frac{d\phi}{d\lambda}\right)^2 &= 0 \\ \frac{d^2 \phi}{d\lambda^2} + \frac{2}{r} \frac{d\phi}{d\lambda} \frac{dr}{d\lambda} + 2 \frac{\cos \theta}{\sin \theta} \frac{d\theta}{d\lambda} \frac{d\phi}{d\lambda} &= 0 \end{aligned} \quad (4.17)$$

Eq. 4.17 is a set of coupled second-order ODEs and at face value looks unpleasant to solve directly. However, the Schwarzschild spacetime has continuous symmetries, and these symmetries provide *first integrals* of the geodesic equation. This reduces the problem from four coupled second-order ODEs to (essentially) a single first-order radial equation as shown next.

4.2.1 Killing vectors and constants of motion

A vector field K^μ is a Killing vector if it generates an isometry of the metric, equivalently

$$\nabla_{(\mu} K_{\nu)} = 0. \quad (4.18)$$

Let $u^\mu \equiv dx^\mu/d\lambda$ be the tangent to an affinely-parametrized geodesic, so that our geodesic equation in covariant form is,

$$u^\nu \nabla_\nu u^\mu = 0. \quad (4.19)$$

Then along any geodesic the scalar quantity

$$C_K \equiv K_\mu u^\mu \quad (4.20)$$

is conserved – to see this take the derivative along the geodesic, and since $u^\mu = dx^\mu/d\lambda$, the derivative along the curve is the directional covariant derivative along the tangent $d/d\lambda = u^\nu \nabla_\nu$:

$$\frac{d}{d\lambda}(K_\mu u^\mu) = u^\nu \nabla_\nu (K_\mu u^\mu) = u^\nu u^\mu \nabla_\nu K_\mu + K_\mu u^\nu \nabla_\nu u^\mu = u^\nu u^\mu \nabla_{(\nu} K_{\mu)} = 0, \quad (4.21)$$

where we used Eqs. (4.18) and (4.19).

Additionally, we also used symmetrization,

$$\nabla_{(\nu} K_{\mu)} \equiv \frac{1}{2}(\nabla_\nu K_\mu + \nabla_\mu K_\nu)$$

Because $u^\nu u^\mu = u^\mu u^\nu$ (symmetric), we know that when you contract something with $u^\nu u^\mu$ only the symmetric part of the other tensor contributes. So decomposing ∇K into symmetric and antisymmetric parts,

$$\nabla_\nu K_\mu = \nabla_{(\nu} K_{\mu)} + \nabla_{[\nu} K_{\mu]}$$

and contracting it with our symmetric object $u^\nu u^\mu$,

$$u^\nu u^\mu \nabla_\nu K_\mu = u^\nu u^\mu \nabla_\nu \nabla_{(\nu} K_{\mu)} + u^\nu u^\mu \nabla_\nu \nabla_{[\nu} K_{\mu]} = u^\nu u^\mu \nabla_\nu \nabla_{(\nu} K_{\mu)}$$

because the anti-symmetric part is zero.

To recap, we found that,

$$\frac{d}{d\lambda}(K_\mu u^\mu) = 0$$

Thus, each Killing vector produces a constant of motion.

Time-translation symmetry: conserved energy

Because the Schwarzschild metric is independent of t , the coordinate vector

$$K_{(t)}^\mu = (\partial_t)^\mu \quad (4.22)$$

is a Killing vector.

From Eq. (4.21), for any Killing vector K^μ we have

$$\frac{d}{d\lambda}(K_\mu u^\mu) = 0 \quad \Rightarrow \quad K_\mu u^\mu = \text{const.} \quad (4.23)$$

For the timelike Killing vector $K_{(t)}^\mu = (\partial_t)^\mu$, we define the *specific energy* (energy per unit mass)

$$\mathcal{E} \equiv -K_{(t)\mu}^\mu u^\mu = -g_{\mu\nu} K_{(t)}^\nu u^\mu = \text{const.} \quad (4.24)$$

If instead we use the four-momentum $p^\mu = mu^\mu$, the corresponding dimensionful conserved energy is

$$E \equiv -K_\mu^{(t)} p^\mu = -m K_\mu^{(t)} u^\mu = m\mathcal{E}. \quad (4.25)$$

Since in these coordinates

$$K_{(t)}^\nu = (1, 0, 0, 0), \quad (4.26)$$

only the $\nu = t$ component contributes, and therefore

$$K_\mu^{(t)} = g_{\mu\nu} K_{(t)}^\nu = g_{\mu t}. \quad (4.27)$$

Contracting with u^μ gives

$$K_\mu^{(t)} u^\mu = g_{\mu t} u^\mu = g_{t\nu} u^\nu. \quad (4.28)$$

Since the Schwarzschild metric is diagonal, this reduces to

$$K_\mu^{(t)} u^\mu = g_{tt} \frac{dt}{d\lambda}. \quad (4.29)$$

Thus,

$$E = -g_{tt} \frac{dt}{d\lambda}. \quad (4.30)$$

For the Schwarzschild metric with $g_{tt} = -(1 - 2GM/r)$, this gives

$$E = \left(1 - \frac{2GM}{r}\right) \frac{dt}{d\lambda}. \quad (4.31)$$

Equivalently,

$$\frac{dt}{d\lambda} = \frac{E}{1 - \frac{2GM}{r}}. \quad (4.32)$$

(One can verify that Eq. (4.31) is the first integral of the t -geodesic equation in Eq. (4.17).)

Rotational symmetry: conserved angular momentum

Similarly, axisymmetry implies $(\partial_\phi)^\mu$ is a Killing vector:

$$K_{(\phi)}^\mu = (\partial_\phi)^\mu. \quad (4.33)$$

The conserved quantity is

$$L \equiv K_\mu^{(\phi)} u^\mu = g_{\phi\nu} u^\nu = g_{\phi\phi} \frac{d\phi}{d\lambda} = r^2 \sin^2 \theta \frac{d\phi}{d\lambda}. \quad (4.34)$$

Restriction to the equatorial plane

Spherical symmetry allows us to choose initial conditions such that the motion lies in a plane through the origin. By rotating coordinates, we may take that plane to be the equatorial plane,

$$\theta = \frac{\pi}{2}, \quad \frac{d\theta}{d\lambda} = 0, \quad (4.35)$$

which is consistent with the θ -geodesic equation.¹ Then Eq. (4.34) simplifies to

$$L = r^2 \frac{d\phi}{d\lambda}, \quad \text{so that} \quad \frac{d\phi}{d\lambda} = \frac{L}{r^2}. \quad (4.36)$$

¹In particular, with $\theta = \pi/2$ we have $\sin \theta \cos \theta = 0$, and if $d\theta/d\lambda = 0$ initially then $d^2\theta/d\lambda^2 = 0$ as well.

Radial equation and effective potential

In addition to the Killing integrals, the tangent vector satisfies the normalization condition

$$g_{\mu\nu}u^\mu u^\nu = -\epsilon, \quad (4.37)$$

where $\epsilon = 1$ for timelike geodesics (massive particles) and $\epsilon = 0$ for null geodesics (photons). In the equatorial plane, this becomes

$$-\left(1 - \frac{2GM}{r}\right) \left(\frac{dt}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{r}\right)^{-1} \left(\frac{dr}{d\lambda}\right)^2 + r^2 \left(\frac{d\phi}{d\lambda}\right)^2 = -\epsilon. \quad (4.38)$$

Substituting Eqs. (4.32) and (4.36) yields a single first-order equation for $r(\lambda)$:

$$\left(\frac{dr}{d\lambda}\right)^2 = E^2 - \left(1 - \frac{2GM}{r}\right) \left(\epsilon + \frac{L^2}{r^2}\right). \quad (4.39)$$

It is convenient to rewrite this in “energy–potential” form by defining the effective potential

$$V_{\text{eff}}(r) \equiv \frac{1}{2} \left(1 - \frac{2GM}{r}\right) \left(\epsilon + \frac{L^2}{r^2}\right). \quad (4.40)$$

Then Eq. (4.39) becomes

$$\frac{1}{2} \left(\frac{dr}{d\lambda}\right)^2 + V_{\text{eff}}(r) = \frac{1}{2} E^2 = \mathcal{E} \quad (4.41)$$

Thus the original coupled system (4.17) reduces to quadratures once the constants of motion (E, L) are fixed.

Eq. 4.41 is the equation for a classical particle of unit mass and “energy” $\mathcal{E}$ moving in a one-dimensional potential given by $V(r)$. The conserved energy per unit mass is E , but the effective potential for the coordinate r responds to $\mathcal{E} = E^2/2$. In classical mechanics,

$$\frac{1}{2} \left(\frac{dr}{d\lambda}\right)^2 + \frac{V(r)}{m} = \frac{E_{\text{Newton}}}{m} = \mathcal{E}_{\text{Newton}}$$

where $\mathcal{E}_{\text{Newton}}$ is the linear physical energy. From Schwarzschild we get Eq. 4.41, where $V(r)$ is not an effective potential of the form Eq. 4.40 and the Newton-style “mechanical energy parameter” is $\mathcal{E}_{\text{mech}} = \frac{1}{2} E^2$. This is neat!

