The Geometry of General Relativity

Sam King

King Monographs on Mathematical Physics

King Monographs on Mathematical Physics

The Geometry of General Relativity

This page begrudgingly left blank.

Introduction

This document develops general relativity from the ground up, targeting physicists who want genuine understanding rather than just computational facility. The niche it occupies is narrow but real: most physics-oriented texts (Carroll, Wald) treat the geometry as a tool to reach the physics quickly and sacrifice either rigor or intuition to get there. Most mathematics-oriented texts (Lee, Spivak) are rigorous but don't connect the formalism to the physical and classical intuitions that make it meaningful. Existing treatments in the physics and math literature tend to be encyclopedic in nature, containing examples and topics which, while of interest to practitioners, may obscure the elegant theoretical spine of the material. This document tries to address all of these issues.

Our construction is based upon a philosophy of rigorous minimalism. We carefully introduce and develop layers of structure, and nothing is included for completeness. Every topic is present because a later part relies on it, or because it earns its place by genuinely illuminating the geometry. Part I builds the metric-free foundation: manifolds, the tangent and cotangent spaces, tensors, differential forms, the exterior derivative, and the generalized Stokes theorem. A metric is conspicuously absent throughout, and deliberately so. Working without one forces every definition to rest on smooth structure alone. Part II then adds the metric, and from that single addition the entire geometric apparatus follows: a unique connection, covariant derivatives, parallel transport, geodesics, and curvature. Part III discusses essential aspects of physics in curved spacetime, culminating with a detailed discussion of general relativity. A cohesive set of appendices illuminates various aspects of the underlying classical field theoretic apparatus.

On rigor: we aim for the level of a careful first graduate course in mathematical physics, with precise definitions, explicit calculations, and real proofs where they illuminate rather than obscure. We do not aim for full measure-theoretic or topological generality.

Contents

Part I

Smooth Structure

Introduction to Part I

Before we begin developing the mathematics, it is worth being clear about what differential geometry is and how it relates to other things called "rigorous calculus." Classical calculus was powerful but imprecise, built on infinitesimals and geometric intuition that resisted formal justification. Real analysis, developed in the 19th century, made calculus rigorous on $\mathbb{R}^n$. It gave precise definitions of limits, continuity, and differentiability, and put integration on solid measure-theoretic foundations. Real analysis, however, is tied to flat space. It tells you how to do calculus on open subsets of $\mathbb{R}^n$, not on a sphere or a curved spacetime.

Differential geometry takes real analysis as its foundation and asks: what does calculus look like on a space that is only locally flat? All the smoothness conditions in differential geometry are real analysis conditions applied in coordinate charts. The exterior derivative is computed in coordinates using real analysis. The proof of Stokes' theorem reduces to the fundamental theorem of calculus. The "differential geometry" part is the insistence on coordinate-independence: finding definitions of vectors, covectors, derivatives, and integrals that make sense intrinsically on the manifold, not just in a particular chart. This is the sense in which DG extends and clarifies classical calculus, not by providing the ε-δ foundations that real analysis provides, but by identifying the correct geometric homes for objects like df that classical calculus described only informally.

Part I covers the metric-free aspects of calculus on manifolds: tensors, differential forms, the exterior derivative, pullbacks, integration, and a proof of the generalized Stokes theorem. Everything here requires only a smooth structure and an orientation.

---

1 Manifolds

1.1 Motivation

Physics takes place on spaces that are not always flat. The surface of a sphere, the spacetime of general relativity, the configuration space of a mechanical system: these are curved spaces that cannot be described by a single global coordinate system. The theory of manifolds is the precise framework for doing calculus on such spaces.

The key insight is that even though a curved space has no global coordinates, it looks flat in small enough regions. The surface of the Earth is curved globally, but locally it looks like a flat plane (which is why maps work). A manifold is a space that is locally flat in exactly this sense, with enough structure to do calculus consistently across overlapping local regions.

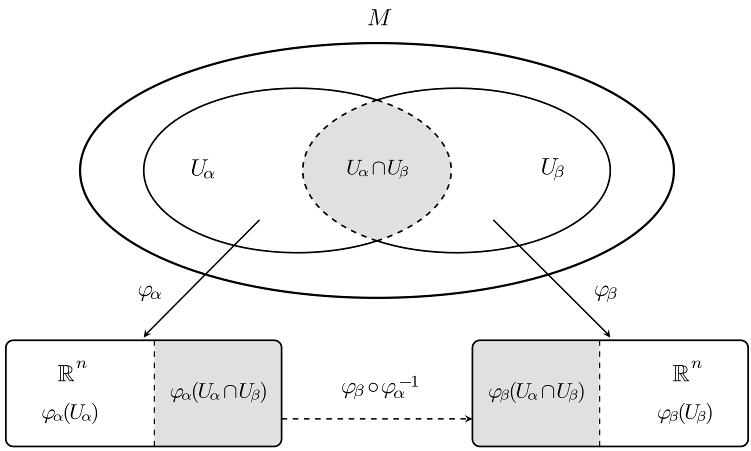
1.2 Definition

A smooth n-dimensional **manifold** is a (Hausdorff and second countable) topological space M equipped with a collection of **coordinate charts** $\{(U_\alpha, \phi_\alpha)\}$, where each $U_\alpha \subset M$ is an open set and $\phi_\alpha \colon U_\alpha \to \mathbb{R}^n$ is a homeomorphism onto an open subset of $\mathbb{R}^n$, such that:

1. The collection of charts $\{U_\alpha\}$, called an **atlas**, covers M, i.e. $M = \bigcup_\alpha U_\alpha$, and
2. All maps $\phi_\beta \circ \phi_\alpha^{-1} \colon \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$ between overlapping charts, called **transition functions**, are infinitely differentiable ("smooth").

A **homeomorphism** is a continuous bijection with a continuous inverse. It provides the topological notion of "same shape," meaning the two spaces look identical from a purely continuous perspective, without any notion of smoothness or distance. Requiring infinitely differentiable transition functions upgrades the homeomorphisms to **diffeomorphisms**.

The coordinates $x^1, \ldots, x^n$ of a point $p \in U_\alpha$ are the components of $\phi_\alpha(p) \in \mathbb{R}^n$. Different charts give different coordinates for the same point. The transition functions give us a way to pass between the coordinate systems of overlapping charts U_α and U_β: $\phi_\beta \circ \phi_\alpha^{-1}$ is simply a map between open subsets of $\mathbb{R}^n$, ordinary multivariable calculus territory. This provides a consistent notion of "smoothness" on M, because changing charts involves a smooth transition function.



1.3 Examples

Euclidean space $\mathbb{R}^n$ is a manifold covered by a single chart, the identity map. It is the trivial case.

The sphere S^2 is a 2-manifold. No single coordinate chart covers it (any map of the Earth has at least one point of distortion or discontinuity), but two charts suffice: stereographic projection from the north and south poles. The transition function between them is smooth.

Spacetime in general relativity (GR) is a 4-dimensional Lorentzian manifold. It is locally flat (special relativity holds in small enough regions) but globally curved by the distribution of matter and energy.

1.4 Smooth Functions and the Smooth Structure

A function $f\colon M \to \mathbb{R}$ is **smooth** if for every chart (U_α, ϕ_α), the composition $f \circ \phi_\alpha^{-1}\colon \mathbb{R}^n \to \mathbb{R}$ is smooth in the ordinary sense. The smoothness of transition functions ensures this definition is chart-independent. The collection of all smooth functions on M is denoted $C^\infty(M)$.

The manifold itself has no preferred coordinates, no metric, no notion of distance. Just this smooth structure, which tells you what it means for a function to be differentiable.

1.5 Compactness

A topological space is **compact** if every open cover has a finite subcover. Intuitively this means the space does not escape to infinity and has no missing boundary points. In $\mathbb{R}^n$ this reduces to closed and bounded (a sphere is compact, a plane is not). A manifold *with* boundary can also be compact. The closed interval $[a, b]$ is the simplest example, and the generalized Stokes theorem applies to compact manifolds with boundary, as stated in §11.

2 The Tangent Space

2.1 Tangent Vectors

At each point $p \in M$ we want to define "directions you can move." The naive definition of an arrow pointing somewhere does not make sense on an abstract manifold because there is no ambient space to point into. The correct definition: a **tangent vector** at p is a **derivation** on smooth functions, i.e. a linear map $v\colon C^\infty(M) \to \mathbb{R}$ satisfying the Leibniz rule:

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

This captures the idea of "differentiate in a direction" without needing an ambient space. In local coordinates $\{x^\mu\}$, the partial derivatives $\partial/\partial x^\mu$ are derivations, and they form a **basis** for the vector space of derivations at p. We adopt the **Einstein summation convention**, where repeated indices, one upper and one lower, are implicitly summed over. A general tangent vector is then:

$$v = v^\mu \frac{\partial}{\partial x^\mu} =: v^\mu \partial_\mu,$$

where the RHS definition establishes standard notation. The coefficients v^μ are the components of the tangent vector in this coordinate basis. The **tangent space** T_pM is the vector space of all tangent vectors at p, with dimension n. The set of all tangent spaces is known as the **tangent bundle** $TM := \{(p, v) \mid p \in M,\ v \in T_pM\}$.

Before going further we must prove the intuitive fact that $\{\partial_\mu\}$ indeed forms a good basis for the vector space of derivations. Let v be any derivation and work WLOG in a chart with $x^\mu(p) = 0$. The proof proceeds in four steps.

Step 1: constants vanish. From Leibniz, $v(1) = v(1 \cdot 1) = 1 \cdot v(1) + 1 \cdot v(1) = 2v(1) \implies v(1) = 0$, and by linearity $v(c) = cv(1) = 0 \quad \forall c \in \mathbb{R}$.

Step 2: smooth decomposition. For any $f \in C^\infty(M)$ and any $q \in U$, let $\tilde{f} := f \circ \phi^{-1}$ and write $\phi(q) = (x^1(q), \ldots, x^n(q))$ for the coordinate vector of q. Consider the straight-line path $s \mapsto s\phi(q)$ in $\mathbb{R}^n$ from $\phi(p) = 0$ to $\phi(q)$, for $s \in [0, 1]$. The fundamental theorem of calculus gives:

$$\tilde{f}(\phi(q)) - \tilde{f}(0) = \int_0^1 ds\, \frac{d}{ds} \tilde{f}(s\phi(q)),$$

and the chain rule gives $\frac{d}{ds}\tilde{f}(s\phi(q)) = \partial_\mu \tilde{f}\big|_{s\phi(q)} \cdot x^\mu(q)$, so:

$$f(q) = f(p) + x^\mu(q) \int_0^1 ds\, \partial_\mu \tilde{f}\big|_{s\phi(q)}.$$

Define $h_\mu(q) := \int_0^1 ds\, \partial_\mu \tilde{f}\big|_{s\phi(q)}$, which are smooth functions on U satisfying $h_\mu(p) = \partial_\mu f\big|_p$. Then $f(q) = f(p) + x^\mu(q)\, h_\mu(q) \quad \forall q \in U$.

Step 3: compute the derivation. Now apply v to our expansion of f from step 2. By linearity and step 1:

$$v(f) = \underbrace{v(f(p))}_{=0} + v(x^\mu)\, h_\mu(p) + \underbrace{x^\mu(p)}_{=0}\, v(h_\mu) = v(x^\mu)\, \partial_\mu f\big|_p.$$

Setting $v^\mu := v(x^\mu)$, we get $v(f) = v^\mu \partial_\mu f\big|_p$ for every f, so $v = v^\mu \partial_\mu$.

Step 4: establish linear independence. Suppose $c^\mu \partial_\mu \equiv 0$ as a derivation (for some constants $c^\mu \in \mathbb{R}$), meaning it returns zero on every smooth function. Evaluating on the coordinate function x^ν:

$$(c^\mu \partial_\mu)(x^\nu) = c^\mu \cdot \partial_\mu(x^\nu) = c^\mu \cdot \frac{\partial x^\nu}{\partial x^\mu} = c^\mu \delta^\nu_\mu = c^\nu = 0.$$

We see that $\{\partial_\mu\}$ is both spanning and linearly independent and is therefore a basis for T_pM. $\square$

2.2 Cotangent Vectors

The **cotangent space** T_p^*M is the **dual space** to T_pM: the vector space of all linear maps $\omega\colon T_pM \to \mathbb{R}$. Given a smooth function $f\colon M \to \mathbb{R}$, its **differential** df at p is the cotangent vector defined by:

$$df(v) = v(f),$$

for any tangent vector v. This takes a tangent vector and returns the directional derivative of f in that direction, a number. Applied to the coordinate functions x^μ themselves, the differentials dx^μ form the **dual basis** to $\{\partial/\partial x^\mu\}$:

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

A general cotangent vector (also called a **1-form**) is $\omega = \omega_\mu \, dx^\mu$, and its action on a tangent vector $v = v^\nu \partial_\nu$ is:

$$\begin{aligned}\omega(v) &= \omega_\mu \, dx^\mu(v^\nu \partial_\nu) = \omega_\mu v^\nu \, dx^\mu(\partial_\nu) = \omega_\mu v^\nu \delta^\mu{}_\nu \\ &= \omega_\mu v^\mu \in \mathbb{R}.\end{aligned}$$

3 Flat Space: Classical Calculus Revisited

3.1 Basis Vectors as Differential Operators

Before going further it is worth understanding what is happening in ordinary $\mathbb{R}^n$, because this is where the classical and modern pictures meet, and where physicists are usually not shown the connection.

In $\mathbb{R}^n$ you have a completely canonical way to differentiate a function f in the direction of a vector v:

$$D_v f = \lim_{\epsilon \to 0} \frac{f(p + \epsilon v) - f(p)}{\epsilon} = v^\mu \partial_\mu f.$$

The map $f \mapsto v^\mu \partial_\mu f$ is linear and satisfies Leibniz, making it a derivation. Crucially, different vectors give different derivations, and every derivation arises this way. So there is a perfect correspondence:

$$v^\mu \hat{e}_\mu \longleftrightarrow v^\mu \partial_\mu.$$

This is why the coordinate basis vectors are identified with partial derivatives. It is not saying basis vectors *are* differential operators in some mysterious sense. It is saying that **the only intrinsic thing a tangent vector does is differentiate functions**, and in $\mathbb{R}^n$ that action is in perfect bijection with the naive arrow picture. The arrow v and the directional derivative D_v carry exactly the same information.

The reason this feels strange is that in $\mathbb{R}^n$, we are accustomed to tangent vectors having a life independent of functions. They point somewhere in space, but on a general manifold there is no ambient space to point into, so the directional derivative action is all there is. The $\mathbb{R}^n$ picture is secretly already this, just disguised by the familiar geometry.

3.2 The Differential as a Cotangent Vector

Now look at the classical expression $df = \frac{\partial f}{\partial x^\mu} dx^\mu$. In classical calculus this is treated as an infinitesimal, something vaguely defined that becomes rigorous only when you integrate it or cancel it against another d-something. The modern picture gives it a precise identity: df is a **cotangent vector**, a linear map that takes a tangent vector and returns a number. Feed it $v = v^\mu \partial_\mu$:

$$df(v) = \frac{\partial f}{\partial x^\mu} v^\mu.$$

This is simply the directional derivative of f in the direction v, a perfectly concrete number: if v is a velocity vector, $df(v)$ is the rate of change of f along that motion. The mysterious infinitesimal df was always this linear map. It just needed a tangent vector to act on.

3.3 The Differential vs. the Gradient

In flat space with Euclidean metric the components of df are $\partial_\mu f$ and the components of ∇f are also $\partial_\mu f$, numerically identical. But df is a cotangent vector that acts on a tangent vector to give a number, while ∇f is itself a tangent vector that you dot with another vector to give a number. The metric $g_{\mu\nu} = \delta_{\mu\nu}$ is doing the conversion between them invisibly, which is why classical calculus never needed to distinguish the two. In curved space $g_{\mu\nu}$ is nontrivial and the distinction becomes essential.

3.4 The Modern Picture: The Chain Rule as the Duality Pairing

With this in place, the chain rule becomes transparent. Consider a curve $\gamma\colon \mathbb{R} \to M$ with parameter λ, and let $D_\lambda = d/d\lambda$ denote its tangent vector:

$$D_\lambda = \frac{dx^\mu}{d\lambda}\frac{\partial}{\partial x^\mu},$$

with components $dx^\mu/d\lambda$, the coordinate velocities along the curve. We can then write:

$$\frac{df}{d\lambda} = \frac{\partial f}{\partial x^\mu}\frac{dx^\mu}{d\lambda}.$$

In the modern picture this is just the duality pairing between the cotangent vector df and the tangent vector D_λ:

$$\boxed{\frac{df}{d\lambda} = df(D_\lambda) = \frac{\partial f}{\partial x^\mu}\frac{dx^\mu}{d\lambda}.}$$

The right side is the pairing $\omega_\mu v^\mu$ with $\omega_\mu = \partial_\mu f$ and $v^\mu = dx^\mu/d\lambda$. The chain rule **is** the duality pairing between T_pM and T_p^*M.

3.5 What Classical Calculus Was Doing

In classical calculus this same result is sometimes hand-waved by writing $df = \frac{\partial f}{\partial x^\mu} dx^\mu$ and then "dividing by $d\lambda$," treating the dx^μ as numbers that cancel. The modern picture explains why it works: you are not canceling infinitesimals, you are evaluating the cotangent vector df on the tangent vector D_λ. The classical df was always a cotangent vector. The modern framework just gives it a precise home.

---

4 Tensors

4.1 Definition

Given any vector space V and its dual V^*, a **tensor of type** (r, s) is a multilinear map taking r elements of V^* and s elements of V and returning a number. There is nothing specific to manifolds here; this is a construction in linear algebra that applies to any vector space.

On a manifold, we apply this at each point p with $V = T_pM$:

$$T\colon \underbrace{T_p^*M \times \cdots \times T_p^*M}_{r} \times \underbrace{T_pM \times \cdots \times T_pM}_{s} \to \mathbb{R}.$$

Tangent vectors are $(1, 0)$ tensors and cotangent vectors are $(0, 1)$ tensors.

4.2 The Tensor Product

Given an (r, s) tensor S and a (p, q) tensor T, their **tensor product** $S \otimes T$ is an $(r+p, s+q)$ tensor defined by evaluating each on its respective arguments and multiplying the results:

$$\begin{aligned}&(S \otimes T)(\omega^1, \ldots, \omega^r, \alpha^1, \ldots, \alpha^p,\, v_1, \ldots, v_s, u_1, \ldots, u_q)\\&\quad = S(\omega^1, \ldots, \omega^r, v_1, \ldots, v_s) \cdot T(\alpha^1, \ldots, \alpha^p, u_1, \ldots, u_q),\end{aligned}$$

where $\omega^i, \alpha^i \in T_p^*M$ and $v_i, u_i \in T_pM$. This has no symmetry, so $S \otimes T \neq T \otimes S$ in general.

The tensor products of basis vectors and covectors, $\partial_{\mu_1} \otimes \cdots \otimes \partial_{\mu_r} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_s}$, themselves form a basis for the space of all (r, s) tensors, which is a vector space in the usual sense. So the coordinate expansion:

$$\begin{aligned}T = T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s}\, &\partial_{\mu_1} \otimes \cdots \otimes \partial_{\mu_r}\\&\otimes\, dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_s}\end{aligned}$$

is exactly a basis expansion, with the components $T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s}$ playing the role of coordinates in that vector space.

A tensor at a single point p is a purely algebraic object. A **tensor field** is a smooth assignment of an (r, s) tensor to every point of the manifold. At each $p \in M$ one has a multilinear map on T_pM and T_p^*M, varying smoothly from point to point. In coordinates, this means the components $T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s}$ are smooth functions of the local coordinates x^μ. This is the object that typically appears in physics and geometry.

Note also that the pairing between a vector and a covector is symmetric in the sense that it does not matter which one you think of as "acting." A $(1,0)$ tensor acting on a cotangent vector gives the same number as the cotangent vector acting on that tangent vector. Both are simply the duality pairing $\langle \omega, v \rangle = \omega_\mu v^\mu$. The distinction between who is acting and who is being acted on is a matter of perspective, not mathematics.

---

5 Coordinate Transformations

A tensor is a coordinate-independent geometric object. It is the same map between vectors and numbers regardless of which chart you use. This forces a specific relationship between how components in different coordinate systems must relate to each other.

5.1 Tangent Vectors

A tangent vector v has components $v^\mu = dx^\mu/d\lambda$ in coordinates $\{x^\mu\}$. Under a change to coordinates $\{x'^\mu\}$, the chain rule gives:

$$v'^\mu = \frac{dx'^\mu}{d\lambda} = \frac{\partial x'^\mu}{\partial x^\nu}\frac{dx^\nu}{d\lambda} = \frac{\partial x'^\mu}{\partial x^\nu} v^\nu.$$

So tangent vector components transform by multiplying by the Jacobian $\partial x'^\mu/\partial x^\nu$ of the coordinate change. In general, components that transform this way (with the forward Jacobian) are called **contravariant**.

5.2 Cotangent Vectors

A cotangent vector ω has components $\omega_\mu = \partial_\mu f$ (for example). Under the same coordinate change:

$$\omega'_\mu = \frac{\partial f}{\partial x'^\mu} = \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial f}{\partial x^\nu} = \frac{\partial x^\nu}{\partial x'^\mu}\omega_\nu.$$

Cotangent components transform by the inverse Jacobian $\partial x^\nu/\partial x'^\mu$. In general, components that transform this way (with the inverse Jacobian) are called **covariant**.

5.3 Why the Pairing Is Invariant

The duality pairing $\omega_\mu v^\mu$ is coordinate-independent because the two Jacobians cancel:

$$\omega'_\mu v'^\mu = \frac{\partial x^\nu}{\partial x'^\mu}\omega_\nu \cdot \frac{\partial x'^\mu}{\partial x^\rho} v^\rho = \delta^\nu{}_\rho\, \omega_\nu v^\rho = \omega_\nu v^\nu.$$

This is not a coincidence; it is forced by the requirement that the pairing return the same number in every coordinate system. The covariant and contravariant transformation laws are precisely defined to make this true.

5.4 General Tensors

A general (r, s) tensor has components that transform with r forward Jacobians (one for each upper index) and s inverse Jacobians (one for each lower index):

$$T'^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s} = \frac{\partial x'^{\mu_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial x'^{\mu_r}}{\partial x^{\alpha_r}} \cdot \frac{\partial x^{\beta_1}}{\partial x'^{\nu_1}} \cdots \frac{\partial x^{\beta_s}}{\partial x'^{\nu_s}} \cdot T^{\alpha_1 \cdots \alpha_r}{}_{\beta_1 \cdots \beta_s}.$$

In GR this is often taken as the definition of a tensor: a tensor is precisely an object whose components transform this way. The geometric definition (a multilinear map on T_pM and T_p^*M) and the transformation law definition are equivalent.

