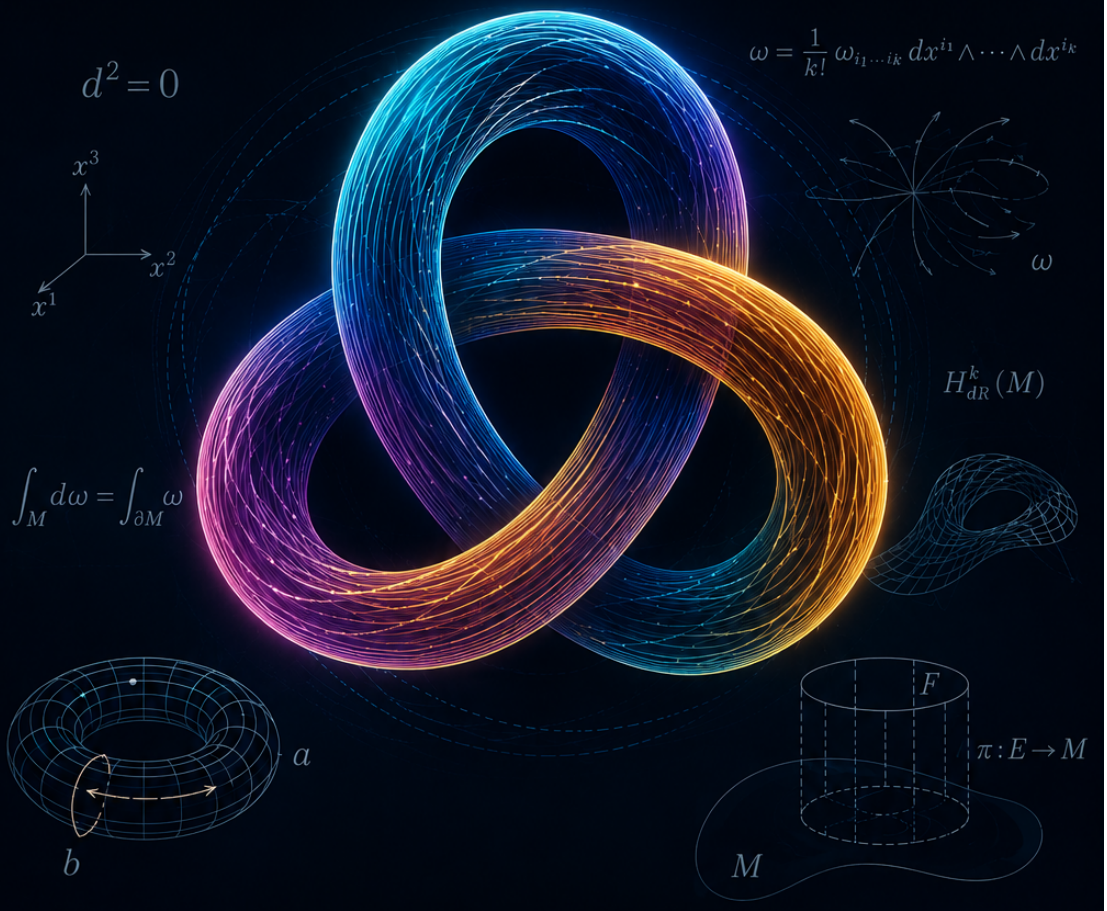
MATHEMATICAL BASICS OF TOPOLOGICAL QUANTUM THEORY



YING NIAN WU

DEPARTMENT OF STATISTICS AND DATA SCIENCE

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Mathematical Basics of Topological Quantum Theory

Ying Nian Wu

April 2026, written with the help of Claude 4 (book cover by GPT 5.5)

Contents

Chapter 1

Differential Forms

1.1 The Problem: Integrals That Do Not Belong to Any Coordinate System

A curve in space is a geometric object. It does not know or care what coordinate system we use to describe it, or how we choose to parametrize it. The work done by a force along that curve is a single number — a geometric fact about the curve and the force field. It does not depend on whether we use Cartesian or polar coordinates, or whether we traverse the curve quickly or slowly in our parametrization.

Yet every concrete computation of that number requires us to choose coordinates and a parametrization. The moment we choose, the formula changes its appearance: different coordinates give different-looking expressions, different parametrizations give different integrands. The geometry is fixed; the formulas are not. This mismatch between the invariant geometry and the variant formulas is the central problem that differential forms solve.

The situation is the same for surface integrals and volume integrals. The flux of a magnetic field through a surface is a single number. The mass of a region is a single number. These numbers depend on the surface or the region and on the field or density, but not on the coordinates we use to describe them or the parametrizations we use to compute them.

Our goal is to write integrals in a form that makes this independence manifest. We want to write

$$\int_M \omega \tag{1.1}$$

on the left-hand side — a single expression involving only the geometric domain M and the integrand ω, with no reference to any coordinate system or parametrization — and then, when computation is needed, to *concretize* this expression by choosing a parametrization and obtaining a standard calculus integral on the right-hand side:

$$\int_M \omega = \int_D \omega\Big(\frac{\partial \phi}{\partial u^1}, \ldots, \frac{\partial \phi}{\partial u^k}\Big)\, du^1 \cdots du^k. \tag{1.2}$$

Here $\phi : D \to M$ is a parametrization and $\partial\phi/\partial u^j$ are the tangent vectors it produces. The left-hand side is *abstract*: it refers only to the geometry. The right-hand side is *concrete*: it is a specific integral in specific coordinates, ready to be evaluated by the techniques of ordinary calculus.

The object ω is what makes this possible. It is a **differential form**: a device that accepts the tangent vectors produced by any parametrization and outputs the scalar density appropriate for integration in that parametrization. If the parametrization changes, the tangent vectors change,

and ω automatically adjusts its output so that the integral comes out the same. The form is, in the language of programming, a *function*: it takes the raw data of a parametrization (the infinitesimal tangent vectors) as input and returns the correctly scaled integrand as output. Different inputs, same final answer.

This chapter defines what a form is, shows concretely how to compute integrals with forms, and demonstrates that the integral is indeed invariant. Everything that follows — the wedge product, the pullback, the antisymmetry — exists to make the passage between (1.1) and (1.2) work correctly.

1.2 What a Form Must Be

Before writing any formulas, we can deduce what properties ω must have, purely from the requirement that the integral be independent of parametrization.

1.2.1 Linearity: Because Reparametrization Rescales Tangent Vectors

Suppose we reparametrize a curve: instead of $\gamma(t)$ on $[a, b]$, we write $\gamma(\psi(s))$ on $[c, d]$, where ψ is a smooth bijection. The tangent vector changes from $\gamma'(t)$ to $\gamma'(\psi(s))\,\psi'(s)$: it gets rescaled by $\psi'(s)$. Meanwhile, the integration element changes from dt to $\psi'(s)\,ds$. For the integral to be invariant, the integrand $\omega(\gamma'(t))$ must scale in a way that exactly absorbs into the new ds:

$$\omega(\gamma'(\psi(s))\,\psi'(s))\,ds = \omega(\gamma'(\psi(s)))\,\psi'(s)\,ds. \tag{1.3}$$

This says $\omega(c\,\mathbf{v}) = c\,\omega(\mathbf{v})$: the form must be **linear** in the tangent vector. Linearity is not an arbitrary axiom; it is forced by the requirement that reparametrization does not change the integral.

For surface integrals (two tangent vectors) and higher, the same argument applies to each tangent vector independently: the form must be **multilinear**, i.e., linear in each input separately.

1.2.2 Antisymmetry: Because Orientation Is Geometric

For a surface parametrized by (s, t), swapping the two parameters $s \leftrightarrow t$ reverses the orientation of the surface (it flips which side is "up"). The two tangent vectors $\mathbf{r}_s$ and $\mathbf{r}_t$ are exchanged. Since the oriented integral should change sign when orientation is reversed (this is what "oriented" means — a flux measured from the opposite side gives the opposite sign), the form must satisfy

$$\omega(\mathbf{r}_t, \mathbf{r}_s) = -\omega(\mathbf{r}_s, \mathbf{r}_t). \tag{1.4}$$

Swapping two inputs negates the output. This is **antisymmetry**, and it is again forced by the geometry: orientation reversal must negate the integral.

Antisymmetry implies that if two inputs are equal, $\omega(\mathbf{v}, \mathbf{v}) = -\omega(\mathbf{v}, \mathbf{v}) = 0$: a "parallelogram" with two identical sides is degenerate and contributes nothing.

1.2.3 The Determinant Connection

A function of k vectors that is multilinear and antisymmetric computes a **determinant**. This is a theorem of linear algebra: the determinant is the unique (up to scale) multilinear alternating function of the columns of a matrix. So a k-form, applied to k vectors, must produce a determinant of the components of those vectors.

This is exactly the property we need. The change-of-variables theorem for multiple integrals says: when you change parametrization, the integration element picks up a factor of the Jacobian

determinant. A k-form, being a determinant-computing machine, automatically produces that Jacobian factor when fed reparametrized tangent vectors. The Jacobian from the form cancels the Jacobian from the change of variables, and the integral is invariant.

Antisymmetry exists because it produces determinants. Determinants exist in this story because the change-of-variables theorem demands them. The entire algebraic structure of forms is reverse-engineered from the requirement of parametrization-independent integration.

1.3 1-Forms: Concretizing Line Integrals

We now make the general picture concrete, starting with the simplest case.

1.3.1 Definition

A **1-form** ω on $\mathbb{R}^n$ is a rule that, at each point, accepts one tangent vector and returns a number, linearly. In coordinates $(x^1, \dots, x^n)$:

$$\omega = f_1\, dx^1 + f_2\, dx^2 + \cdots + f_n\, dx^n, \tag{1.5}$$

where $f_1, \dots, f_n$ are smooth functions of position. The symbols $dx^1, \dots, dx^n$ are the elementary 1-forms that read off components of a vector:

$$dx^i(\mathbf{v}) = v^i. \tag{1.6}$$

That is, dx^i extracts the i-th component of whatever vector is fed to it. The action of a general 1-form on a vector $\mathbf{v}$ is then:

$$\omega(\mathbf{v}) = f_1\, v^1 + f_2\, v^2 + \cdots + f_n\, v^n. \tag{1.7}$$

Why does $dx^i(\mathbf{v}) = v^i$ make sense? Because x^i is the coordinate function that assigns to each point its i-th coordinate. The derivative of x^i in the direction $\mathbf{v}$ is v^i: if you move in the direction $\mathbf{v}$, the i-th coordinate changes at rate v^i. So (1.6) is not a definition imposed from outside but a consequence of calculus we already know.

1.3.2 Concretization: The Line Integral

Given a curve $\gamma : [a, b] \to \mathbb{R}^n$ and a 1-form ω, we concretize the abstract integral $\int_\gamma \omega$ by feeding the tangent vector $\gamma'(t)$ into ω:

$$\int_\gamma \omega = \int_a^b \omega(\gamma'(t))\, dt = \int_a^b \left[\sum_i f_i(\gamma(t))\, \frac{d\gamma^i}{dt} \right] dt. \tag{1.8}$$

The abstract object $\int_\gamma \omega$ has become an ordinary single-variable integral. The form ω has done its job: it accepted the parametrization-dependent tangent vector $\gamma'(t)$ and produced a scalar density ready for integration against dt.

1.3.3 Where Is the Smallness?

In a Riemann sum, we partition the curve into small segments. Each segment is a small displacement vector $\Delta\mathbf{r}_i$. We feed this *small* vector into ω:

$$\omega(\Delta\mathbf{r}_i) = f_1\, \Delta x_i^1 + \cdots + f_n\, \Delta x_i^n, \tag{1.9}$$

which is small because $\Delta\mathbf{r}_i$ is small and ω is linear. Sum these small numbers, take the limit, and we get the integral.

The form itself is a perfectly finite linear function. It does not contain anything infinitesimal. The infinitesimal character of integration comes from feeding it infinitesimal vectors. Linearity is what makes this work: if $\Delta\mathbf{r}_i \approx \gamma'(t_i)\,\Delta t_i$, then

$$\omega(\Delta\mathbf{r}_i) \approx \omega(\gamma'(t_i))\,\Delta t_i. \tag{1.10}$$

The smallness migrates from the vector into the scalar Δt_i, and the finite quantity $\omega(\gamma'(t_i))$ becomes the integrand of an ordinary Riemann sum in t.

1.3.4 Example: Work Along a Helix

Let $\omega = y\,dx - x\,dy + dz$ on $\mathbb{R}^3$, and let $\gamma(t) = (\cos t, \sin t, t)$ for $0 \le t \le 2\pi$.

$$\begin{aligned} \gamma'(t) &= (-\sin t, \cos t, 1), \\ \omega(\gamma'(t)) &= \sin t \cdot (-\sin t) - \cos t \cdot \cos t + 1 = -1 + 1 = 0. \end{aligned} \tag{1.11}$$

Therefore $\int_\gamma \omega = \int_0^{2\pi} 0\,dt = 0$. The form concretized the abstract integral into a computation we can execute in a few lines.