Why? Because $E = -K_\mu u^\mu$ is a geometric conserved quantity. The radial equation derived from normalization makes it enter quadratically and the quadratic structure produces a clean Newtonian-style energy equation. So the “energy” in the effective potential picture is not the conserved Killing energy itself, but a simple function of it. This is not arbitrary – it comes directly from the Lorentzian metric constraint.

Notice what happened structurally. In Newtonian mechanics, energy appears linearly in the Hamiltonian. But in GR, the normalization $u^\mu u_\mu = -1$ is quadratic! So conserved quantities naturally appear squared. That’s fundamentally relativistic. The fact that it reorganizes into a Newtonian-looking equation is actually quite elegant.

For large r and slow motion,

$$E \approx 1 + \frac{v^2}{2} - \frac{GM}{r}. \quad (4.42)$$

Then

$$\frac{1}{2}E^2 \approx \frac{1}{2} + \left(\frac{v^2}{2} - \frac{GM}{r} \right) \quad (4.43)$$

The $1/2$ is irrelevant (just shifts zero of energy), and what remains is exactly the Newtonian specific energy. That's why the analogy works so beautifully. It shows how Newtonian mechanics is embedded inside the relativistic structure.

4.2.2 Effective Potentials: Newtonian vs. Schwarzschild

In the large- r , slow-motion limit we found Eqs. 4.42 and 4.43. The additive constant $1/2$ merely shifts the zero of energy. The remaining term is precisely the Newtonian specific energy,

$$\mathcal{E}_{\text{Newton}} = \frac{v^2}{2} - \frac{GM}{r}. \quad (4.44)$$

This demonstrates explicitly how Newtonian mechanics is embedded inside the relativistic structure.

Newtonian effective potentials. In Newtonian gravity the effective potential for a particle of angular momentum per unit mass L is

$$V_N(r) = \frac{L^2}{2r^2} - \frac{GM}{r}. \quad (4.45)$$

For massless particles (which in Newtonian theory feel no gravitational attraction), only the centrifugal term remains,

$$V_N^{(\text{massless})}(r) = \frac{L^2}{2r^2}. \quad (4.46)$$

To facilitate direct comparison with the relativistic case, where $V_{\text{eff}} \rightarrow 1/2$ as $r \rightarrow \infty$ for timelike geodesics, we shift the Newtonian massive potential by an additive constant $+1/2$. This does not alter the dynamics but aligns the asymptotic normalization:

$$V_{N,\text{shifted}}(r) = \frac{1}{2} + \frac{L^2}{2r^2} - \frac{GM}{r}. \quad (4.47)$$

The resulting Newtonian effective potentials are shown in Fig. 4.1. The code used to generate these plots is provided in Appendix B.1.

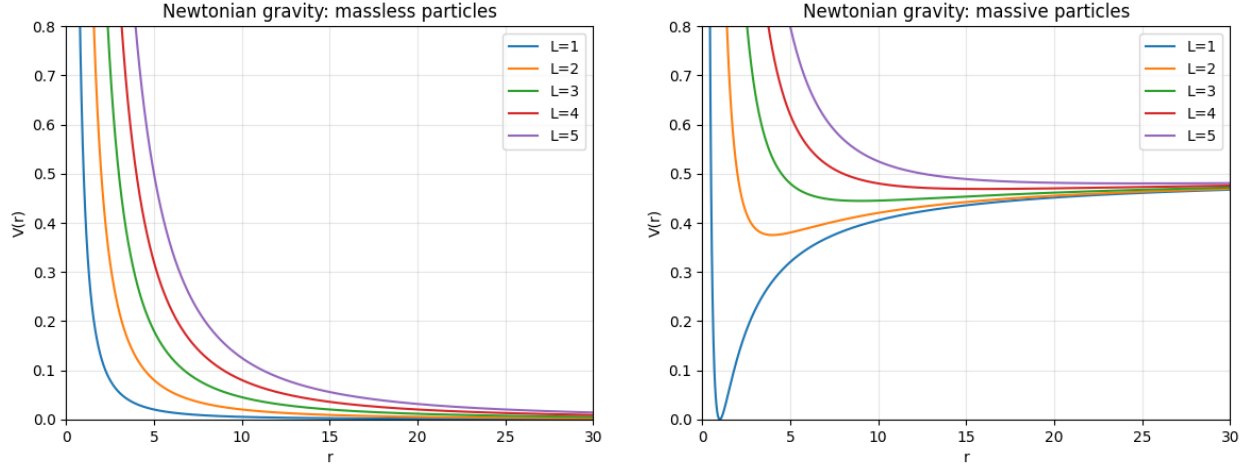


Figure 4.1: Effective potentials in Newtonian gravity ($GM = 1$). The left panel shows massless particles, for which the potential reduces to the centrifugal barrier $V(r) = L^2/(2r^2)$. The right panel shows massive particles with angular momentum per unit mass $L = 1, \dots, 5$, using the effective potential $V(r) = \frac{1}{2} + \frac{L^2}{2r^2} - \frac{1}{r}$. The additive constant $1/2$ shifts the zero of energy to facilitate comparison with the relativistic case; it does not affect the dynamics.

Schwarzschild effective potentials. For Schwarzschild geodesics, the radial equation may be written in Newtonian form,

$$\frac{1}{2}\dot{r}^2 + V_{\text{eff}}(r) = \frac{1}{2}E^2, \quad (4.48)$$

where

$$V_{\text{eff}}(r) = \frac{1}{2} \left(1 - \frac{2GM}{r} \right) \left(\epsilon + \frac{L^2}{r^2} \right), \quad (4.49)$$

with $\epsilon = 1$ for timelike geodesics (massive particles) and $\epsilon = 0$ for null geodesics (massless particles).

The corresponding relativistic effective potentials are shown in Fig. 4.2. The numerical code used to generate these figures appears in Appendix B.2.

Circular Orbits and Their Stability. Circular orbits occur at extrema of the effective potential. Differentiating Eq. 4.41 with respect to r and setting $dV_{\text{eff}}/dr = 0$ yields the algebraic condition

$$\epsilon GM r_c^2 - L^2 r_c + 3GML^2\gamma = 0, \quad (4.50)$$

where $\gamma = 0$ for Newtonian gravity and $\gamma = 1$ for the Schwarzschild case.

Newtonian gravity ($\gamma = 0$). Equation (4.50) reduces to

$$\epsilon GM r_c^2 - L^2 r_c = 0.$$

For massive particles ($\epsilon = 1$),

$$r_c = \frac{L^2}{GM},$$

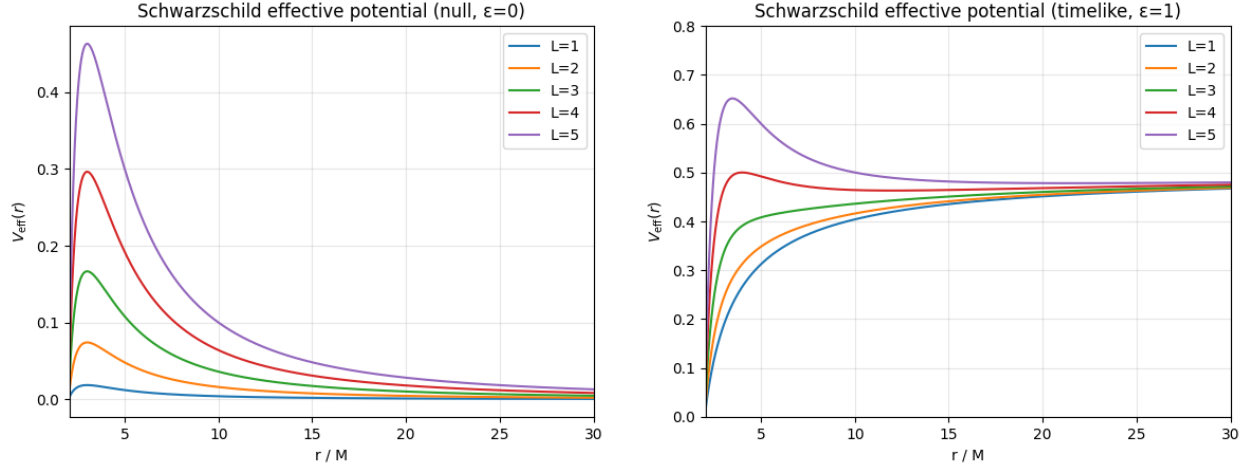


Figure 4.2: Effective potentials for Schwarzschild geodesics ($GM = 1$). The upper panel shows null geodesics ($\epsilon = 0$), exhibiting an unstable maximum corresponding to the photon sphere at $r = 3M$. The lower panel shows timelike geodesics ($\epsilon = 1$) for $L = 1, \dots, 5$, where the potential develops both stable minima and unstable maxima. For sufficiently small L , the minimum disappears, signaling the innermost stable circular orbit (ISCO). In both cases the effective potential is defined by $V_{\text{eff}}(r) = \frac{1}{2} \left(1 - \frac{2M}{r}\right) \left(\epsilon + \frac{L^2}{r^2}\right)$.

corresponding to the single minimum visible in the lower panel of Fig. 4.1. Each value of L determines one stable circular orbit; no unstable circular orbit appears in Newtonian theory.