5.5 Lorentz Transformations as a Special Case

The general Jacobian $\partial x'^{\mu}/\partial x^{\nu}$ can be anything, varying from point to point. Lorentz transformations are the special case where the coordinate change is linear, $x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}$, so the Jacobian is constant everywhere and equals $\Lambda^{\mu}{}_{\nu}$ itself. The tensor transformation law then corresponds to multiplication by Λ and its inverse, which is exactly how tensors transform in special relativity.

6 The Wedge Product

Among all $(0,k)$ tensors, the **antisymmetric** ones change sign whenever any two arguments are swapped:

$$\omega(v_1,\dots,v_i,\dots,v_j,\dots,v_k) = -\omega(v_1,\dots,v_j,\dots,v_i,\dots,v_k).$$

These are called **k-forms**: antisymmetric multilinear maps at a single point, a purely algebraic object. A differential k-form, defined in §7, is a smooth assignment of a k-form to every point of the manifold. Given a k-form α and an l-form β, the **wedge product** $\alpha\wedge\beta$ is their antisymmetrized tensor product:

$$\alpha\wedge\beta = \frac{(k+l)!}{k!\,l!}\,\mathrm{Alt}(\alpha\otimes\beta),$$

where Alt antisymmetrizes over all arguments by summing over all permutations with signs:

$$\begin{aligned}&\mathrm{Alt}(\alpha\otimes\beta)(v_1,\dots,v_{k+l})\\ &\qquad= \frac{1}{(k+l)!}\sum_{\sigma}\mathrm{sgn}(\sigma)\,(\alpha\otimes\beta)(v_{\sigma(1)},\dots,v_{\sigma(k+l)}),\end{aligned}$$

where the sum is over all $(k+l)!$ permutations σ of $\{1,\dots,k+l\}$, and $\mathrm{sgn}(\sigma)=\pm1$ is the sign of σ. The combinatorial prefactor corrects for overcounting: Alt sums over all $(k+l)!$ permutations, but the result is already antisymmetric in the first k and last l slots separately, so each distinct term appears $k!\,l!$ times. The prefactor removes this redundancy, leaving a correctly normalized $(k+l)$-form. The wedge product provides a standard mechanism for combining lower-rank forms into higher-rank antisymmetric forms, a construction whose utility will become clear shortly.

On basis covectors the antisymmetry gives:

$$dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu, \qquad dx^\mu \wedge dx^\mu = 0.$$

The wedge extends to multiple factors by iteration, with each new factor wedged on in turn.

$$dx^\mu \wedge dx^\nu \wedge dx^\rho = (dx^\mu \wedge dx^\nu) \wedge dx^\rho,$$

and so on for any number of factors. For general k- and l-forms α and β the antisymmetry generalizes to graded commutativity:

$$\alpha \wedge \beta = (-1)^{kl}\, \beta \wedge \alpha.$$

The action of a basis k-form on k vectors is a determinant:

$$\boxed{\begin{aligned}&(dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k})(v_1, \ldots, v_k)\\ &\quad = \det \begin{pmatrix} v_1^{\mu_1} & \cdots & v_1^{\mu_k} \\ \vdots & & \vdots \\ v_k^{\mu_1} & \cdots & v_k^{\mu_k} \end{pmatrix}.\end{aligned}}$$

The entries are the $\{\mu_1, \ldots, \mu_k\}$-components of each vector, giving the projection onto the chosen coordinate k-plane. The determinant gives the signed k-volume of the parallelepiped they span. This is not an analogy: the definition of the wedge product via Alt is precisely the Leibniz formula for the determinant. The exterior algebra is the algebra of determinants, which is why the wedge product is also called the **exterior product**. Note also that this determinant interpretation provides a simple way to see that $dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu$ and $dx^\mu \wedge dx^\mu = 0$. The table below provides several explicit examples.

k	**Example**	**Geometric Content**
1	$dx^\mu(v) = v^\mu$	Component of v in the μ direction; a 1×1 determinant.
2	$(dy \wedge dz)(v, w) = v^y w^z - v^z w^y$	Signed projected area of (v, w) onto the yz-plane.
3	$(dx \wedge dy \wedge dz)(u, v, w)$	Full 3×3 determinant; signed volume of the parallelepiped spanned by u, v, w.

7 Differential Forms

7.1 Definition

A **differential k-form** on M is a smooth assignment of a k-form to each point of M, specifically a smooth choice at every p of an antisymmetric multilinear map $T_pM \times \cdots \times$

$T_pM \to \mathbb{R}$. A differential k-form is a special case of a tensor field. It is a $(0, k)$ tensor field whose components are fully antisymmetric.

7.2 Smoothness

It is worth being precise about what this means. The tangent space T_pM is a different vector space at each point, so the basis covectors $dx^\mu|_p$ are literally different objects at different points. There is a separate dual basis at each p, and a differential k-form ω assigns to each p a k-form built from that point's basis covectors, with real-number coefficients that depend on p. In a coordinate chart, ω has component functions $\omega_{\mu_1 \cdots \mu_k}(p)$, and smoothness means these are smooth functions of the coordinates $x^1, \ldots, x^n$ on that chart, in the usual multivariable calculus sense. This condition is chart-independent. If the components are smooth in one chart, they are smooth in any overlapping chart, because the transition functions between charts are themselves smooth, which is exactly what it means for M to be a smooth manifold. Equivalently, for any smooth vector fields $v_1, \ldots, v_k$ on M, the function $p \mapsto \omega_p(v_1(p), \ldots, v_k(p))$ is a smooth function $M \to \mathbb{R}$.

Just as a vector field is a linear combination of basis vector fields ∂_μ with smooth coefficient functions, a differential k-form is a linear combination of **basis k-forms** $dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$ with smooth coefficient functions. The space of all differential k-forms on M is denoted $\Omega^k(M)$: concretely, all possible choices of smooth coefficient functions. It is a vector space over $\mathbb{R}$ (you can add forms and scale by constants), and infinite-dimensional globally since the coefficients can be any smooth functions. But at each individual point p, the space of k-forms is merely a finite-dimensional vector space of dimension $\binom{n}{k}$.

7.3 Examples in $\mathbb{R}^3$

In $\mathbb{R}^3$ the independent basis forms of each degree, and the corresponding general forms, are:

0-forms: a smooth function f. No arguments, just a number at each point.

1-forms: $f_1\,dx + f_2\,dy + f_3\,dz$, with f_1, f_2, f_3 arbitrary smooth functions. Takes one tangent vector, returns a number.

2-forms: $f_1\,dy \wedge dz + f_2\,dz \wedge dx + f_3\,dx \wedge dy$. Takes two tangent vectors, returns a weighted sum of projected areas.

3-forms: $f\,dx \wedge dy \wedge dz$. Takes three tangent vectors, returns their signed volume weighted by f.

7.4 Counting Basis Forms

The number of independent basis k-forms in $\mathbb{R}^n$ is $\binom{n}{k}$. You choose k coordinate directions from n, and antisymmetry means order does not matter up to sign. In $\mathbb{R}^3$: 0-forms have 1 component, 1-forms have 3, 2-forms have 3, and 3-forms have 1. The coincidence that 1-forms and 2-forms have the same number of components is special to 3 dimensions, and has significant consequences for how vector calculus works, as we will soon see. There are no nonzero k-forms for $k > n$. Repeating a basis element gives a repeated row in the determinant, which is zero.

7.5 Why Forms Are the Right Objects to Integrate

The fact that a k-form returns signed k-volumes is not incidental; it is precisely what makes forms the right objects to integrate. A length element, an area element, a volume element are all instances of the same thing: a k-form evaluated on k infinitesimal displacement vectors. This gives a single unified definition of integration over curves, surfaces, volumes, and arbitrary manifolds, as we will see in §8.

---

8 Integration

8.1 Orientation

An **orientation** on a manifold M is a consistent choice of "handedness" across all coordinate charts, concretely a choice of which basis orderings $(\partial_1, \ldots, \partial_k)$ count as positively oriented. On a curve this means a direction of travel; on a surface, which side is up; on a volume, a choice of right- vs. left-handed axes. Two charts are orientation-compatible if the Jacobian of their transition function has positive determinant. An orientable manifold is one that admits a consistent such choice globally. A Möbius strip is the classic example of a manifold that does not. The orientation of M also induces a natural orientation on its boundary ∂M via the outward normal convention, which we make precise in the proof of the generalized Stokes theorem in §11.

8.2 Definition

To integrate a k-form ω over an oriented k-dimensional submanifold M, work in a positively oriented coordinate chart with coordinates $x^1, \ldots, x^k$. In this chart ω has a single independent coefficient:

$$\omega = \omega_{12\cdots k}(x)\, dx^1 \wedge \cdots \wedge dx^k.$$

We define the integral:

$$\boxed{\int_M \omega := \int dx^1 \cdots dx^k \, \omega_{12\cdots k}(x),}$$

where the right-hand side is a standard multivariable integral, where the dx^i are ordinary Lebesgue measure factors, not wedge products. Orientation enters through the sign: swapping two coordinate directions flips the sign of the determinant and hence the sign of the integral, corresponding to integrating over the oppositely oriented manifold. On a non-orientable manifold no global sign convention exists, so forms cannot be integrated globally. No metric is required anywhere in this construction.

8.3 Why This Definition Makes Sense

Recall the intuitive heuristic underlying a Riemann integral: divide M into tiny coordinate boxes, evaluate something on each box, and sum. Each tiny box has edge vectors $\epsilon_1\partial_1, \ldots, \epsilon_k\partial_k$ where ϵ_i is the side length in the i-th direction (an ordinary real number, not a form). The natural thing to assign to this box is the value of ω on its edge vectors:

$$\omega(\epsilon_1\partial_1, \ldots, \epsilon_k\partial_k) = \omega_{12\cdots k}(x)\ \det\begin{pmatrix} \epsilon_1 & & \\ & \ddots & \\ & & \epsilon_k \end{pmatrix}$$
$$= \omega_{12\cdots k}(x)\,\epsilon_1 \cdots \epsilon_k,$$

where we used the determinant rule from §6 and the fact that the edge vectors point in orthogonal coordinate directions. Summing over all boxes and taking the limit $\epsilon_i \to dx^i$ (now Lebesgue measure factors) gives exactly the standard Riemann integral of $\omega_{12\cdots k}$. So the definition is the unique natural way to sum up what the form assigns to infinitesimal pieces of M.

8.4 Coordinate Independence

This might seem anticlimactic. After all that machinery, integrating a form just reduces to a standard integral in $\mathbb{R}^n$. But that is precisely the point. On a bare manifold there is no canonical way to integrate a plain function. If you pick different coordinates, you get a different number because you forgot the Jacobian. A differential form fixes this. As shown in §5, a k-form coefficient $\omega_{12\cdots k}$ has k lower indices, so it transforms with k inverse Jacobians; these k factors combine into the single determinant $\det(\partial x/\partial x')$ by the same antisymmetry of the wedge product that collapses k Jacobian factors into a determinant in the first place:

$$\omega'_{1\cdots k} = \frac{\partial x^{\mu_1}}{\partial x'^1} \cdots \frac{\partial x^{\mu_k}}{\partial x'^k}\,\omega_{\mu_1\cdots\mu_k} = \det\left(\frac{\partial x}{\partial x'}\right)\omega_{1\cdots k}.$$

The last step uses the fact that $\omega_{\mu_1\cdots\mu_k}$ is fully antisymmetric. The sum over all $\mu_1, \dots, \mu_k$ kills every repeated-index term and flips signs on transpositions, leaving exactly the alternating sum of products of Jacobian entries (i.e. the determinant). The Lebesgue measure $dx^1 \cdots dx^k$ transforms by the change of variables formula with the reciprocal factor $\det(\partial x'/\partial x)$. These cancel exactly, leaving $\int_M \omega$ the same number in every chart. The entire apparatus exists to guarantee that the reduction to a standard integral is coordinate-independent.

---

9 The Exterior Derivative

9.1 Definition

The **exterior derivative** d is a map $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$, defined on a k-form $\omega = \frac{1}{k!}\omega_{\mu_1\cdots\mu_k}\, dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$ by:

$$\boxed{d\omega = \frac{1}{k!}\frac{\partial \omega_{\mu_1\cdots\mu_k}}{\partial x^\nu}\, dx^\nu \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}.}$$

Differentiate the coefficient and wedge on a new basis covector. On a 0-form this is just the differential $df = \partial_\mu f\, dx^\mu$. On a 1-form $\omega = \omega_\mu\, dx^\mu$, applying the definition:

$$d\omega = \partial_\nu \omega_\mu\, dx^\nu \wedge dx^\mu.$$

Writing the sum as the average of itself with dummy indices $\mu \leftrightarrow \nu$ swapped, then using $dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu$:

$$\begin{aligned} d\omega &= \frac{1}{2}\left(\partial_\nu \omega_\mu \, dx^\nu \wedge dx^\mu + \partial_\mu \omega_\nu \, dx^\mu \wedge dx^\nu\right) \\ &= \frac{1}{2}\left(\partial_\nu \omega_\mu \, dx^\nu \wedge dx^\mu - \partial_\mu \omega_\nu \, dx^\nu \wedge dx^\mu\right) \\ &= \frac{1}{2}\left(\partial_\nu \omega_\mu - \partial_\mu \omega_\nu\right) dx^\nu \wedge dx^\mu. \end{aligned}$$

The antisymmetric combination $\partial_\nu \omega_\mu - \partial_\mu \omega_\nu$ is the curl of the 1-form coefficients, which is exactly where curl comes from, as we will see below.

9.2 Nilpotency

The antisymmetry of the wedge product automatically antisymmetrizes the derivatives, which is why $d^2 = 0$. Applying d twice gives $\partial_\rho \partial_\nu \omega_\mu$ contracted with $dx^\rho \wedge dx^\nu \wedge \cdots$, and since $\partial_\rho \partial_\nu = \partial_\nu \partial_\rho$ (equality of mixed partials) but $dx^\rho \wedge dx^\nu = -dx^\nu \wedge dx^\rho$ (antisymmetry of wedge), every term cancels with its swap. Symmetry in the indices is killed by antisymmetry in the basis.

9.3 Grad, Curl, and Div

All of grad, curl, and div are special cases of d. To see this explicitly, associate to any vector field $F = (F_x, F_y, F_z)$ in $\mathbb{R}^3$ a 1-form $\omega_F = F_x \, dx + F_y \, dy + F_z \, dz$ and a 2-form $\sigma_F = F_x \, dy \wedge dz + F_y \, dz \wedge dx + F_z \, dx \wedge dy$.

Grad: apply d to a 0-form f:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

The components of df are exactly the components of ∇f, packaged as a 1-form.

Curl: apply d to the 1-form ω_F:

$$\begin{aligned} d\omega_F = &\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) dy \wedge dz + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) dz \wedge dx \\ &+ \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) dx \wedge dy. \end{aligned}$$

The three antisymmetric combinations of partial derivatives are exactly the components of $\nabla \times F$, packaged as a 2-form.

Div: apply d to the 2-form σ_F:

$$d\sigma_F = \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\right) dx \wedge dy \wedge dz.$$

The coefficient is exactly $\nabla \cdot F$, packaged as a 3-form.

The single operation d encodes all three classical operations. The reason they look different in vector calculus is that d acts on different types of forms each time, and the 3D coincidence $\binom{3}{1} = \binom{3}{2} = 3$ allows both 1-forms and 2-forms to masquerade as vector fields. Curl is the special case that relies on this coincidence, which is why it does not generalize beyond 3 dimensions. The single elegant identity

$d^2 = 0$ unifies what in vector calculus are two separate identities:

$$\nabla \times (\nabla f) = 0, \qquad \nabla \cdot (\nabla \times F) = 0.$$

---

10 Pullbacks and Pushforwards

A smooth map $\phi\colon M \to N$ between manifolds induces natural maps on tangent and cotangent spaces: a **pushforward** on vectors and a **pullback** on forms.

10.1 Pushforward

The **pushforward** $\phi_*\colon T_pM \to T_{\phi(p)}N$ takes a tangent vector at $p \in M$ to a tangent vector at $\phi(p) \in N$. Recall that a tangent vector is a derivation that acts on smooth functions and returns a number. So $\phi_* v$ must be a derivation on functions $g\colon N \to \mathbb{R}$. The question is: given such a g, how do we use v, which only knows how to differentiate functions on M?

The answer is to compose: $g \circ \phi\colon M \to \mathbb{R}$ pulls g back to M, where v can act on it. So the pushforward is defined by:

$$(\phi_* v)(g) = v(g \circ \phi).$$

This is the only obvious definition. It is the unique way to turn a function on N into a function on M using ϕ. To get the coordinate formula, apply $\phi_* v$ to the coordinate function x'^μ on N and use the chain rule:

$$(\phi_* v)^\mu = (\phi_* v)(x'^\mu) = v(x'^\mu \circ \phi) = v(\phi^\mu) = v^\nu \frac{\partial \phi^\mu}{\partial x^\nu},$$

where the last step uses $v = v^\nu \partial_\nu$ acting on ϕ^μ as a function on M. So the pushforward acts on components by the Jacobian $\partial \phi^\mu / \partial x^\nu$.

10.2 Pullback

The **pullback** $\phi^* \colon \Omega^k(N) \to \Omega^k(M)$ takes a k-form on N back to a k-form on M. It is defined by evaluating the form on pushed-forward vectors:

$$(\phi^* \omega)(v_1, \ldots, v_k) = \omega(\phi_* v_1, \ldots, \phi_* v_k).$$

This is again the only reasonable definition. ω lives on N and needs tangent vectors at $\phi(p) \in N$ to act on, and the pushforward supplies exactly those. For a 0-form (smooth function) $f \colon N \to \mathbb{R}$, the pullback is simply precomposition: $\phi^* f = f \circ \phi$. There are no vectors to push forward, just the function composed with the map.

10.3 Coordinate Changes as a Special Case

A coordinate change $x \to x'$ is a smooth map $\phi \colon U \to U'$ between open sets. The transformation laws of §5 are exactly the pullback and pushforward of ϕ: contravariant components transform by the pushforward Jacobian $\partial x'^\mu / \partial x^\nu$, and covariant components transform by the pullback inverse Jacobian $\partial x^\nu / \partial x'^\mu$.

10.4 Pullback Commutes with d

A fundamental property of the exterior derivative is that it commutes with pullback. For any smooth map $\phi\colon M \to N$ and any differential form ω on N,

$$\boxed{\phi^*(d\omega) = d(\phi^*\omega).}$$

This says that differentiating and then pulling back gives the same result as pulling back and then differentiating. It reflects the fact that d is a purely smooth operation, intrinsic to the manifold structure and independent of any particular choice of map or coordinates.

The proof reduces to the 0-form case. For a smooth function $f\colon N \to \mathbb{R}$ and a tangent vector $v \in T_pM$:

$$\begin{aligned} \phi^*(df)(v) &= df(\phi_* v) = (\phi_* v)(f) = v(f \circ \phi) \\ &= d(f \circ \phi)(v) = d(\phi^* f)(v), \end{aligned}$$

where each step uses in turn: the definition of pullback, the definition of df, the definition of pushforward, the definition of d on a 0-form, and $\phi^* f = f \circ \phi$. So $\phi^*(df) = d(\phi^* f)$. For a general k-form, any form can be written locally as a sum of terms $f\, dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$. Since ϕ^* commutes with $\wedge$ (which follows directly from its definition) and with sums, and d satisfies the Leibniz rule, the identity extends to all forms by the 0-form case plus linearity.

10.5 Restriction to a Submanifold

A **submanifold** $S \subset M$ is a subset that is itself a manifold, with the smooth structure inherited from M. For example, a curve or surface sitting inside $\mathbb{R}^3$. The natural map is the **inclusion** $\zeta\colon S \hookrightarrow M$, which sends each point of S to itself viewed as a point of M. The pullback $\zeta^*\omega$ is then the restriction of ω to S. It takes vectors tangent to S, pushes them forward into M via ζ_*, and evaluates ω on the result.

For a coordinate held fixed on S, say $x^i = b_i$, the inclusion map sends:

$$(x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^k) \mapsto (x^1, \dots, x^{i-1}, b_i, x^{i+1}, \dots, x^k).$$

Applying the commutation identity with $\phi = \zeta$ and $\omega = x^i$ (a 0-form):

$$\zeta^*(dx^i) = d(x^i \circ \zeta).$$

Now $x^i \circ \zeta$ is the function that takes a point on the face and returns its i-th coordinate, but on the face $x^i = b_i$, that is always the constant b_i:

$$d(x^i \circ \zeta) = d(b_i) = 0.$$

This fact will be useful when we prove the generalized Stokes theorem.

10.6 Terminology

There is an asymmetry in direction: vectors go forward with ϕ while forms go backward against ϕ, reflecting a fundamental contrast in their nature. A form is a linear functional on vectors, so to evaluate a form from N at a point in M, you first need to produce a vector in N, which you do by pushing forward. This makes the pullback of forms always well-defined for any smooth map. The pushforward of forms is not generally well-defined, because ϕ would need to be invertible to pull vectors back from N to M. This directional asymmetry is why the word "pullback" appears so often in differential geometry. Forms, metrics, and connections all pull back naturally, while vectors only push forward.

---

11 The Generalized Stokes Theorem

If M is a compact oriented k-dimensional manifold with boundary ∂M, and ω is a $(k-1)$-form:

$$\boxed{\int_{\partial M} \omega = \int_M d\omega.}$$

This single statement contains all the classical integral theorems as special cases:

M	ω	**Classical Name**
$[a, b]$	0-form f	Fundamental theorem of calculus
Region in $\mathbb{R}^2$	1-form	Green's theorem
Surface in $\mathbb{R}^3$	1-form	Classical Stokes' theorem
Volume in $\mathbb{R}^3$	2-form	Divergence theorem

They were never different theorems. The proof makes this obvious.

11.1 Proof

Step 1: reduce to a coordinate patch. A partition of unity is a collection of smooth functions $\{\rho_i\}$, one for each coordinate chart, satisfying $0 \le \rho_i \le 1$, each ρ_i vanishing outside its chart, and $\sum_i \rho_i = 1$ everywhere. This lets us write $\omega = \sum_i \rho_i \, \omega$, where each piece $\rho_i \, \omega$ is supported entirely within a single chart. Since the integral is linear, it suffices to prove the theorem for each piece, so assume ω is supported in one chart with coordinates $x^1, \ldots, x^k$.