1.3.5 Example: The Same Integral, Two Parametrizations

Let $\omega = x\,dy$ and let C be the upper half of the unit circle, traversed counterclockwise.

Parametrization 1: $\gamma_1(t) = (\cos t, \sin t)$, $0 \le t \le \pi$.

$$\begin{aligned} \gamma_1'(t) &= (-\sin t, \cos t), \\ \omega(\gamma_1'(t)) &= \cos t \cdot \cos t = \cos^2 t, \\ \int_C \omega &= \int_0^\pi \cos^2 t\,dt = \frac{\pi}{2}. \end{aligned} \tag{1.12}$$

Parametrization 2: $\gamma_2(s) = (\cos 2s, \sin 2s)$, $0 \le s \le \pi/2$.

$$\begin{aligned} \gamma_2'(s) &= (-2\sin 2s, 2\cos 2s), \\ \omega(\gamma_2'(s)) &= \cos 2s \cdot 2\cos 2s = 2\cos^2 2s, \\ \int_C \omega &= \int_0^{\pi/2} 2\cos^2 2s\,ds = \frac{\pi}{2}. \end{aligned} \tag{1.13}$$

The second parametrization traverses the curve twice as fast. The tangent vector is twice as large, so ω outputs twice the value. But the parameter interval is half as long. The two effects cancel: ω produced the integrand appropriate for each parametrization, and the abstract integral $\int_C \omega$ came out the same.

1.4 2-Forms: Concretizing Surface Integrals

1.4.1 Definition via Determinants

A surface integral requires two tangent vectors (the surface has two independent directions at each point). A **2-form** is the device that accepts two vectors and returns a number, multilinearly and antisymmetrically.

The elementary 2-form $dx^i \wedge dx^j$ is defined by:

$$(dx^i \wedge dx^j)(\mathbf{u}, \mathbf{v}) = \det \begin{pmatrix} u^i & v^i \\ u^j & v^j \end{pmatrix} = u^i v^j - u^j v^i. \tag{1.14}$$

This is the signed area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$, projected onto the (x^i, x^j)-coordinate plane. The 2×2 determinant ensures multilinearity and antisymmetry automatically.

A general 2-form in $\mathbb{R}^3$ is:

$$\omega = A\, dy \wedge dz + B\, dz \wedge dx + C\, dx \wedge dy, \tag{1.15}$$

with three components (A, B, C), one for each coordinate plane. Its action on two vectors is:

$$\omega(\mathbf{u}, \mathbf{v}) = A(u_y v_z - u_z v_y) + B(u_z v_x - u_x v_z) + C(u_x v_y - u_y v_x). \tag{1.16}$$

The reader familiar with the cross product will recognize this as $\omega(\mathbf{u}, \mathbf{v}) = (A, B, C) \cdot (\mathbf{u} \times \mathbf{v})$: the 2-form computes the same thing as the flux integral $\mathbf{F} \cdot d\mathbf{A}$, but without invoking a cross product or a normal vector. The determinant does the work that the cross product did in multivariable calculus.

1.4.2 The Wedge Product: Algebra of Determinants

The notation $dx^i \wedge dx^j$ is the **wedge product** of two 1-forms. Its algebraic rules follow directly from properties of 2×2 determinants:

$$\begin{aligned} &dx^i \wedge dx^j = -dx^j \wedge dx^i && \text{(swapping columns negates a determinant),} \\ &dx^i \wedge dx^i = 0 && \text{(a determinant with two equal columns is zero).} \end{aligned} \tag{1.17}$$

These two rules, together with bilinearity, are all we need to compute any wedge product.

1.4.3 Concretization: The Surface Integral

Given a surface S parametrized by $\mathbf{r}(s, t)$ over a domain D, the abstract integral $\int_S \omega$ concretizes to:

$$\int_S \omega = \iint_D \omega(\mathbf{r}_s, \mathbf{r}_t)\, ds\, dt. \tag{1.18}$$

The 2-form accepts the two parametrization-dependent tangent vectors $\mathbf{r}_s = \partial\mathbf{r}/\partial s$ and $\mathbf{r}_t = \partial\mathbf{r}/\partial t$, and outputs the scalar density appropriate for integration in the (s, t) parametrization. If we reparametrize, the tangent vectors change, ω adjusts its output, and the integral remains the same.

1.4.4 Where Is the Smallness for Surfaces?

The same mechanism as for curves. In the Riemann sum, each small surface patch is spanned by two small vectors $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$. Feeding them to the 2-form:

$$\omega(\Delta\mathbf{u}, \Delta\mathbf{v}) = A(\Delta u_y\, \Delta v_z - \Delta u_z\, \Delta v_y) + \cdots \tag{1.19}$$

Each term is a product of two small quantities, so the result is small of second order — exactly as the area of the patch is second order. The bilinearity of ω is what makes this work. If $\Delta\mathbf{u} \approx \mathbf{r}_s\, \Delta s$ and $\Delta\mathbf{v} \approx \mathbf{r}_t\, \Delta t$, then bilinearity gives:

$$\omega(\Delta\mathbf{u}, \Delta\mathbf{v}) \approx \omega(\mathbf{r}_s, \mathbf{r}_t)\, \Delta s\, \Delta t. \tag{1.20}$$

The smallness migrates into $\Delta s\, \Delta t$, and the finite quantity $\omega(\mathbf{r}_s, \mathbf{r}_t)$ becomes the integrand of an ordinary double integral.

1.4.5 Example: Flux Through a Hemisphere

Let $\omega = x\,dy\wedge dz + y\,dz\wedge dx + z\,dx\wedge dy$ (this is the 2-form corresponding to the radial vector field $\mathbf{F} = (x, y, z)$). Let S be the upper unit hemisphere, parametrized by $\mathbf{r}(\theta, \phi) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$, $0 \le \theta \le 2\pi$, $0 \le \phi \le \pi/2$.

The tangent vectors are:

$$\begin{aligned} \mathbf{r}_\theta &= (-\sin\phi\sin\theta,\ \sin\phi\cos\theta,\ 0), \\ \mathbf{r}_\phi &= (\cos\phi\cos\theta,\ \cos\phi\sin\theta,\ -\sin\phi). \end{aligned} \tag{1.21}$$

Evaluate each component of ω:

$$\begin{aligned} (dy\wedge dz)(\mathbf{r}_\theta, \mathbf{r}_\phi) &= \sin\phi\cos\theta\cdot(-\sin\phi) - 0\cdot\cos\phi\sin\theta = -\sin^2\phi\cos\theta, \\ (dz\wedge dx)(\mathbf{r}_\theta, \mathbf{r}_\phi) &= 0\cdot\cos\phi\cos\theta - (-\sin\phi\sin\theta)(-\sin\phi) = -\sin^2\phi\sin\theta, \\ (dx\wedge dy)(\mathbf{r}_\theta, \mathbf{r}_\phi) &= (-\sin\phi\sin\theta)\cos\phi\sin\theta - \sin\phi\cos\theta\cdot\cos\phi\cos\theta \\ &= -\sin\phi\cos\phi. \end{aligned} \tag{1.22}$$

So:

$$\begin{aligned} \omega(\mathbf{r}_\theta, \mathbf{r}_\phi) &= \sin\phi\cos\theta\cdot(-\sin^2\phi\cos\theta) + \sin\phi\sin\theta\cdot(-\sin^2\phi\sin\theta) + \cos\phi\cdot(-\sin\phi\cos\phi) \\ &= -\sin^3\phi - \sin\phi\cos^2\phi = -\sin\phi. \end{aligned} \tag{1.23}$$

The integral is:

$$\int_S \omega = \int_0^{2\pi}\int_0^{\pi/2} (-\sin\phi)\,d\phi\,d\theta = -2\pi\big[-\cos\phi\big]_0^{\pi/2} = -2\pi(0-(-1)) = -2\pi. \tag{1.24}$$

The negative sign reflects our orientation convention; with the outward orientation, the flux is 2π. At every step the procedure is mechanical: parametrize, differentiate, feed into the form, integrate.

1.5 3-Forms: Concretizing Volume Integrals

1.5.1 Definition

A **3-form** in $\mathbb{R}^3$ accepts three vectors and returns a signed volume via a 3×3 determinant:

$$(dx\wedge dy\wedge dz)(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \det\begin{pmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{pmatrix}. \tag{1.25}$$

A general 3-form is $\omega = f\,dx\wedge dy\wedge dz$, where f is a scalar density (charge, mass, energy per unit volume).

In $\mathbb{R}^3$ there is only one independent 3-form up to a scalar factor, because there is only one way to choose three directions out of three. In $\mathbb{R}^n$ with $n > 3$, a 3-form would have $\binom{n}{3}$ independent components.

1.5.2 Concretization

For a region V parametrized by $\phi(u, v, w)$:

$$\int_V f\,dx \wedge dy \wedge dz = \iiint_D f(\phi(u,v,w))\,\det\left(\frac{\partial\phi}{\partial u}, \frac{\partial\phi}{\partial v}, \frac{\partial\phi}{\partial w}\right) du\,dv\,dw. \tag{1.26}$$

The 3×3 determinant of the three tangent vectors is the Jacobian $\partial(x, y, z)/\partial(u, v, w)$. When ϕ is the identity (Cartesian coordinates), the Jacobian is 1, and we recover $\iiint f\,dV$. When ϕ is the spherical coordinate map, the Jacobian is $r^2 \sin\phi$, produced automatically by the determinant.

The 3-form absorbed the tangent vectors of the parametrization and output the integrand (density times Jacobian) ready for integration in the parameter coordinates. This is the same concretization pattern as for 1-forms and 2-forms.

1.6 General k-Forms: The Unified Pattern

1.6.1 The Definition

The three cases above (1-forms, 2-forms, 3-forms) are instances of a single construction. A **k-form** on $\mathbb{R}^n$ is a machine that, at each point, accepts k tangent vectors and returns a number, via a $k \times k$ determinant.

The elementary k-form $dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ (with $i_1 < i_2 < \cdots < i_k$) acts on k vectors by:

$$(dx^{i_1} \wedge \cdots \wedge dx^{i_k})(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \det \begin{pmatrix} v_1^{i_1} & v_2^{i_1} & \cdots & v_k^{i_1} \\ v_1^{i_2} & v_2^{i_2} & \cdots & v_k^{i_2} \\ \vdots & \vdots & \ddots & \vdots \\ v_1^{i_k} & v_2^{i_k} & \cdots & v_k^{i_k} \end{pmatrix}. \tag{1.27}$$

This is the signed volume of the parallelepiped $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ projected onto the $(i_1, \ldots, i_k)$-coordinate plane. A general k-form is a linear combination:

$$\omega = \sum_{i_1 < \cdots < i_k} \omega_{i_1 \cdots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \tag{1.28}$$

with smooth coefficient functions $\omega_{i_1 \cdots i_k}$. The number of independent components is $\binom{n}{k}$. For $k > n$, every k-form is identically zero.

A 0-form is simply a smooth function $f : \mathbb{R}^n \to \mathbb{R}$. It requires no input vectors and returns a number at each point.

1.6.2 The Wedge Product in General

Given a p-form α and a q-form β, their **wedge product** $\alpha \wedge \beta$ is a $(p + q)$-form, computed by expanding in basis elements and using the rules (1.17). For example, with $\alpha = 3\,dx + 2\,dy$ and $\beta = x\,dy - z\,dz$:

$$\begin{aligned} \alpha \wedge \beta &= 3x\,dx \wedge dy - 3z\,dx \wedge dz + 2x\underbrace{dy \wedge dy}_{=0} - 2z\,dy \wedge dz \\ &= 3x\,dx \wedge dy - 3z\,dx \wedge dz - 2z\,dy \wedge dz. \end{aligned} \tag{1.29}$$

The general sign rule for swapping two forms past each other:

$$\alpha \wedge \beta = (-1)^{pq}\,\beta \wedge \alpha. \tag{1.30}$$

1.6.3 The Concretization Recipe, Stated Once for All k

Let ω be a k-form on $\mathbb{R}^n$. Let M be a k-dimensional surface parametrized by $\phi : D \to \mathbb{R}^n$, where $D \subset \mathbb{R}^k$. The integral of ω over M is:

$$\boxed{\int_M \omega = \int_D \omega\Big(\frac{\partial \phi}{\partial u^1}, \ldots, \frac{\partial \phi}{\partial u^k}\Big)\, du^1 \cdots du^k.} \tag{1.31}$$

The recipe in four steps:

1. **Parametrize**: write $\phi(u^1, \ldots, u^k)$ mapping a flat domain D to the geometric surface M.
2. **Differentiate**: compute the k tangent vectors $\partial\phi/\partial u^j$.
3. **Feed**: plug these tangent vectors into the k-form. The determinant structure of the form converts parametrization-dependent vectors into a parametrization-appropriate scalar density.
4. **Integrate**: evaluate the resulting ordinary integral over D.