For the Newtonian “massless” case ($\epsilon = 0$), the equation admits no finite solution. The effective potential reduces to the purely centrifugal term $V = L^2/(2r^2)$, shown in the upper panel of Fig. 4.1. Since Newtonian gravity couples only to mass, no attractive $-GM/r$ term is present. The motion is inertial, and the curves represent straight lines in flat space expressed in polar coordinates. The effective potential encodes only angular-momentum kinematics, not gravitational dynamics.

Schwarzschild spacetime ($\gamma = 1$). Equation (4.50) becomes

$$\epsilon GM r_c^2 - L^2 r_c + 3GML^2 = 0.$$

Null geodesics ($\epsilon = 0$). The circular-orbit condition reduces to

$$-L^2 r_c + 3GML^2 = 0,$$

giving

$$r_c = 3GM.$$

This orbit corresponds to the maximum of the effective potential shown in the upper panel of Fig. 4.2. It is therefore unstable. The radius $r = 3GM$ defines the photon sphere, a purely relativistic feature arising from the curvature correction proportional to $3GML^2$. In contrast to Newtonian theory, spacetime curvature modifies the centrifugal barrier itself, allowing null circular geodesics to exist.

Timelike geodesics ($\epsilon = 1$). For massive particles,

$$GM r_c^2 - L^2 r_c + 3GML^2 = 0.$$

This quadratic equation admits two solutions when L is sufficiently large, corresponding to a minimum and a maximum of the effective potential in the lower panel of Fig. 4.2. The larger root represents a stable circular orbit ($d^2V/dr^2 > 0$), while the smaller root represents an unstable circular orbit ($d^2V/dr^2 < 0$).

The two extrema merge when the discriminant vanishes,

$$L^4 - 12G^2M^2L^2 = 0,$$

which occurs at

$$r_c = 6GM.$$

This radius defines the innermost stable circular orbit (ISCO). For $r > 6GM$ stable circular geodesics exist; for $3GM < r < 6GM$ only unstable circular geodesics exist; for $r < 3GM$ no circular timelike geodesics exist at all.

The horizon at $r = 2GM$ marks a separate geometric boundary: inside this radius no future-directed timelike or null geodesic can return to larger r . However, there is no “barrier” at $r = 3GM$ or $r = 6GM$. These radii characterize extrema of the effective potential, not hard surfaces. A particle with sufficient energy may pass through $r = 6GM$ or even dip below $r = 3GM$ and re-emerge, provided it remains outside the horizon.

Comparison of Fig. 4.1 and Fig. 4.2 therefore reveals both continuity and qualitative change. In the weak-field limit ($r \gg 2GM$), the Schwarzschild potentials reduce to their Newtonian counterparts. In the strong-field regime, the additional curvature term $3GML^2$ introduces new dynamical structure: an unstable photon sphere at $3GM$ and the ISCO at $6GM$. These features play a central role in black-hole accretion physics and gravitational-wave dynamics.

4.3 Experimental Tests

To test general relativity, Einstein suggested three tests: the deflection of light, the precession of perihelia, and gravitational redshift. The deflection of light will not be discussed in this chapter. However using the Schwarzschild radial equation of motion in Eq. 4.41 will investigate the latter two.

4.3.1 The Precession of Perihelia

Orbits in GR are not perfect closed ellipses. To a good approximation they are ellipses that precess as shown in Fig. 4.3.1.

For orbital motion our goal is to write the radial equation of motion given in Eq. 4.41 in terms of $r(\phi)$ rather than $r(\lambda)$. Start with the Killing vector from Eq. 4.36,

$$\frac{d\phi}{d\lambda} = \frac{L}{r^2},$$

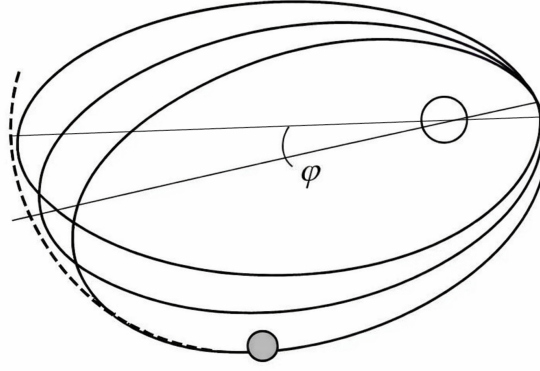


Figure 4.3: Orbits in general relativity describe precessing ellipses.

and square it,

$$\left(\frac{d\phi}{d\lambda}\right)^2 = \frac{L^2}{r^4},$$

and invert

$$\left(\frac{d\phi}{d\lambda}\right)^{-2} = \frac{r^4}{L^2}.$$

Using the chain rule,

$$\frac{dr}{d\lambda} = \frac{dr}{d\phi} \frac{d\phi}{d\lambda}$$

and square it,

$$\left(\frac{dr}{d\lambda}\right)^2 = \left(\frac{dr}{d\phi}\right)^2 \left(\frac{d\phi}{d\lambda}\right)^2 = \left(\frac{dr}{d\phi}\right)^2 \frac{L^2}{r^4}.$$

Substitute this into the radial equation of motion for the timelike case ($\epsilon = 1$),

$$\frac{1}{2} \left(\frac{dr}{d\phi}\right)^2 \frac{L^2}{r^4} + \frac{1}{2} \left(1 - \frac{2GM}{r}\right) \left(1 + \frac{L^2}{r^2}\right) = \mathcal{E},$$

and solve for $(dr/d\phi)^2$,

$$\left(\frac{dr}{d\phi}\right)^2 + \frac{r^4}{L^2} \left(1 - \frac{2GM}{r}\right) \left(1 + \frac{L^2}{r^2}\right) = \frac{2\mathcal{E}}{L^2} r^4.$$

Define a new variable,

$$x = \frac{L^2}{GM r},$$

notice that $x = 1$ at the Newtonian circular orbit $r_c = L^2/GM$. Using the chain rule,

$$\frac{dx}{d\phi} = \frac{dx}{dr} \frac{dr}{d\phi} = -\frac{GM x^2}{L^2} \frac{dr}{d\phi} \quad \text{or} \quad \left(\frac{dx}{d\phi}\right)^2 = \left(\frac{GM}{L^2}\right)^2 x^4 \left(\frac{dr}{d\phi}\right)^2.$$

Using this and $r = L^2/GM x$,

$$\left(\frac{dx}{d\phi}\right)^2 + \frac{L^2}{G^2 M^2} - 2x + x^2 - \frac{2G^2 M^2}{L^2} x^3 = \frac{2\mathcal{E} L^2}{G^2 M^2}.$$

Now differentiate with respect to ϕ ,

$$\frac{d^2x}{d\phi^2} - 1 + x = \frac{3G^2M^2}{L^2}x^2 \quad (4.51)$$

In the Newtonian limit the last term would be absent and we could solve for x exactly. Expand x into a Newtonian solution plus a small deviation,

$$x = x_0 + x_1.$$

Then the zeroth-order part of Eq. 4.51 is,

$$\frac{d^2x_0}{d\phi^2} - 1 + x_0 = 0,$$

and the first order part is,

$$\frac{d^2x_1}{d\phi^2} + x_1 = \frac{3G^2M^2}{L^2}x_0^2.$$

The solution for the zeroth-order equation can be written,

$$x_0 = 1 + e \cos \phi. \quad (4.52)$$

This is the standard result of Newton or Kepler describing a perfect ellipse, with e eccentricity! Plugging in x_0 into the first-order equation,

$$\begin{aligned} \frac{d^2x_1}{d\phi^2} + x_1 &= \frac{3G^2M^2}{L^2}(1 + e \cos \phi)^2 \\ &= \frac{3G^2M^2}{L^2} \left[\left(1 + \frac{1}{2}e^2\right) + 2e \cos \phi + \frac{1}{2}e^2 \cos 2\phi \right]. \end{aligned} \quad (4.53)$$

Let $D \equiv d/d\phi$ and notice that the LHS is a linear operator on x_1 ,

$$(D^2 + 1)x_1 = \frac{3G^2M^2}{L^2} \left[\left(1 + \frac{1}{2}e^2\right) + 2e \cos \phi + \frac{1}{2}e^2 \cos 2\phi \right],$$

which means we can solve for each piece separately and add them.