Step 2: write out both sides. We define a $(k-1)$-form on a k-dimensional manifold as:

$$\omega = \sum_i (-1)^{i-1} f_i \, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k,$$

where the hat means that factor is omitted. The sign $(-1)^{i-1}$ is inserted into an otherwise standard expansion (its purpose will become clear momentarily). When we apply d, the new dx^i lands at the left and must be moved

$i-1$ positions rightward to reach its desired place, picking up a factor of $(-1)^{i-1}$ from the wedge antisymmetry. The inserted sign exactly cancels this. To see this explicitly, the i-th term after applying d is:

$$(-1)^{i-1}\frac{\partial f_i}{\partial x^j}dx^j \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k.$$

For $j \neq i$, the factor dx^j already appears in the sequence, giving zero by antisymmetry. Only $j = i$ survives, and moving dx^i past $i-1$ factors costs $(-1)^{i-1}$, which cancels the prefactor:

$$d\omega = \sum_i \frac{\partial f_i}{\partial x^i}\, dx^1 \wedge \cdots \wedge dx^k.$$

The antisymmetry of the wedge product kills all mixed partial terms; only the diagonal $\partial_i f_i$ terms survive.

Step 3: apply the fundamental theorem. Integrating $d\omega$ over M, we handle each term in the sum separately. For the i-th term, integrate over x^i first using the fundamental theorem of calculus:

$$\int dx^i\, \frac{\partial f_i}{\partial x^i} = f_i\big|_{x^i=b_i} - f_i\big|_{x^i=a_i}.$$

Integrating the remaining $k-1$ variables over their ranges gives:

$$\int_M d\omega = \sum_i \int dx^1 \cdots \widehat{dx^i} \cdots dx^k\, \left[f_i\big|_{x^i=b_i} - f_i\big|_{x^i=a_i}\right].$$

Each term is now an integral over a $(k-1)$-dimensional face of the boundary, specifically the two faces where x^i takes its endpoint values.

Step 4: recognize the boundary. Recall from Step 1 that we reduced to a single coordinate patch, which locally looks like a box in $\mathbb{R}^k$ with coordinates $x^1, \dots, x^k$ each ranging over some interval $[a_i, b_i]$. The boundary of this box consists of $2k$ faces, which are pairs of $(k-1)$-dimensional sides, where one coordinate is held fixed at its endpoint. In $\mathbb{R}^3$ for example: if M is a 2D patch (a piece of surface), its boundary faces are 4 edges; if M is a 1D patch, its boundary faces are 2 endpoints.

The restriction of ω to the face $x^i = b_i$ is precisely the pullback $\zeta^*\omega$, where ζ is the inclusion of that face into M. By the commutation identity of §10, $\zeta^*(dx^j) = d(x^j \circ \zeta)$. For $j \neq i$, $x^j \circ \zeta = x^j$ (those coordinates are free on the face), so $\zeta^*(dx^j) = dx^j$. For $j = i$, $x^i \circ \zeta = b_i$ is constant, so $\zeta^*(dx^i) = d(b_i) = 0$. Every term in ω except the i-th contains dx^i and is therefore killed by ζ^*. The surviving term is:

$$\zeta^*\omega\big|_{x^i=b_i} = (-1)^{i-1} f_i(x^1, \dots, b_i, \dots, x^k) \\ \cdot\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k.$$

Step 5: induced orientation. The boundary ∂M inherits an orientation from M by the following convention: a basis $(w_1, \dots, w_{k-1})$ for $T_p(\partial M)$ is positively oriented if $(n, w_1, \dots, w_{k-1})$ is positively oriented in T_pM, where n is the outward-pointing normal. On the face $x^i = b_i$, the outward normal points in the $+x^i$ direction, so the induced orientation agrees with $dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k$ up to the sign $(-1)^{i-1}$, exactly the sign already present in $\zeta^*\omega$. On the face $x^i = a_i$, the outward normal points in the $-x^i$ direction, reversing the orientation and contributing a minus sign. This is precisely the $\pm$ structure in Step 3.

Integrating $\zeta^*\omega$ over all $2k$ faces and summing gives $\int_{\partial M}\omega$, which is term-by-term identical to what Step 3 produced. $\square$

The whole proof is the fundamental theorem of calculus, applied once per coordinate direction. The exterior derivative is built precisely so that its antisymmetry kills interior terms and leaves only the boundary, which is why $d^2 = 0$ and Stokes' theorem are two sides of the same coin.

We have assumed throughout that the local pieces from the partition of unity sum back up cleanly to give the full global picture, specifically that boundary contributions from overlapping charts cancel consistently. This is intuitively clear but requires careful verification; a rigorous treatment is given in Lee, *Introduction to Smooth Manifolds*.

Part II

The Metric

Introduction to Part II

Part I developed the metric-free story: manifolds, tensors, differential forms, the exterior derivative, and the generalized Stokes theorem. Everything there required only a smooth structure and an orientation. A metric was conspicuously (and deliberately) absent. Working without one makes its role visible rather than assumed. The reader arrives here having felt the absence of distances, angles, and a canonical way to compare vectors.

Part II adds exactly one object: a metric $g_{\mu\nu}$. From this single addition the geometry of the manifold follows without further arbitrary choices. The metric uniquely determines a connection (forced by metric compatibility and torsion-freeness), which determines covariant derivatives, parallel transport, geodesics, and curvature. Part III will add matter and make the metric itself dynamical.

---

12 The Metric

12.1 Definition

In Part I, a manifold came equipped with only a smooth structure. You could do calculus on it, but could not measure distances or angles, and there was no way to compare the lengths of tangent vectors. The metric supplies this structure and is thus the centrally important ingredient required to build physics models in curved space.

A **metric** on a manifold M is a smooth, symmetric, nondegenerate $(0,2)$ tensor field $g_{\mu\nu}$. At each point $p \in M$ it is a symmetric bilinear form on T_pM: a map $g_p\colon T_pM \times T_pM \to \mathbb{R}$ that is linear in each argument and satisfies $g_p(v,w) = g_p(w,v)$. **Nondegeneracy** means that if $g_p(v,w) = 0$ for all $w \in T_pM$, then $v = 0$. This is the condition that the matrix of components $g_{\mu\nu}$ is invertible.

In coordinates the metric is specified by its components $g_{\mu\nu} = g(\partial_\mu, \partial_\nu)$, a symmetric matrix of smooth functions. We define the following (abuses of) notation for the metric:

$$\boxed{\quad g = g_{\mu\nu}\, dx^\mu \otimes dx^\nu =: g_{\mu\nu}\, dx^\mu\, dx^\nu =: ds^2, \quad}$$

where the so-called **invariant line element** ds^2 notation hand-wavily alludes to “squared infinitesimal length.” This is just notation: s is not a function on the manifold and d here is not the exterior derivative of Part I. It is also conventional in physics to omit the tensor product symbol as shown above and we will do this when the context is clear. Note also that the metric determines a canonical volume form $\sqrt{|g|}\,dx^1 \wedge \cdots \wedge dx^n$ on any oriented manifold, where $|g| = |\det(g_{\mu\nu})|$.

12.2 Signature and Causality

A symmetric bilinear form can be diagonalized by a linear change of basis, producing a standard form with $+1$s and -1s on the diagonal. The number of positive and negative entries is a property of the form itself, not the choice of basis: this is the **signature**. A **Riemannian** metric has all positive entries, written $(+, \ldots, +)$. A **Lorentzian** metric has exactly one negative entry, written $(-, +, +, +)$ in four dimensions. A **Riemannian manifold** is a pair (M, g) where M is a smooth manifold and g is a Riemannian metric on it. A **Lorentzian manifold** is the analogous pair (M, g) with g Lorentzian. The metric is not intrinsic to M. The same smooth manifold can carry many different metrics, and specifying one is an additional choice. The term **pseudo-Riemannian** covers any metric with indefinite signature and will be used when statements apply to both cases. For a Riemannian metric, g is a genuine inner product on T_pM — symmetric, bilinear, and positive definite. A Lorentzian metric drops positive definiteness, so that g is an indefinite inner product (sometimes called a pseudo-inner product). In either case, the metric is the structure that allows one to measure lengths and angles on the manifold.

General relativity uses a Lorentzian metric because special relativity already tells us spacetime has this structure. The invariant interval of special relativity,

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

is precisely the **Minkowski metric** $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$. At each point of a Lorentzian manifold, tangent vectors split into three classes according to the sign of $g(v, v) = g_{\mu\nu} v^\mu v^\nu$. Vectors with $g(v, v) < 0$ are **timelike**, vectors with $g(v, v) = 0$ are **null**, and vectors with $g(v, v) > 0$ are **spacelike**.

12.3 Examples

Metric	Line Element	Description
Euclidean $\mathbb{R}^3$	$+dx^2 + dy^2 + dz^2$	Flat 3D space; $g_{\mu\nu} = \delta_{\mu\nu}$.
Minkowski $\mathbb{R}^{3,1}$	$-dt^2 + dx^2 + dy^2 + dz^2$	Flat spacetime of special relativity; $g_{\mu\nu} = \eta_{\mu\nu}$. The local approximation to any Lorentzian manifold at small scales.
Round S^2	$+d\theta^2 + \sin^2\theta\, d\varphi^2$	The unit 2-sphere in spherical coordinates. Note $g_{\varphi\varphi} = \sin^2\theta$ vanishes at the poles: a coordinate singularity, not a geometric one.

12.4 Raising and Lowering Indices

In Part I, tangent vectors and covectors lived in separate worlds. A vector $V^\mu \in T_pM$ and a covector $\omega_\mu \in T_p^*M$ are dual to each other (one acts on the other to return a number), but there was no canonical way to convert between them. Any such conversion would require choosing a basis, making it coordinate-dependent. The metric provides this identification intrinsically.

Since $g_{\mu\nu}$ is nondegenerate, it has an inverse $g^{\mu\nu}$, defined by $g^{\mu\rho} g_{\rho\nu} = \delta^\mu{}_\nu$. Given a tangent vector V^μ, the metric produces a covector $V_\mu = g_{\mu\nu}V^\nu$. Given a covector ω_μ, the inverse metric produces a tangent vector $\omega^\mu = g^{\mu\nu}\omega_\nu$. These operations are called **lowering** and **raising** indices ("index gymnastics") for obvious reasons. The operations are inverse to each other and extend to arbitrary tensors: one factor of $g_{\mu\nu}$ per upper index lowered, one factor of $g^{\mu\nu}$ per lower index raised.

Geometrically, this means that if V^μ is a tangent vector, then V_μ is not a different object but the same vector re-expressed as a linear functional on other vectors. Specifically, $V_\mu W^\mu = g_{\mu\nu}V^\nu W^\mu = g(V, W)$, i.e. acting V_μ on any vector W returns the inner product of V with W. The covector V_μ is the operation "take the inner product with V." Raising and lowering is therefore the metric's way of saying that vectors and covectors, while technically distinct, are two representations of the same underlying geometric object.

12.5 Why the Metric Is the Right Object

It is worth reflecting on why the metric, among all $(0, 2)$ tensor fields, occupies such a special role. A generic $(0, 2)$ tensor is just a bilinear map. It eats two tangent vectors and returns a number. The metric does this too, but its symmetry and nondegeneracy give it three properties that no other structure provides without additional arbitrary choices:

1. Nondegeneracy and symmetry $\Rightarrow$ g is an inner product (or at least pseudo-inner product) at each T_pM, supplying the manifold with notions of length and angle, hence the name "metric." The length of a tangent vector v is $\sqrt{|g(v, v)|}$, and the angle between v and another vector w at the same point is given by $\cos\theta = g(v, w) \,/\, (|v|\,|w|)$. A generic $(0, 2)$ tensor cannot do this because it may be degenerate or asymmetric.

2. Nondegeneracy gives a canonical isomorphism between T_pM and T_p^*M, as discussed above ("raising and lowering"). Without a metric, no such identification exists.

3. Nondegeneracy and symmetry uniquely determine a connection out of infinitely many possibilities on M. Since the connection, as we will see, determines parallel transport, geodesics, and curvature, the metric is not just one tensor among many but the root of the entire geometric edifice of GR.

13 The Covariant Derivative and the Levi-Civita Connection

13.1 The Problem

Physical laws in curved spacetime must be coordinate-independent. A law that holds in one coordinate system must hold in all of them, or it is not a law but a coordinate artifact. This forces us to write physics as tensor equations, which in turn requires a way to differentiate tensor fields and get tensors back. The exterior derivative of Part I gave a coordinate-independent derivative for differential forms, but the naive generalization to arbitrary tensor fields fails. We need something more general. Indeed, even taking partial derivatives of a simple vector field, $\partial_\mu V^\nu$, does not produce a tensor. Under a coordinate change $x \to x'$:

$$\begin{aligned}
\partial'_\mu V'^\nu &= \left(\frac{\partial x^\sigma}{\partial x'^\mu}\partial_\sigma\right)\left(\frac{\partial x'^\nu}{\partial x^\rho}V^\rho\right) \\
&= \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\rho}\,\partial_\sigma V^\rho + \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial^2 x'^\nu}{\partial x^\sigma \partial x^\rho}\,V^\rho.
\end{aligned}$$

The first term is correct tensor behavior. The second is the obstruction. It vanishes only for linear coordinate changes, which is why partial derivatives work in flat space with Cartesian coordinates. On a curved manifold with general coordinates it does not vanish, and $\partial_\mu V^\nu$ is coordinate-dependent.

13.2 The Covariant Derivative

We define a **covariant derivative** ∇ to be an operator mapping (p, q) tensor fields to $(p, q+1)$ tensor fields satisfying:

1. Linearity for tensors S, T and constants α, β: $\nabla(\alpha S + \beta T) = \alpha \nabla S + \beta \nabla T$
2. Leibniz rule: $\nabla(S \otimes T) = (\nabla S) \otimes T + S \otimes (\nabla T)$
3. Reduces to exterior derivative for any scalar field f: $\nabla f = df$
4. Commutes with contraction: $\nabla(\operatorname{tr} T) = \operatorname{tr}(\nabla T)$.

These axioms should seem plausible for differential operators, but they do not uniquely determine ∇. Specifying a covariant derivative is additional structure on the manifold and many valid choices exist. We seek to specify a particular covariant derivative that resolves the problem identified in the previous subsection.

Since ∇ is linear and satisfies Leibniz, its action on any tensor field is completely determined by its action on basis vector fields ∂_ν. The result $\nabla(\partial_\nu)$ is a $(1, 1)$ tensor, and we define the **connection coefficients** (or **Christoffel symbols**) $\Gamma^\rho_{\mu\nu}$ via its components in the μ-direction:

$$[\nabla(\partial_\nu)]_\mu =: \Gamma^\rho_{\mu\nu}\, \partial_\rho.$$

We henceforth adopt the notation $\nabla_\mu(\cdot) \equiv [\nabla(\cdot)]_\mu$. This is an unfortunate abuse of notation since it incorrectly seems to suggest that ∇_μ is a covector. It is calculationally useful, though, and therefore ubiquitous, so we are stuck with it. The previous expression then reads:

$$\boxed{\nabla_\mu \partial_\nu =: \Gamma^\rho_{\mu\nu}\, \partial_\rho.}$$

13.3 The Covariant Derivative is Coordinate-Free

To see how the covariant derivative resolves the problem identified above, we derive its component expression for a vector field and verify tensorial transformation. Write $V = V^\nu \partial_\nu$ and apply ∇ via Leibniz:

$$\begin{aligned}
(\nabla V)_\mu &= [\nabla(V^\nu \partial_\nu)]_\mu \\
&= (\nabla V^\nu)_\mu\, \partial_\nu + V^\nu\, [\nabla(\partial_\nu)]_\mu \\
&= (\partial_\mu V^\nu)\, \partial_\nu + V^\nu\, \Gamma^\rho_{\mu\nu}\, \partial_\rho \\
&= (\partial_\mu V^\nu)\, \partial_\nu + V^\rho\, \Gamma^\nu_{\mu\rho}\, \partial_\nu \\
&= \left(\partial_\mu V^\nu + \Gamma^\nu_{\mu\rho}\, V^\rho\right) \partial_\nu,
\end{aligned}$$

where we've used the fact that $(\nabla V^\nu)_\mu = \partial_\mu V^\nu$ for the scalar coefficients V^ν, and we also relabeled the dummy index $\rho \leftrightarrow \nu$ in the penultimate expression. Reading off the ν-component and further desecrating notation (in an unfortunately standard way) so that $(\nabla V)^\nu_\mu \equiv \nabla_\mu V^\nu$, we obtain the following classic coordinate-space expression for the covariant derivative of a vector field:

$$\boxed{\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\rho}\, V^\rho.}$$

For this to be a genuine $(1,1)$ tensor it must satisfy:

$$\begin{aligned}\nabla'_\mu V'^\nu &= \partial'_\mu V'^\nu + \Gamma'^\nu_{\mu\rho} V'^\rho \\ &= \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\lambda}\,\nabla_\sigma V^\lambda.\end{aligned}$$

Using the transformation of $\partial'_\mu V'^\nu$ derived above, one can verify that this is satisfied if and only if Γ transforms as:

$$\Gamma'^\nu_{\mu\rho} = \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\lambda}\frac{\partial x^\tau}{\partial x'^\rho}\,\Gamma^\lambda_{\sigma\tau} - \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x^\tau}{\partial x'^\rho}\frac{\partial^2 x'^\nu}{\partial x^\sigma \partial x^\tau}.$$

The second term is inhomogeneous: a pure second-derivative piece with no Γ in it. It cancels precisely the obstruction that appeared in $\partial'_\mu V'^\nu$. The non-tensorial behavior of Γ is a requirement, not a defect. Γ must transform this way to kill the obstruction.

The action of ∇ extends to arbitrary tensors by the Leibniz rule. Each upper index contributes a $+\Gamma$ term and each lower index a $-\Gamma$ term. For example, two common cases are a covector and a $(2,0)$ tensor:

$$\begin{aligned}\nabla_\rho\, \omega_\mu &= \partial_\rho\, \omega_\mu - \Gamma^\lambda_{\rho\mu}\, \omega_\lambda, \\ \nabla_\rho\, T^{\mu\nu} &= \partial_\rho\, T^{\mu\nu} + \Gamma^\mu_{\rho\lambda}\, T^{\lambda\nu} + \Gamma^\nu_{\rho\lambda}\, T^{\mu\lambda},\end{aligned}$$

where the signs are fixed by the Leibniz rule applied to contractions with other tensors.

13.4 Interpreting the Connection Coefficients

The name *connection* points at something specific. Tangent spaces at different points T_pM and T_qM are separate vector spaces with no built-in relationship. There is no canonical sense in which a vector at p "points the same way" as one at q. The connection is the extra structure that provides this. It connects the tangent spaces, supplying a notion of what it means for a vector to vary (or remain constant) as you move. The $\Gamma^\rho_{\mu\nu}$ are its coordinate expression, encoding the rate at which the tangent spaces twist relative to each other. This is what makes parallel transport possible: it is the subject of the next section.

To see this concretely, notice that the two terms in the component formula $\nabla_\mu V^\rho = \partial_\mu V^\rho + \Gamma^\rho_{\mu\nu} V^\nu$ have different origins: $\partial_\mu V^\rho$ is the change in the components, while $\Gamma^\rho_{\mu\nu} V^\nu$ corrects for the basis vectors themselves changing. A plain ∂_μ sees only the first, which is why it fails as a tensor derivative.

Recall that the expression underlying the component formula is our definition $\nabla_\mu \partial_\nu = \Gamma^\rho_{\mu\nu}\,\partial_\rho$. This says that as you move in the ∂_μ direction, the change in ∂_ν is itself a vector with ρ-th component $\Gamma^\rho_{\mu\nu}$. In flat space with Cartesian coordinates the basis vectors are constant everywhere, so $\Gamma = 0$. In polar coordinates on flat space they tilt as you move (∂_r and ∂_θ rotate) so that $\Gamma \neq 0$ without any curvature. Nonzero Christoffel symbols can therefore reflect a curved coordinate system rather than a curved space. Curvature lives in the derivatives of Γ, which is the content of §16.

13.5 The Levi-Civita Connection

On a bare manifold there are infinitely many valid choices of $\Gamma^\rho_{\mu\nu}$. The metric singles out a unique one by imposing two conditions:

1. **Metric compatibility:** $\nabla_\rho\, g_{\mu\nu} = 0$. The metric is covariantly constant; equivalently, raising and lowering indices commutes with ∇.

2. **Torsion-free:** $\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu}$. The lower two indices are symmetric.

Together these uniquely determine Γ. Expanding metric compatibility using the covariant derivative formula for a $(0, 2)$ tensor:

$$\nabla_\rho\, g_{\mu\nu} = \partial_\rho g_{\mu\nu} - \Gamma^\lambda_{\rho\mu}\, g_{\lambda\nu} - \Gamma^\lambda_{\rho\nu}\, g_{\mu\lambda} = 0.$$

Write this out for three cyclic permutations of (ρ, μ, ν):

$$\begin{aligned} \partial_\mu g_{\nu\rho} &= \Gamma^\lambda_{\mu\nu}\, g_{\lambda\rho} + \Gamma^\lambda_{\mu\rho}\, g_{\nu\lambda}, \\ \partial_\nu g_{\rho\mu} &= \Gamma^\lambda_{\nu\rho}\, g_{\lambda\mu} + \Gamma^\lambda_{\nu\mu}\, g_{\rho\lambda}, \\ \partial_\rho g_{\mu\nu} &= \Gamma^\lambda_{\rho\mu}\, g_{\lambda\nu} + \Gamma^\lambda_{\rho\nu}\, g_{\mu\lambda}. \end{aligned}$$

Add the first two and subtract the third. Applying torsion-freeness $\Gamma^\lambda_{\alpha\beta} = \Gamma^\lambda_{\beta\alpha}$ and symmetry of g, four of the six Γ terms cancel in pairs, leaving only $2\Gamma^\lambda_{\mu\nu} g_{\lambda\rho}$:

$$\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu} = 2\, \Gamma^\lambda_{\mu\nu}\, g_{\lambda\rho}.$$

Contracting both sides with $g^{\rho\sigma}$ and using $g_{\lambda\rho}\, g^{\rho\sigma} = \delta^\sigma_\lambda$ gives the famous **Levi-Civita connection**, which will be our default connection for the remainder of this text:

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

14 Parallel Transport

The connection defines what it means for a vector to remain constant as you move along a curve $\gamma\colon \mathbb{R} \to M$. A **vector field along a curve** is a map V assigning to each $\lambda \in \mathbb{R}$ a tangent vector $V(\lambda) \in T_{\gamma(\lambda)}M$. Note this is defined only along γ, not on a neighborhood of it. Such a field is **parallel-transported** along γ if its covariant derivative in the direction of $\gamma'(\lambda)$ vanishes:

$$\boxed{\frac{DV^\mu}{d\lambda} := \frac{dx^\nu}{d\lambda}\nabla_\nu V^\mu = \frac{dx^\nu}{d\lambda}\big(\partial_\nu V^\mu + \Gamma^\mu_{\nu\rho}V^\rho\big) = 0.}$$

The notation $D/d\lambda$ is the **covariant derivative along the curve**: $\frac{D}{d\lambda} \equiv \frac{dx^\nu}{d\lambda}\nabla_\nu$, the component of ∇ in the direction of $\gamma'(\lambda)$. In flat space with Cartesian coordinates $\Gamma = 0$ and this reduces to ordinary constancy of components. In curved space or curved coordinates, the Γ terms compensate for the tilting basis, keeping the vector geometrically constant even as its components change.