This single formula replaces all the separate integral formulas of multivariable calculus. The dimension of M must equal the degree of ω: a k-form can only be integrated over a k-dimensional domain. This dimensional matching is automatic in the language of forms and prevents the kind of dimensional mismatch errors that are possible in classical notation.

1.7 Why the Integral Is Parametrization-Independent

We have asserted that $\int_M \omega$ does not depend on the choice of ϕ. We now prove this, first by example and then in general.

1.7.1 Line Integral: Two Parametrizations

We already saw this in Section 3 with $\omega = x\,dy$ on the upper semicircle. Two parametrizations with different speeds gave the same integral $\pi/2$. The mechanism: the faster parametrization produced a larger tangent vector, the form (being linear) returned a proportionally larger value, and the shorter parameter interval compensated exactly.

1.7.2 Surface Integral: Reparametrization of a Plane

Consider $\omega = dx \wedge dy$ on the square $[0,1]^2$.

Parametrization 1: $\phi_1(s,t) = (s,t)$. Tangent vectors: $\mathbf{e}_x, \mathbf{e}_y$. $(dx \wedge dy)(\mathbf{e}_x, \mathbf{e}_y) = 1$. Integral: $\int_0^1 \int_0^1 1\, ds\, dt = 1$.

Parametrization 2: $\phi_2(u,v) = (u+v, u-v)$ over the appropriate diamond-shaped domain D', which maps onto the same square. Tangent vectors: $\partial\phi_2/\partial u = (1,1)$, $\partial\phi_2/\partial v = (1,-1)$.

$$(dx \wedge dy)((1,1),(1,-1)) = \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -2. \tag{1.32}$$

The area of D' in (u,v)-space is $1/2$ (each side of the diamond is $1/\sqrt{2}$). And the integral: $\int_{D'} (-2)\, du\, dv = -2 \cdot \frac{1}{2} = -1$.

The sign flip to -1 tells us the parametrization ϕ_2 reverses orientation (the determinant is negative). If we require orientation-preserving parametrizations ($\det > 0$), or account for orientation

by taking the absolute value when appropriate, both parametrizations give the same magnitude. The point is that the form automatically tracked the orientation reversal through the sign of the determinant; we did not need to check by hand.

1.7.3 The General Proof: Why It Always Works

Let $\phi' = \phi \circ \psi$ be a reparametrization, where $\psi : D' \to D$ is a smooth bijection. By the chain rule,

$$\frac{\partial \phi'}{\partial u'^j} = \sum_l \frac{\partial \phi}{\partial u^l} \frac{\partial \psi^l}{\partial u'^j}. \tag{1.33}$$

The k-form is multilinear and antisymmetric, so feeding it these reparametrized tangent vectors produces:

$$\omega\left(\frac{\partial \phi'}{\partial u'^1}, \ldots, \frac{\partial \phi'}{\partial u'^k}\right) = \det\left(\frac{\partial \psi^l}{\partial u'^j}\right) \cdot \omega\left(\frac{\partial \phi}{\partial u^1}, \ldots, \frac{\partial \phi}{\partial u^k}\right). \tag{1.34}$$

The factor $\det\left(\partial \psi^l / \partial u'^j\right)$ is the Jacobian determinant of the reparametrization. This is the key step: the multilinearity and antisymmetry of the form extracted the Jacobian determinant automatically.

Now integrate over D' and apply the standard change-of-variables theorem for the parameter integral:

$$\begin{aligned} \int_{D'} \omega\left(\frac{\partial \phi'}{\partial u'^1}, \ldots\right) du'^1 \cdots du'^k &= \int_{D'} \det(J_\psi) \cdot \omega\left(\frac{\partial \phi}{\partial u^1}, \ldots\right) du'^1 \cdots du'^k \\ &= \int_D \omega\left(\frac{\partial \phi}{\partial u^1}, \ldots\right) du^1 \cdots du^k. \end{aligned} \tag{1.35}$$

The Jacobian determinant from the form combines with $du'^1 \cdots du'^k$ to produce $du^1 \cdots du^k$, and the two integrals are equal. This is parametrization independence.

The argument works for any k and any n. It uses only two things: (i) the form is multilinear and antisymmetric, so it produces the Jacobian determinant; (ii) the change-of-variables theorem absorbs that determinant. The algebraic properties of forms are precisely tailored to produce the factor that the change-of-variables theorem requires.

1.8 Independence of Ambient Coordinates

Parametrization independence says the integral does not depend on how we label points on the surface. But we also want independence from how we label points in the *ambient space*. If the surface sits in $\mathbb{R}^3$, the integral should be the same whether we describe $\mathbb{R}^3$ in Cartesian, polar, or any other coordinates.

1.8.1 Example: Area of a Disk

Consider $\omega = dx \wedge dy$ and the unit disk $D = \{x^2 + y^2 \leq 1\}$.

In Cartesian coordinates: $(dx \wedge dy)(\mathbf{e}_x, \mathbf{e}_y) = 1$, and $\int_D dx \wedge dy = \pi$.

In polar coordinates: Since $x = r\cos\theta$ and $y = r\sin\theta$, we have $dx = \cos\theta\, dr - r\sin\theta\, d\theta$ and $dy = \sin\theta\, dr + r\cos\theta\, d\theta$. Wedging:

$$\begin{aligned} dx \wedge dy &= (\cos\theta\, dr - r\sin\theta\, d\theta) \wedge (\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= r\cos^2\theta\, dr \wedge d\theta + r\sin^2\theta\, dr \wedge d\theta \\ &= r\, dr \wedge d\theta. \end{aligned} \tag{1.36}$$

(We used $dr \wedge dr = 0$, $d\theta \wedge d\theta = 0$, and $d\theta \wedge dr = -dr \wedge d\theta$.) The factor r appeared automatically. Integrating:

$$\int_D dx \wedge dy = \int_0^{2\pi} \int_0^1 r \, dr \, d\theta = \pi. \tag{1.37}$$

The same answer. The polar Jacobian r was not inserted by hand; it was produced by the algebraic rules of the wedge product applied to dx and dy expressed in polar form.

This is the concrete mechanism of coordinate independence: the abstract form ω is one object, but when we concretize it in different coordinate systems, the wedge product algebra automatically generates the appropriate Jacobian. We write the form once; it handles every coordinate system.

1.9 Pullbacks: The Mechanism of Concretization

The passage from the abstract integral to the concrete one — expressing ω in terms of the parametrization coordinates and computing — has a precise name: the **pullback**.

1.9.1 The Definition

Let $\phi : U \to \mathbb{R}^n$ be a smooth map (a parametrization, a coordinate change, or any smooth map). If ω is a k-form on $\mathbb{R}^n$, the **pullback** $\phi^*\omega$ is the k-form on U defined by:

$$(\phi^*\omega)_p(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \omega_{\phi(p)}(d\phi_p(\mathbf{v}_1), \ldots, d\phi_p(\mathbf{v}_k)), \tag{1.38}$$

where $d\phi_p$ is the Jacobian matrix of ϕ at p, acting on vectors by matrix multiplication. In words: to evaluate $\phi^*\omega$ on vectors in the domain, push those vectors forward to the target via $d\phi$, then evaluate ω.

The pullback is the formal expression of concretization. The abstract form ω lives on $\mathbb{R}^n$; the pullback $\phi^*\omega$ lives on the parameter domain U and is ready for integration there:

$$\int_{\phi(U)} \omega = \int_U \phi^*\omega. \tag{1.39}$$

This is the master formula (1.31) restated: integrating ω over the surface is the same as integrating its pullback over the flat parameter domain.

1.9.2 The Jacobian Falls Out

For the volume form $\omega = dx^1 \wedge \cdots \wedge dx^k$, the pullback is:

$$\phi^*(dx^1 \wedge \cdots \wedge dx^k) = \det\left(\frac{\partial \phi^i}{\partial u^j}\right) du^1 \wedge \cdots \wedge du^k. \tag{1.40}$$

The Jacobian determinant appears automatically from the determinant structure of the k-form evaluated on the k tangent vectors $\partial\phi/\partial u^j$. The change-of-variables theorem is not an additional fact to be remembered; it is a consequence of how forms eat vectors.

1.9.3 Pullback of the Polar Coordinate Change

The calculation from Section 8 is a pullback. The coordinate change $\phi : (r, \theta) \mapsto (r\cos\theta, r\sin\theta)$ has

$$\phi^*(dx \wedge dy) = r \, dr \wedge d\theta. \tag{1.41}$$

This is $\phi^*\omega$ expressed in the parameter coordinates (r, θ), and it is the concretization of $dx \wedge dy$ in polar coordinates.

1.10 Forms and Tensors: Abstraction for Different Purposes

Students of physics will have encountered **tensors**: objects with multiple indices that "transform under coordinate changes." Forms are also objects that transform under coordinate changes. The natural question is: how do they differ, and when should we use which?

1.10.1 What They Share

Both tensors and forms are multilinear objects on manifolds. A rank-$(0, k)$ tensor at a point is a multilinear function $T : T_pM \times \cdots \times T_pM \to \mathbb{R}$. A k-form is the same thing with the requirement of antisymmetry added. Both transform predictably under coordinate changes: if we know the components in one system, we can compute them in any other using the Jacobian matrix.

Both are abstractions: they encode geometric content in a coordinate-free way, and they can be concretized in any coordinate system to produce specific component expressions.

1.10.2 How They Differ: Symmetry and Purpose

The crucial difference is **symmetry type**, and the difference in symmetry reflects a difference in purpose.

A general tensor T_{ij} has no required symmetry. The metric tensor g_{ij} is symmetric. A differential form ω_{ij} is antisymmetric. Each symmetry type is adapted to a different task.

Tensors are abstraction for differential equations. When we write a differential equation like Einstein's equation $G_{\mu\nu} = 8\pi T_{\mu\nu}$, we need the equation to hold in every coordinate system. Tensors provide this: if a tensor equation holds in one coordinate system, it holds in all, because both sides transform the same way under the Jacobian matrix. The transformation involves the Jacobian matrix itself (not its determinant), and the components can have any symmetry (symmetric, antisymmetric, or none).

Forms are abstraction for integration. When we write an integral $\int_M \omega$, we need the integral to be independent of parametrization. Forms provide this: their antisymmetry produces the Jacobian *determinant*, which is exactly what the change-of-variables theorem demands. A general tensor, fed k vectors, does *not* produce a determinant (unless it happens to be antisymmetric), and so a general tensor does not integrate correctly.

The division is clean:

	Tensor	**Form**
Symmetry	any (symmetric, none, ...)	antisymmetric
Transforms via	Jacobian matrix	Jacobian determinant
Purpose	differential equations	integration
Key theorem	covariance of equations	change of variables
Needs metric?	often (raising/lowering)	no

1.10.3 Where the Two Worlds Meet

The electromagnetic field $F_{\mu\nu}$ is an antisymmetric rank-2 tensor, which means it is simultaneously a tensor and a 2-form. Its differential equation (Maxwell's equations) can be written in tensor form:

$$\partial_\mu F^{\mu\nu} = J^\nu, \tag{1.42}$$

which is natural for solving PDEs in specific coordinate systems. Its integral formulation (Gauss's law, Faraday's law) is natural in the language of forms:

$$dF = 0, \qquad d\star F = J. \tag{1.43}$$

Both descriptions are valid; each is adapted to its task.

A **metric** g_{ij} bridges the two worlds. Given a metric, we can convert a 1-form ω_i into a vector $v^i = g^{ij}\omega_j$ and vice versa (the "musical isomorphism"). Without a metric, forms exist but cannot be canonically identified with vectors. This is why forms are more fundamental for integration: integrals make sense on any manifold, with or without a metric. Tensor equations involving the metric (like Einstein's equation) require the additional structure of a distance measurement.

1.11 Essential Examples

1.11.1 Exact Forms and Path Independence

For any smooth function f, the 1-form $df = \sum_i (\partial f/\partial x^i)\, dx^i$ satisfies

$$\int_\gamma df = f(\gamma(b)) - f(\gamma(a)). \tag{1.44}$$

The integral depends only on the endpoints, not on the path or its parametrization or the ambient coordinates. This is the fundamental theorem of calculus for line integrals, and it is a statement purely about the abstract objects df and γ.

1.11.2 The Angle Form and Topology

The 1-form $\eta = (-y\, dx + x\, dy)/(x^2 + y^2)$ on $\mathbb{R}^2 \setminus \{0\}$ satisfies $d\eta = 0$ (closed), yet $\oint_{S^1} \eta = 2\pi \neq 0$ (not exact). No smooth function f with $df = \eta$ exists globally, because the domain has a hole at the origin. This is a topological obstruction, detectable purely through the abstract integral $\oint \eta$: no parametrization or coordinate system can remove the hole, and the nonzero integral reflects this.