For the first term, let

$$A \equiv \frac{3G^2M^2}{L^2} \left(1 + \frac{1}{2}e^2\right)$$

Try a constant particular solution $x_{1,A} = C$, then,

$$(D^2 + 1)C = 0 + C = C$$

So choose $C = A$, hence,

$$x_{1,A} = A.$$

For the second term let,

$$B \equiv \frac{3G^2M^2}{L^2}(2e) = \frac{6eG^2M^2}{L^2}.$$

If you tried, $x_{1,B} \propto \cos \phi$, it would fail because $\cos \phi$ is a homogeneous solution of $x'' + x = 0$ (resonance). So multiply by ϕ and try the standard resonant ansatz,

$$x_{1,B} = a\phi \sin \phi.$$

No use the identity,

$$(D^2 + 1)(\phi \sin \phi) = 2a \cos \phi.$$

We want this to equal $B \cos \phi$, so set $2a = B$,

$$a = \frac{B}{2} = \frac{3eG^2M^2}{L^2}.$$

Thus a particular solution for this piece is,

$$x_{1,B} = \frac{B}{2} \phi \sin \phi = \frac{3eG^2M^2}{L^2} \phi \sin \phi.$$

Now for the final term, let

$$C \equiv \frac{3G^2M^2}{L^2} \left(\frac{1}{2} e^2 \right) = \frac{3e^2G^2M^2}{2L^2}.$$

Try $x_{1,c} = b \cos 2\phi$. Then,

$$(D^2 + 1) \cos 2\phi = -4 \cos 2\phi + \cos 2\phi = -3 \cos 2\phi.$$

We want $-3b \cos 2\phi = C \cos 2\phi$, so

$$b = -\frac{C}{3},$$

and

$$x_{1,c} = -\frac{C}{3} \cos 2\phi = -\frac{1}{3} \cdot \frac{3e^2G^2M^2}{2L^2} \cos 2\phi = -\frac{e^2G^2M^2}{2L^2} \cos 2\phi.$$

Now put it all together,

$$x_1 = \frac{3G^2M^2}{L^2} \left[\left(1 + \frac{1}{2} e^2 \right) + e\phi \sin \phi - \frac{1}{6} e^2 \cos 2\phi \right]. \quad (4.54)$$

The first term is simply a constant displacement, while the third oscillates around zero. The second term, $e\phi \sin \phi$, accumulates over successive orbits. So we combine it with the zeroth-order solution in Eq. 4.52. To first order, the combined solution takes the schematic form

$$x(\phi) \approx 1 + e \cos \phi + \eta \phi \sin \phi, \quad (4.55)$$

where η is small. The term proportional to $\phi \sin \phi$ is secular: it grows linearly with ϕ and signals a slow drift in the orbital phase.

To interpret this term, consider a cosine with a slightly shifted angular frequency,

$$\cos((1 - \delta)\phi). \quad (4.56)$$

Expanding to first order in small δ ,

$$\begin{aligned} \cos((1 - \delta)\phi) &= \cos(\phi - \delta\phi) \\ &\approx \cos \phi + (\delta\phi) \sin \phi. \end{aligned} \quad (4.57)$$

Multiplying by e gives

$$e \cos((1 - \delta)\phi) \approx e \cos \phi + e \delta \phi \sin \phi. \quad (4.58)$$

Comparing with the perturbative solution,

$$e \cos \phi + \eta \phi \sin \phi,$$

we identify

$$\eta = e \delta \quad \Rightarrow \quad \delta = \frac{\eta}{e}. \quad (4.59)$$

In the present case,

$$\eta = \frac{3eG^2M^2}{L^2},$$

so

$$\delta = \frac{3G^2M^2}{L^2}. \quad (4.60)$$

One radial period occurs when

$$(1 - \delta)\phi = 2\pi,$$

so the angular advance after one complete orbit is

$$\phi = \frac{2\pi}{1 - \delta} \approx 2\pi(1 + \delta). \quad (4.61)$$

The perihelion advance per orbit is therefore

$$\Delta\phi = \phi - 2\pi \approx 2\pi\delta = \frac{6\pi G^2M^2}{L^2}. \quad (4.62)$$

Using the Newtonian relation for an ellipse,

$$L^2 = GM a(1 - e^2),$$

we obtain the standard result

$$\boxed{\Delta\phi = \frac{6\pi GM}{a(1 - e^2)}} \quad (c = 1). \quad (4.63)$$

Restoring factors of c ,

$$\boxed{\Delta\phi = \frac{6\pi GM}{a(1 - e^2)c^2}}. \quad (4.64)$$

Application to Mercury. For Mercury orbiting the Sun, the relevant parameters are

$$\frac{GM_\odot}{c^2} = 1.48 \times 10^5 \text{ cm}, \quad (4.65)$$

$$a = 5.79 \times 10^{12} \text{ cm}, \quad (4.66)$$

$$e = 0.2056, \quad (4.67)$$

$$c = 3.00 \times 10^{10} \text{ cm/s}. \quad (4.68)$$

Using

$$\Delta\phi = \frac{6\pi GM}{a(1-e^2)c^2},$$

we obtain

$$\Delta\phi_{\text{Mercury}} = 5.01 \times 10^{-7} \text{ rad/orbit} = 0.103''/\text{orbit}. \quad (4.69)$$

Mercury completes one orbit in approximately 88 days, corresponding to

$$\frac{365.25}{88} \approx 4.15 \text{ orbits per year},$$

or about 415 orbits per century. Therefore,

$$\Delta\phi_{\text{Mercury}} = (0.103'') \times 415 \approx 43.0''/\text{century}. \quad (4.70)$$

The total observed anomalous precession of Mercury is

$$5601''/\text{century}.$$

However,

$$5025''/\text{century}$$

is attributable to the precession of the equinoxes in the geocentric reference frame, and

$$532''/\text{century}$$

arises from Newtonian gravitational perturbations due to the other planets. Subtracting these contributions,

$$5601 - 5025 - 532 = 44''/\text{century},$$

leaving approximately

$$43''/\text{century}$$

to be explained by general relativity.

The agreement between prediction and observation was one of the earliest quantitative confirmations of Einstein's theory.

4.3.2 Gravitational Redshift

Consider an observer with four-velocity U^μ who is stationary in Schwarzschild coordinates, so that

$$U^i = 0 \quad \Rightarrow \quad dr = d\theta = d\phi = 0.$$

We choose the observer to be stationary in order to isolate the gravitational frequency shift without superimposing an ordinary Doppler shift due to relative motion.

For such an observer, the worldline is

$$x^\mu(\tau) = (t(\tau), r = \text{const}, \theta = \text{const}, \phi = \text{const}),$$

so the four-velocity takes the form

$$U^\mu = \left(\frac{dt}{d\tau}, 0, 0, 0 \right) \equiv (U^0, 0, 0, 0).$$

By definition, the four-velocity satisfies

$$U_\mu U^\mu = -1.$$

Using the Schwarzschild metric (Eq. 4.15),

$$U_\mu U^\mu = g_{\mu\nu} U^\mu U^\nu = g_{tt} (U^0)^2 = - \left(1 - \frac{2GM}{r} \right) (U^0)^2.$$

Setting this equal to -1 gives

$$\left(1 - \frac{2GM}{r} \right) (U^0)^2 = 1,$$

and hence

$$U^0 = \left(1 - \frac{2GM}{r} \right)^{-1/2}.$$

Since $U^0 = dt/d\tau$, we obtain

$$d\tau = \sqrt{1 - \frac{2GM}{r}} dt.$$

This is gravitational time dilation: a stationary clock deeper in the gravitational field runs more slowly relative to coordinate time t , which coincides with proper time at spatial infinity.