Parallel transport is path-dependent in the sense that moving a vector from p to q along different curves generally produces different results. This path-dependence is the geometric signature of curvature, made precise by the Riemann tensor in a later section.

15 Geodesics

15.1 The Geodesic Equation

A **geodesic** is a curve that parallel-transports its own tangent vector. This is the natural generalization of a straight line to curved space. Setting $V^\mu = dx^\mu/d\lambda$ in the parallel transport equation:

$$\begin{aligned} 0 &= \frac{dx^\nu}{d\lambda}\,\partial_\nu\left(\frac{dx^\mu}{d\lambda}\right) + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda} \\ &= \frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda}, \end{aligned}$$

where the last step uses $(dx^\nu/d\lambda)\,\partial_\nu = d/d\lambda$ along the curve. This second-order ODE is the **geodesic equation**, whose solutions are geodesics:

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

Writing $u^\mu = dx^\mu/d\lambda$ for the tangent vector, the quantity $g_{\mu\nu}u^\mu u^\nu$ is constant along any geodesic. This can be seen by noticing that, since $D/d\lambda$ agrees with $d/d\lambda$ on scalars, the Leibniz rule gives:

$$\frac{d}{d\lambda}(g_{\mu\nu}u^\mu u^\nu) = \frac{Dg_{\mu\nu}}{d\lambda}\,u^\mu u^\nu + g_{\mu\nu}\frac{Du^\mu}{d\lambda}\,u^\nu + g_{\mu\nu}\,u^\mu\frac{Du^\nu}{d\lambda} = 0,$$

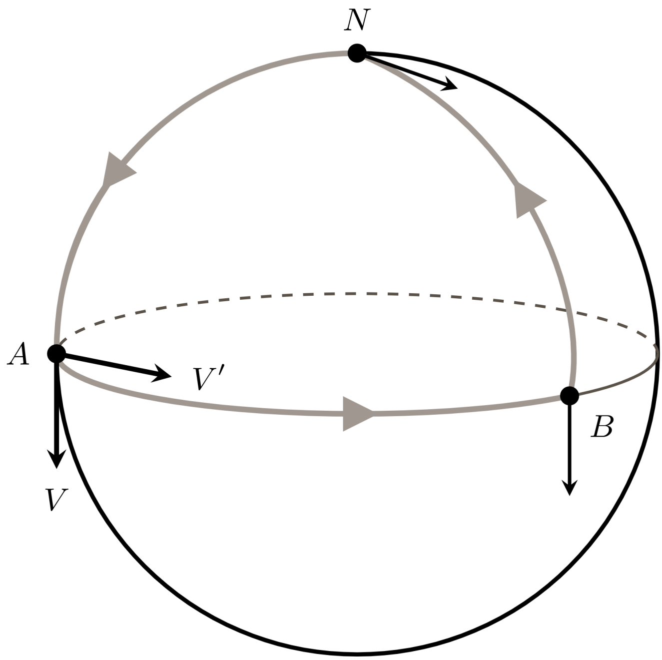
where the first term vanishes by metric compatibility and each of the last two vanish by the parallel transport condition applied to the tangent vector u^μ. The constant value of $g_{\mu\nu}u^\mu u^\nu$ classifies the geodesic: negative for timelike curves, zero for null, positive for spacelike. The physical interpretation of these classes is taken up in Part III.

15.2 Parallel Transport on the Sphere

The conservation argument above is not specific to tangent vectors. For any two parallel-transported vectors V and W along the same curve, the Leibniz rule and metric compatibility give

$$\begin{aligned}\frac{d}{d\lambda}g(V,W) &= \frac{Dg_{\mu\nu}}{d\lambda}\,V^\mu W^\nu + g_{\mu\nu}\frac{DV^\mu}{d\lambda}\,W^\nu + g_{\mu\nu}\,V^\mu\frac{DW^\nu}{d\lambda}\\ &= 0.\end{aligned}$$

For example, if γ is a geodesic with tangent u, then u is itself parallel-transported by definition, so $g(V,u)$ is constant for any parallel-transported V. The angle between V and the geodesic's tangent is preserved throughout the arc.



The meridians and equator of a round sphere are geodesics. Take the North Pole N and two equatorial points A, B $90°$ apart (see figure), and parallel-transport a vector V, initially pointing south at A, around the triangle they form. First leg $A \to B$: V starts perpendicular to the arc and this is preserved throughout, so V arrives at B still pointing south. Second leg $B \to N$: V is directed along the meridian, opposite the direction of travel. Parallel transport preserves this, so V arrives at N pointing back tangentially along the meridian toward B. Third leg $N \to A$: V is perpendicular to the arc and arrives back at A pointing due-East. We see that the parallel transport loop has rotated the vector by $90°$, in contrast to flat space where the same closed path returns any vector unchanged.

16 The Riemann Tensor

The closing observation of §14 was that path-dependence is the geometric signature of curvature. Take a vector V at a point p and parallel-transport it around a closed loop, returning to p. In flat space, you recover V exactly. In curved space, you recover a rotated vector. The Riemann tensor makes this precise.

16.1 The Definition

Partial derivatives commute, but covariant derivatives need not. Whether they do depends on the geometry. To understand this more deeply, we must compute the commutator $[\nabla_\mu, \nabla_\nu] \equiv \nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu$. Applying this operator to a vector V gives:

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

where we used the definition of the covariant derivative for (1,0) and (1,1) tensors in the second and third lines, respectively. Performing the antisymmetrization, three classes of terms cancel: $\partial_\mu \partial_\nu V^\rho - \partial_\nu \partial_\mu V^\rho = 0$ by symmetry of partial derivatives; the first-derivative-of-V terms cancel pairwise after relabeling dummy indices; and the $(\Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu})$ terms vanish by torsion-freeness. Gather the remaining terms so that

$$[\nabla_\mu, \nabla_\nu]\, V^\rho =: R^\rho{}_{\sigma\mu\nu}\, V^\sigma,$$

where we have defined the **Riemann tensor**:

$$\boxed{\begin{aligned} R^\rho{}_{\sigma\mu\nu} := \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} \\ + \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda} \Gamma^\lambda_{\mu\sigma}. \end{aligned}}$$

This object is, perhaps surprisingly given its origins as a commutator of differential operators, algebraic in V with no derivatives. Covariant derivatives generally fail to commute, and their failure (the Riemann tensor) acts directly on V as a $(1,3)$ tensor field. The commutator $[\nabla_\mu, \nabla_\nu]$ is the difference between transporting in the μ-then-ν order versus ν-then-μ: the infinitesimal version of going around a loop, and the Riemann tensor is what that loop does to a vector, which is an intuitive way to quantify curvature. Note that $R^\rho{}_{\sigma\mu\nu}$ is built from Christoffel symbols (two derivative-of-Γ terms and two Γ^2 terms). The connection coefficients, specifically their derivatives and self-interactions, encode curvature.

For a general (non-Levi-Civita) connection, notice that the $(\Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu})$ terms do not vanish. They combine into $-T^\lambda{}_{\mu\nu} \nabla_\lambda V^\rho$, where $T^\lambda{}_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}$ is the torsion tensor.

That term involves a derivative of V, making the commutator a differential operator rather than a tensor acting pointwise on V. Torsion-freeness is what makes the commutator algebraic in V and hence a genuine tensor.

16.2 The Bianchi Identity

At any point p there is a coordinate system, called **Riemann normal coordinates**, in which $\Gamma^{\rho}_{\mu\nu}(p) = 0$. To construct these coordinates, begin by fixing a basis $\{e_\mu\}$ for T_pM. Every vector $v = v^\mu e_\mu \in T_pM$, by standard existence and uniqueness of ODE solutions applied to the geodesic equation, determines a unique geodesic γ_v with $\gamma_v(0) = p$ and $\dot{\gamma}_v(0) = v$. Define the coordinates of a nearby point q via $\gamma_v(1) = q$. The smooth structure of M guarantees that we can always find a q "close enough" to p that the same coordinate chart is valid for both.

In these coordinates (by construction), γ_v has coordinate representation $x^\mu(\lambda) = \lambda v^\mu$. That is, the point $\gamma_v(\lambda)$ is reached from p by traveling along a linear ray from the origin within the coordinate chart. Substituting into the geodesic equation:

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

Setting $\lambda = 0$: $\Gamma^\mu_{\nu\rho}(p)\,v^\nu v^\rho = 0$ for every $v \in T_pM$. Replace v by $v+w$ and subtract the pure-v and pure-w cases. What remains is $\Gamma^\mu_{\nu\rho}(p)(v^\nu w^\rho + w^\nu v^\rho) = 0$ for all v, w. Since $\Gamma^\mu_{\nu\rho} = \Gamma^\mu_{\rho\nu}$ by torsion-freeness, this is $2\Gamma^\mu_{\nu\rho}(p)\,v^\nu w^\rho = 0$ for all v, w, hence $\Gamma(p) = 0$, as we claimed. At p, the Riemann tensor then reduces to $R^\rho{}_{\sigma\mu\nu} = \partial_\mu\Gamma^\rho_{\nu\sigma} - \partial_\nu\Gamma^\rho_{\mu\sigma}$, and we also see $\nabla = \partial$ at p. This construction is a coordinate trick, not a geometric fact. Γ is not a tensor, so its vanishing at p is always achievable by a suitable chart and says nothing about the curvature.

These coordinates are the tool for proving both Bianchi identities. We work at an arbitrary p in normal coordinates, derive each identity by pure partial-derivative algebra, and then promote the result to all coordinates everywhere by tensoriality. For the first, antisymmetrize $R^\rho{}_{\sigma\mu\nu}$ over its three lower indices and regroup by which Γ is being differentiated:

$$\begin{aligned} R^\rho{}_{[\sigma\mu\nu]} := \tfrac{1}{6}\big(&R^\rho{}_{\sigma\mu\nu} + R^\rho{}_{\mu\nu\sigma} + R^\rho{}_{\nu\sigma\mu} \\ &- R^\rho{}_{\mu\sigma\nu} - R^\rho{}_{\nu\mu\sigma} - R^\rho{}_{\sigma\nu\mu}\big) \\ = \tfrac{1}{3}\big(&R^\rho{}_{\sigma\mu\nu} + R^\rho{}_{\mu\nu\sigma} + R^\rho{}_{\nu\sigma\mu}\big) \\ = \tfrac{1}{3}\big[&\partial_\mu(\Gamma^\rho_{\nu\sigma} - \Gamma^\rho_{\sigma\nu}) + \partial_\nu(\Gamma^\rho_{\sigma\mu} - \Gamma^\rho_{\mu\sigma}) \\ &+ \partial_\sigma(\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu})\big]\Big|_p = 0, \end{aligned}$$

where the first line defines the antisymmetrization, the second uses antisymmetry of R in its last pair (easy to see from the definition, especially in normal coordinates) to collapse the six terms to the cyclic sum, and each bracket on the third line vanishes by torsion-freeness. Generalizing from p to everywhere by tensoriality then gives the **algebraic Bianchi identity**:

$$\boxed{R^\rho{}_{[\sigma\mu\nu]} = 0.}$$

Now antisymmetrize $\nabla_\lambda R^\rho{}_{\sigma\mu\nu}$ over $\{\lambda, \mu, \nu\}$ and regroup by which Γ is being doubly differentiated:

$$\begin{aligned}&\left[\nabla_\lambda R^\rho{}_{\sigma\mu\nu} + \nabla_\mu R^\rho{}_{\sigma\nu\lambda} + \nabla_\nu R^\rho{}_{\sigma\lambda\mu}\right]\Big|_p \\ &\qquad\qquad = [\partial_\lambda, \partial_\mu]\Gamma^\rho_{\nu\sigma} + [\partial_\mu, \partial_\nu]\Gamma^\rho_{\lambda\sigma} + [\partial_\nu, \partial_\lambda]\Gamma^\rho_{\mu\sigma} = 0,\end{aligned}$$

with each commutator vanishing due to equality of mixed partials. The **second Bianchi identity** is then:

$$\boxed{\nabla_\lambda R^\rho{}_{\sigma\mu\nu} + \nabla_\mu R^\rho{}_{\sigma\nu\lambda} + \nabla_\nu R^\rho{}_{\sigma\lambda\mu} = 0.}$$

16.3 Symmetries of the Riemann Tensor

Normal coordinates give a clean proof of the algebraic symmetries of $R_{\rho\sigma\mu\nu} = g_{\rho\lambda} R^\lambda{}_{\sigma\mu\nu}$. Since $\nabla g = 0$, the covariant derivative component formula says $\partial_\mu g_{\rho\sigma} = \Gamma^\lambda_{\mu\rho} g_{\lambda\sigma} + \Gamma^\lambda_{\mu\sigma} g_{\rho\lambda}$ so that $\partial_\mu g_{\rho\sigma}\big|_p = 0$ when $\Gamma(p) = 0$. Lowering the first index of $R^\lambda{}_{\sigma\mu\nu}\big|_p = \partial_\mu \Gamma^\lambda_{\nu\sigma} - \partial_\nu \Gamma^\lambda_{\mu\sigma}$ and expanding the Levi-Civita connection coefficients:

$$\begin{aligned}R_{\rho\sigma\mu\nu}\Big|_p &= \partial_\mu \Gamma_{\rho\nu\sigma} - \partial_\nu \Gamma_{\rho\mu\sigma} \\ &= \tfrac{1}{2}\big(\partial_\mu\partial_\nu g_{\rho\sigma} + \partial_\mu\partial_\sigma g_{\rho\nu} - \partial_\mu\partial_\rho g_{\nu\sigma}\big) \\ &\quad - \tfrac{1}{2}\big(\partial_\nu\partial_\mu g_{\rho\sigma} + \partial_\nu\partial_\sigma g_{\rho\mu} - \partial_\nu\partial_\rho g_{\mu\sigma}\big) \\ &= \tfrac{1}{2}\big(\partial_\sigma\partial_\mu g_{\rho\nu} - \partial_\rho\partial_\mu g_{\sigma\nu} \\ &\qquad - \partial_\sigma\partial_\nu g_{\rho\mu} + \partial_\rho\partial_\nu g_{\sigma\mu}\big),\end{aligned}$$

where the $\partial_\mu\partial_\nu g_{\rho\sigma}$ second derivative terms cancel by symmetry of mixed partials. Two symmetries are immediate, both of which hold everywhere by tensoriality. Swapping

$\rho \leftrightarrow \sigma$ negates the expression (by $g_{\rho\sigma} = g_{\sigma\rho}$), giving first-pair antisymmetry:

$$\boxed{R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}.}$$

Swapping both pairs simultaneously, $(\rho\sigma) \leftrightarrow (\mu\nu)$, leaves it unchanged, giving pair symmetry:

$$\boxed{R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma}.}$$

16.4 Contractions and the Einstein Tensor

Several contractions of the Riemann tensor will prove essential in what follows. The **Ricci tensor** is the trace on the first and third indices:

$$R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}.$$

It is symmetric and has 10 independent components. Substituting the Riemann tensor expression in terms of Γ into $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$ gives a useful explicit formula:

$$R_{\mu\nu} = \partial_{\rho}\Gamma^{\rho}_{\nu\mu} - \partial_{\nu}\Gamma^{\rho}_{\rho\mu} + \Gamma^{\rho}_{\rho\lambda}\Gamma^{\lambda}_{\nu\mu} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\rho\mu}.$$

The **Ricci scalar** is its metric trace: $R = g^{\mu\nu}R_{\mu\nu}$. These combine into the symmetric 10 component **Einstein tensor**:

$$\boxed{G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R.}$$

The algebraic Bianchi identity is the cyclic sum ${R^\rho}_{\sigma\mu\nu} + {R^\rho}_{\mu\nu\sigma} + {R^\rho}_{\nu\sigma\mu} = 0$. Isolating the first term and contracting with $g^{\sigma\nu}$ gives:

$$g^{\sigma\nu}{R^\rho}_{\sigma\mu\nu} = -g^{\sigma\nu}{R^\rho}_{\mu\nu\sigma} - g^{\sigma\nu}{R^\rho}_{\nu\sigma\mu}.$$

The RHS first term vanishes since $g^{\sigma\nu}$ is symmetric in $\sigma\nu$ while ${R^\rho}_{\mu\nu\sigma}$ is antisymmetric in $\nu\sigma$. For the second, lower ρ, apply antisymmetry in the first pair ($R_{abcd} = -R_{bacd}$), then pair symmetry ($R_{abcd} = R_{cdab}$):

$$\begin{aligned} g^{\sigma\nu}{R^\rho}_{\nu\sigma\mu} &= g^{\rho\tau}g^{\sigma\nu}R_{\tau\nu\sigma\mu} \\ &= -g^{\rho\tau}g^{\sigma\nu}R_{\nu\tau\sigma\mu} \\ &= -g^{\rho\tau}g^{\sigma\nu}R_{\sigma\mu\nu\tau} \\ &= -g^{\rho\tau}{R^\nu}_{\mu\nu\tau} = -g^{\rho\tau}R_{\mu\tau} = -{R^\rho}_\mu. \end{aligned}$$

Therefore $g^{\sigma\nu}{R^\rho}_{\sigma\mu\nu} = {R^\rho}_\mu$. Contracting ρ with λ in the second Bianchi identity gives $\nabla_\rho {R^\rho}_{\sigma\mu\nu} - \nabla_\mu R_{\sigma\nu} + \nabla_\nu R_{\sigma\mu} = 0$. Contracting this result with $g^{\sigma\nu}$ (where $\nabla^\sigma := g^{\sigma\nu}\nabla_\nu$ raises the derivative index by the usual convention) and inserting $g^{\sigma\nu}{R^\rho}_{\sigma\mu\nu} = {R^\rho}_\mu$ then yields:

$$\begin{aligned} 0 &= g^{\sigma\nu}\big(\nabla_\rho {R^\rho}_{\sigma\mu\nu} - \nabla_\mu R_{\sigma\nu} + \nabla_\nu R_{\sigma\mu}\big) \\ &= \nabla^\sigma R_{\sigma\mu} - \nabla_\mu R + \nabla^\sigma R_{\sigma\mu} \\ &= 2\nabla^\sigma R_{\sigma\mu} - \nabla_\mu R. \end{aligned}$$

Relabeling indices gives $2\nabla^\mu R_{\mu\nu} = \nabla_\nu R$. Since $\nabla^\mu(g_{\mu\nu}R) = \nabla_\nu R$, this yields $\nabla^\mu G_{\mu\nu} = 0$ where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$:

$$\boxed{\nabla^\mu G_{\mu\nu} = 0.}$$

This statement about the Einstein tensor is so far purely geometric with no physics assumed. Its significance becomes clear in the Einstein field equations.

The contracted Bianchi identity $2\nabla^\mu R_{\mu\nu} = \nabla_\nu R$ also shows that $G_{\mu\nu}$ is the unique divergence-free combination of $R_{\mu\nu}$ and $g_{\mu\nu}$ linear in curvature. To see this, note that any such combination takes the form $aR_{\mu\nu} + bRg_{\mu\nu}$. Its divergence is $\nabla^\mu(aR_{\mu\nu} + bRg_{\mu\nu}) = (a/2 + b)\nabla_\nu R$, which vanishes identically only when $b = -a/2$. Up to an overall scale, this forces $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$.

Part III

Physics

Introduction to Part III

Parts I and II built the geometric language of general relativity. Part I required only a smooth structure: manifolds, tensors, differential forms, and the generalized Stokes theorem, all without a notion of distance or angle. Part II added a single object, the metric $g_{\mu\nu}$, from which the entire geometry of the manifold followed without further arbitrary choices: a unique connection, covariant derivatives, parallel transport, geodesics, and curvature. The Einstein tensor $G_{\mu\nu}$ emerged from contracting the Riemann tensor, and the Bianchi identity established $\nabla^\mu G_{\mu\nu} = 0$ as a purely geometric fact.

Part III adds the final ingredient: matter. The energy-momentum tensor $T^{\mu\nu}$ encodes how matter and energy are distributed in spacetime. The Einstein field equations then assert $G_{\mu\nu} = 8\pi G\, T_{\mu\nu}$. This is not a definition but a dynamical law. Geometry and matter are not independent. The distribution of matter determines the curvature, and the curvature governs how matter moves. The metric, which Parts I and II treated as given structure on a manifold, becomes dynamical. It is determined by, and in turn determines, everything.

17 Physics on Manifolds

The geometric structure of Parts I and II is purely abstract. Connecting it to the physical world requires postulates that the geometry alone cannot supply. This section covers the kinematic postulates: how observers move, what clocks measure, and how causality is encoded in the metric. The dynamical postulate, how matter sources the curvature of spacetime, is deferred to §19, where it appears as the Einstein field equations.

17.1 The Geodesic Postulate

The **equivalence principle** states that in any sufficiently small region of spacetime, the effects of gravity are locally indistinguishable from those of uniform acceleration. All bodies fall identically in a gravitational field regardless of their mass or composition. The **geodesic postulate** is its precise geometric formulation: a freely falling massive particle traces a timelike geodesic of the spacetime metric. No force acts on it. It simply follows the geometry.