1.11.3 The Electromagnetic 2-Form

In spacetime with coordinates (t, x, y, z), the electromagnetic field is the 2-form

$$F = E_x\, dx \wedge dt + E_y\, dy \wedge dt + E_z\, dz \wedge dt + B_x\, dy \wedge dz + B_y\, dz \wedge dx + B_z\, dx \wedge dy. \tag{1.45}$$

The electric and magnetic fields are not separate objects but components of one 2-form, mixed by Lorentz boosts (which are just coordinate changes in spacetime). Maxwell's equations become $dF = 0$ and $d\star F = J$: two equations valid in every coordinate system and on any spacetime manifold.

1.11.4 Thermodynamics

The first law $dU = T\, dS - P\, dV$ is an equation between 1-forms on thermodynamic state space. The exactness of dU is a coordinate-free statement: it holds regardless of which thermodynamic variables we choose as coordinates. The Maxwell relations follow from $d(dU) = 0$, the nilpotency of the exterior derivative.

1.12 Summary: Abstraction and Concretization

The entire theory of differential forms rests on one idea: separate the abstract from the concrete.

The abstract side. The integral $\int_M \omega$ is a geometric quantity: it depends only on the geometric domain M and the form ω. It does not refer to any parametrization or coordinate system. Properties discovered on the abstract side — Stokes' theorem, de Rham cohomology, topological invariance — are automatically true in every coordinate system, because they were never stated in any particular one.

The concrete side. To compute $\int_M \omega$, we choose a parametrization ϕ and concretize:

$$\int_M \omega = \int_D \omega\Big(\frac{\partial \phi}{\partial u^1}, \ldots, \frac{\partial \phi}{\partial u^k}\Big)\, du^1 \cdots du^k. \tag{1.46}$$

The form is the concretization device: it accepts the tangent vectors of the parametrization (the raw, coordinate-dependent data) and outputs the scalar density appropriate for integration in that parametrization (the processed, ready-to-integrate data). Different parametrizations feed the form different vectors; the form adjusts its output so the integral is invariant.

Why antisymmetry. The form must be multilinear (because reparametrization rescales tangent vectors) and antisymmetric (because reparametrization can reverse orientation). These two properties force the form to compute determinants. Determinants are exactly the Jacobian factors that the change-of-variables theorem demands. The algebraic structure of forms is not arbitrary; it is uniquely determined by the requirement of parametrization-independent integration.

How forms differ from tensors. Both are abstractions that can be concretized in any coordinate system. Tensors are the right abstraction for differential equations: their transformation law (via the Jacobian matrix) ensures covariance of equations. Forms are the right abstraction for integration: their transformation law (via the Jacobian determinant) ensures invariance of integrals.

The reward. By working on the abstract side, we can discover properties — such as the general Stokes' theorem $\int_M d\omega = \int_{\partial M} \omega$ — that are deeper and more unified than anything visible in the concrete formulas of any single coordinate system. The forms are the means; the coordinate-free understanding is the end.

Chapter 2

The Exterior Derivative

2.1 Boundary and Bulk: The Idea That Generates Everything

In calculus, the derivative of a function f on an interval $[a, b]$ satisfies the fundamental theorem:

$$\int_a^b f'(x)\,dx = f(b) - f(a). \tag{2.1}$$

The right side is the *boundary effect*: the values of f at the two endpoints. The left side is the accumulated *bulk source*: the total rate of change inside the interval. The equation says they are equal.

This relationship — boundary effect equals bulk source — is the single idea that generates everything in this chapter and, ultimately, everything in this book. The exterior derivative is the operation that generalises (2.1) to all dimensions: it takes a form that lives on boundaries and constructs the form that lives in the bulk, such that the two integrals agree.

In the language of Chapter 1, the abstract statement is

$$\int_M d\omega = \int_{\partial M} \omega. \tag{2.2}$$

This is Stokes' theorem, and we do *not* begin by proving it. We begin by *requiring* it, and asking what operation d must be in order for it to hold. The coordinate formula for d is forced on us by this requirement. There is no freedom and no guesswork.

2.2 From 0-Forms to 1-Forms: The Derivative as a Boundary-to-Bulk Device

Let f be a smooth function (a 0-form) on $\mathbb{R}$. The "boundary" of an interval $[a, b]$ consists of two points, and the "integral" of a 0-form over this boundary is the signed sum $f(b) - f(a)$. We want a 1-form df whose integral over the interval equals this boundary effect:

$$\int_a^b df = f(b) - f(a). \tag{2.3}$$

We are reading the fundamental theorem of calculus not as a theorem to be proved but as a *requirement to be satisfied*. If $df = g(x)\,dx$ for some function g, then we need $\int_a^b g(x)\,dx = f(b) - f(a)$ for all a, b. The only function g satisfying this is $g = f'$. So:

$$df = f'(x)\,dx. \tag{2.4}$$

The exterior derivative of a 0-form is the ordinary derivative, packaged as a 1-form. We did not define df and then prove the fundamental theorem; we required the fundamental theorem and read off df.

In higher dimensions, if $f : \mathbb{R}^n \to \mathbb{R}$, the same argument applies along each coordinate direction:

$$d_0 f = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\, dx^i. \tag{2.5}$$

The 1-form $d_0 f$ acts on any tangent vector $\mathbf{v}$ by $(d_0 f)(\mathbf{v}) = \sum_i (\partial f/\partial x^i)\, v^i = \nabla f \cdot \mathbf{v}$: the directional derivative of f in the direction $\mathbf{v}$. This is the gradient, now understood as the unique 1-form that converts boundary data (endpoint values of f) into bulk data (an integrand along the path).

boundary	$\longleftrightarrow$	**bulk**
two endpoints (0-dim)		interval (1-dim)
$f(b) - f(a)$	$=$	$\displaystyle\int_a^b df$

2.3 From 1-Forms to 2-Forms: The Small Rectangle Derives the Curl

Now we step up one dimension. Let $\omega = P\, dx + Q\, dy$ be a 1-form on $\mathbb{R}^2$. We want a 2-form $d\omega$ such that

$$\int_R d\omega = \oint_{\partial R} \omega \tag{2.6}$$

for every region R. We require this and ask what $d\omega$ must be.

Take R to be a small axis-aligned rectangle with one corner at (x_0, y_0) and sides Δx, Δy. The boundary ∂R is traversed counterclockwise: bottom, right, top, left. The 1-form ω is integrated along each edge by feeding it the edge's tangent vector, exactly as in Chapter 1.

Bottom edge ($y = y_0$, traversed rightward): tangent $(1, 0)$, so $\omega((1,0)) = P$. Contribution: $P(x_0, y_0)\, \Delta x$.

Right edge ($x = x_0 + \Delta x$, traversed upward): tangent $(0, 1)$, so $\omega((0,1)) = Q$. Contribution: $Q(x_0 + \Delta x, y_0)\, \Delta y$.

Top edge ($y = y_0 + \Delta y$, traversed leftward): contribution $-P(x_0, y_0 + \Delta y)\, \Delta x$.

Left edge ($x = x_0$, traversed downward): contribution $-Q(x_0, y_0)\, \Delta y$.

Summing, the terms pair up by component:

$$\begin{aligned} \oint_{\partial R} \omega &= [Q(x_0 + \Delta x, y_0) - Q(x_0, y_0)]\Delta y - [P(x_0, y_0 + \Delta y) - P(x_0, y_0)]\Delta x \\ &\approx \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \Delta x\, \Delta y. \end{aligned} \tag{2.7}$$

On opposite edges, the *same* component of ω is evaluated at two positions separated in the *perpendicular* direction. On the bottom and top, P is evaluated at y_0 and $y_0 + \Delta y$; the difference produces $\partial_y P$. On the left and right, Q is evaluated at x_0 and $x_0 + \Delta x$; the difference produces $\partial_x Q$. Each component is differentiated in the direction orthogonal to its own.

Now, if $d\omega = h\,dx \wedge dy$, its integral over the small rectangle is $h\,\Delta x\,\Delta y$. Comparing with (2.7):

$$\boxed{d(P\,dx + Q\,dy) = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy.} \tag{2.8}$$

This is the curl in two dimensions (Green's theorem integrand). It was not postulated; it was derived from the boundary-equals-bulk requirement applied to a small rectangle.

boundary	$\longleftrightarrow$	**bulk**
four edges (1-dim)		rectangle (2-dim)
$\oint_{\partial R} \omega$	$=$	$\displaystyle\int_R d\omega$

Worked example. The rotation 1-form $\omega = -y\,dx + x\,dy$ has $P = -y$, $Q = x$, so $d\omega = (1 - (-1))\,dx \wedge dy = 2\,dx \wedge dy$: uniform vorticity of strength 2 everywhere. Integrating $d\omega$ over any region D gives $2 \times \text{area}(D)$; integrating ω around ∂D gives the same, by Stokes / Green.

In $\mathbb{R}^3$: the same derivation, applied to $\omega = P\,dx + Q\,dy + R\,dz$ using three families of small rectangles (one in each coordinate plane), yields

$$d_1\omega = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy. \tag{2.9}$$

The three components are $\nabla \times \mathbf{F}$, the curl. Stokes' theorem for this case is the classical Stokes' theorem: $\int_S d_1\omega = \oint_{\partial S} \omega$.

2.4 From 2-Forms to 3-Forms: The Small Cube Derives the Divergence

Let $\eta = A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy$ be a 2-form on $\mathbb{R}^3$. We require $\int_V d\eta = \oiint_{\partial V} \eta$ and take V to be a small axis-aligned cube with sides Δx, Δy, Δz.

The boundary has six faces in three pairs. Consider the pair perpendicular to the x-axis. On these faces only the $dy \wedge dz$ component contributes (the other components contain dx, which vanishes where x is constant). The net flux through this pair is:

$$A(x_0 + \Delta x)\,\Delta y\,\Delta z - A(x_0)\,\Delta y\,\Delta z \approx \frac{\partial A}{\partial x}\,\Delta x\,\Delta y\,\Delta z. \tag{2.10}$$

The y- and z-pairs contribute $(\partial_y B)$ and $(\partial_z C)$ times the volume, respectively. Summing:

$$\boxed{d(A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy) = \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right) dx \wedge dy \wedge dz.} \tag{2.11}$$

This is the divergence. Stokes' theorem for this case is the divergence theorem: $\iiint_V d_2\eta = \oiint_{\partial V} \eta$.

boundary	$\longleftrightarrow$	**bulk**
six faces (2-dim)		cube (3-dim)
$\oiint_{\partial V} \eta$	$=$	$\displaystyle\int_V d\eta$

2.5 The General Formula and the Orthogonality Principle

The pattern is now clear. For any k-form $\omega = f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, requiring $\int_R d\omega = \int_{\partial R} \omega$ on a small $(k+1)$-dimensional box, we find: a face perpendicular to the x^j-axis contributes to the boundary integral only if $j \notin \{i_1, \ldots, i_k\}$; otherwise the form vanishes on that face. For each contributing direction j, the pair of opposite faces yields $(\partial f/\partial x^j)\,\Delta x^j \prod \Delta x^{i_\ell}$. Summing:

$$\boxed{d(f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\,dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.} \tag{2.12}$$

The sum runs over all j, but terms with $j \in \{i_1, \ldots, i_k\}$ vanish automatically because $dx^j \wedge dx^j = 0$. This algebraic vanishing is the coordinate expression of the geometric fact we just established: faces in directions already present in ω contribute nothing to the boundary integral.

This is the **orthogonality principle**. The exterior derivative differentiates only in directions *not already used* by the form. A k-form already measures k-dimensional content; its derivative $d\omega$ asks how that measurement changes as we step into a *new* dimension. Each such step augments the degree by one and produces one component of the $(k+1)$-form $d\omega$.

The orthogonality principle is not an additional axiom; it is a consequence of the boundary geometry. Opposite faces of a rectangular box lie in perpendicular directions, so the boundary integral naturally produces derivatives in the directions orthogonal to the form's support. The antisymmetry of the wedge product ($dx^j \wedge dx^j = 0$) is the algebraic encoding of this geometric fact.

A general k-form is a sum of such terms, and d extends by linearity:

$$d\omega = \sum_{i_1 < \cdots < i_k} \sum_{j=1}^{n} \frac{\partial \omega_{i_1 \cdots i_k}}{\partial x^j}\,dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \tag{2.13}$$

2.5.1 Concrete illustration

Consider $\omega = f(x,y)\,dx$ in $\mathbb{R}^2$. Although f depends on both x and y, the formula gives:

$$d(f\,dx) = \frac{\partial f}{\partial x}\underbrace{dx \wedge dx}_{=0} + \frac{\partial f}{\partial y}\,dy \wedge dx = -\frac{\partial f}{\partial y}\,dx \wedge dy. \tag{2.14}$$

Only the y-derivative survives. Geometrically: $f\,dx$ measures flow along x. Its exterior derivative detects how this x-flow varies as we step in the y-direction. If $\partial_y f > 0$, the flow is stronger above than below, creating a clockwise vortex. The vortex density is the 2-form $d\omega$, the bulk source of the circulation observed on the boundary.