In special relativity a photon has four-momentum p^μ , an observer has four-velocity U^μ , and the energy measured by that observer is $E_{\text{obs}} = -p_\mu U^\mu$. Frequency is $\omega = E/\hbar$ so for photons, $\omega = -p_\mu U^\mu$ in units where $\hbar = 1$. To recall why $p_\mu U^\mu$ gives measured energy, in flat spacetime, if the observer is at rest, $U^\mu = (1, 0, 0, 0)$ so $-p_\mu U^\mu = -(E, \vec{p}) \cdot (1, 0, 0, 0) = E$. In curved spacetime, since photons are massless, their four-momentum is tangent to the null geodesic, so up to an overall constant $p^\mu = dx^\mu/d\lambda$. So a photon traveling along a null geodesic $x^\mu(\lambda)$ has a frequency measured by an observer with four-velocity U^μ ,

$$\omega = -g_{\mu\nu} U^\mu \frac{dx^\nu}{d\lambda}.$$

This tells us that the frequency measured by an observer is the projection of the photon's four-momentum onto the observer's four-velocity. For a stationary observer,

$$\omega = -g_{tt} U^0 \frac{dt}{d\lambda}.$$

Using $g_{tt} = -(1 - 2GM/r)$ and the expression for U^0 , we find

$$\omega = \left(1 - \frac{2GM}{r} \right)^{1/2} \frac{dt}{d\lambda}.$$

The quantity

$$E \equiv \left(1 - \frac{2GM}{r} \right) \frac{dt}{d\lambda}$$

is conserved along the null geodesic (Eq. 4.31), since it arises from the timelike Killing vector. Therefore,

$$\frac{dt}{d\lambda} = \frac{E}{1 - \frac{2GM}{r}}.$$

Substituting,

$$\omega = \left(1 - \frac{2GM}{r}\right)^{-1/2} E.$$

Since E is constant along the photon trajectory, the locally measured frequency depends only on radius. For emission at r_1 and observation at r_2 ,

$$\frac{\omega_2}{\omega_1} = \left(\frac{1 - \frac{2GM}{r_1}}{1 - \frac{2GM}{r_2}}\right)^{1/2}.$$

This is the gravitational frequency shift.

In the weak-field limit $r \gg 2GM$,

$$\left(1 - \frac{2GM}{r}\right)^{1/2} \approx 1 - \frac{GM}{r}.$$

Thus

$$\frac{\omega_2}{\omega_1} \approx 1 - \frac{GM}{r_1} + \frac{GM}{r_2} = 1 + \Phi_1 - \Phi_2,$$

where $\Phi = -GM/r$ is the Newtonian gravitational potential.

If $r_2 > r_1$ (the photon climbs outward), then $\Phi_2 > \Phi_1$, so

$$\omega_2 < \omega_1,$$

corresponding to a redshift. Conversely, photons falling inward are blueshifted.

The effect is purely geometric: it arises from the variation of proper time with radius and does not depend on any force acting on the photon.

4.4 Causal Structure of the Schwarzschild Geometry

One way to understand the geometry of spacetime is to explore its causal structure as defined by light cones. Consider radial null curves - that is, those for which θ and ϕ are constant and $ds^2 = 0$,

$$ds^2 = 0 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2$$

So,

$$\frac{dt}{dr} = \pm \left(1 - \frac{2GM}{r}\right)^{-1}. \quad (4.71)$$

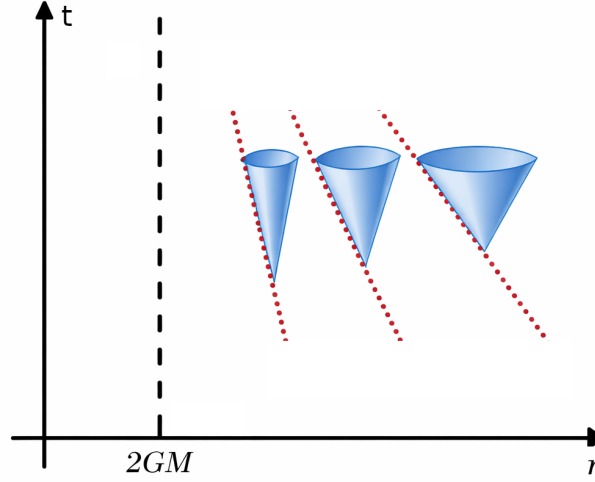


Figure 4.4: In Schwarzschild coordinates, light cones appear to close up as we approach $r = 2GM$.

Eq. 4.71 measures the slope of the light cones on a spacetime diagram of the $t - r$ plane as shown in Fig. 4.4. For large r the slope tends to ± 1 , as it would be in flat space. As we approach $r = 2GM$ we get $dt/dr \rightarrow \pm\infty$, and the light cones "close up", as shown in Fig. 4.4. The light ray approaches $r = 2GM$ but seems to asymptote this radius in this coordinate system.

Consider launching a beacon sending back signals as it moves toward $r = 2GM$ depicted in diagram form in Fig. 4.5. And fixed interval $\Delta\tau_1$ of its proper time corresponds to a longer and longer interval $\Delta\tau_2$ from our point of view sitting some distance $r > 2GM$ away. This continues forever; we would never see the observer cross $r = 2GM$, just asymptotically approach it.

The fact that we never see the infalling beacon reach $r = 2GM$ is a meaningful statement, but the fact that their trajectory in the $t - r$ plane never reaches there is not. It is dependent on our chosen coordinate system. The problem with our current coordinates is that $dt/dr \rightarrow \infty$ along radial null geodesics that approach $r = 2GM$; progress in the r direction becomes slower and slower with respect to the coordinate time t . We now try to fix this problem by replacing t with a coordinate that moves more slowly along null geodesics. If we solve Eq. 4.71,

$$\frac{dt}{dr} = \pm \left(1 - \frac{2GM}{r}\right)^{-1},$$

we first rewrite the expression algebraically:

$$\frac{dt}{dr} = \pm \frac{1}{1 - \frac{2GM}{r}} = \pm \frac{r}{r - 2GM}.$$

Thus,

$$dt = \pm \frac{r}{r - 2GM} dr = \pm \left(1 + \frac{2GM}{r - 2GM}\right) dr.$$

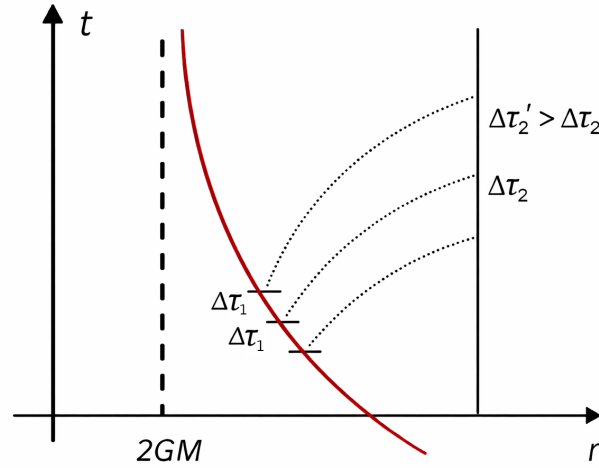


Figure 4.5: A beacon falling freely into a black hole emits signals at intervals of constant proper time $\Delta\tau_1$. An observer at fixed r receives the signals at successively longer time intervals $\Delta\tau_2$.

Integrating both terms separately,

$$t(r) = \pm \left(\int dr + 2GM \int \frac{dr}{r - 2GM} \right) + C \quad (4.72)$$

$$= \pm (r + 2GM \ln |r - 2GM|) + C. \quad (4.73)$$

It is often convenient to factor out $2GM$ inside the logarithm,

$$t(r) = \pm \left(r + 2GM \ln \left| \frac{r}{2GM} - 1 \right| \right) + C.$$

The logarithmic term diverges as $r \rightarrow 2GM$, reflecting the coordinate singularity of Schwarzschild time at the horizon. If we restrict ourselves only to $r > 2GM$ we can drop the absolute value bars and write as,

$$t = \pm r^* + C \quad (4.74)$$

where the **tortoise coordinate**, r^* , is defined by,

$$r^* = r + 2GM \ln \left(\frac{r}{2GM} - 1 \right), \quad r > 2GM. \quad (4.75)$$

Using the tortoise coordinate in the Schwarzschild metric 4.15 gives,

$$ds^2 = \left(1 - \frac{2GM}{r} \right) (-dt^2 + dr^{*2}) + r^2 d\Omega^2. \quad (4.76)$$

Here $r^* = f(r)$. Fig. 4.6 shows the result of using the Schwarzschild metric with tortoise coordinates – the light cones don't close up and none of the metric coefficients become infinite at $r = 2GM$ (although both $g_{tt} = g_{r^*r^*} = 0$). However, the surface $r = 2GM$ has been pushed to infinity.