17.2 Proper Time

A **worldline** is the curve in spacetime traced by a particle. The **proper time** elapsed along a worldline is the time measured by a clock carried along it:

$$\boxed{d\tau^2 = -g_{\mu\nu}\, dx^\mu\, dx^\nu.}$$

This is positive along any timelike curve by the signature $(-,+,+,+)$. Equivalently, $d\tau^2 = -ds^2$ where $ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu$ is the line element of Part II. Proper time is that same object, sign-flipped so the arc length is real and positive along timelike curves. The only new ingredient here is physical interpretation. $d\tau^2$ is the coordinate-independent Lorentzian arc length of the worldline. Two observers following different worldlines between the same two events accumulate different proper times regardless of what coordinate system is used.

It may seem odd that the spatial displacements dx^i appear in a quantity called time. In the rest frame of the clock $(dx^i = 0)$, they vanish. In flat spacetime $d\tau = dt$ and the clock simply reads coordinate time. When the clock moves, the spatial terms subtract from dt^2, giving $d\tau < dt$ between fixed coordinate endpoints. This is time dilation. Moving clocks tick slower, and the formula captures the effect in a single invariant. That this arc length is what physical clocks actually read is the **clock hypothesis**: an ideal clock measures the proper time elapsed along its worldline, regardless of its acceleration. This is an imposed condition rather than something derived, and its validation is experimental.

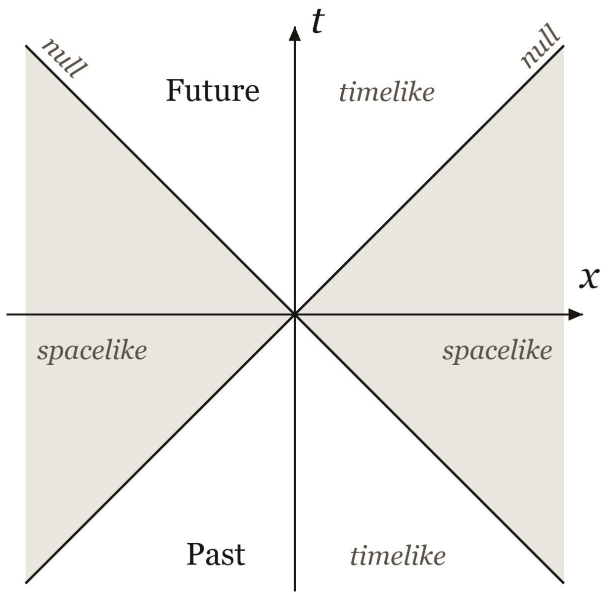
For a freely falling massive particle we write $u^\mu = dx^\mu/d\tau$ for the 4-velocity. Its time component $u^0 = dt/d\tau$ is the **Lorentz factor** γ. Since proper time is the Lorentzian arc length of the worldline, it is given by the standard arc length integral. Writing $\dot{x}^\mu = dx^\mu/dt$ for ordinary derivatives along the worldline:

$$\tau = \int_0^T dt\,\sqrt{-g_{\mu\nu}\,\dot{x}^\mu\dot{x}^\nu}.$$

In flat spacetime $-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 1 - v^2$ where $v^2 = \sum_i(\dot{x}^i)^2$, so $\tau = \int_0^T dt\,\sqrt{1-v^2}$, and thus $d\tau/dt = \sqrt{1-v^2}$, so $\gamma = dt/d\tau = 1/\sqrt{1-v^2}$. A particle at rest has $\gamma = 1$. A moving particle has $\gamma > 1$, with the excess above 1 encoding kinetic energy per unit rest mass. Switching to τ as the curve parameter, the arc length integral becomes $\tau = \int_0^\tau d\tau'\,\sqrt{-g_{\mu\nu}\,u^\mu u^\nu}$. For this to hold $\forall\,\tau$, the integrand must equal 1, giving $g_{\mu\nu}u^\mu u^\nu = -1$ for any worldline parametrized by proper time. For a photon the worldline is null, so $d\tau = 0$ and proper time cannot serve as a parameter. Instead, one uses an affine parameter λ, and the tangent $p^\mu = dx^\mu/d\lambda$ satisfies $g_{\mu\nu}p^\mu p^\nu = 0$.

17.3 Causal Structure

Recall from Part II that the metric partitions tangent vectors at every point into timelike ($g(V,V) < 0$), null ($g(V,V) = 0$), and spacelike ($g(V,V) > 0$) classes. In the physical setting of Part III this classification acquires direct meaning: timelike vectors describe directions available to massive particles, null vectors describe directions available to photons and other massless particles, and no physical particle travels in a spacelike direction.



At each point the null vectors form the **light cone**, whose walls represent the speed of light barrier. The signature $(-,+,+,+)$ is what creates this structure: the negative sign makes one direction geometrically distinct from the three spatial ones, partitioning tangent vectors into timelike, null, and spacelike classes. But the light cone has two halves, loosely "future" and "past," and the metric alone cannot distinguish them. A **time orientation** is a continuous global choice of which half is future at every point. In adapted coordinates, this amounts to designating the half with $u^t > 0$ as future-directed at each point. This is an independent structure, not determined by the metric. With a time orientation in hand, a **future-directed** curve is one whose tangent u^μ lies in the future half of the light cone at every point or, equivalently, $u^t > 0$ in coordinates adapted to the orientation. Physical worldlines are required to be future-directed. This is a postulate, not a theorem. Nothing in the geometry forbids backward-in-time motion. The restriction is the physical assumption that matter and energy propagate forward in time.

Together, the metric and the time orientation impose a causal structure on spacetime. The metric defines the light cone and the time orientation picks which half is the future. Which events can influence which others is determined by whether they can be connected by a future-directed causal curve.

---

18 The Energy-Momentum Tensor

18.1 Energy and Momentum in Special Relativity

In Newtonian physics, energy E is a scalar and momentum p a 3-vector. These are treated as separate quantities with separate **conservation laws**: $\frac{dE}{dt} = 0$ and $\frac{dp}{dt} = 0$ (for an isolated system). Special relativity (SR) unifies them. The 4-momentum of a particle of rest mass m is $p^\mu = mu^\mu$, with $p^0 = E$ and p^i the spatial momentum. For a single particle this is enough. For a *continuous medium*, we need to track the density and flux of this quantity throughout spacetime.

A scalar conserved charge, such as electric charge, is associated with a 4-current J^μ where a single index is sufficient to encode the four directions of flow. This satisfies the conservation law $\partial_\mu J^\mu = 0$. Energy-momentum is already a

4-vector, so tracking its flow in each spacetime direction requires a second index. The result is a rank-$(2,0)$ tensor. The **energy-momentum tensor** $T^{\mu\nu}$ is defined as the flux of the μ-th component of 4-momentum through surfaces of constant x^ν. In a local inertial frame:

- T^{00}: energy density.
- T^{0i}: energy flux in the x^i-direction, equal to the i-th momentum density (these coincide in SR).
- T^{ij}: flux of the i-th momentum component through surfaces of constant x^j, i.e. stress.

Symmetry $T^{\mu\nu} = T^{\nu\mu}$ follows: $T^{0i} = T^{i0}$ is the energy-flux/momentum-density equality, and $T^{ij} = T^{ji}$ from angular momentum conservation. The SR conservation law is $\partial_\mu T^{\mu\nu} = 0$, which expands as:

$$\partial_t T^{00} + \partial_i T^{i0} = 0,$$
$$\partial_t T^{0j} + \partial_i T^{ij} = 0.$$

The first is energy continuity. The second is momentum continuity for each spatial component.

18.2 Minimal Coupling

To lift this law to curved spacetime, we use the equivalence principle: at any point p of a curved spacetime, one can choose coordinates in which the metric is $\eta_{\mu\nu}$ at p and the Christoffel symbols vanish at p. In this local inertial frame, spacetime is instantaneously flat and SR holds. Any law that holds in SR must therefore hold locally in curved spacetime. The **minimal coupling prescription** makes this systematic: take any SR equation, replace $\eta_{\mu\nu} \to g_{\mu\nu}$, and replace partial derivatives $\partial_\mu \to \nabla_\mu$. The word "minimal" means we introduce no additional curvature terms, such as a coupling $\xi R T^{\mu\nu}$, that vanish in flat space but could in principle appear. We emphasize that, despite its elegance, there is no "magic" reason that minimal coupling should be true. Its ultimate validation is an experimental matter. Applied to energy-momentum conservation:

$$\boxed{\nabla_\mu T^{\mu\nu} = 0.}$$

This is local energy-momentum conservation in curved spacetime. No global version exists, since integrating a local conservation law into a global charge requires a background symmetry to sum over, and a general curved spacetime has no globally defined time-translation symmetry.

18.3 Perfect Fluids

A **perfect fluid** has no viscosity and no heat conduction. It is characterized at each point by three fields: an energy density ρ, an isotropic pressure p, and a 4-velocity u^μ. The 4-velocity is a timelike unit vector. In the rest frame of the fluid, $u^\mu = (1, 0, 0, 0)$, where $u^i = 0$ says the fluid has no spatial motion and $u^0 = 1$ follows from the normalization $u^\mu u_\mu = g_{00} = -1$.

The energy-momentum tensor is a separate object. In the rest frame, $T^{0i} = 0$. There is no bulk flow (rest frame) and no heat conduction (perfect fluid), so neither momentum density nor energy flux are present. The spatial block T^{ij}

is diagonal with equal entries by isotropy, with the proportionality constant equal to the pressure p. In total:

$$T^{\mu\nu} = \mathrm{diag}(\rho,\, p,\, p,\, p).$$

To write this covariantly, we need a symmetric $(2,0)$ tensor built from the available local data u^μ and $g^{\mu\nu}$. The only symmetric rank-2 tensors constructible from these are $u^\mu u^\nu$ and $g^{\mu\nu}$, so the general form is

$$T^{\mu\nu} = \alpha\, u^\mu u^\nu + \beta\, g^{\mu\nu}$$

for scalars α and β. Matching to the rest frame with $g^{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$ and $u^0 = 1$:

$$\begin{aligned} T^{00} &= \alpha (u^0)^2 + \beta\, g^{00} = \alpha - \beta = \rho, \\ T^{11} &= \beta\, g^{11} = \beta = p. \end{aligned}$$

So $\beta = p$ and $\alpha = \rho + p$:

$$\boxed{T^{\mu\nu} = (\rho + p)\, u^\mu u^\nu + p\, g^{\mu\nu}.}$$

This is the form we will use in the cosmology section, where the universe is modeled as a perfect fluid. Note also that $T^{\mu\nu} = 0$ in a vacuum (the Schwarzschild solution is the leading example here).

19 The Einstein Field Equations

Two objects are now in hand. From Part II: the Einstein tensor $G_{\mu\nu}$, a symmetric rank-$(0,2)$ tensor with $\nabla^\mu G_{\mu\nu} = 0$ proved as a purely geometric fact. From the previous section: the energy-momentum tensor $T_{\mu\nu}$, also symmetric and rank-$(0,2)$, with $\nabla^\mu T_{\mu\nu} = 0$ established by minimal coupling. The **Einstein field equations** assert they are proportional:

$$\boxed{G_{\mu\nu} = 8\pi G\, T_{\mu\nu},}$$

where G is Newton's gravitational constant. The factor of 8π is not obvious and is fixed by requiring the equations to reproduce Newtonian gravity in the appropriate limit, carried out below.

The LHS of the EFE is geometry, and the RHS is matter. However, this is not a "one-way" statement: $T_{\mu\nu}$ itself generally depends on $g_{\mu\nu}$. For example, the perfect fluid formula $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\, g^{\mu\nu}$ makes this explicit, since $g^{\mu\nu}$ appears directly, and lowering indices to obtain $T_{\mu\nu}$ uses g again. The EFE is therefore a coupled system, and there is no stage at which one side is fixed and the other solved. This is the deepest break from special relativity, where $\eta_{\mu\nu}$ is a fixed background stage on which matter moves. In general relativity the stage is dynamical. It is shaped by the matter on it, and simultaneously governs how that matter moves. Concretely, the EFE is a system of ten coupled nonlinear second-order PDEs for the ten independent components of $g_{\mu\nu}$. There are ten equations because $g_{\mu\nu}$ is symmetric, second-order because $G_{\mu\nu}$ involves second derivatives of g, and nonlinear because the Riemann tensor contains $\Gamma\Gamma$ terms quadratic in first derivatives. The geometric machinery of Parts I and II makes it possible to write this system in a single line.

The choice of $G_{\mu\nu}$ on the left is not arbitrary. The right side $T_{\mu\nu}$ is symmetric and divergence-free; the left side must match. The Bianchi identity (proved in Part II) already singles out $G_{\mu\nu}$. It is the unique divergence-free combination of $R_{\mu\nu}$ and $g_{\mu\nu}$ linear in curvature, as shown there. We further restrict to terms at most linear in second derivatives of $g_{\mu\nu}$, so that the field equations for the metric are second-order. This is the standard structure for physical field equations (Maxwell is second-order in the potential, Newtonian gravity second-order in Φ). Under this restriction one further freedom remains: a multiple of $g_{\mu\nu}$ itself, which is also divergence-free ($\nabla^\mu g_{\mu\nu} = 0$ identically). This is the cosmological constant, addressed below. Despite the naturalness which is implied by all of these comments, the EFE is ultimately a postulate, not a derivation. Its validation is the Newtonian limit and experiment.

19.1 Vacuum Equations

When $T_{\mu\nu} = 0$, the EFE becomes $G_{\mu\nu} = 0$. Taking the metric trace: $g^{\mu\nu}G_{\mu\nu} = -R = 0$, so $R = 0$, and the vacuum equations reduce to:

$$\boxed{R_{\mu\nu} = 0.}$$

Note that this is not strong enough to imply flat spacetime. Flatness requires $R^\rho{}_{\sigma\mu\nu} = 0$ while the Ricci tensor is only a contraction of the full Riemann tensor. The Schwarzschild solution satisfies $R_{\mu\nu} = 0$ everywhere outside its source yet is unmistakably curved. Light deflects, orbits precess, and clocks at different radii run at different rates.

19.2 The Newtonian Limit

To fix $8\pi G$, we require that the EFE reproduce Newtonian gravity when gravity is weak and motion is slow. Write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1,$$

and take the source to be pressureless dust moving slowly, so $T^{00} \approx \rho$ dominates, where ρ is the mass density in the rest frame. For non-relativistic matter, kinetic and internal energies are negligible compared to rest mass energy, so T^{00} approximately reduces to the mass density.

The Newtonian potential Φ enters through the geodesic equation. The equivalence principle says that freely falling particles follow geodesics. In the slow-motion limit $u^i \ll u^0$, the spatial components of the geodesic equation give $\ddot{x}^i \approx -\Gamma^i_{00}$. Comparing to Newton's second law $\ddot{x}^i = -\partial^i \Phi$ requires $\Gamma^i_{00} \equiv \partial^i \Phi$. The Levi-Civita connection formula from Part II gives $\Gamma^i_{00} = \frac{1}{2} g^{ij}(2\partial_0 g_{0j} - \partial_j g_{00})$. In a static spacetime the metric has no time dependence, so $\partial_0 g_{\mu\nu} = 0$ and the first term vanishes. Using $g^{ij} \approx \delta^{ij}$ to first order in h:

$$\Gamma^i_{00} = -\tfrac{1}{2}\delta^{ij}\partial_j g_{00} = -\tfrac{1}{2}\partial^i g_{00} \equiv \partial^i \Phi,$$

so $\partial_i g_{00} = -2\partial_i \Phi$, giving $g_{00} = -2\Phi + C$ for some constant C. Far from any source $\Phi \to 0$ and spacetime must be flat, so $g_{00} \to \eta_{00} = -1$, fixing $C = -1$.

Therefore $g_{00} = -(1 + 2\Phi)$, or equivalently $h_{00} = -2\Phi$. The explicit formula for the Ricci tensor (Part II) gives $R_{00} = \partial_\rho \Gamma^\rho_{00} - \partial_0 \Gamma^\rho_{\rho 0} + \Gamma \cdot \Gamma$ terms. The ∂_0 term vanishes by the static assumption, and the $\Gamma \cdot \Gamma$ terms are $O(h^2)$, i.e. second order in the perturbation. To first order, then:

$$R_{00} = \partial_i \Gamma^i_{00} + O(h^2) \approx \nabla^2 \Phi.$$

Taking the trace of the EFE: $-R = 8\pi G T$, and in this limit $T = g^{\mu\nu} T_{\mu\nu} \approx -\rho$, so $R \approx 8\pi G \rho$. Substituting into the $(0, 0)$ component of the EFE yields $R_{00} + \frac{1}{2} R = 8\pi G \rho \Rightarrow R_{00} = 4\pi G \rho$, whence:

$$\boxed{\nabla^2 \Phi = 4\pi G \rho,}$$

which we recognize as Poisson's equation of Newtonian gravity. The factor $8\pi G$ is precisely what makes this work.

The contracted Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$, proved in Part II as a geometric fact, now carries physical content. The EFE gives immediately $\nabla^\mu T_{\mu\nu} = 0$. Energy-momentum conservation is not an independent postulate but a consequence of the field equations. The geometry enforces what the physics requires.

19.3 The Cosmological Constant

As noted above, restricting to second-order field equations leaves one remaining freedom: a multiple of $g_{\mu\nu}$ itself. Including it gives the most general EFE consistent with the symmetry and divergence-free constraints:

$$\boxed{G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu}.}$$

The term $\Lambda g_{\mu\nu}$ is automatically consistent since $\nabla^\mu(\Lambda g_{\mu\nu}) = 0$ identically. Moving it to the right defines a contribution $T^{(\Lambda)}_{\mu\nu} := -(\Lambda/8\pi G)g_{\mu\nu}$, a perfect fluid with $p = -\rho$. Whether Λ belongs to the geometry or to the matter content is a matter of interpretation. Observationally, $\Lambda > 0$ drives the accelerating expansion of the universe as we will see below.

---

20 Killing Fields

In physics, symmetry produces conservation laws. If the metric components do not depend on some coordinate, say $\partial_t g_{\alpha\beta} = 0$, then nothing in the geometry changes as t advances, and the spacetime has a time-translation symmetry.

However, this statement depends on the choice of coordinates. A **Killing field** is the covariant version; it is a vector field K^μ satisfying **Killing's equation**:

$$\boxed{\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0.}$$

To see that this is indeed the coordinate-independent statement of metric independence, take $K = \partial_\sigma$ where σ is a fixed coordinate, so $K^\mu = \delta^\mu_\sigma$ and $K_\nu = g_{\nu\sigma}$. Expanding $\nabla_\mu K_\nu = \partial_\mu g_{\nu\sigma} - \Gamma^\lambda_{\mu\nu} g_{\lambda\sigma}$ and symmetrizing:

$$\nabla_\mu K_\nu + \nabla_\nu K_\mu = \partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - 2\Gamma^\lambda_{\mu\nu} g_{\lambda\sigma}.$$

Substituting the Levi-Civita connection formula from Part II and using $g^{\lambda\rho} g_{\lambda\sigma} = \delta^\rho_\sigma$:

$$\begin{aligned} \Gamma^\lambda_{\mu\nu} g_{\lambda\sigma} &= \tfrac{1}{2} g^{\lambda\rho} g_{\lambda\sigma} (\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}) \\ &= \tfrac{1}{2} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}). \end{aligned}$$

Substituting back:

$$\begin{aligned} \nabla_\mu K_\nu + \nabla_\nu K_\mu = \partial_\mu g_{\nu\sigma} &+ \partial_\nu g_{\mu\sigma} \\ &- (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) = \partial_\sigma g_{\mu\nu}. \end{aligned}$$

Killing's equation for $K = \partial_\sigma$ is therefore exactly $\partial_\sigma g_{\mu\nu} = 0$, which shows that the metric is independent of x^σ. Because Killing's equation is a tensor equation, our above example can be generalized to all coordinate systems. It captures the same symmetry without requiring coordinates adapted to it.

The conserved quantity follows. Let $u^\mu = dx^\mu/d\lambda$ be the tangent to a geodesic. Then:

$$\frac{D}{d\lambda}(K_\mu u^\mu) = u^\nu \nabla_\nu (K_\mu u^\mu) = u^\mu u^\nu \nabla_\nu K_\mu + K_\mu \, u^\nu \nabla_\nu u^\mu.$$

The second term vanishes by the geodesic equation $u^\nu \nabla_\nu u^\mu = 0$. For the first, symmetrize in μ and ν:

$$u^\mu u^\nu \nabla_\nu K_\mu = \tfrac{1}{2}\, u^\mu u^\nu (\nabla_\mu K_\nu + \nabla_\nu K_\mu) = 0,$$

using Killing's equation. Therefore:

$$\boxed{K_\mu u^\mu \text{ is constant along any geodesic.}}$$

The result has a clean interpretation: a Killing field is a direction the metric cannot see, and every freely falling particle carries a conserved quantity for each such symmetry of the spacetime. This is the Riemannian analogue of Noether's theorem. Each continuous symmetry of the metric produces a conserved charge along geodesics. As we will soon see, the Schwarzschild metric has two Killing fields, ∂_t and ∂_ϕ, and the corresponding conserved quantities are the energy and angular momentum of a freely falling particle.