2.6 The Dimension Ladder

The derivations of Sections 2–4 are all instances of one structure, ascending a ladder of dimensions:

Form	**Boundary**	**Bulk**	**Classical name**	d **produces**
0-form f	2 points (0-dim)	interval (1-dim)	fund. thm. of calculus	gradient
1-form ω	loop (1-dim)	surface (2-dim)	Green / Stokes	curl
2-form η	closed surface (2-dim)	solid (3-dim)	divergence theorem	divergence
$(k-1)$-form	$(k-1)$-surface	k-region	general Stokes	d_{k-1}

Each row is the same logical structure: a form on the boundary, its exterior derivative in the bulk, and the equality of their integrals. The classical integral theorems — fundamental theorem of calculus, Green's theorem, Stokes' theorem, divergence theorem — are not four separate results but one result applied at four rungs of the ladder.

In multivariable calculus, these are proved separately, each with its own notation and apparatus. In the language of forms, the *definition* of d is uniform at every rung. At each level, d is the operation that converts boundary data into bulk data, determined uniquely by requiring the two integrals to agree. This is why the formula (2.13) takes the form it does: it is not an algebraic construction that happens to have a geometric interpretation; it is a geometric requirement that happens to have an algebraic expression.

2.6.1 Gradient, curl, divergence: three faces of one operator

In $\mathbb{R}^3$, the exterior derivative takes three different appearances depending on degree: d_0 is the gradient, d_1 is the curl, d_2 is the divergence. Classical vector calculus treats these as three separate operations with three separate formulas. The reason they look different is an accident of three dimensions: in $\mathbb{R}^3$, both 1-forms and 2-forms have three independent components (since $\binom{3}{1} = \binom{3}{2} = 3$), so both can be identified with vectors via the metric. This disguises the fact that d_1 maps 1-forms (three components) to 2-forms (three components), making it look like a vector-to-vector operation.

In $\mathbb{R}^{10}$, there are $\binom{10}{1} = 10$ independent 1-forms and $\binom{10}{2} = 45$ independent 2-forms. The map d_1 takes a 10-component object to a 45-component object. There is no possibility of calling this "the curl" or thinking of it as a vector operation. The only honest description is: d_1 raises degree by one, differentiating in the new orthogonal directions. The uniformity of d becomes unavoidable.

2.6.2 Notation: the family d_k

The symbol d denotes a family of operators, one per degree:

$$d_k : \Omega^k(M) \to \Omega^{k+1}(M). \tag{2.15}$$

When we write simply d, we mean the operator appropriate for the form it acts on. The chain of spaces

$$0 \to \Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \Omega^2(M) \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} \Omega^n(M) \to 0 \tag{2.16}$$

is the **de Rham complex**. The key property — nilpotency $d_{k+1} \circ d_k = 0$, whose proof we give below — means the image of each d_k lies inside the kernel of d_{k+1}: every exact form is closed. The quotient $\ker d_k / \operatorname{im} d_{k-1}$ is the de Rham cohomology $H^k_{\mathrm{dR}}(M)$, the subject of Chapter 4.

2.7 Properties of the Exterior Derivative

Having derived the formula (2.13) from the boundary-equals-bulk requirement, we now verify the properties that make d a powerful tool for abstract reasoning. These properties can alternatively be taken as axioms that uniquely determine d; the fact that our derived formula satisfies them confirms that the boundary-bulk derivation and the axiomatic approach give the same operator.

2.7.1 Linearity

$d(a\omega + b\eta) = a\,d\omega + b\,d\eta$ for constants a, b. This is immediate from (2.13).

2.7.2 The Leibniz rule

For a p-form α and a q-form β:

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \, \alpha \wedge d\beta. \tag{2.17}$$

The sign $(-1)^p$ arises because d must commute past the p-form α to reach β, picking up one sign flip per degree. This is the product rule for forms, and it can be verified directly from (2.13) using the Leibniz rule for ordinary partial derivatives.

Example. For $\omega = P\,dx + Q\,dy + R\,dz$, the Leibniz rule gives $d(P\,dx) = dP \wedge dx + P\,d(dx)$. Now $d(dx) = d(d_0 x) = 0$ (by nilpotency, proved below), so $d(P\,dx) = dP \wedge dx$, and expanding $dP = P_x\,dx + P_y\,dy + P_z\,dz$ and using $dx \wedge dx = 0$ reproduces the curl formula (2.9). The Leibniz rule provides an efficient alternative route to the same formulas we derived from boundary integrals.

2.7.3 Nilpotency: $d^2 = 0$

The deepest property of the exterior derivative is $d_{k+1} \circ d_k = 0$: applying d twice always gives zero. This can be verified directly from (2.13): applying d twice introduces second partial derivatives with wedge products $dx^j \wedge dx^l$, and the symmetry of mixed partials ($\partial^2 f/\partial x^j \partial x^l = \partial^2 f/\partial x^l \partial x^j$) combined with the antisymmetry of the wedge product ($dx^j \wedge dx^l = -dx^l \wedge dx^j$) forces every term to cancel against its partner with j and l swapped.

But the geometric meaning is more illuminating: $d^2 = 0$ because **the boundary of a boundary is empty**. If M is a solid region, its boundary ∂M is a closed surface, which has no boundary of its own. If S is a surface with boundary loop $C = \partial S$, then C is a closed curve with no boundary. In general, $\partial(\partial M) = \emptyset$. Since d was defined to convert boundary integrals to bulk integrals, applying it twice would ask for the boundary of a boundary — which is empty, giving zero.

In the language of the de Rham complex, nilpotency means $\operatorname{im} d_k \subset \ker d_{k+1}$: every exact form (something of the form $d\alpha$) is automatically closed ($d(d\alpha) = 0$). The converse — whether every closed form is exact — is the central question of cohomology, and its answer depends on the topology of the manifold.

Classical consequences. In $\mathbb{R}^3$: $d_1 \circ d_0 = 0$ is $\nabla \times (\nabla f) = 0$ (curl of gradient is zero); $d_2 \circ d_1 = 0$ is $\nabla \cdot (\nabla \times \mathbf{F}) = 0$ (divergence of curl is zero). These identities, proved by tedious coordinate computation in vector calculus, are both instances of one principle: the boundary of a boundary is empty.

2.7.4 Coordinate independence

The formula (2.13) was derived in a specific coordinate system (using axis-aligned boxes). But $d\omega$ is a geometric object that does not depend on the choice of coordinates. This can be verified by transforming to new coordinates $(y^1, \ldots, y^n)$ and checking that the chain rule reproduces the same form — or, more elegantly, by observing that the boundary-bulk requirement (2.2) is itself coordinate-free: neither "boundary" nor "integral" refers to any coordinate system.

2.7.5 Uniqueness

The four properties — linearity, the Leibniz rule, nilpotency, and the action $d_0 f = \sum_i (\partial f/\partial x^i)\,dx^i$ on functions — uniquely determine d on all forms. Every smooth manifold has exactly one exterior derivative, requiring no metric, no connection, and no structure beyond smoothness.

2.8 Stokes' Theorem

We have used Stokes' theorem as the motivation for defining d. Now we verify that the derived formula does indeed satisfy $\int_M d\omega = \int_{\partial M} \omega$ in full generality.

Theorem 2.1 (Stokes' Theorem)**.** *Let M be a compact oriented $(k+1)$-dimensional manifold with boundary, and let ω be a smooth k-form on M. Then*

$$\int_M d\omega = \int_{\partial M} \omega. \tag{2.18}$$

Every classical integral theorem is a special case. The fundamental theorem of calculus ($k = 0$, $M = [a, b]$, boundary = two points). Green's theorem ($k = 1$, M a planar region, boundary = curve). The classical Stokes' theorem ($k = 1$, M a surface in $\mathbb{R}^3$, boundary = loop). The divergence theorem ($k = 2$, M a solid in $\mathbb{R}^3$, boundary = closed surface). All are one theorem, proved once, valid in any dimension.

2.8.1 Proof sketch

Step 1: Half-space. Let $\mathbb{H}^{k+1} = \{x^{k+1} \geq 0\}$ with boundary $\{x^{k+1} = 0\} \cong \mathbb{R}^k$. For a compactly supported k-form $\omega = f\, dx^1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^{k+1}$ (hat omits dx^j), there are two cases.

If $j < k+1$: the form contains dx^{k+1}, so it restricts to zero on the boundary $\{x^{k+1} = 0\}$; meanwhile $d\omega$ integrates to zero over $\mathbb{H}^{k+1}$ by the fundamental theorem in the x^j direction (compact support gives zero at infinity). Both sides vanish.

If $j = k+1$: $\omega = f\, dx^1 \wedge \cdots \wedge dx^k$ and $d\omega = (-1)^k \partial_{x^{k+1}} f\, dx^1 \wedge \cdots \wedge dx^{k+1}$. Integrating over $\mathbb{H}^{k+1}$ and applying the fundamental theorem in x^{k+1}:

$$\int_{\mathbb{H}^{k+1}} d\omega = (-1)^k \int_{\mathbb{R}^k} \left[f(x', \infty) - f(x', 0) \right] dx^1 \cdots dx^k = (-1)^{k+1} \int_{\mathbb{R}^k} f(x', 0)\, dx^1 \cdots dx^k. \tag{2.19}$$

With the induced orientation on the boundary (outward normal $-\partial_{x^{k+1}}$ giving sign $(-1)^{k+1}$), this equals $\int_{\partial \mathbb{H}^{k+1}} \omega$. ✓

Step 2: Extension to manifolds. Cover M by coordinate charts $(U_\alpha, \varphi_\alpha)$ with a subordinate partition of unity $\{\rho_\alpha\}$, $\sum_\alpha \rho_\alpha = 1$. Each $\rho_\alpha \omega$ has compact support in U_α, so Step 1 applies chart by chart. Summing:

$$\int_M d\omega = \sum_\alpha \int_M d(\rho_\alpha \omega) = \sum_\alpha \int_{\partial M} \rho_\alpha \omega = \int_{\partial M} \omega. \qquad \square \tag{2.20}$$

2.8.2 The boundary/bulk interpretation, revisited

With Stokes' theorem now established, we can read the boundary-bulk relationship as a physical principle.

Degree 0: Temperature. Let f be temperature. The boundary effect $f(B) - f(A)$ (temperature difference between endpoints) equals the bulk source $\int_\gamma df$ (accumulated gradient along the path).

Degree 1: Vorticity. Let ω be the velocity 1-form of a fluid. The boundary effect $\oint_C \omega$ (circulation around a loop) equals the bulk source $\int_S d\omega$ (total vorticity through any spanning surface).

Degree 2: Flux. Let η be the magnetic flux 2-form. The boundary effect $\oiint_{\partial V} \eta$ (net outward flux) equals the bulk source $\int_V d\eta$ (total divergence inside). In nature, $d\eta = 0$ (no magnetic monopoles), so the net outward flux is always zero.

In each case: the exterior derivative $d\omega$ is the bulk source that explains the boundary effect $\int_{\partial M} \omega$. The causal direction runs from boundary observation to bulk inference. Given what we measure on the boundary, d tells us what source density inside is responsible.

2.9 Closed and Exact Forms

Two classes of forms are defined by their relationship to d, and the gap between them is one of the deepest ideas in geometry.

2.9.1 Exact forms: gradient flows and potentials

A k-form ω is **exact** if $\omega = d_{k-1}\alpha$ for some $(k-1)$-form α. For 1-forms: $\omega = df$ for some function f, the gradient of a potential.

The canonical image is water flowing downhill. The gradient flow $\omega = df$ always decreases f; a particle following it can never return to its starting height. Therefore the integral around any closed loop is zero: $\oint_\gamma df = f(\gamma(T)) - f(\gamma(0)) = 0$. An exact form cannot circulate. More generally, the integral of an exact form depends only on the endpoints, not on the path.

2.9.2 Closed forms: local balance without global potential

A k-form ω is **closed** if $d_k\omega = 0$. For a 1-form: zero curl, no local vortices.

Every exact form is closed (by nilpotency: $d(d\alpha) = 0$). But a closed form need not be exact, and this is where topology enters.

The circular canal. Water flows at constant speed around a ring-shaped channel. At every point the flow is smooth and uniform: a small paddle wheel would not rotate, so $d\omega = 0$ (closed). But this is not a gradient flow: no height function on the ring can decrease at every point and return to the starting height after a full circuit. The total circulation $\oint \omega \neq 0$, so ω is not exact.

The topology of the domain — the hole at the center of the ring — is what permits a closed form to fail to be exact. On a simply connected domain (no holes), the Poincaré lemma guarantees that every closed form is exact. The failure of exactness is a topological obstruction.

2.9.3 The torus

The torus T^2 with angular coordinates (θ, ϕ) carries the 1-form $\omega = d\theta$, which is closed ($d(d\theta) = 0$) but not exact: if $d\theta = df$ for some smooth function f, then $f = \theta + g(\phi)$, but θ is not a well-defined function on the torus (it jumps by 2π after a full circuit). The integral $\oint d\theta = 2\pi \neq 0$ confirms non-exactness.