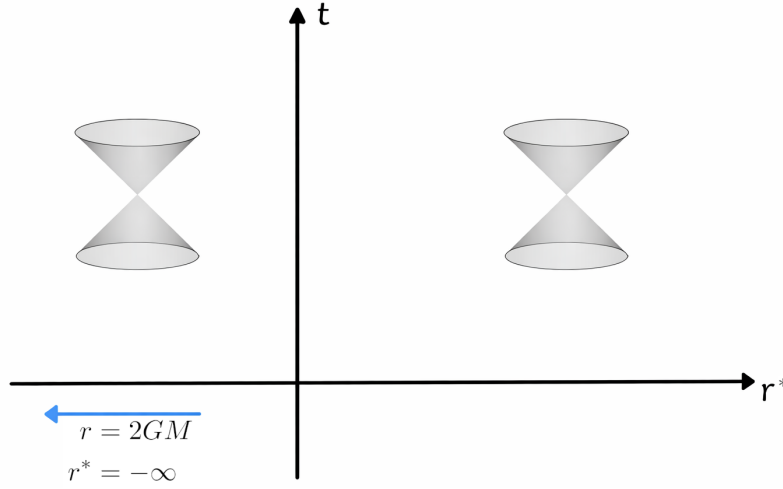


Figure 4.6: Schwarzschild light cones in tortoise coordinates, eq 4.76. Light cones remain nondegenerate, but the surface $r = 2GM$ has been pushed to infinity.

If we let,

$$\begin{aligned} v &= t + r^* \\ u &= t - r^* \end{aligned} \quad (4.77)$$

the the infalling radial null geodesics are characterized by $v = \text{constant} = t + r^*$, while the outgoing ones satisfy $u = \text{constant} = t - r^*$ from Eq. 4.74. If we go back to the original radial coordinate, r , but replace the timelike coordinate t with the new coordinate v , we get the **Eddington-Finkelstein coordinates** and apply to the Schwarzschild metric,

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \quad (4.78)$$

define

$$f(r) \equiv 1 - \frac{2GM}{r}.$$

Then the metric may be written

$$ds^2 = -f dt^2 + f^{-1} dr^2 + r^2 d\Omega^2.$$

We now introduce the ingoing Eddington-Finkelstein coordinate

$$v \equiv t + r^*,$$

where the tortoise coordinate r^* satisfies

$$\frac{dr^*}{dr} = f^{-1}.$$

Taking the differential,

$$dv = dt + dr^* = dt + f^{-1} dr,$$

so that

$$dt = dv - f^{-1} dr.$$

Compute dt^2 :

$$dt^2 = (dv - f^{-1} dr)^2 \quad (4.79)$$

$$= dv^2 - 2f^{-1} dv dr + f^{-2} dr^2. \quad (4.80)$$

Substituting into the metric,

$$ds^2 = -f (dv^2 - 2f^{-1} dv dr + f^{-2} dr^2) + f^{-1} dr^2 + r^2 d\Omega^2. \quad (4.81)$$

Distributing the factor of $-f$,

$$ds^2 = -f dv^2 + 2 dv dr - f f^{-2} dr^2 + f^{-1} dr^2 + r^2 d\Omega^2. \quad (4.82)$$

Since $f f^{-2} = f^{-1}$, the dr^2 terms cancel:

$$-f^{-1} dr^2 + f^{-1} dr^2 = 0.$$

Thus we obtain the Eddington–Finkelstein form of the metric,

$$ds^2 = - \left(1 - \frac{2GM}{r} \right) dv^2 + 2 dv dr + r^2 d\Omega^2. \quad (4.83)$$

or writing out the cross term symmetrically $2 dv dr = dv dr + dr dv$ to match Carroll,

$$ds^2 = - \left(1 - \frac{2GM}{r} \right) dv^2 + (dv dr + dr dv) + r^2 d\Omega. \quad (4.84)$$

Notice that there is not f^{-1} anymore, so nothing blows up at $r = 2GM$ even though g_{vv} vanishes! Progress! The horizon singularity was purely coordinate, and v coordinates are regular there.

To compute the determinant of the ingoing Eddington–Finkelstein metric,

$$ds^2 = - \left(1 - \frac{2GM}{r} \right) dv^2 + 2 dv dr + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2,$$

work in coordinates $(x^\mu) = (v, r, \theta, \phi)$ and define

$$f(r) \equiv 1 - \frac{2GM}{r}.$$

The nonzero metric components are

$$g_{vv} = -f, \quad g_{vr} = g_{rv} = 1, \quad g_{\theta\theta} = r^2, \quad g_{\phi\phi} = r^2 \sin^2 \theta,$$

with $g_{rr} = 0$.

Thus the metric matrix is block diagonal between the (v, r) sector and the angular sector:

$$(g_{\mu\nu}) = \begin{pmatrix} -f & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix} = \begin{pmatrix} \boxed{\begin{matrix} -f & 1 \\ 1 & 0 \end{matrix}} & 0 \\ 0 & \boxed{\begin{matrix} r^2 & 0 \\ 0 & r^2 \sin^2 \theta \end{matrix}} \end{pmatrix}.$$

Therefore,

$$\det(g_{\mu\nu}) = \det \begin{pmatrix} -f & 1 \\ 1 & 0 \end{pmatrix} \cdot \det \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2 \theta \end{pmatrix}.$$

Compute each determinant:

$$\det \begin{pmatrix} -f & 1 \\ 1 & 0 \end{pmatrix} = (-f)(0) - (1)(1) = -1,$$

and

$$\det \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2 \theta \end{pmatrix} = r^2 \cdot r^2 \sin^2 \theta = r^4 \sin^2 \theta.$$

Multiplying,

$$\det(g_{\mu\nu}) = -r^4 \sin^2 \theta.$$

Notice at $r = 2GM$ the determinant is perfectly regular. Therefore the metric is invertible, and we see once and for all that $r = 2GM$ is simply a coordinate singularity in our original (t, r, θ, ϕ) system. In the Eddington-Finkelstein coordinates the condition for radial null curves is solved by,

$$\frac{dv}{dr} = \begin{cases} 0 & \text{(infalling)} \\ 2 \left(1 - \frac{2GM}{r}\right)^{-1} & \text{(outgoing)} \end{cases}$$

Notice that $dv/dr = 0 \rightarrow dv = 0$; meaning that infalling rays are just constant- v lines and outgoing rays are distorted by the metric.

We can now interpret the result. The key observation is that the metric in Eddington-Finkelstein coordinates is completely regular at $r = 2GM$. Nothing singular occurs in the coefficients of the line element. The light cones therefore do not collapse or “pinch off” at $r = 2GM$. However, their orientation changes in an essential way.

For $r > 2GM$, the function $f(r) = 1 - 2GM/r$ is positive, and the outgoing null rays satisfy

$$\frac{dv}{dr} > 0.$$

The future-directed light cone therefore contains directions of both increasing and decreasing r , just as in flat spacetime. Note that in these diagrams we are plotting v vertically and r horizontally, so

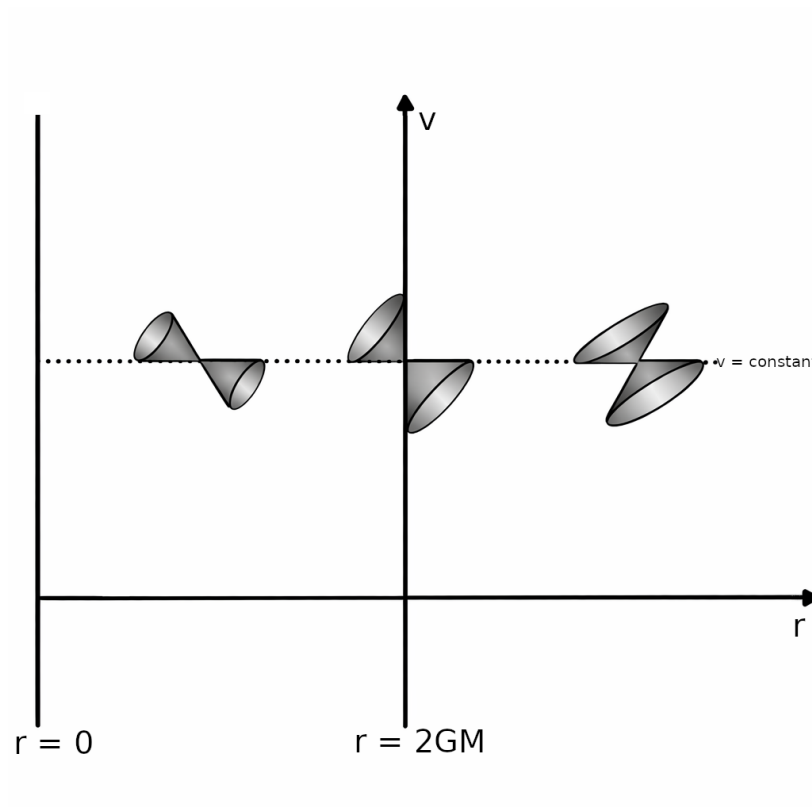


Figure 4.7: Schwarzschild light cones in the (v, r) coordinates of (4.84). In these coordinates we can follow future-directed timelike paths past $r = 2GM$.

the visual slope corresponds to dv/dr rather than the more physically intuitive dr/dv ; this reverses the usual convention outside of spacetime diagrams, where one typically thinks in terms of how r evolves with increasing time. This means that even the outgoing null direction now points toward decreasing r .