21 Black Holes

21.1 Spherical Symmetry and Birkhoff's Theorem

Spherical symmetry is the simplest physically meaningful constraint one can impose on a solution to the EFE. The most general metric possessing spherical symmetry can be written, in suitable coordinates, as

$$ds^2 = -e^{2\alpha(t,r)}dt^2 + e^{2\beta(t,r)}dr^2 + r^2\,d\Omega^2,$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2$ is the round metric on the unit 2-sphere, for two unknown functions α and β. Inserting the Christoffel symbols (computed from the Levi-Civita connection formula) into $R_{\mu\nu} = \partial_\rho\Gamma^\rho_{\nu\mu} - \partial_\nu\Gamma^\rho_{\rho\mu} + \Gamma^\rho_{\rho\lambda}\Gamma^\lambda_{\nu\mu} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\rho\mu}$ yields four nontrivial components:

$$\begin{aligned}
R_{tt} &= e^{2(\alpha-\beta)}\left[\alpha'' + (\alpha')^2 - \alpha'\beta' + \tfrac{2\alpha'}{r}\right] \\
&\quad - \left[\ddot\beta + (\dot\beta)^2 - \dot\alpha\dot\beta\right], \\
R_{rr} &= -\alpha'' - (\alpha')^2 + \alpha'\beta' + \tfrac{2\beta'}{r} \\
&\quad + e^{2(\beta-\alpha)}\left[\ddot\beta + (\dot\beta)^2 - \dot\alpha\dot\beta\right], \\
R_{tr} &= \frac{2\dot\beta}{r}, \\
R_{\theta\theta} &= e^{-2\beta}\left[r(\alpha' - \beta') + 1\right] - 1,
\end{aligned}$$

where primes denote ∂_r and dots ∂_t. The vacuum condition $R_{\mu\nu} = 0$ requires each of these to vanish. Setting $R_{tr} = 0$

gives immediately $\dot{\beta} = 0$, so $\beta = \beta(r)$ only. With $\dot{\beta} = 0$, the R_{tt} and R_{rr} equations reduce to:

$$\alpha'' + (\alpha')^2 - \alpha'\beta' + \tfrac{2\alpha'}{r} = 0,$$
$$-\alpha'' - (\alpha')^2 + \alpha'\beta' + \tfrac{2\beta'}{r} = 0.$$

Adding these two equations gives $\frac{2}{r}(\alpha' + \beta') = 0$, whence $\alpha + \beta = f(t)$ for some function of t alone. Since $\alpha = f(t) - \beta$, the metric reads $-e^{2f(t)}e^{-2\beta}dt^2 + e^{2\beta}dr^2 + r^2\, d\Omega^2$. Redefining the time coordinate $d\tilde{t}^2 := e^{2f(t)}dt^2$ absorbs the time-dependent factor, leaving $-e^{-2\beta}d\tilde{t}^2 + e^{2\beta}dr^2 + r^2\, d\Omega^2$. In the new coordinate the metric has $\alpha = -\beta$. The remaining equation is $R_{\theta\theta} = 0$, which becomes

$$e^{-2\beta}(1 - 2r\beta') = 1,$$

which rearranges to

$$\frac{d}{dr}\left[r\, e^{-2\beta}\right] = 1,$$

which integrates to $e^{-2\beta} = 1 - r_s/r$ for a constant of integration r_s, where we choose the sign to be negative for reasons that will be clear in the next subsection. Since $\alpha = -\beta$, we have $e^{2\alpha} = 1 - r_s/r$ as well. The metric is static (no t-dependence in any component, no $dt\, dr$ cross terms) and uniquely determined by r_s. This is **Birkhoff's theorem**: spherical symmetry alone forces the vacuum solution to be static. A pulsating but spherically symmetric star has the same exterior geometry as a static one. Note that the metric is a vacuum EFE $R_{\mu\nu} = 0$ solution by construction, having been derived directly from those equations.

21.2 The Schwarzschild Metric

A **black hole** is a region of spacetime from which no future-directed causal curve escapes to any point outside the region. The boundary of that region is the **event horizon**. The Schwarzschild metric describes the unique spherically symmetric vacuum geometry outside such an object. By Birkhoff's theorem, the exterior is the same regardless of how the black hole formed or what collapsed to produce it. The mass M is the only parameter — the geometry carries no other information about the source. Inside the horizon, as we will see, the causal structure of the metric forces every future-directed worldline toward the singularity. The impossibility of escape is not a matter of speed but of geometry.

The spherically symmetric metric determined by Birkhoff's theorem is called the **Schwarzschild metric**:

$$\boxed{ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2\, d\Omega^2,}$$

where $r_s = 2GM$ is the **Schwarzschild radius**. The constant M is the mass of the source, identified by comparison with the Newtonian limit of §19. The metric satisfies $R_{\mu\nu} = 0$ for all $r > 0$.

Recall from §17 that proper time τ is the time measured by a clock carried along a worldline. For a static observer in Schwarzschild spacetime ($dr = d\theta = d\phi = 0$), this gives $d\tau = \sqrt{1 - r_s/r}\, dt$. The coordinate time t is the time registered by a clock at $r \to \infty$. A static clock at finite $r > r_s$ runs slow by the factor $\sqrt{1 - r_s/r}$. At $r = r_s$ the factor vanishes, so $d\tau = 0$, and time appears to freeze here, a pathology explained below. For $r < r_s$ the factor is negative and $d\tau^2 < 0$ for any static trajectory. This means such trajectories are spacelike, and there are no static observers inside the horizon at all (this is also addressed below).

21.3 Energy and Angular Momentum

The metric is independent of both t and ϕ, so by §20, ∂_t and ∂_ϕ are Killing fields. For any geodesic, spherical symmetry allows us to rotate coordinates so that both the initial position and initial tangent vector lie in the equatorial plane $\theta = \pi/2$. The map $\theta \to \pi - \theta$ is a reflection of the sphere through the equatorial plane, and it is a symmetry of the Schwarzschild metric. Applying it to our geodesic produces another curve with the same initial position and tangent vector. By uniqueness of geodesics, the two curves must be identical, so the geodesic stays in the plane $\theta = \pi/2$. To compute the conserved charges associated with our Killing fields, consider a timelike geodesic with tangent $u^\mu = dx^\mu/d\tau$ (we include a conventional minus sign in the definition of E to ensure positive energy):

$$\boxed{\begin{aligned} E &= -(\partial_t)_\mu\, u^\mu = -g_{\mu\nu}(\partial_t)^\nu\, u^\mu = -g_{\mu\nu}\delta^\nu_t\, u^\mu \\ &= -g_{tt}\frac{dt}{d\tau} = \left(1 - \frac{r_s}{r}\right)\frac{dt}{d\tau}, \\ L &= (\partial_\phi)_\mu\, u^\mu = g_{\phi\phi}\frac{d\phi}{d\tau} = r^2\frac{d\phi}{d\tau}. \end{aligned}}$$

E and L are the energy and angular momentum per unit mass, conserved along any geodesic. E is the product of the Lorentz factor $\gamma = dt/d\tau$ and the factor $(1 - r_s/r)$ from

g_{tt}, which encodes depth in the gravitational field. The first grows as the particle moves faster. The second shrinks as the particle falls deeper. Their product is conserved: kinetic energy gained equals gravitational potential energy lost, just as in Newtonian mechanics. $L = r^2\, d\phi/d\tau$ is the GR analogue of the Newtonian angular momentum $r^2\dot{\phi}$, with $L = 0$ for purely radial motion. Both are conserved because of the symmetries of the metric, exactly as in §20.

21.4 Redshift

Gravitational redshift follows from the same Killing field applied to a photon. A photon travels on a null geodesic. Since $d\tau = 0$, it must be parametrized by an affine parameter λ instead. Its tangent vector $p^\mu = dx^\mu/d\lambda$ is the photon 4-momentum. Because it travels on a geodesic, the Killing field conservation law applies, giving the conserved quantity $(\partial_t)_\mu p^\mu = -(1 - r_s/r)\, p^t \equiv -\varepsilon$. A static observer at radius r has no spatial motion, so $u^\mu_{\text{obs}} = (u^t, 0, 0, 0)$. The proper time normalization condition $g_{\mu\nu}u^\mu u^\nu = -1$ reduces to $g_{tt}(u^t)^2 = -1$, and since $g_{tt} = -(1 - r_s/r)$, we get $u^t = (1 - r_s/r)^{-1/2}$. In the observer's local rest frame, $u^\mu_{\text{obs}} = (1, 0, 0, 0)$ and the metric is locally $\eta_{\mu\nu}$. The photon 4-momentum is $p^\mu = (E, \vec{p})$, so $-p_\mu u^\mu_{\text{obs}} = -p_t = E$, using $p_t = \eta_{tt}p^t = -E$ from the signature. The contraction projects the 4-momentum onto the observer's time direction, picking out the energy. It's clear that $-g_{\mu\nu}p^\mu u^\nu_{\text{obs}}$ is a scalar, since a full contraction of tensors takes the same numerical value in every coordinate system. Because it equals E in the rest frame, it equals E in every frame. Since frequency is proportional to energy via $E = hf$, the observed frequency of the photon is proportional to $-g_{\mu\nu}p^\mu u^\nu_{\text{obs}}$. Evaluating with our static observer:

$$\begin{aligned} -g_{\mu\nu}p^\mu u^\nu_{\text{obs}} &= -g_{tt}p^t u^t_{\text{obs}} \\ &= (1 - r_s/r)\cdot p^t \cdot (1 - r_s/r)^{-1/2} \\ &= \frac{\varepsilon}{\sqrt{1 - r_s/r}}, \end{aligned}$$

so the observed frequency is proportional to $\varepsilon/\sqrt{1 - r_s/r}$. Since ε is the same constant at emission and reception, the ratio of observed to emitted frequency is:

$$\boxed{\frac{f_{\text{o}}}{f_{\text{e}}} = \sqrt{\frac{1 - r_s/r_e}{1 - r_s/r_o}},}$$

where r_e is the radius of emission and r_o the radius of observation. A photon climbing outward ($r_e < r_o$) arrives redshifted ($f_{\text{o}} < f_{\text{e}}$). As $r_e \to r_s$, the ratio tends to zero. The photon's frequency as measured by any outside observer vanishes and no signal from the horizon is ever received.

21.5 Light Deflection

A photon passing close to a massive body is deflected from its straight-line path. We can compute the deflection angle for a null geodesic in Schwarzschild spacetime. By the same symmetry argument as in §21.3, we may restrict to the equatorial plane $\theta = \pi/2$. As in §21.4, the photon is parametrized by an affine parameter λ, and its tangent vector is $p^\mu = dx^\mu/d\lambda$, with non-zero components $p^t = dt/d\lambda$,

$p^r = dr/d\lambda$, $p^\phi = d\phi/d\lambda$. The null condition $g_{\mu\nu}p^\mu p^\nu = 0$ is then evaluated by substituting the Schwarzschild metric components $g_{tt} = -(1-r_s/r)$, $g_{rr} = (1-r_s/r)^{-1}$, $g_{\phi\phi} = r^2$, giving:

$$-\left(1-\frac{r_s}{r}\right)(p^t)^2 + \left(1-\frac{r_s}{r}\right)^{-1}(p^r)^2 + r^2(p^\phi)^2 = 0.$$

Using the Killing conserved quantities $\varepsilon = (1-r_s/r)\,p^t$ and $L = r^2\,p^\phi$, and defining the **impact parameter** $b = L/\varepsilon$, substitution gives:

$$(p^r)^2 = \varepsilon^2\left[1-\frac{b^2}{r^2}\left(1-\frac{r_s}{r}\right)\right].$$

To find the shape of the trajectory $r(\phi)$, we eliminate λ by noting that the chain rule applied to $r(\phi(\lambda))$ gives $p^r = p^\phi(dr/d\phi)$, so $(p^r)^2 = (p^\phi)^2(dr/d\phi)^2 = (L^2/r^4)(dr/d\phi)^2 = (\varepsilon^2 b^2/r^4)(dr/d\phi)^2$, and thus:

$$\left(\frac{dr}{d\phi}\right)^2 = \frac{r^4}{b^2} - r^2 + r_s r.$$

Now substitute $u = 1/r$, so $dr/d\phi = -(1/u^2)\,du/d\phi$, and:

$$\left(\frac{du}{d\phi}\right)^2 = \frac{1}{b^2} - u^2 + r_s u^3.$$

Differentiating each side with respect to ϕ and dividing by $2\,du/d\phi$ gives:

$$\frac{d^2u}{d\phi^2} + u = \frac{3}{2}r_s\,u^2.$$

In the flat spacetime limit (i.e. when $M \to 0$ and therefore $r_s = 2GM \to 0$), the right-hand side vanishes, leaving $u'' + u = 0$, with solution $u_0 = \frac{1}{b}\sin\phi$, i.e. $r = b/\sin\phi$, a horizontal straight line at perpendicular distance b from the origin. For the full GR equation, we substitute $u = u_0 + u_1$ where u_1 is a small correction of order r_s. Keeping only first-order terms gives the equation for the correction:

$$u_1'' + u_1 = \tfrac{3}{2} r_s u_0^2 = \frac{3r_s}{2b^2}\sin^2\phi = \frac{3r_s}{4b^2}(1 - \cos 2\phi).$$

This is a driven harmonic oscillator with solution:

$$u_1 = \frac{r_s}{2b^2}(1 + \cos^2\phi).$$

The full solution is $u = u_0 + u_1 = \frac{1}{b}\sin\phi + \frac{r_s}{2b^2}(1 + \cos^2\phi)$. In asymptotic terms, the photon travels from $u = 0$ at $\phi = 0$ (incoming) to $u = 0$ at $\phi = \pi + \delta\phi/2$ (outgoing), where $\delta\phi/2$ is the deviation from the flat-space asymptote on the outgoing side. By symmetry the incoming leg contributes an equal $\delta\phi/2$, for a total deflection $\delta\phi$. Setting $u = 0$ at $\phi = \pi + \delta\phi/2$ and substituting into the full solution:

$$0 = \frac{1}{b}\sin\left(\pi + \frac{\delta\phi}{2}\right) + \frac{r_s}{2b^2}\left(1 + \cos^2\left(\pi + \frac{\delta\phi}{2}\right)\right).$$

Expanding to first order in $\delta\phi$ and r_s, and using $\sin(\pi + \delta\phi/2) \approx -\delta\phi/2$ and $\cos^2(\pi + \delta\phi/2) \approx 1$, we obtain:

$$0 = -\frac{\delta\phi/2}{b} + \frac{r_s}{b^2},$$

whence $\delta\phi/2 = r_s/b$, giving

$$\boxed{\delta\phi = \frac{2r_s}{b} = \frac{4GM}{b},}$$

twice the value obtained from a Newtonian calculation treating light as a particle. This was confirmed by Eddington's measurement of starlight deflection during the 1919 solar eclipse.

21.6 The Event Horizon and the Singularity

At $r = r_s$, the Schwarzschild coordinates break down: $g_{tt} = 0$ and $g_{rr} \to \infty$. To determine whether this is a genuine curvature singularity, one inspects a coordinate-independent quantity. The **Kretschmann scalar** $\mathcal{K} = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ for the Schwarzschild solution evaluates to:

$$\mathcal{K} = \frac{48\, G^2 M^2}{r^6}.$$

At $r = r_s$ this is finite, which means the geometry is regular there. The apparent singularity is an artifact of the coordinate system. A change of coordinates, such as Eddington-Finkelstein or Kruskal-Szekeres, covers the manifold smoothly through $r = r_s$.

The surface $r = r_s$ is the event horizon, a null surface. Radial outgoing light emitted exactly at $r = r_s$ stays there, neither escaping nor falling in. For $r < r_s$, the metric components g_{tt} and g_{rr} exchange signs. g_{tt} becomes positive and g_{rr} negative, so r is now the timelike coordinate and t

spacelike. This is not a metaphor. Outside the horizon you can remain stationary in r and move forward in t. Inside, r plays the role t played outside. Just as you cannot avoid moving forward in time, you cannot avoid moving toward smaller r. Every future-directed causal curve inside the horizon points toward decreasing r. Escape is impossible. It would require moving backward in the timelike direction, which is causally forbidden by the same logic that forbids moving backward in time. Outside the horizon, the impossibility of moving backward in time is a postulate. We impose $u^t > 0$ as the time orientation, and physical worldlines must respect it. Inside, the same time orientation, extended smoothly through the horizon, designates decreasing r as the future direction. The geometry has squeezed the future light cone so that it points entirely toward smaller r. The direction of increasing r is not merely forbidden but spacelike. It has ceased to be a timelike direction at all. The horizon is the precise surface where "outward" stops being a meaningful notion for a future-directed curve.

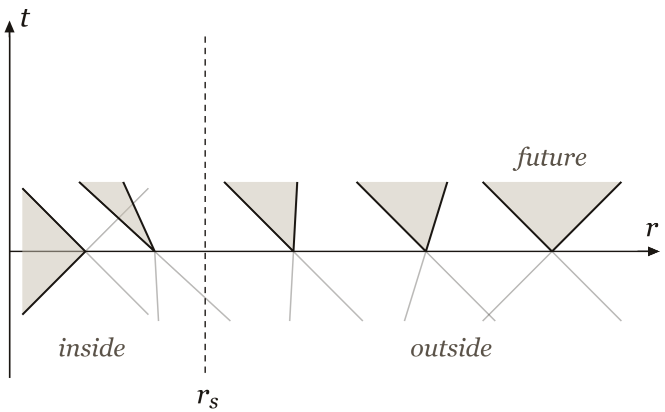
The behavior of radial null geodesics makes this clear. Setting $ds^2 = 0$ in the Schwarzschild metric and, for simplicity, $d\Omega^2 = 0$:

$$0 = -\left(1 - \tfrac{r_s}{r}\right) dt^2 + \left(1 - \tfrac{r_s}{r}\right)^{-1} dr^2.$$

The coordinate velocity of radial light is then:

$$\frac{dr}{dt} = \pm\left(1 - \frac{r_s}{r}\right).$$

The factor $(1 - r_s/r)$ determines everything. Outside the horizon it is positive and the light cone opens in the direction of increasing time, as usual. At $r = r_s$ it vanishes. Both branches give $dr/dt = 0$, and outgoing light makes no radial progress in coordinate time. Inside $(1 - r_s/r) < 0$, the $+$ branch now gives $dr/dt < 0$, and both null rays move toward smaller r. The sign flip at $r = r_s$ is the entire mechanism.



An external observer watching an infalling body sees it slow, redshift, and asymptotically freeze at the horizon, but not linger there visibly. The slowing is a coordinate effect. In Schwarzschild coordinates, successive events near the horizon are stretched over ever-longer intervals of coordinate time t, so the body appears to decelerate as it approaches r_s. The redshift drops the energy of each arriving photon exponentially, and the rate of arrival drops too. The body fades to invisibility on a very short timescale. The infalling observer, meanwhile, crosses the horizon in finite proper time and notices nothing locally special at $r = r_s$. At $r = 0$, the Kretschmann scalar diverges. This is a genuine curvature singularity. Geodesics reaching $r = 0$ cannot be extended; the spacetime is **geodesically incomplete**. In general relativity, singularities are defined by this incompleteness rather than by any coordinate pathology. The singularity at $r = 0$ is not a location in space but a moment in the future of every interior worldline as inevitable as tomorrow.

21.7 Black Hole Categorization

A black hole is not simply a very massive object. Every massive object has a Schwarzschild radius $r_s = 2GM$, and any matter compressed to a size smaller than its own r_s is a black hole. There is nothing else to it. A star of mass M and a black hole of the same mass produce identical gravitational fields in their exteriors. A distant observer cannot distinguish them by any gravitational measurement. What makes a black hole different is purely structural. All the mass lies within r_s, the Schwarzschild metric holds all the way to $r = 0$, and the causal structure forbids any signal from escaping. The horizon is the defining feature, not the mass.

The Schwarzschild solution describes a black hole that is uncharged and non-rotating. Two exact generalizations are known. The **Reissner-Nordström metric** describes a charged, non-rotating black hole, parameterized by mass M and electric charge Q. The **Kerr metric** describes an uncharged, rotating black hole, parameterized by M and angular momentum J. It reduces to Schwarzschild when $J = 0$. The fully general case, with both charge and rotation, is the **Kerr-Newman metric**, parameterized by M, Q, and J.

The **no-hair theorem** states that these three parameters are the only ones needed. Any black hole that forms from gravitational collapse, regardless of the complexity of the initial matter distribution, settles into a Kerr-Newman state. All other information about the collapsing matter is lost to an outside observer. A black hole has no hair.

---

22 Cosmology

22.1 The Cosmological Principle

On sufficiently large scales, the universe is isotropic at every point. The metric looks the same in every direction from every location. This imposed requirement, the **cosmological principle**, severely constrains the metric. Isotropy at a point p means the spatial metric at p is symmetric under the action of SO(3), the rotation group in three spatial dimensions.

The **sectional curvature** of a 2-plane $\Pi \subset T_pM$ spanned by vectors u, v is:

$$K(\Pi) := \frac{R_{ijkl}\, u^i v^j u^k v^l}{(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk})\, u^i v^j u^k v^l},$$

where γ_{ij} is the metric on the 3-dimensional spatial section and u^i, v^j are the components of the spanning vectors. The denominator is $|u|^2|v|^2(1 - \cos^2\theta) = |u|^2|v|^2 \sin^2\theta$, where θ is the angle between u and v. This is the squared area of the parallelogram spanned by u and v, which normalizes K so that it is independent of the choice of spanning vectors. The numerator equals $\langle R(u, v)v,\, u\rangle$ for any u, v, where $R(u, v) = [\nabla_u, \nabla_v]$ is the curvature operator from Part II. For orthonormal u, v the denominator is 1 and

$K = \langle R(u,v)v,\, u\rangle$ directly. Recall that $R(u,v)$ measures the failure of parallel transport around an infinitesimal loop with sides in the directions u and v. Parallel transporting any vector w around that loop returns it with a deviation of $R(u,v)w$ from its original orientation. The sectional curvature $K(\Pi)$ is the component of this rotation acting within Π. It measures how much v rotates toward u. On a flat space $R(u,v) = 0$ everywhere, so $K = 0$. On a general manifold K varies across both points and planes.

Since any rotation in SO(3) can carry any 2-plane to any other, every 2-plane at p has the same sectional curvature. Isotropy at p reduces it to a single number $K(p)$. It is not obvious that K must be the same at every point. The argument has two stages: first, isotropy forces the Riemann tensor into a specific algebraic form. Then the contracted Bianchi identity forces K to be constant.

Isotropy at p forces a specific form on the Riemann tensor at p. Taking p fixed for the moment, define $T_{ijkl} := R_{ijkl} - K(p)(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk})$, where all quantities are evaluated at p. Since every 2-plane at p has curvature $K(p)$, we have $T_{ijkl}u^i v^j u^k v^l = 0 \;\forall\, u, v \in T_pM$. This does not immediately imply $T = 0$, since contractions of the form $u^i v^j u^k v^l$ are a restricted class (not all possible index combinations). The algebraic Bianchi identity, however, allows us to upgrade this to $T_{ijkl}u^i v^j w^k z^l = 0 \;\forall$ independent u, v, w, z, which does imply $T = 0$. The γ-combination shares three symmetries with R_{ijkl}, so T inherits them: pair-swap $T_{ijkl} = T_{klij}$, antisymmetry in the last pair $T_{ijkl} = -T_{ijlk}$, and the algebraic Bianchi identity $T_{i[jkl]} = 0$. Three steps now reduce T to zero:

1. Substitute $u \to u+w$: expanding $T_{ijkl}(u^i+w^i)v^j(u^k+w^k)v^l = 0$ and canceling the vanishing $T(\cdot, v, \cdot, v)$

terms leaves $T_{ijkl}u^i v^j w^k v^l + T_{ijkl}w^i v^j u^k v^l = 0$. Pair-swap makes both terms equal, so $T_{ijkl}u^i v^j w^k v^l = 0$.