2.9.4 Closed 2-forms: the magnetic monopole

On the sphere S^2 surrounding a magnetic monopole, the magnetic flux 2-form $B = (g/4\pi)\sin\theta\, d\theta \wedge d\phi$ is closed ($d_2 B = 0$, trivially on the 2-sphere) but not exact: if $B = dA$ globally, then by Stokes $\int_{S^2} B = \int_{\partial S^2} A = 0$ (since S^2 has no boundary), contradicting $\int_{S^2} B = g \neq 0$. No globally consistent vector potential exists. The Dirac quantization condition $qg \in 2\pi\mathbb{Z}$ arises from requiring the local potentials (which do exist on patches) to be consistent where they overlap.

2.9.5 The obstruction principle

The general pattern: a closed form is always locally exact (Poincaré lemma), and local potentials always exist on small contractible patches. The question is whether these local potentials can be assembled into a single global potential. When they cannot — when extending a local potential around a non-contractible loop accumulates a discrepancy — the form is closed but not exact, and the discrepancy is a **topological obstruction**.

This obstruction is detected by integrating the form over the non-contractible cycle: $\oint \omega \neq 0$ for the angle form around a puncture; $\int_{S^2} B \neq 0$ for a monopole field. The obstruction depends only on the cohomology class $[\omega] \in H^k_{\mathrm{dR}}(M)$, not on the specific representative form.

The same structure appears throughout physics: the Aharonov-Bohm phase (enclosed magnetic flux detected by integrating A around a solenoid), gauge bundle topology (the first Chern class as obstruction to global gauge choice), and topological phases of matter (the Chern number as obstruction to a global Bloch basis on the Brillouin zone) are all instances of a closed form that is locally exact but globally obstructed, with the obstruction detected by integration over non-contractible cycles.

2.10 The Exterior Derivative in Physics

Maxwell's equations. The electromagnetic field strength is a 2-form

$$F = E_x\, dx \wedge dt + E_y\, dy \wedge dt + E_z\, dz \wedge dt + B_x\, dy \wedge dz + B_y\, dz \wedge dx + B_z\, dx \wedge dy. \tag{2.21}$$

Maxwell's equations become $dF = 0$ (Faraday + no monopoles) and $d(\star F) = J$ (Gauss + Ampère). The first says F is closed: its boundary integral over any closed surface is zero, meaning no net electromagnetic charge is enclosed by any closed surface in spacetime. The second says the Hodge dual $\star F$ has a bulk source J (the current 3-form): the boundary effect of $\star F$ on any closed 3-surface is the total charge inside.

Gauge invariance. Since $dF = 0$, locally $F = dA$ for a 1-form A (the vector potential). But A is not unique: $A' = A + d\Lambda$ gives $dA' = dA + d(d\Lambda) = F + 0 = F$ by nilpotency. The gauge freedom is exactly the freedom to add exact 1-forms to A, and $d^2 = 0$ is what makes this work.

Thermodynamics. The first law $dU = T\, dS - P\, dV$ is an equation between 1-forms. Applying d and using nilpotency $d(dU) = 0$ together with the Leibniz rule yields all Maxwell thermodynamic relations simultaneously.

Fluid mechanics. For a velocity 1-form V, the vorticity is $\Omega = dV$. Kelvin's circulation theorem ($\frac{d}{dt}\oint_C V = 0$ for material loops) says the bulk source of circulation is conserved. The exterior derivative d turns the boundary measurement (circulation) into the bulk source (vorticity) responsible for it.

2.11 Abstraction and Concretization, Revisited

We close by placing the exterior derivative in the framework of Chapter 1.

The abstract side. Stokes' theorem $\int_M d\omega = \int_{\partial M} \omega$ is a relationship between geometric objects: a form, its derivative, a manifold, and its boundary. It refers to no coordinates. Properties discovered on the abstract side — nilpotency, cohomology, topological obstructions — hold in every coordinate system because they were never stated in any particular one.

The concrete side. To compute $d\omega$, we choose coordinates and apply (2.13). The formula was derived by concretizing Stokes' theorem on small coordinate boxes: decomposing the boundary into

faces, evaluating the form on each face, and comparing opposite faces to extract partial derivatives. Different coordinate systems give different-looking formulas for $d\omega$, but the geometric object is the same.

The parallel with Chapter 1. In Chapter 1, the form ω was the concretization device for integrals: it accepted parametrization-dependent tangent vectors and output the parametrization-independent integrand. Here, the exterior derivative d is the concretization device for the boundary-bulk relationship: it converts the abstract statement "boundary effect equals bulk source" into a specific coordinate formula. Just as the form's antisymmetry was forced by parametrization independence, the formula for d is forced by the boundary-equals-bulk requirement.

Why d is the right notion of derivative. In multivariable calculus, gradient, curl, and divergence are three separate operations connected by the mysterious identities $\nabla \times (\nabla f) = 0$ and $\nabla \cdot (\nabla \times \mathbf{F}) = 0$. In the language of forms, all three are d, and both identities are $d^2 = 0$: the boundary of a boundary is empty. The exterior derivative is the coordinate-free, dimension-independent, metric-free generalisation that reveals gradient, curl, and divergence as three faces of one operation, and their identities as one geometric principle.

Chapter 3

Nilpotency: Why $d_{k+1} \circ d_k = 0$

3.1 The Operators and the Slogan

In the previous chapter we wrote $d^2 = 0$ as a shorthand for the family of identities $d_{k+1} \circ d_k = 0$, one for each degree k. The shorthand is convenient — it is the "insider notation" of differential geometry, and you will see it everywhere in the literature. But it is important to remember what it abbreviates. The operators $d_0, d_1, d_2, \ldots$ are *different operators* acting on different spaces:

$$\Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \Omega^2(M) \xrightarrow{d_2} \Omega^3(M) \xrightarrow{d_3} \cdots \tag{3.1}$$

The nilpotency statement $d_{k+1} \circ d_k = 0$ says: for every k, applying the $(k+1)$-th operator immediately after the k-th operator always produces zero. This is not one fact but a whole sequence of facts. When we write $d^2 = 0$ we are compressing all of them into one symbol.

This chapter is devoted to understanding, from as many angles as possible, *why* $d_{k+1} \circ d_k = 0$. We will see four perspectives — the water-flow intuition, the geometric boundary-of-a-boundary principle, the algebraic cancellation in coordinates, and the linear-algebraic column-space / null-space relationship — and then show how they are all faces of the same underlying truth.

3.2 Water Flowing Downhill: Why Gradient Flows Cannot Circulate

The simplest instance of nilpotency is $d_1 \circ d_0 = 0$: the curl of any gradient is zero. We encountered this in Chapter 2 as an algebraic cancellation of mixed partial derivatives. Here we build the geometric intuition first.

3.2.1 The Water-Flow Picture

Imagine a landscape with a height function $f(x, y)$: the altitude at each point. The gradient 1-form $d_0 f$ points in the direction of steepest ascent at every point. The corresponding flow is water running *downhill*: always moving in the direction opposite to the gradient, always descending.

Water flowing strictly downhill can never return to where it started.

This single observation contains $d_1 \circ d_0 = 0$ in its entirety. If a water molecule starts at a point p of height $h = f(p)$ and flows along the gradient (or against it — downhill), its height changes

monotonically. It cannot complete a closed loop and return to p without at some point flowing uphill. But a gradient flow never flows uphill by definition.

Mathematically: the integral of $d_0 f$ along any path from A to B gives exactly $f(B) - f(A)$. For a closed loop starting and ending at the same point, $f(\text{end}) - f(\text{start}) = 0$. No circulation. The vorticity 2-form $(d_1 \circ d_0)f$ measures the circulation density per unit area. If circulation is zero around every loop, the density is zero at every point. That is $d_1(d_0 f) = 0$.

The deepest version of this insight: a gradient flow is generated by a potential function f. The potential is like a height. Water always flows from high potential to low potential. It can never accumulate circulation, because that would require a loop where the potential both increases and decreases by the same total amount — but the two paths around the loop cannot both be downhill.

3.2.2 The Algebraic Verification

In coordinates, $d_0 f = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy$. Applying d_1:

$$d_1(d_0 f) = d_1\left(\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) dy \wedge dx + \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) dx \wedge dy. \tag{3.2}$$

Since mixed partial derivatives commute ($\partial_y \partial_x f = \partial_x \partial_y f$ by Clairaut's theorem) and $dy \wedge dx = -dx \wedge dy$:

$$d_1(d_0 f) = \frac{\partial^2 f}{\partial y \partial x}\,dy \wedge dx + \frac{\partial^2 f}{\partial x \partial y}\,dx \wedge dy = \frac{\partial^2 f}{\partial x \partial y}(-dx \wedge dy + dx \wedge dy) = 0. \tag{3.3}$$

Two symmetries cancel each other: the *commutativity* of mixed partials (an analytic symmetry, even under swapping the order of differentiation) and the *antisymmetry* of the wedge product (an algebraic symmetry, odd under swapping). One even symmetry plus one odd symmetry equals zero.

3.2.3 Discrete Version: The Telescoping Sum

On five vertices arranged in a circle with values $v_0 = 3$, $v_1 = 7$, $v_2 = 2$, $v_3 = 9$, $v_4 = 5$, assign to each directed edge the difference of its endpoint values (the discrete gradient):

$$e_{01} = 4, \quad e_{12} = -5, \quad e_{23} = 7, \quad e_{34} = -4, \quad e_{40} = -2.$$

Sum all edge values around the loop: $4 + (-5) + 7 + (-4) + (-2) = 0$. This is always zero, regardless of the vertex values, because the sum telescopes: each v_i appears once with a $+$ sign and once with a $-$ sign and they cancel. No matter how the potential function assigns heights to vertices, the total circulation around a closed loop is zero. The discrete gradient has zero discrete curl.

3.3 Vorticity Does Not Expand: Why Curl Fields Are Divergence-Free

One dimension higher: $d_2 \circ d_1 = 0$, the divergence of any curl is zero.

3.3.1 The Vortex Picture

Consider a fluid whose velocity 1-form is ω. The vorticity 2-form $d_1\omega$ measures how much the fluid spins around each point. Now ask: can a pure vorticity field — one that is itself the curl of something — have net outward flux through any closed surface?

The answer is no, and here is why. Flux through a closed surface measures net outflow: how much more fluid leaves than enters. But a vorticity field is composed entirely of spinning motion. Spinning does not transport fluid systematically outward. The fluid goes around in circles; on average, as much crosses any small surface element in one direction as the other.

More precisely: if $d_1\omega$ is the vorticity of ω, then by Stokes' theorem applied twice,

$$\iiint_V d_2(d_1\omega) = \oiint_{\partial V} d_1\omega = \oint_{\partial(\partial V)} \omega = 0, \tag{3.4}$$

because the boundary of the boundary is empty: $\partial(\partial V) = \emptyset$. This is true for any volume V, so $d_2(d_1\omega) = 0$ pointwise. The vorticity field has no divergence because it has no source to create outward flux.

3.3.2 The Discrete Version: Edge Cancellation on Surfaces

Take a triangulated closed surface (a triangulated sphere is the cleanest example). Assign a value to each oriented edge. For each triangular face, sum the values of its three boundary edges (with orientation signs): this is the discrete curl applied to each face. Now sum these face curls over all faces.

The result is always zero. Every internal edge is shared by exactly two faces, appearing with opposite orientations in the two face-boundary computations. When we sum over all faces, each edge's contribution cancels with its partner from the adjacent face. Only boundary edges would not cancel — but a closed surface has no boundary edges. The total sum is zero.

This is the discrete statement: the discrete curl of discrete edge values has zero total flux through any closed surface made of triangles.

3.4 Boundaries Have No Boundaries

3.4.1 The Slogan and What It Means

The single most geometric way to understand $d_{k+1} \circ d_k = 0$ is through the slogan:

The boundary of a boundary is empty. $\partial^2 = 0$.

But let us be precise about what this means, because the slogan sounds almost too simple. A k-dimensional region has a $(k-1)$-dimensional boundary. That boundary is itself a $(k-1)$-dimensional manifold. The statement is that this boundary manifold is always *closed* — it has no boundary of its own. A $(k-1)$-dimensional manifold that arises as the boundary of a k-dimensional region is a closed manifold: no rim, no edge, no further boundary.

Let us see this in examples, going from small to large.

A segment has two endpoints. Take a line segment, a 1-dimensional region. Its boundary is two points, a 0-dimensional manifold. A point has no boundary at all. So the boundary of the boundary of the segment is empty.

A piece of cloth has a rim; a closed surface does not. Take a square piece of cloth: a 2-dimensional region. Its boundary is the four-sided rim around the edge, a 1-dimensional manifold. This rim is a closed loop (a rectangle): it has no endpoints, no gaps. The boundary of the boundary of the cloth is empty.

Now contrast with a sphere. A sphere is a 2-dimensional closed surface. If it arises as the boundary of a 3-dimensional ball, then the sphere itself has no rim. You cannot walk off the edge of

a sphere. The sphere is compact without boundary — it is closed. Taking the boundary of a closed surface gives nothing: $\partial S^2 = \emptyset$.