At $r = 2GM$, $f(r) = 0$ and the outgoing branch formally has infinite slope, so the light cone tips over and becomes tangent to the surface $r = 2GM$.

For $r < 2GM$, we have $f(r) < 0$, and therefore

$$\frac{dv}{dr} < 0.$$

In other words, *all future-directed null curves inside the horizon move toward smaller r* . Since timelike vectors must lie inside the null cone, the same is true for all future-directed timelike worldlines.

Thus the light cones do not “close up,” but they tilt so far inward that increasing r is no longer a future-directed possibility once $r < 2GM$. The causal structure itself forbids escape.

The surface $r = 2GM$ is therefore not merely a coordinate artifact. It is a *null hypersurface*: its normal vector is itself null. Because it is generated by null curves, it forms a causal boundary. No future-directed causal curve (null or timelike) that crosses this surface inward can later return to larger r .

This surface is called the **event horizon**. A black hole is the region of spacetime that cannot send signals to future null infinity. The event horizon is the boundary of that region.

It is important to emphasize that nothing singular happens to the local gravitational field at $r = 2GM$. For a static observer outside a Schwarzschild black hole, the spacetime geometry depends only on M . Replacing the Sun with a black hole of the same mass would not change the orbital motion of the Earth, provided the radius remained outside $2GM$.

The misleading Newtonian analogy arises from the escape-velocity formula

$$v_{\text{esc}} = \sqrt{\frac{2GM}{r}},$$

which suggests that when $v_{\text{esc}} = c$ something catastrophic occurs. In general relativity, however, the horizon is not defined by a force becoming infinite, nor by a particle failing to achieve sufficient speed. It is defined by the global causal structure of spacetime: inside $r = 2GM$, the coordinate r becomes timelike and future evolution necessarily proceeds toward smaller r .

4.5 The Maximally Extended Schwarzschild Solution

In the (v, r) coordinate system we can cross the event horizon on future-directed paths, but not on past-directed ones.

If we had chosen u instead of v , the metric would have been,

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) du^2 - (du dr + dr du) + r^2 d\Omega^2. \quad (4.85)$$

Now we can pass through the event horizon once again, but this time only along past-directed curves as shown in Fig. 4.8.

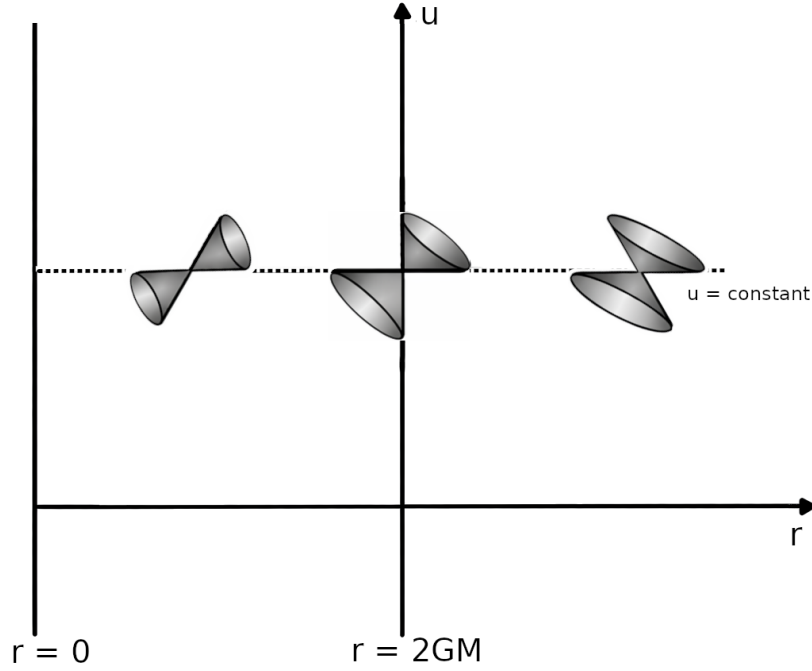


Figure 4.8: Schwarzschild light cones in the (u, r) coordinates of (4.85). In these coordinates we can follow past-directed timelike paths past $r = 2GM$.

We can follow either future-directed or past-directed curves through $r = 2GM$ but we arrive at different places. From the definitions (4.77), if v is held constant and we decrease r we must have $t \rightarrow +\infty$ while if we keep u constant and decrease r we have $t \rightarrow -\infty$. ($r^* \rightarrow -\infty$ as $r \rightarrow 2GM$). So we have extended spacetime in two different directions, one to the future and one to the past.

Let's try to define coordinates that are good all over. A first guess might be to use both u and v at once (in place of t and r) which leads to,

$$ds^2 = -\frac{1}{2} \left(1 - \frac{2GM}{r}\right) (dv du + du dv) + r^2 d\Omega^2,$$

with r defined in terms of v and u by,

$$\frac{1}{2}(v - u) = r + 2GM \ln \left(\frac{r}{2GM} - 1 \right).$$

However, we have reintroduced the degeneracy we started out with. In these coordinates $r = 2GM$ is "infinitely far away" (at either $v = \infty$ or $u = \infty$). Instead let's try,

$$\begin{aligned}v' &= e^{v/4GM} \\ u' &= -e^{-u/4GM}\end{aligned}$$

which in terms of the original (t, r) system,

$$\begin{aligned}v' &= \left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r+t)/4GM} \\ u' &= -\left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r-t)/4GM}.\end{aligned}$$

In the (v', u', θ, ϕ) system the Schwarzschild metric is,

$$ds^2 = -\frac{16G^3M^3}{r} e^{-r/2GM} (dv' du' + du' dv') + r^2 d\Omega^2.$$

Notice that at $r = 2GM$ none of the metric coefficients behave in a special way at the event horizon. Let's go a step further and build a system from this where one coordinate is timelike and the rest are spacelike,

$$T = \frac{1}{2}(v' + u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \sinh\left(\frac{t}{4GM}\right) \quad (4.86)$$

and

$$R = \frac{1}{2}(v' - u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \cosh\left(\frac{t}{4GM}\right). \quad (4.87)$$

Now the metric becomes,

$$ds^2 = \frac{32G^3M^3}{r} e^{-r/2GM} (-dT^2 + dR^2) + r^2 d\Omega^2, \quad (4.88)$$

where r is defined from,

$$T^2 - R^2 = \left(1 - \frac{r}{2GM}\right) e^{r/2GM}. \quad (4.89)$$

These coordinates (T, R, θ, ϕ) are known as **Kruskal coordinates**, or sometimes Kruskal-Szekeres coordinates.

The Kruskal coordinates, like the (t, r^*) coordinates, have radial null curves that look like they do in flat space:

$$T = \pm R + \text{constant}.$$

Unlike the (t, r^*) coordinates, the event horizon $r = 2GM$ is not infinitely far away and is defined by,

$$T = \pm R, \quad (4.90)$$

consistent with it being a null surface. More generally, we can consider the surfaces $r = \text{constant}$. From (4.89) these satisfy,

$$T^2 - R^2 = \text{constant}.$$

Thus they appear as a hyberbolae in the $R - T$ plane and the surfaces of constant t are given by,

$$\frac{T}{R} = \tanh\left(\frac{t}{4GM}\right), \quad (4.91)$$

which defines straight lines through the origin with slope $\tanh(t/4GM)$. As $t \rightarrow \pm\infty$, (4.91) becomes (4.90); therefore $t = \pm\infty$ represents the same surface as $r = 2GM$.

The coordinates (T, R) should be allowed to range over every value they can take without hitting the real singularity at $r = 0$,

$$\begin{aligned} -\infty &\leq R \leq \infty \\ T^2 &< R^2 + 1. \end{aligned}$$

From (4.86) and (4.87), T and R seem to become imaginary for $r < 2GM$, but this is an illusion because in that region (r, t) are no good because $|t| > \infty$. We can draw a spacetime diagram in the $T - R$ plane (with θ and ϕ supressed), known as a **Kruskal diagram** as shown in Fig. 4.9.

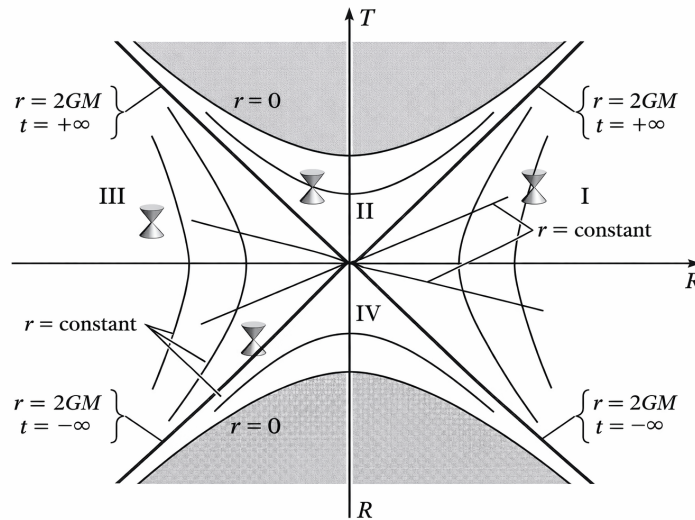


Figure 4.9: The Kruskal diagram – the Schwarzschild solution in Kruskal coordinates, where all light cones are at $\pm 45^\circ$.