2. Substitute $v \to v+z$ in the result of step 1: expanding and canceling the vanishing terms gives

$$T_{ijkl}u^i v^j w^k z^l + T_{ijkl}u^i z^j w^k v^l = 0$$

so $T(u,v,w,z) = -T(u,z,w,v)$.

3. The algebraic Bianchi identity gives $T(u,v,w,z) + T(u,w,z,v) + T(u,z,v,w) = 0$. Applying the result of step 2: $T(u,w,z,v) = -T(u,v,z,w) = T(u,v,w,z)$ (last step by antisymmetry in the last pair); similarly $T(u,z,v,w) = -T(u,z,w,v) = T(u,v,w,z)$. So $3T(u,v,w,z) = 0$, whence $T = 0$.

So $R_{ijkl} = K(p)(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk})$: isotropy forces the Riemann tensor into this form at each point. We now contract to find the Einstein tensor. The Ricci tensor $R_{jl} = \gamma^{ik}R_{ijkl}$ gives:

$$\begin{aligned} R_{jl} &= K\gamma^{ik}(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk}) \\ &= K(3\,\gamma_{jl} - \delta^k_l\,\gamma_{jk}) = 2K\gamma_{jl}. \end{aligned}$$

The Ricci scalar and Einstein tensor follow immediately:

$$\begin{aligned} R &= \gamma^{jl}R_{jl} = 6K, \\ G_{ij} &= R_{ij} - \tfrac{1}{2}R\,\gamma_{ij} = 2K\gamma_{ij} - 3K\gamma_{ij} = -K\gamma_{ij}. \end{aligned}$$

Since $\nabla^i G_{ij} = 0$, proved in Part II from the Bianchi identity, and $\nabla^i \gamma_{ij} = 0$:

$$0 = \nabla^i G_{ij} = -(\partial^i K)\gamma_{ij} = -\partial_j K.$$

So $\partial_j K = 0$ at every point, i.e. K is the same constant everywhere. This is **Schur's theorem**. Simply-connected 3-manifolds of constant sectional curvature are classified by the sign of K. The magnitude can always be normalized away by rescaling the spatial coordinates $r \to r/\sqrt{|K|}$, leaving only $k \in \{+1, 0, -1\}$. The sign of K has a direct reading from the definition $K \propto \langle R(u,v)v,\, u\rangle$. It tells you which way $R(u,v)v$ (the deviation that v picks up from the curvature in the u-v plane) points relative to u. If $K > 0$, the deviation has a component toward u. This means the space folds inward and initially parallel geodesics converge. If $K < 0$, the deviation points away, which means the space opens up and geodesics diverge. If $K = 0$, there is of course no deviation. These cases are conventionally called spherical, hyperbolic, and flat spatial sections.

22.2 The FLRW Metric

The previous subsection established that the spatial sections have constant sectional curvature $k \in \{+1, 0, -1\}$. We now derive the coordinate form of the metric from this constraint. Isotropy forces the spatial metric to be spherically symmetric. Define r as the area-radius, so the angular part is exactly $r^2\, d\Omega^2$ by convention. The most general spherically symmetric Riemannian 3-metric is then

$$d\sigma^2 = B(r)\, dr^2 + r^2\, d\Omega^2$$

for some $B(r)$ to be determined. There are two independent sectional curvatures: $K_{r\theta}$ for a radial-angular 2-plane, and $K_{\theta\phi}$ for a purely tangential 2-plane. For the radial curvature, take $u = \partial_r$, $v = \partial_\theta$ as spanning vectors. The numerator $R_{ijkl}u^i v^j u^k v^l$ collapses to $R_{r\theta r\theta}$, and the denominator

$(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk})u^i v^j u^k v^l$ to $g_{rr}g_{\theta\theta} - g_{r\theta}^2 = g_{rr}g_{\theta\theta}$, since the metric is diagonal. So $K_{r\theta} = R_{r\theta r\theta}/(g_{rr}g_{\theta\theta})$. The Christoffel symbols can be computed directly from the metric, from which $R_{r\theta r\theta} = rB'/(2B)$ and therefore $K_{r\theta} = B'/(2rB^2)$. Following a similar procedure for the tangential curvature, taking $u = \partial_\theta$, $v = \partial_\phi$, one finds $K_{\theta\phi} = (1 - B^{-1})r^{-2}$.

Setting $K_{\theta\phi} \equiv k$ gives $B(r) = (1 - kr^2)^{-1}$, and one then checks that $K_{r\theta} = B'/(2rB^2) = k$ as well (with $B' = 2krB^2$). The spatial metric with constant sectional curvature k is therefore:

$$d\sigma^2 = \frac{dr^2}{1 - kr^2} + r^2\, d\Omega^2.$$

The cosmological principle fixes the spatial geometry at each moment of time to be of this form, but allows the overall scale to vary. A time-dependent conformal factor $a(t)^2$ stretches or compresses the spatial sections uniformly without changing their curvature type. Observers to whom the universe looks isotropic, i.e. observers at rest with respect to the expansion with fixed spatial coordinates (r, θ, ϕ), are called **comoving**. They measure proper time t, called **cosmic time**. Assembling these gives the **Friedmann-Lemaître-Robertson-Walker (FLRW) metric**:

$$\boxed{ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2\, d\Omega^2 \right].}$$

The **scale factor** $a(t)$ encodes how physical distances evolve, with the proper distance between any two comoving points scaling as $a(t)$. Conventionally $a(t_0) \equiv 1$ at the present time t_0. This recession is not motion through space, but rather the stretching of space itself. A comoving observer has no peculiar velocity. They are simply carried along as the spatial sections expand. The cosmological principle fixes the form of the metric, and the requirement that it solves the EFE then imposes additional constraints on $a(t)$, which we derive in the next subsection.

22.3 The Friedmann Equations

Cosmological matter and radiation are modeled as a perfect fluid, so $T^{\mu\nu} = (\rho+p)u^\mu u^\nu + pg^{\mu\nu}$. For comoving observers $u^\mu = (1,0,0,0)$, the mixed components are simply:

$$T^\mu{}_\nu = \mathrm{diag}(-\rho,\, p,\, p,\, p).$$

The connection coefficients can be computed directly. For Γ^0_{ij}, for example, the definition gives

$$\begin{aligned}\Gamma^0_{ij} &= \tfrac{1}{2}g^{00}(\partial_i g_{j0} + \partial_j g_{i0} - \partial_0 g_{ij}) \\ &= \tfrac{1}{2}(-1)(0 + 0 - 2a\dot{a}\,\gamma_{ij}) = a\dot{a}\,\gamma_{ij}.\end{aligned}$$

The factor γ_{ij} survives because it is the time-independent part of $g_{ij} = a(t)^2\gamma_{ij}$. Similarly, one finds:

$$\Gamma^i_{0j} = \tfrac{1}{2}g^{ik}\,\partial_t g_{jk} = \tfrac{1}{2}\cdot\frac{\gamma^{ik}}{a^2}\cdot 2a\dot{a}\,\gamma_{jk} = \frac{\dot{a}}{a}\,\delta^i_j.$$

The spatial Christoffels Γ^i_{jk} of γ_{ij} could also of course be computed directly, but for our purposes we only need the Riemann tensor of γ_{ij}, which satisfies $\tilde{R}_{ijkl} = k(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk})$ as established in §22.1. Inserting everything into the Riemann formula, the nonzero contributions to $R^0{}_0$ come from $\partial_t \Gamma^i_{i0}$ and the product $\Gamma^i_{0j}\Gamma^j_{i0}$ (with factors of 3 arising from the sum over three spatial dimensions):

$$R_{00} = -3\,\partial_t\big(\tfrac{\dot a}{a}\big) - 3\big(\tfrac{\dot a}{a}\big)^2 = -3\Big(\tfrac{\ddot a}{a} - \big(\tfrac{\dot a}{a}\big)^2\Big) - 3\big(\tfrac{\dot a}{a}\big)^2 = -3\frac{\ddot a}{a},$$

and raising with $g^{00} = -1$ gives $R^0{}_0 = 3\ddot a/a$.

For $R^i{}_j$, the Riemann contraction splits as $R^i{}_j = R^0{}_{i0j} + R^k{}_{ikj}$. The first term involves $\partial_t \Gamma^i_{0j}$ and products of Γ^i_{0j}, giving the $\ddot a/a$ and $(\dot a/a)^2$ pieces. The second involves the spatial block $\tilde R^k{}_{ikj} = \tilde R_{ij}$, which was computed in §22.1: $\tilde R_{ij} = 2k\gamma_{ij}$. This contributes $2k/a^2\,\delta^i_j$ after raising with $g^{ik} = \gamma^{ik}/a^2$:

$$R^i{}_j = \left[\frac{\ddot a}{a} + 2\left(\frac{\dot a}{a}\right)^2\right]\delta^i_j + \frac{2k}{a^2}\,\delta^i_j = \left[\frac{\ddot a}{a} + 2\left(\frac{\dot a}{a}\right)^2 + \frac{2k}{a^2}\right]\delta^i_j.$$

The Ricci scalar $R = R^\mu{}_\mu$ follows by tracing:

$$R = 6\left[\frac{\ddot a}{a} + \left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2}\right].$$

The Einstein tensor $G^\mu{}_\nu = R^\mu{}_\nu - \frac{1}{2}\delta^\mu_\nu R$ then gives:

$$G^0{}_0 = -3\left[\left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2}\right],$$

$$G^i{}_j = -\left[2\frac{\ddot a}{a} + \left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2}\right]\delta^i_j.$$

Applying the EFE with cosmological constant, $G^{\mu}{}_{\nu} + \Lambda\delta^{\mu}_{\nu} = 8\pi G\, T^{\mu}{}_{\nu}$, the 00 component gives the first **Friedmann equation** directly:

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}.}$$

The spatial component gives $-[2\ddot{a}/a + (\dot{a}/a)^2 + k/a^2] + \Lambda = 8\pi G p$, which still contains $(\dot{a}/a)^2$ and k/a^2. Substituting $(\dot{a}/a)^2 + k/a^2 = 8\pi G\rho/3 + \Lambda/3$ from the first equation and rearranging yields the **acceleration equation**:

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}.}$$

These are the conditions that $a(t)$ must satisfy for the FLRW metric to be a solution of the Einstein field equations. The cosmological principle fixes the form of the metric. The EFE then constrain how the scale factor evolves.

Covariant conservation $\nabla_\mu T^{\mu\nu} = 0$ holds in any solution to the EFE, as established in §19. Applying it to the FLRW case, the covariant divergence $\nabla_\mu T^{\mu}{}_{\nu} = 0$ for $\nu = 0$ reads:

$$\partial_0 T^0{}_0 + \Gamma^{\mu}_{\mu 0} T^0{}_0 - \Gamma^{i}_{j0} T^j{}_i = 0.$$

With $T^\mu{}_\nu = \mathrm{diag}(-\rho, p, p, p)$, isotropy killing spatial gradients, and $\Gamma^i_{i0} = \dot{a}/a$, this becomes $-\dot{\rho} - 3(\dot{a}/a)\rho - 3(\dot{a}/a)p = 0$, giving the **continuity equation**:

$$\boxed{\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0.}$$

The three equations are not independent. Conservation is already encoded in the geometry, so the continuity equation follows from the Friedmann and acceleration equations and does not add new information. The Friedmann equation and the continuity equation together determine the evolution of $a(t)$ once p is specified as a function of ρ.

22.4 Equation of State and Cosmic Expansion

The Friedmann and continuity equations relate three unknowns: a, ρ, and p. To close the system, one typically specifies how pressure relates to energy density. This is a constitutive relation that encodes the empirical material properties of whatever fills the universe. The isotropy of FLRW forces both ρ and p to depend only on t. If either varied spatially, its gradient would define a preferred spatial direction, contradicting isotropy. We therefore see that p is parametrically related to ρ through their common time dependence. The simplest such relation is linear, so a classic choice for the **equation of state** $p(\rho)$ is $p \equiv w\rho$, where

w is a constant characterizing how much pressure a substance generates per unit energy density. Substituting into the continuity equation:

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a},$$

which integrates to $\rho \propto a^{-3(1+w)}$. Three cases cover the main constituents of the universe:

- **Pressureless matter** ($w = 0$): $\rho \propto a^{-3}$. Energy density dilutes inversely with volume, the expected behavior for non-relativistic particles.
- **Radiation** ($w = \frac{1}{3}$): $\rho \propto a^{-4}$. One power beyond volume dilution accounts for the redshifting of each photon's energy as space expands.
- **Cosmological constant** ($w = -1$): $\rho = \text{const}$. Moving Λ to the right-hand side of the EFE identifies an effective energy density $\rho_\Lambda = \Lambda/8\pi G$ (established in §19), which is a fixed constant of nature, not carried by particles. As space expands, each new region carries the same ρ_Λ, so it does not dilute.

Spatial flatness $k = 0$ is not assumed but observed. Measurements of the temperature anisotropies of the cosmic microwave background (CMB) constrain the spatial curvature to be consistent with zero to high precision. This is a statement about the large-scale geometry of the universe, not local curvature (spacetime is of course highly curved near black holes, stars, and galaxies). We take $k = 0$ henceforth. Each constituent is then solved for in isolation (setting the others to zero). For $w = 0$ and $w = \frac{1}{3}$, one sets $\Lambda = 0$.

For $w = -1$, $\rho = \text{const}$ so the Friedmann equation gives $\dot{a}/a = \sqrt{\Lambda/3} = \text{const}$, whence exponential growth:

$$a(t) \propto \begin{cases} t^{2/3} & (w = 0,\ \Lambda = 0), \\ t^{1/2} & (w = \frac{1}{3},\ \Lambda = 0), \\ e^{\sqrt{\Lambda/3}\,t} & (w = -1). \end{cases}$$

With $\Lambda = 0$, matter and radiation both decelerate expansion. The acceleration equation gives $\ddot{a} < 0$ whenever $\rho + 3p > 0$, i.e. $w > -\frac{1}{3}$. A positive Λ adds a competing $+\Lambda/3$ term to $\ddot{a}/a$, which can reverse the deceleration and drive exponential expansion.

The reason Λ is included here is observational. Measurements of Type Ia supernovae in 1998 established that the universe's expansion is accelerating: $\ddot{a} > 0$. The acceleration equation requires $\rho + 3p < 0$ for this, which no combination of ordinary matter or radiation can provide. A positive Λ, however, can lead to positive acceleration. As matter and radiation dilute away with expansion, the $\Lambda/3$ term in the Friedmann equation remains constant and eventually dominates, locking in exponential growth. Whether Λ represents a geometric property of spacetime or a physical energy density of the vacuum is an open question. As a matter contribution, it goes by the name dark energy. Either way, $\Lambda > 0$ is the simplest model consistent with observation.

The **Hubble parameter** $H(t) = \dot{a}/a$ measures the instantaneous rate of expansion. Its present value $H_0 \approx 67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ is the Hubble constant. For a comoving object at coordinate r, the proper distance is

$D = a(t)\, r$ (using $k = 0$, so the spatial metric is flat and r is the Euclidean radial coordinate). Differentiating:

$$\dot{D} = \dot{a}\, r = \frac{\dot{a}}{a} \cdot a\, r = H\, D.$$

Recession velocity is proportional to proper distance. At the present epoch this is **Hubble's law**, $v = H_0 D$. No peculiar motion is assumed: the recession is expansion of the space between observers, not motion through it.

Not everything expands. The FLRW metric describes the universe on scales large enough that matter is distributed homogeneously. On smaller scales, local gravitational and electromagnetic binding overwhelms the cosmological expansion entirely. Solar systems, galaxies, and even galaxy clusters are held together by forces that dominate locally. They do not participate in the expansion. It is only on the scale of the separation between galaxy clusters and beyond that the expansion is the dominant effect, and comoving coordinates faithfully track physical separations.

Part IV

Appendices

Introduction to the Appendices

Parts I, II, and III tell a single story: from smooth manifolds to the Einstein field equations. The appendices run a parallel track which develops certain aspects of classical field theory in curved spacetime more generally.

§A.1 develops the covariant action principle in full generality: the canonical volume form, the Christoffel trace identity, and the covariant Euler-Lagrange equations, with the Klein-Gordon field as a concrete example. §A.2 applies this framework to classical electrodynamics, using the Hodge star to write Maxwell's equations in differential forms notation and verify they reproduce the classical equations in curved spacetime. §A.3 derives the Einstein field equations from the Einstein-Hilbert action by explicit variation, with a careful treatment of why the Euler-Lagrange formalism of §A.1 does not directly apply and how the Palatini identity resolves the difficulty.

These appendices are not required to follow the main text, but they are not merely supplementary either. A reader who has finished Part III and wants to understand the generic underlying classical field theoretic machinery will find this a natural next step.

A.1 Classical Field Theory

A.1.1 The Canonical Volume Form

Fix an oriented pseudo-Riemannian n-manifold (M, g). The metric g determines a canonical n-form, the **volume form**:

$$\boxed{\mathrm{vol} = \sqrt{|g|}\, dx^1 \wedge \cdots \wedge dx^n,}$$

where $g = \det(g_{\mu\nu})$. Under a coordinate change $x \to x'$ the metric determinant transforms as $|g| \to (\det \partial x/\partial x')^2 |g|$. Since $g_{\mu\nu}$ is a $(0, 2)$ tensor it picks up two Jacobian factors, one per index, and taking the determinant of the full transformation law $g'_{\mu\nu} = (\partial x^\alpha/\partial x'^\mu)(\partial x^\beta/\partial x'^\nu) g_{\alpha\beta}$ yields a square. Taking the square root, $\sqrt{|g|}$ transforms with exactly one factor of $\det(\partial x/\partial x')$. Meanwhile $dx^1 \wedge \cdots \wedge dx^n \to (\det \partial x'/\partial x)^{-1} dx'^1 \wedge \cdots \wedge dx'^n$. The two factors cancel, and orientation-preservation ensures the sign is correct. Note that any nondegenerate $(0, 2)$ tensor would work by the same argument; we use the metric because it is the geometric structure already in hand. With our signature $(-, +, +, +)$ the determinant satisfies $g < 0$, so $\sqrt{|g|} = \sqrt{-g}$. As a concrete check: in flat spacetime with spherical coordinates, $g_{\mu\nu} = \mathrm{diag}(-1, 1, r^2, r^2 \sin^2\theta)$, so $g = -r^4 \sin^2\theta$ and $\sqrt{-g} = r^2 \sin\theta$, recovering the familiar volume factor in $d^4x = r^2 \sin\theta\, dt\, dr\, d\theta\, d\phi$.

A.1.2 The Christoffel Trace Identity

A key identity used throughout these appendices is $\Gamma^{\mu}_{\mu\lambda} = \partial_\lambda \ln\sqrt{-g}$. Computing $\Gamma^{\mu}_{\mu\lambda}$ explicitly from the Christoffel definition, the first and third terms cancel on tracing (relabeling $\mu \leftrightarrow \rho$ in the third shows it equals minus the first), leaving:

$$\Gamma^{\mu}_{\mu\lambda} = \tfrac{1}{2}\, g^{\mu\rho}\partial_\lambda g_{\mu\rho}.$$

To evaluate $g^{\mu\rho}\partial_\lambda g_{\mu\rho}$, we need the matrix identity $\partial_\lambda(\det A) = \det(A)\cdot \mathrm{tr}(A^{-1}\partial_\lambda A)$, which we prove in three steps, assuming the well-known Laplace expansion $\det(A) = \sum_j A_{ij}C_{ij}$, where i is any fixed row (the sum gives the same result for every choice of i), and C_{ij} is the determinant of A with the ith row and jth column removed.

Step 1: cofactor formula. Since C_{ij} involves no entry from row i, differentiating the Laplace expansion gives $\partial(\det A)/\partial A_{ij} = C_{ij}$.

Step 2: adjugate identity. The (i,k) entry of $A\cdot \mathrm{adj}(A)$, where $\mathrm{adj}(A)_{ij} = C_{ji}$, is $\sum_j A_{ij}C_{kj}$. For $i = k$ this is the Laplace expansion, giving $\det(A)$; for $i \neq k$ it is the determinant of a matrix with two identical rows (i.e., A with row k replaced by row i), giving 0. Therefore:

$$A\cdot \mathrm{adj}(A) = \det(A)\cdot I \implies A\cdot\frac{\mathrm{adj}(A)}{\det A} = I$$
$$\implies A^{-1} = \frac{\mathrm{adj}(A)}{\det A},$$

hence $C_{ij} = \det(A)\cdot(A^{-1})_{ji}$.

Step 3: contract and conclude. Combining and contracting with $\partial_\lambda A_{ij}$:

$$\begin{aligned}\partial_\lambda(\det A) &= \frac{\partial \det(A)}{\partial A_{ij}}\,\partial_\lambda A_{ij} \\ &= \det(A)\cdot(A^{-1})_{ji}\,\partial_\lambda A_{ij} \\ &= \det(A)\cdot \mathrm{tr}(A^{-1}\partial_\lambda A).\end{aligned}$$

Dividing by $\det(A)$ gives $\mathrm{tr}(A^{-1}\partial_\lambda A) = \partial_\lambda \ln|\det A|$. Applied to $g_{\mu\nu}$: $g^{\mu\rho}\partial_\lambda g_{\mu\rho} = \partial_\lambda \ln|g|$, so:

$$\Gamma^\mu_{\mu\lambda} = \tfrac{1}{2}\,\partial_\lambda \ln|g| = \partial_\lambda \ln\sqrt{-g}.$$

This is the desired identity. □

A.1.3 The Least Action Principle

The action principle extends naturally to curved spacetime for any classical field. The general framework developed here underlies the specific applications in Appendices A.2 and A.3. For a scalar field ϕ on a curved manifold, the action takes the form

$$\boxed{S := \int_M \mathcal{L}(\phi, \nabla_\mu\phi)\,\mathrm{vol} = \int d^4x\,\sqrt{-g}\,\mathcal{L}(\phi, \nabla_\mu\phi),}$$

where $\mathcal{L}$ is a Lorentz scalar Lagrangian and $\sqrt{-g}\,d^4x$ is the covariant volume element. Varying with respect to ϕ:

$$\delta S = \int d^4x\,\sqrt{-g}\left[\frac{\partial\mathcal{L}}{\partial\phi}\,\delta\phi + \frac{\partial\mathcal{L}}{\partial(\nabla_\mu\phi)}\,\nabla_\mu(\delta\phi)\right].$$

Setting $V^\mu := \partial\mathcal{L}/\partial(\nabla_\mu\phi)$ and applying the Leibniz rule gives $V^\mu\nabla_\mu(\delta\phi) = \nabla_\mu(V^\mu\delta\phi) - (\nabla_\mu V^\mu)\delta\phi$. Writing $W^\mu := V^\mu\delta\phi$, we can calculate the covariant divergence:

$$\begin{aligned}
\nabla_\mu W^\mu &= \partial_\mu W^\mu + \Gamma^\mu_{\mu\lambda} W^\lambda \\
&= \partial_\mu W^\mu + (\partial_\lambda \ln\sqrt{-g})\, W^\lambda \\
&= \frac{1}{\sqrt{-g}}\,\partial_\mu(\sqrt{-g}\, W^\mu),
\end{aligned}$$

where $\Gamma^\mu_{\mu\lambda} = \partial_\lambda \ln\sqrt{-g}$ is proved above, so $\int d^4x\,\sqrt{-g}\,\nabla_\mu W^\mu = \int d^4x\,\partial_\mu(\sqrt{-g}\,W^\mu)$ is a surface term and vanishes by the generalized Stokes theorem:

$$\delta S = \int d^4x\,\sqrt{-g}\left[\frac{\partial\mathcal{L}}{\partial\phi} - \nabla_\mu\left(\frac{\partial\mathcal{L}}{\partial(\nabla_\mu\phi)}\right)\right]\delta\phi = 0.$$

Since this holds $\forall\,\delta\phi$, the **covariant Euler-Lagrange equations** are

$$\boxed{\frac{\partial\mathcal{L}}{\partial\phi} - \nabla_\mu\left(\frac{\partial\mathcal{L}}{\partial(\nabla_\mu\phi)}\right) = 0.}$$

In flat spacetime $\nabla_\mu \to \partial_\mu$ and $\sqrt{-g} \to 1$, recovering the standard result.