A solid ball has a spherical shell; the shell has no rim. A solid ball is a 3-dimensional region. Its boundary is the spherical shell (a 2-sphere), a closed surface. The sphere has no boundary (you cannot reach the "edge" of a sphere by walking on it). So the boundary of the boundary of the ball is empty.

A tetrahedron. A solid tetrahedron (3D) has as its boundary four triangular faces (2D). Those four triangles are glued edge-to-edge to form a closed surface. Every edge of every triangle is shared by exactly two triangles. Going around the boundary of the boundary, every edge is traversed once in each direction and they cancel. Empty.

The pattern is general and is one of the deepest facts about manifolds:

$$\partial(\partial M) = \emptyset \quad \text{for any compact manifold } M. \tag{3.5}$$

3.4.2 A Bulk Cannot Be the Boundary of a Higher-Dimensional Bulk

The companion slogan is:

A bulk has no bulk.

Let us be precise about this too. A "bulk" is a region that fills in a boundary: if $N = \partial M$, then M is the bulk and N is its boundary. The statement is that a bulk M cannot *itself* be the boundary of some higher-dimensional region, unless it is a very special situation.

More concretely: the boundary ∂M of any region M is always a closed manifold (no boundary of its own, by the previous slogan). But a closed manifold cannot be the boundary of a higher-dimensional region in the same space, because a boundary must be a closed manifold, and here we already have one. The point is that ∂M is "used up" as a boundary; it is closed, and a closed manifold cannot be further bounded.

Example. The sphere $S^2 = \partial B^3$ is the boundary of the 3-ball. The sphere is itself a closed 2-manifold: $\partial S^2 = \emptyset$. It has no interior dimension left to be a boundary of something 3-dimensional. It cannot be the boundary of a ball *in the way the boundary is defined*: a boundary of a 3-dimensional region is 2-dimensional and closed, and S^2 is already that.

Example. A circular loop $S^1 = \partial D^2$ is the boundary of a disk. The loop is a closed 1-manifold. The "bulk" behind it (the disk) is 2-dimensional. You cannot then ask: "what is the 3-dimensional region whose boundary is this disk?" The disk is not a closed manifold (it has a boundary, the circle), so it cannot itself be a boundary. The circle is the boundary; the disk is the bulk; and there is no higher-dimensional bulk of the disk in this chain.

3.4.3 The Bridge to $d^2 = 0$ via Stokes

The connection between $\partial^2 = 0$ and $d_{k+1} \circ d_k = 0$ is provided by Stokes' theorem. For any smooth k-form ω and any compact oriented $(k+2)$-manifold M:

$$\int_M d_{k+1}(d_k\omega) = \int_{\partial M} d_k\omega = \int_{\partial(\partial M)} \omega = \int_{\emptyset} \omega = 0. \tag{3.6}$$

The first equality is Stokes applied to $d_k\omega$ over M. The second is Stokes applied to ω over ∂M. The third uses $\partial(\partial M) = \emptyset$. Since the integral of $d_{k+1}(d_k\omega)$ over any region is zero, the form itself must be zero: $d_{k+1}(d_k\omega) = 0$.

This argument reveals that $d^2 = 0$ is the analytic shadow of $\partial^2 = 0$. The boundary operator and the exterior derivative are dual to each other (via Stokes), and their nilpotency is one fact, not two.

3.5 The Algebraic Proof in Coordinates

Having built the intuition, we now give the direct coordinate proof which makes no reference to integration.

Theorem 3.1 (Nilpotency)**.** *For any smooth k-form ω on a smooth manifold, $d_{k+1}(d_k\omega) = 0$.*

Proof. By linearity it suffices to prove this for a single coordinate term $\omega = f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}$. Then:

$$d_k\omega = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \tag{3.7}$$

Applying d_{k+1}:

$$d_{k+1}(d_k\omega) = \sum_{j=1}^{n}\sum_{\ell=1}^{n} \frac{\partial^2 f}{\partial x^\ell \partial x^j}\, dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \tag{3.8}$$

The double sum contains paired terms (ℓ, j) and (j, ℓ) with $\ell \neq j$. In these pairs:

$$\frac{\partial^2 f}{\partial x^\ell \partial x^j}\, dx^\ell \wedge dx^j + \frac{\partial^2 f}{\partial x^j \partial x^\ell}\, dx^j \wedge dx^\ell \tag{3.9}$$

$$= \frac{\partial^2 f}{\partial x^\ell \partial x^j}\, dx^\ell \wedge dx^j - \frac{\partial^2 f}{\partial x^j \partial x^\ell}\, dx^\ell \wedge dx^j \quad (\text{using } dx^j \wedge dx^\ell = -dx^\ell \wedge dx^j) \tag{3.10}$$

$$= \left(\frac{\partial^2 f}{\partial x^\ell \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^\ell} \right) dx^\ell \wedge dx^j = 0, \tag{3.11}$$

because mixed partial derivatives commute for smooth f. Diagonal terms with $\ell = j$ give $dx^\ell \wedge dx^\ell = 0$. Every term vanishes, so $d_{k+1}(d_k\omega) = 0$. □

The proof has two ingredients: the *symmetry* of mixed partials ($\partial_\ell \partial_j f = \partial_j \partial_\ell f$, which is even under swapping ℓ and j) and the *antisymmetry* of the wedge product ($dx^\ell \wedge dx^j = -dx^j \wedge dx^\ell$, which is odd under swapping). A symmetric factor times an antisymmetric factor, summed over all pairs, is always zero.

3.6 The Linear Algebra View: Column Space Inside Null Space

The most computationally transparent way to understand nilpotency is through the linear algebra of matrices. In discrete settings (simplicial complexes, finite graphs), the exterior derivative operators d_k become literal matrices, and the identity $d_{k+1} \circ d_k = 0$ becomes a concrete statement about the relationship between two matrices.

3.6.1 The Exterior Derivative as a Matrix

Consider a simplicial complex K: a triangulated space with N_0 vertices, N_1 edges, N_2 triangles, and so on. A k-cochain is a column vector in $\mathbb{R}^{N_k}$, assigning one real number to each oriented k-simplex. The exterior derivative d_k is represented by the $N_{k+1} \times N_k$ **coboundary matrix** $\mathbf{D}_k$: the matrix whose (j, i) entry is $+1$ if the j-th $(k+1)$-simplex contains the i-th k-simplex as a boundary face with agreeing orientation, -1 with disagreeing orientation, and 0 otherwise.

Explicitly, for the triangle with vertices A, B, C and edges e_{AB}, e_{BC}, e_{CA}:

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}, \qquad \mathbf{D}_1 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}. \tag{3.12}$$

The matrix $\mathbf{D}_0$ has three rows (one per edge) and three columns (one per vertex). The matrix $\mathbf{D}_1$ has one row (the triangle) and three columns (the edges). Acting on a vertex cochain $\mathbf{f} = (f_A, f_B, f_C)^\top$ gives the edge differences; acting on that with $\mathbf{D}_1$ sums the edge differences around the triangle.

3.6.2 Nilpotency as a Matrix Equation

The nilpotency $d_1 \circ d_0 = 0$ becomes:

$$\mathbf{D}_1\mathbf{D}_0 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}. \tag{3.13}$$

The product of the two matrices is the zero matrix. Every column of $\mathbf{D}_0$ (which describes one edge difference, the gradient of a vertex basis function) is in the null space of $\mathbf{D}_1$ (which sums edge values around the triangle boundary): every gradient field has zero circulation around every face.

3.6.3 Column Space Inside Null Space

For any two consecutive matrices $\mathbf{D}_k$ and $\mathbf{D}_{k+1}$ in the cochain complex:

$$\mathbf{D}_{k+1}\mathbf{D}_k = \mathbf{0} \tag{3.14}$$

is equivalent to the statement:

$$\operatorname{col}(\mathbf{D}_k) \subseteq \ker(\mathbf{D}_{k+1}). \tag{3.15}$$

Every vector in the column space of $\mathbf{D}_k$ (i.e., every exact $(k+1)$-cochain, every "coboundary") lies in the null space of $\mathbf{D}_{k+1}$ (i.e., is a closed $(k+1)$-cochain, a "cocycle"). In words: *every exact form is closed*.

This is not an equality. In general:

$$\operatorname{col}(\mathbf{D}_k) \subsetneq \ker(\mathbf{D}_{k+1}). \tag{3.16}$$

The column space is a *proper* subspace of the null space — unless the space has no topological holes in degree $k+1$. The quotient vector space

$$H^{k+1} = \ker(\mathbf{D}_{k+1}) \,/\, \operatorname{col}(\mathbf{D}_k) \tag{3.17}$$

is the cohomology group in degree $k+1$: it measures the gap, counting how many independent closed forms are not exact. If $H^{k+1} = 0$, the column space fills the entire null space and there is no gap. If $\dim H^{k+1} = r$, there are r independent directions in the null space that escape the column space — r independent topological obstructions.

3.6.4 Concrete Example: The Circle

For the circle triangulated as three vertices (v_0, v_1, v_2), three edges (e_{01}, e_{12}, e_{20}), no triangles:

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}, \qquad \mathbf{D}_1 = \mathbf{0} \quad \text{(no triangles)}. \tag{3.18}$$

Column space of $\mathbf{D}_0$: what are all possible edge-difference vectors? A vertex cochain $(a, b, c)^\top$ gives edge differences $(b-a, c-b, a-c)^\top$. These always satisfy the constraint $(b-a)+(c-b)+(a-c) = 0$: the three differences sum to zero. So the column space is the 2-dimensional subspace $\{(x, y, z) : x + y + z = 0\} \subset \mathbb{R}^3$.

Null space of $\mathbf{D}_1$: since $\mathbf{D}_1 = \mathbf{0}$, every vector in $\mathbb{R}^3$ is in the null space. $\ker(\mathbf{D}_1) = \mathbb{R}^3$.

The gap: $\ker(\mathbf{D}_1) = \mathbb{R}^3$ has dimension 3; $\mathrm{col}(\mathbf{D}_0)$ has dimension 2. The gap is 1-dimensional:

$$H^1 = \mathbb{R}^3 / \{x + y + z = 0\} \cong \mathbb{R}. \tag{3.19}$$

The generator of this 1-dimensional quotient is (any representative of) the cochain $(1, 1, 1)^\top$ — the uniform flow around the circle. This cochain is in $\ker(\mathbf{D}_1)$ (closed: no triangles to have zero circulation over) but not in $\mathrm{col}(\mathbf{D}_0)$ (not exact: its values do not sum to zero, so no global potential exists). The single dimension of H^1 records the one topological hole: the circle has one non-contractible loop.

Geometrically: the column space (exact cochains) consists of all flows with zero total circulation around the circle. The null space (closed cochains, here all cochains) is the whole space. The quotient identifies all cochains that differ by a gradient. A single number — the total circulation $x + y + z$ — is the invariant, and it determines the cohomology class.

3.6.5 The General Picture

For any cochain complex $\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2, \ldots$:

$$\mathbb{R}^{N_0} \xrightarrow{\mathbf{D}_0} \mathbb{R}^{N_1} \xrightarrow{\mathbf{D}_1} \mathbb{R}^{N_2} \xrightarrow{\mathbf{D}_2} \mathbb{R}^{N_3} \longrightarrow \cdots \tag{3.20}$$

the nilpotency $\mathbf{D}_{k+1}\mathbf{D}_k = \mathbf{0}$ holds at every stage, enforcing the nested inclusion:

$$\mathrm{col}(\mathbf{D}_0) \subseteq \ker(\mathbf{D}_1), \qquad \mathrm{col}(\mathbf{D}_1) \subseteq \ker(\mathbf{D}_2), \qquad \ldots \tag{3.21}$$

At each stage, the image of the previous map is a subspace of the kernel of the next. The quotients $\ker(\mathbf{D}_{k+1})/\mathrm{col}(\mathbf{D}_k)$ are the cohomology groups H^{k+1}, measuring the gaps. If all gaps are zero, the complex is *exact*: every closed cochain is a coboundary, and there is no topological information in any degree. Non-trivial cohomology means non-trivial topology.

In this matrix language, $d^2 = 0$ (or $\mathbf{D}_{k+1}\mathbf{D}_k = \mathbf{0}$) is the prerequisite for the entire cohomological machinery to make sense. Without it, the notion of "closed but not exact" would be undefined. Nilpotency creates the structure within which obstructions can be measured.

3.7 The Bulk-Boundary Interpretation: A Bulk Has No Bulk

There is a fourth way to see nilpotency, which directly connects the algebraic identity to the physical intuition of boundary effects and bulk sources from Chapter 2.

Recall: d_k converts a k-form (a measurement on k-dimensional objects) into a $(k+1)$-form (the bulk source responsible for that measurement). The question is: can that bulk source itself have a bulk source?

The answer is no. Here is the geometric reason.