4.6 Physical Relalizations: Stars and Black Holes

Appendix A

Mathematical Methods and Tensor Identities

A.1 Derivation Cheat Sheet

A.1.1 Covariant Derivatives in Components

We derive here the component expression for the second covariant derivative of a covector field, used in Eq. 2.48.

Recall the general rule:

If T_{bc} is a $(0, 2)$ tensor, then

$$\nabla_a T_{bc} = \partial_a T_{bc} - \Gamma^e_{ab} T_{ec} - \Gamma^e_{ac} T_{be}.$$

Each lower index contributes one Christoffel correction term.

Now observe that $\nabla_b \omega_c$ has two lower indices (b, c) . Therefore it is a $(0, 2)$ tensor, and differentiating it must produce two correction terms — one for each lower index.

Applying the rule directly gives

$$\nabla_a \nabla_b \omega_c = \partial_a (\nabla_b \omega_c) - \Gamma^e_{ab} \nabla_e \omega_c - \Gamma^e_{ac} \nabla_b \omega_e,$$

which is precisely Eq. 2.48.

Takeaway: *A covariant derivative introduces one Christoffel correction term for each tensor index.*

A.2 Useful Identities and Tensor Manipulations

A.2.1 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.

A.2.2 Minkowski Metric and Raising/Lowering

the textbook uses signature $(-+++)$:

$$\eta_{ab} = \text{diag}(-1, 1, 1, 1), \quad \eta^{ab} = \text{diag}(-1, 1, 1, 1).$$

Lowering a vector:

$$V_a = \eta_{ab} V^b.$$

Raising a covector:

$$W^a = \eta^{ab} W_b.$$

Special cases:

$$V_0 = -V^0, \quad V_i = V^i \quad (i = 1, 2, 3).$$

A.2.3 Contractions

General contractions:

$$T^a{}_a = \eta^{ab} T_{ba}, \quad T_{ab} V^a W^b, \quad R^a{}_{0a0}.$$

Contraction always means: **raise then sum** (or sum directly if mixed).

A.2.4 Trace Reversal (Eq. 3.108)

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:

$$\bar{\gamma} = -\gamma, \quad (\gamma = 0) \Rightarrow \bar{\gamma}_{ab} = \gamma_{ab}.$$

A.2.5 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:

$$\partial^a \bar{\gamma}'_{ab} = \partial^a \bar{\gamma}_{ab} + \square \xi_b.$$

Residual gauge freedom:

$$\square \xi_b = 0.$$

A.2.6 Metric as a Bilinear Map (Geometric Meaning)

Lower a vector:

$$V_b = g_{ba} V^a \quad (\text{'flat' operator}).$$

Raise a covector:

$$\alpha^a = g^{ab} \alpha_b \quad (\text{'sharp' operator}).$$

Used in:

- switching between γ_{ab} , γ^{ab} , $\bar{\gamma}_a{}^b$
- manipulating stress–energy components
- Riemann tensor computations

A.2.7 Linearized Riemann Tensor (Used in Geodesic Deviation)

$$R_{abcd}^{(1)} = \frac{1}{2} (\partial_c \partial_b \gamma_{ad} + \partial_d \partial_a \gamma_{bc} - \partial_c \partial_a \gamma_{bd} - \partial_d \partial_b \gamma_{ac}).$$

Raise first index:

$$R^a{}_{bcd} = \eta^{ae} R_{ebcd}.$$

Used to show:

$$R^i{}_{0j0} = -\frac{1}{2}\ddot{\gamma}^{\text{TT}i}{}_j.$$

A.2.8 Plane-Wave Polarizations

Plane wave:

$$\gamma_{\mu\nu} = A_{\mu\nu}e^{ik_ax^a}.$$

Null condition:

$$k_ak^a = 0.$$

Lorenz transversality:

$$k^a A_{ab} = 0.$$

Radiation gauge:

$$A_{0\mu} = 0, \quad A_{3\mu} = 0.$$

Tracelessness:

$$A^\mu{}_\mu = 0.$$

Remaining components:

$$A_{11}, A_{22}, A_{12} = A_{21}, \quad A_{11} + A_{22} = 0.$$

These form the $+$ and $\times$ polarization tensors.

A.2.9 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 the textbook's equation (4.4.46).

A.2.10 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	$A^\mu{}_\mu = 0, \quad k^\mu A_{\mu\nu} = 0$	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.

Appendix B

Code Listings

B.1 Effective Potentials in Newtonian Gravity

The following Python script generates effective-potential plots for massless and massive particles in Newtonian gravity. See Listing B.1.

```

1  import numpy as np
2  import matplotlib.pyplot as plt
3
4  GM = 1.0
5  Ls = range(1, 6)
6
7  # Start r closer to zero to avoid the artificial cutoff at r=0.5
8  r = np.linspace(0.2, 30, 3000)
9
10 def V_newton_massless(r, L):
11     return L**2/(2*r**2)
12
13 def V_newton_massive(r, L):
14     return 0.5 + L**2/(2*r**2) - GM/r
15
16 # Plot 1: Newtonian gravity, massless particles
17 plt.figure()
18 for L in Ls:
19     plt.plot(r, V_newton_massless(r, L), label=f"L={L}")
20 plt.xlim(0, 30)
21 plt.ylim(0, 0.8)
22 plt.xlabel("r")
23 plt.ylabel("V(r)")
24 plt.title("Newtonian gravity: massless particles")
25 plt.legend()
26 plt.grid(True, alpha=0.3)
27 plt.show()
28
29 # Plot 2: Newtonian gravity, massive particles

```

```

30 plt.figure()
31 for L in Ls:
32     plt.plot(r, V_newton_massive(r, L), label=f"L={L}")
33 plt.xlim(0, 30)
34 plt.ylim(0, 0.8)
35 plt.xlabel("r")
36 plt.ylabel("V(r)")
37 plt.title("Newtonian gravity: massive particles")
38 plt.legend()
39 plt.grid(True, alpha=0.3)
40 plt.show()

```

Listing B.1: Effective Potentials in Newtonian Gravity

B.2 Effective Potentials in General Relativity

The following Python script generates Schwarzschild effective-potential plots for timelike and null geodesics. See Listing B.2.

```

1  import numpy as np
2  import matplotlib.pyplot as plt
3
4  # Geometric units: G = c = 1
5  M = 1.0
6  Ls = range(1, 6)
7
8  # Schwarzschild coordinate radius (avoid the horizon r=2M for plotting)
9  r = np.linspace(2.05*M, 30.0, 3000)
10
11 def V_eff_schwarzschild(r, L, eps, M=1.0):
12     """
13     Carroll-style effective potential for Schwarzschild geodesics:
14      $(1/2) (dr/d\lambda)^2 + V_{\text{eff}}(r) = (1/2) E^2$ 
15
16     eps = 1 for timelike geodesics (massive particles)
17     eps = 0 for null geodesics (massless particles / photons)
18     """
19     return 0.5 * (1.0 - 2.0*M/r) * (eps + (L**2)/(r**2))
20
21 # ---- Plot 1: Massive particles (timelike), eps = 1 ----
22 plt.figure()
23 for L in Ls:
24     plt.plot(r, V_eff_schwarzschild(r, L, eps=1.0, M=M), label=f"L={L}")
25 plt.xlim(2.0, 30.0)
26 plt.ylim(0.0, 0.8)
27 plt.xlabel("r / M")
28 plt.ylabel("V(r)")
29 plt.title("Schwarzschild effective potential: massive particles
30         (eps=1)")
31 plt.legend()

```

```
31 plt.grid(True, alpha=0.3)
32 plt.show()
33
34 # ---- Plot 2: Massless particles (null), eps = 0 ----
35 plt.figure()
36 for L in Ls:
37     plt.plot(r, V_eff_schwarzschild(r, L, eps=0.0, M=M), label=f"L={L}")
38 plt.xlim(2.0, 30.0)
39 plt.ylim(0.0, 0.8)
40 plt.xlabel("r / M")
41 plt.ylabel("V(r)")
42 plt.title("Schwarzschild effective potential: massless particles
43           (eps=0)")
43 plt.legend()
44 plt.grid(True, alpha=0.3)
45 plt.show()
```

Listing B.2: Effective Potentials in General Relativity