A.1.4 Example: Klein-Gordon Field

The Lagrangian for a real scalar field of mass m is

$$\boxed{\mathcal{L} = -\tfrac{1}{2}\left(g^{\mu\nu}\nabla_\mu\phi\,\nabla_\nu\phi + m^2\phi^2\right).}$$

The two terms are $\partial\mathcal{L}/\partial\phi = -m^2\phi$ and $\partial\mathcal{L}/\partial(\nabla_\mu\phi) = -\nabla^\mu\phi$. Substituting:

$$-m^2\phi + \nabla_\mu\nabla^\mu\phi = 0,$$

i.e. $(\Box - m^2)\phi = 0$, the **covariant Klein-Gordon equation**, where $\Box = \nabla_\mu\nabla^\mu$ is the covariant d'Alembertian.

A.2 Classical Electrodynamics in Curved Spacetime

A.2.1 The Gauge Potential

Electromagnetism on a 4-manifold begins with a 1-form A, the **gauge potential** (the relativistic packaging of the scalar and vector potentials). The **field strength** is

$$\boxed{F = dA.}$$

F is a 2-form; its 6 independent components $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ (defined via $F = \frac{1}{2}F_{\mu\nu}\, dx^\mu \wedge dx^\nu$, summing over all ordered pairs μ, ν, with the $\frac{1}{2}$ absorbing the double-counting) encode the electric and magnetic fields: $F_{0i} = -E_i$ and $F_{ij} = \varepsilon_{ijk}B^k$.

The Maxwell Lagrangian $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ fits the framework of §A.1 with the 1-form A_μ in place of ϕ. Since $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ depends only on first derivatives of A, the Euler-Lagrange equations give $\partial\mathcal{L}/\partial A_\nu = 0$ and

$\partial\mathcal{L}/\partial(\partial_\mu A_\nu) = -F^{\mu\nu}$, yielding $\nabla_\mu F^{\mu\nu} = 0$. Adding the coupling term $A_\mu J^\mu$ to the Lagrangian shifts this to $\nabla_\mu F^{\mu\nu} = J^\nu$, the curved-space Maxwell equation developed below.

A.2.2 The Hodge Star

The language of differential forms gives a strikingly clean formulation of Maxwell's equations which lays bare their geometric content. The essential new ingredient is the Hodge star, an operation that uses the metric to convert k-forms into $(4-k)$-forms on spacetime.

Let $\varepsilon_{\mu_1\cdots\mu_n}$ denote the Levi-Civita symbol: $+1$ if $(\mu_1,\ldots,\mu_n)$ is an even permutation of $(0,\ldots,n-1)$, -1 if odd, and 0 if any two indices coincide. The **Hodge star** $\star : \Omega^k \to \Omega^{n-k}$ is defined in coordinates by

$$\boxed{(\star\beta)_{\mu_{k+1}\cdots\mu_n} = \frac{\sqrt{|g|}}{k!}\,\varepsilon_{\mu_1\cdots\mu_n}\,\beta^{\mu_1\cdots\mu_k},}$$

where $\beta^{\mu_1\cdots\mu_k} = g^{\mu_1\nu_1}\cdots g^{\mu_k\nu_k}\beta_{\nu_1\cdots\nu_k}$ has indices raised by the metric, and $1/k!$ corrects for summing over all orderings of the contracted indices. The Hodge star therefore replaces a k-form with a specific $(n-k)$-form whose meaning we can interpret geometrically. Raising the indices of β via g identifies the k directions β points in; $\varepsilon_{\mu_1\cdots\mu_n}$ then antisymmetrizes over all n indices and selects the complementary $(n-k)$ directions. The antisymmetry of $\varepsilon_{\mu_1\cdots\mu_n}$ forces $\mu_{k+1},\ldots,\mu_n$ to be exactly the indices not contracted against β (the directions β does not occupy).

A.2.3 Maxwell's Equations in Differential Forms Notation

The first pair of Maxwell equations follow immediately from $d^2 = 0$:

$$\boxed{dF = 0.}$$

No metric is needed. This is a statement about smooth structure alone. In components $dF = 0$ reads $\partial_{[\mu} F_{\nu\rho]} = 0$, i.e.

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0.$$

Taking $(\mu, \nu, \rho) = (1, 2, 3)$ gives $\partial_1 F_{23} + \partial_2 F_{31} + \partial_3 F_{12} = 0$, which in vector notation is $\nabla \cdot \mathbf{B} = 0$. Taking one index to be 0 gives $\partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i} = 0$, i.e. $\partial_t B^k + (\nabla \times \mathbf{E})^k = 0$, Faraday's law.

The second pair requires the metric. Let J be the **current 1-form**, with components $J_\mu = g_{\mu\nu} J^\nu$ where $J^\mu = (\rho, \mathbf{j})$ is the 4-current. The dynamical Maxwell equations are

$$\boxed{d{\star}F = {\star}J.}$$

To unpack this, apply the Hodge star definition to $F = \frac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu$. With $k = 2$ and $n = 4$, the component formula gives:

$$({\star}F)_{\rho\sigma} = \frac{1}{2} \sqrt{-g}\, \varepsilon_{\mu\nu\rho\sigma}\, F^{\mu\nu}.$$

For J a 1-form ($k = 1$), the same formula gives:

$$(\star J)_{\lambda\rho\sigma} = \sqrt{-g}\,\varepsilon_{\mu\lambda\rho\sigma}\,J^{\mu}.$$

Since $d\star F$ is the 3-form with components $\partial_{[\lambda}(\star F)_{\rho\sigma]}$, the component equation $d\star F = \star J$ reads:

$$\partial_\lambda(\star F)_{\rho\sigma} + \partial_\rho(\star F)_{\sigma\lambda} + \partial_\sigma(\star F)_{\lambda\rho} = (\star J)_{\lambda\rho\sigma}.$$

Substituting $(\star F)_{\rho\sigma} = \frac{1}{2}\sqrt{-g}\,\varepsilon_{\alpha\beta\rho\sigma}F^{\alpha\beta}$ and contracting both sides with $\varepsilon^{\nu\lambda\rho\sigma}$, the three cyclic terms on the left are equal by relabeling dummies, giving a factor of 3:

$$\frac{3}{2}\,\varepsilon^{\nu\lambda\rho\sigma}\varepsilon_{\alpha\beta\rho\sigma}\,\partial_\lambda\Big(\sqrt{-g}\,F^{\alpha\beta}\Big)\,.$$

Applying $\varepsilon^{\nu\lambda\rho\sigma}\varepsilon_{\alpha\beta\rho\sigma} = -2\Big(\delta^\nu_\alpha\delta^\lambda_\beta - \delta^\nu_\beta\delta^\lambda_\alpha\Big)$ and contracting the deltas:

$$-3\,\Big[\partial_\beta\Big(\sqrt{-g}\,F^{\nu\beta}\Big) - \partial_\alpha\big(\sqrt{-g}\,F^{\alpha\nu}\big)\Big] = -6\,\partial_\mu\big(\sqrt{-g}\,F^{\mu\nu}\big)\,.$$

On the other hand, we also have $\varepsilon^{\nu\lambda\rho\sigma}\varepsilon_{\alpha\lambda\rho\sigma} = -6\,\delta^\nu_\alpha$, so $\varepsilon^{\nu\lambda\rho\sigma}(\star J)_{\lambda\rho\sigma} = -6\sqrt{-g}\,J^\nu$. Canceling -6 gives:

$$\partial_\mu(\sqrt{-g}\,F^{\mu\nu}) = \sqrt{-g}\,J^\nu.$$

Taking $\nu = 0$: $F^{i0} = g^{i\rho}g^{0\sigma}F_{\rho\sigma} = -E^i$, so $\partial_i(\sqrt{-g}\,E^i) = \sqrt{-g}\,\rho$. In Minkowski space, $\sqrt{-g} = 1$ giving $\nabla\cdot\mathbf{E} = \rho$, Gauss's law. Taking $\nu = i$: the spatial components give (again in Minkowski space) $\partial_t E^i - (\nabla\times\mathbf{B})^i = -J^i$, Ampère's law with the displacement current.

The forms notation is certainly elegant and concise, but its real benefit is making the geometric structure manifest. $dF = 0$ and $d\star F = \star J$ are the Maxwell equations on any pseudo-Riemannian manifold. In the component form $\partial_\mu(\sqrt{-g}\,F^{\mu\nu}) = \sqrt{-g}\,J^\nu$, the $\sqrt{-g}$ inside the derivative is what makes the expression coordinate-independent, accounting for a volume element that varies across spacetime. In Minkowski space $\sqrt{-g} = 1$ throughout, and the equation reduces immediately to $\partial_\mu F^{\mu\nu} = J^\nu$, the classical form. Geometrically, F is a 2-form: it measures flux through 2-surfaces. The equation $dF = 0$ says this flux has no boundary. This is a purely topological statement requiring no metric, encoding the absence of magnetic monopoles and Faraday's law. The second equation $d{\star}F = {\star}J$ introduces geometry via $\star$. To define the directions complementary to those F occupies, you need a notion of perpendicularity, which is exactly what the metric provides. It is the divergence of this complementary form that is sourced by the current, encoding Gauss's law and Ampère's law.

A.2.4 Minimal Coupling

The prescription of minimal coupling promotes the flat-space equation $\partial_\mu F^{\mu\nu} = J^\nu$ to curved spacetime by replacing $\partial_\mu \to \nabla_\mu$, giving $\nabla_\mu F^{\mu\nu} = J^\nu$. Expanding the covariant derivative:

$$\nabla_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \Gamma^\mu_{\mu\lambda} F^{\lambda\nu} + \Gamma^\nu_{\mu\lambda} F^{\mu\lambda}.$$

The last term vanishes: $\Gamma^\nu_{\mu\lambda}$ is symmetric in μ, λ while $F^{\mu\lambda}$ is antisymmetric. By the Christoffel trace identity of §A.1, $\Gamma^\mu_{\mu\lambda} = \partial_\lambda \ln\sqrt{-g}$. Substituting and applying the product rule:

$$\nabla_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \left(\partial_\lambda \ln\sqrt{-g}\right) F^{\lambda\nu} = \frac{1}{\sqrt{-g}}\,\partial_\mu\left(\sqrt{-g}\,F^{\mu\nu}\right).$$

So $\nabla_\mu F^{\mu\nu} = J^\nu$ is equivalent to $\partial_\mu(\sqrt{-g}\, F^{\mu\nu}) = \sqrt{-g}\, J^\nu$, exactly the component form of $d{\star}F = {\star}J$ derived above. The forms equation is the minimally coupled curved-space Maxwell equation.

A.3 The Einstein-Hilbert Action

A.3.1 The Action

The EFE can be derived from an action principle. The **Einstein-Hilbert action** is:

$$\boxed{S = \frac{1}{16\pi G}\int d^4x\,\sqrt{-g}\,R + S_{\text{matter}}.}$$

Here S_{matter} is the action for whatever matter fields are present; its variation with respect to the metric defines $T_{\mu\nu}$, as detailed in §A.3.4. Among diffeomorphism-invariant scalars built from $g_{\mu\nu}$ and at most two derivatives, R is a simple, natural choice, and we will see that it indeed produces the EFE. Setting $\delta S/\delta g^{\mu\nu} = 0$ requires three ingredients which we develop below.

Before proceeding, it is worth asking why we do not simply apply the Euler-Lagrange formalism of §A.1. The answer is that the framework there assumes a Lagrangian depending on a field and its first derivatives. Here the field is $g_{\mu\nu}$ and the Lagrangian is R, which contains second derivatives of the metric through the Christoffel symbols. The Euler-Lagrange equations of §A.1 therefore do not apply without modification, and the explicit variation of S is the cleaner route for our purposes. The Palatini identity handles the problematic second-derivative terms by showing they form a total divergence, disposing of them as a boundary contribution.

Turning to the variation, write $\delta S = \delta S_H + \delta S_{\text{matter}}$, where $\delta S_H = \frac{1}{16\pi G}\int d^4x\, \delta(\sqrt{-g}\,R)$. Expanding $\delta(\sqrt{-g}\,R) = \delta\sqrt{-g}\cdot R + \sqrt{-g}\cdot\delta(g^{\mu\nu}R_{\mu\nu})$ and splitting the last factor:

$$\delta S_H = \frac{1}{16\pi G}\int d^4x\big[R\,\delta\sqrt{-g} + \sqrt{-g}\,R_{\mu\nu}\,\delta g^{\mu\nu} \\ + \sqrt{-g}\,g^{\mu\nu}\delta R_{\mu\nu}\big].$$

Each of the three terms requires separate treatment, handled in the subsections below.

A.3.2 Variation of $\sqrt{-g}$

The matrix identity proved in §A.1 gives $\delta(\det A) = (\det A)\,\mathrm{tr}(A^{-1}\delta A)$ for any invertible A. Applied to $g_{\mu\nu}$:

$$\delta g = g\, g^{\mu\nu}\delta g_{\mu\nu}.$$

To express $\delta g_{\mu\nu}$ in terms of $\delta g^{\mu\nu}$, vary the identity $g^{\mu\alpha}g_{\alpha\nu} = \delta^{\mu}_{\nu}$. The right side is constant, so $\delta g^{\mu\alpha}\cdot g_{\alpha\nu} + g^{\mu\alpha}\cdot\delta g_{\alpha\nu} = 0$, giving $g^{\mu\alpha}\delta g_{\alpha\nu} = -g_{\alpha\nu}\delta g^{\mu\alpha}$. Multiplying both sides by $g_{\mu\beta}$:

$$\delta g_{\beta\nu} = -g_{\mu\beta}\,g_{\alpha\nu}\,\delta g^{\mu\alpha}.$$

Contracting: $g^{\mu\nu}\delta g_{\mu\nu} = g^{\mu\nu}(-g_{\mu\alpha}g_{\nu\beta}\delta g^{\alpha\beta}) = -\delta^{\nu}_{\alpha}g_{\nu\beta}\delta g^{\alpha\beta} = -g_{\alpha\beta}\delta g^{\alpha\beta} = -g_{\mu\nu}\delta g^{\mu\nu}$. Since $g < 0$:

$$\delta\sqrt{-g} = -\tfrac{1}{2}\sqrt{-g}\, g_{\mu\nu}\,\delta g^{\mu\nu}.$$

A.3.3 The Palatini Identity

We need $g^{\mu\nu}\delta R_{\mu\nu}$. Christoffel symbols are not tensors; to see why, start from the definition $\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu})$. Under $x \to x'$ the metric transforms as $g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}g_{\alpha\beta}$, so differentiating with the chain and product rules gives:

$$\begin{aligned}
\partial'_\lambda g'_{\mu\nu} &= \partial'_\lambda\left[\frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}g_{\alpha\beta}\right] \\
&= \frac{\partial x^\gamma}{\partial x'^\lambda}\frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}\,\partial_\gamma g_{\alpha\beta} \\
&\quad + g_{\alpha\beta}\left[\frac{\partial^2 x^\alpha}{\partial x'^\lambda \partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu} + \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial^2 x^\beta}{\partial x'^\lambda \partial x'^\nu}\right].
\end{aligned}$$

Forming the Christoffel combination $\partial'_\mu g'_{\nu\lambda} + \partial'_\nu g'_{\mu\lambda} - \partial'_\lambda g'_{\mu\nu}$: the single-derivative terms assemble into $2\frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}\frac{\partial x^\gamma}{\partial x'^\lambda}g_{\gamma\sigma}\Gamma^{\sigma}_{\alpha\beta}$, while four of the six second-derivative terms cancel in pairs, leaving $2g_{\alpha\beta}\frac{\partial x^\beta}{\partial x'^\lambda}\frac{\partial^2 x^\alpha}{\partial x'^\mu \partial x'^\nu}$. Multiplying by $\frac{1}{2}g'^{\rho\lambda}$ and using the inverse metric transformation together with the Jacobian identity $\frac{\partial x'^\lambda}{\partial x^\nu}\frac{\partial x^\beta}{\partial x'^\lambda} = \delta^\beta_\nu$, both contractions reduce the same way: $g'^{\rho\lambda}\frac{\partial x^\gamma}{\partial x'^\lambda}g_{\gamma\sigma} = \frac{\partial x'^\rho}{\partial x^\mu}g^{\mu\nu}g_{\nu\sigma} = \frac{\partial x'^\rho}{\partial x^\sigma}$, and similarly

$g'^{\rho\lambda} g_{\alpha\beta} \frac{\partial x^\beta}{\partial x'^\lambda} = \frac{\partial x'^\rho}{\partial x^\alpha}$. Putting these pieces together, we get the connection transformation:

$$\Gamma'^{\rho}_{\mu\nu} = \frac{\partial x'^\rho}{\partial x^\sigma} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \Gamma^\sigma_{\alpha\beta} + \frac{\partial x'^\rho}{\partial x^\alpha} \frac{\partial^2 x^\alpha}{\partial x'^\mu \, \partial x'^\nu}.$$

The second term depends only on the coordinate transformation, not on the metric. So for any two connections Γ and $\tilde{\Gamma}$, it is identical in both and cancels in the difference $\Gamma - \tilde{\Gamma}$, which therefore transforms as a $(1, 2)$ tensor.

We now vary R. At any point, choose Riemann normal coordinates so $\Gamma^\lambda_{\mu\nu} = 0$ there. We also know that partial and covariant derivatives then coincide at that point. The quadratic $\Gamma\Gamma$ terms in the Riemann definition vanish, leaving $R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma}$. Varying and replacing ∂ with ∇:

$$\delta R^\rho{}_{\sigma\mu\nu} = \nabla_\mu(\delta\Gamma^\rho_{\nu\sigma}) - \nabla_\nu(\delta\Gamma^\rho_{\mu\sigma}).$$

Here $\delta\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu}[g + \delta g] - \Gamma^\lambda_{\mu\nu}[g]$ is exactly a difference of two connections, so by the result above it is a $(1, 2)$ tensor, justifying the application of ∇. Since this is a tensor equation it holds in all coordinate systems. Setting $\rho = \mu$ gives $\delta R_{\sigma\nu} = \nabla_\mu(\delta\Gamma^\mu_{\nu\sigma}) - \nabla_\nu(\delta\Gamma^\mu_{\mu\sigma})$. Contracting with $g^{\sigma\nu}$ and using $\nabla g = 0$ to pull $g^{\sigma\nu}$ inside ∇:

$$g^{\sigma\nu} \delta R_{\sigma\nu} = \nabla_\mu(g^{\sigma\nu} \delta\Gamma^\mu_{\nu\sigma}) - \nabla_\nu(g^{\sigma\nu} \delta\Gamma^\mu_{\mu\sigma}).$$

Relabeling $\nu \to \mu$ in the second term (both are dummy indices) combines the two into a single divergence:

$$g^{\mu\nu} \delta R_{\mu\nu} = \nabla_\mu Z^\mu, \qquad Z^\mu := g^{\sigma\nu} \delta\Gamma^\mu_{\nu\sigma} - g^{\mu\nu} \delta\Gamma^\lambda_{\nu\lambda}.$$

Using the identity $\nabla_\mu Z^\mu = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\, Z^\mu)$ proved in §A.1, the integral $\int d^4x\, \sqrt{-g}\, \nabla_\mu Z^\mu = \int d^4x\, \partial_\mu(\sqrt{-g}\, Z^\mu)$ is a boundary term and vanishes for fields decaying at infinity.

A.3.4 The Matter Variation

The variation of the matter action with respect to $g^{\mu\nu}$ defines the energy-momentum tensor:

$$\delta S_{\text{matter}} =: -\tfrac{1}{2} \int d^4x\, \sqrt{-g}\, T_{\mu\nu}\, \delta g^{\mu\nu}.$$

This is the definition of $T_{\mu\nu}$ in curved spacetime: it is whatever the matter action produces when the metric is varied.

A.3.5 The Einstein Field Equations

Now substitute the three results into the expansion of δS_H from §A.3.1. The first term gives $R\,\delta\sqrt{-g} = -\tfrac{1}{2}\sqrt{-g}\, g_{\mu\nu} R\, \delta g^{\mu\nu}$, the third term $\sqrt{-g}\, g^{\mu\nu}\delta R_{\mu\nu} = \sqrt{-g}\, \nabla_\mu Z^\mu$ is the Palatini boundary term and vanishes, leaving:

$$\begin{aligned}\delta S_H &= \frac{1}{16\pi G}\int d^4x\, \sqrt{-g}\, \left(R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R\right) \delta g^{\mu\nu} \\ &= \frac{1}{16\pi G}\int d^4x\, \sqrt{-g}\, G_{\mu\nu}\, \delta g^{\mu\nu}.\end{aligned}$$

Setting $\delta(S_H + S_{\text{matter}})/\delta g^{\mu\nu} = 0$ and factoring out $\sqrt{-g}\,\delta g^{\mu\nu}$:

$$\boxed{G_{\mu\nu} = 8\pi G\, T_{\mu\nu}.}$$