A bulk source $d_k\omega$ explains the boundary data of ω: by Stokes, $\int_{\partial V}\omega = \int_V d_k\omega$ for any region V. The $(k+1)$-form $d_k\omega$ fills the $(k+1)$-dimensional bulk V. Now ask whether $d_k\omega$ can itself be a boundary effect that requires a $(k+2)$-dimensional bulk source to explain it.

That would require a region W of dimension $k+2$ such that $\int_{\partial W} d_k\omega \neq 0$. But by Stokes applied twice:

$$\int_{\partial W} d_k\omega = \int_{\partial(\partial W)} \omega = 0, \tag{3.22}$$

because $\partial(\partial W) = \emptyset$ (the boundary of a boundary is empty). The bulk $d_k\omega$ integrates to zero over every closed surface it could possibly be the boundary effect of. It has no "room" to be a boundary effect, because the boundaries available to it (∂W) always have empty further boundary. The bulk $d_k\omega$ is closed; it cannot be an exact form in the next dimension.

In other words:

A boundary has a bulk. (That is what d_k finds.) **But a bulk has no bulk.** (The bulk $d_k\omega$ cannot be a boundary effect of a still-higher source, because $\partial^2 = \emptyset$.)

This is $d_{k+1} \circ d_k = 0$: the bulk source of ω is not itself a boundary effect of any higher source. The chain of causation — from boundary observation to bulk source — can take exactly one step and no further.

Summary: Four Faces of One Truth

The nilpotency $d_{k+1} \circ d_k = 0$ is a single mathematical fact that can be seen from four complementary angles.

Water flows downhill: gradient flows (exact 1-forms) cannot circulate because a height function cannot increase and decrease along the same loop. No circulation means zero curl. This is $d_1 \circ d_0 = 0$ in physical language.

Boundaries have no boundaries: the boundary of any compact manifold is a closed manifold — a manifold without boundary. The boundary ∂M cannot itself be the boundary of a lower-dimensional piece inside M. Equivalently, a bulk cannot be the boundary of a higher-dimensional bulk. This geometric fact $\partial^2 = 0$ implies $d^2 = 0$ via Stokes.

Mixed partials commute, wedge is antisymmetric: a symmetric factor (mixed partials) paired with an antisymmetric factor (the wedge product sign) always cancels when summed over all pairs (ℓ, j). This is the algebraic proof, which works in coordinates on any smooth manifold.

Column space inside null space: in the discrete matrix picture, $\mathbf{D}_{k+1}\mathbf{D}_k = \mathbf{0}$ means the column space of $\mathbf{D}_k$ (the exact $(k+1)$-cochains) is a subspace of the null space of $\mathbf{D}_{k+1}$ (the closed $(k+1)$-cochains). The gap between these two subspaces is the cohomology group, measuring topological obstructions. If there is no gap, there is no topology to detect.

All four perspectives are facets of the same gem. In the chapters ahead, the cohomological machinery built on $d^2 = 0$ — de Rham cohomology, characteristic classes, the BRST formalism of gauge theory — is an elaboration of the nested inclusion $\mathrm{col}(d_k) \subseteq \ker(d_{k+1})$ and the study of when and why the inclusion is strict.

Chapter 4

Simplicial Homology and Cohomology

4.1 Why Finite Matrices Can Capture Infinite Topology

The smooth theory of differential forms operates on infinite-dimensional spaces: the space $\Omega^k(M)$ of all smooth k-forms is infinite-dimensional, the exterior derivative d_k is a differential operator, and computing cohomology in that setting requires functional analysis. Yet the topological invariants we are after — the Betti numbers b_k, the Euler characteristic χ, the number of holes in each dimension — are finite integers. Why should an infinite-dimensional computation produce a finite answer?

The answer is that the topological content is already present in any finite triangulation of the manifold. Replace a smooth manifold by a triangulation: a finite assembly of vertices, edges, triangles, and tetrahedra. Now the exterior derivative becomes a finite matrix (the coboundary matrix $\mathbf{D}_k$), forms become column vectors, and integration becomes a matrix-vector product. The infinite analysis compresses to finite linear algebra, and the topological invariants emerge from ranks and nullities of matrices you can write down on paper.

This is the strategy of the chapter. We will build the simplicial theory from scratch, emphasizing what every operation means concretely. The boundary operator ∂_k will be a matrix whose entries are $+1$, -1, and 0. The coboundary operator $D_k = \partial_{k+1}^\top$ will be its transpose. Homology groups will be kernels modulo images, computable by Gaussian elimination. The Betti numbers will follow from rank-nullity.

By the end, the reader will understand the following table, in which every row is a concrete numerical example:

Space	N_0	N_1	N_2	b_0	b_1	b_2
Single triangle (filled)	3	3	1	1	0	0
Triangle boundary (circle)	3	3	0	1	1	0
Square boundary	4	4	0	1	1	0
Torus T^2	9	27	18	1	2	1
Sphere S^2				1	0	1

Every entry is computable by matrix arithmetic alone.

4.2 Simplicial Complexes: Spaces Built from Pieces

A **k-simplex** is the convex hull of $k+1$ affinely independent points: a 0-simplex is a vertex, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron. The **faces** of a simplex are all sub-simplices obtained by omitting one or more vertices.

A **simplicial complex** K is a finite collection of simplices such that every face of a simplex in K is also in K, and any two simplices meet only along a common face. In other words: the pieces must be glued cleanly, face to face, with no interior overlaps.

Any smooth manifold M admits a triangulation: a simplicial complex K homeomorphic to M. Different triangulations exist, but the homology and cohomology we compute are independent of the choice — they are topological invariants.

4.2.1 Orientation: Which Way the Edges Point

Before defining any algebra, we must orient the simplices. An **orientation** of a k-simplex is a choice of cyclic ordering of its vertices, up to even permutations. Each simplex has two orientations, which we call positive and negative; the negative orientation is the simplex with a minus sign.

For edges: an orientation is a direction (from v_i to v_j or vice versa). For triangles: an orientation is a choice of counterclockwise or clockwise traversal of the three edges.

Why orientation? Because every quantity we care about — work along a path, flux through a surface, circulation around a loop — is signed. Work done in one direction is the negative of work done in the other. Orientation tracks this sign automatically, without requiring any external bookkeeping.

4.3 Two Kinds of Objects: Chains and Cochains

The simplicial theory has two parallel worlds, and understanding the distinction between them is the key to understanding everything that follows.

Chains are the geometric objects: collections of simplices that you can integrate over. A k-chain is a formal linear combination of oriented k-simplices with real coefficients: $c = \sum_i a_i \sigma_i$. The vector space of all k-chains is $C_k(K;\mathbb{R})$, of dimension N_k (the number of k-simplices). Think of a 1-chain as a weighted collection of directed edges — a formal path. Think of a 2-chain as a weighted collection of triangles — a formal surface.

Cochains are the measurement devices: functions that assign numbers to chains. A k-cochain is a linear function $\alpha : C_k \to \mathbb{R}$, determined by its values on each basis simplex. The vector space of all k-cochains is $C^k(K;\mathbb{R}) = C_k(K;\mathbb{R})^*$, also of dimension N_k. Think of a 0-cochain as a scalar field on vertices (height, temperature, potential). Think of a 1-cochain as a flow field on edges (measuring how much flows along each directed edge). Think of a 2-cochain as a flux field on triangles (measuring how much flux passes through each face).

The pairing $\langle \alpha, c \rangle = \sum_i a_i \alpha(\sigma_i)$ is the discrete analog of integration: a cochain α is integrated over a chain c. The cochain is the integrand; the chain is the domain. This is the simplicial version of $\int_M \omega$.

The fundamental duality. Chains are geometric (you can draw them); cochains are analytic (they measure). Homology is the study of chains: which closed chains bound something? Cohomology is the study of cochains: which closed cochains have a potential? These are dual questions, and over $\mathbb{R}$ they have the same answer (the same Betti numbers), but they approach the topology from opposite sides.

4.4 The Boundary Operator: From Region to Its Edge

The boundary operator $\partial_k : C_k \to C_{k-1}$ maps a k-chain to its $(k-1)$-dimensional boundary. For a single oriented simplex:

$$\partial_k[v_0, v_1, \ldots, v_k] = \sum_{m=0}^{k} (-1)^m [v_0, \ldots, \hat{v}_m, \ldots, v_k], \tag{4.1}$$

where $\hat{v}_m$ means "omit vertex v_m". The sign $(-1)^m$ ensures consistent orientation: each boundary face inherits the orientation induced by the original simplex.

For an edge: $\partial_1[v_0, v_1] = v_1 - v_0$. The boundary of a directed edge is the head vertex minus the tail vertex. This encodes direction: the boundary "knows" which end is the start and which is the finish.

For a triangle: $\partial_2[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$. The boundary of a counterclockwise triangle is the three edges traversed counterclockwise. The boundary of the triangle is its perimeter, as expected.

Extended by linearity, ∂_k is a linear map and is therefore a matrix: the **boundary matrix** $\boldsymbol{\partial}_k$, a $N_{k-1} \times N_k$ matrix with entries in $\{+1, -1, 0\}$.

4.4.1 The Nilpotency $\partial^2 = 0$: A Boundary Has No Boundary

The fundamental identity is $\partial_{k-1} \circ \partial_k = 0$: the boundary of a boundary is empty. Geometrically: the boundary of a triangle is its three edges forming a closed loop; a closed loop has no endpoints. The boundary of a solid ball is its spherical surface; a sphere has no rim.

Algebraically: when we compute $\partial_{k-1}(\partial_k \sigma)$ for a k-simplex σ, each $(k-2)$-face of σ appears exactly twice in the double sum, with opposite signs (corresponding to the two $(k-1)$-faces it belongs to), and the contributions cancel.

As matrices: $\boldsymbol{\partial}_{k-1}\boldsymbol{\partial}_k = \mathbf{0}$.

4.5 The Coboundary Operator: The Discrete Exterior Derivative

The coboundary operator $D_k : C^k \to C^{k+1}$ is the transpose of the boundary operator:

$$\mathbf{D}_k = \boldsymbol{\partial}_{k+1}^\top. \tag{4.2}$$

In terms of the pairing, this is the identity: $(D_k\alpha)(\sigma) = \alpha(\partial_{k+1}\sigma)$ for any $(k+1)$-simplex σ. Stokes' theorem for simplices.

The nilpotency $D_{k+1} \circ D_k = 0$ follows immediately:

$$\mathbf{D}_{k+1}\mathbf{D}_k = \boldsymbol{\partial}_{k+2}^\top \boldsymbol{\partial}_{k+1}^\top = (\boldsymbol{\partial}_{k+1}\boldsymbol{\partial}_{k+2})^\top = \mathbf{0}^\top = \mathbf{0}. \tag{4.3}$$

The coboundary operators have natural geometric meanings:

D_0: takes a 0-cochain (vertex values) and produces a 1-cochain (edge values) by $(D_0 f)(e_{ij}) = f(v_j) - f(v_i)$. This is the **discrete gradient**: the difference in potential across each edge. If f is height, $D_0 f$ measures the rise along each edge.

D_1: takes a 1-cochain (edge values) and produces a 2-cochain (triangle values) by $(D_1\alpha)(t) = \sum_{e \in \partial t} \alpha(e)$. This is the **discrete curl**: the circulation around each triangular face.

D_2: takes a 2-cochain (triangle values) and produces a 3-cochain (tetrahedron values) by $(D_2\omega)(T) = \sum_{f \in \partial T} \omega(f)$. This is the **discrete divergence**: the total outward flux through each tetrahedron's faces.

4.6 A Fully Worked Example: The Square

Before studying cohomology in general, let us compute everything explicitly for a simple example: a filled square.

4.6.1 Setting Up the Complex

Label the vertices of the square as v_0 (bottom-left), v_1 (bottom-right), v_2 (top-right), v_3 (top-left). The four edges are e_{01} (bottom), e_{12} (right), e_{23} (top, going right to left), e_{30} (left, going top to bottom). The square is divided into two triangles: $t_A = [v_0, v_1, v_2]$ (lower-right) and $t_B = [v_0, v_2, v_3]$ (upper-left), both with counterclockwise orientation.

Wait — we also need a diagonal. Let us add edge e_{02} (diagonal from v_0 to v_2). Now $t_A = [v_0, v_1, v_2]$ and $t_B = [v_0, v_2, v_3]$ are the two triangles.

Summary: $N_0 = 4$ vertices, $N_1 = 5$ edges, $N_2 = 2$ triangles.

4.6.2 The Boundary Matrices

$\boldsymbol{\partial}_1 : \mathbb{R}^5 \to \mathbb{R}^4$ (rows = vertices, cols = edges in order $e_{01}, e_{12}, e_{02}, e_{23}, e_{30}$):

$$\boldsymbol{\partial}_1 = \begin{pmatrix} -1 & 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \tag{4.4}$$

Reading column 1 (e_{01}): -1 at v_0 (tail), $+1$ at v_1 (head), 0 elsewhere. Column 3 (e_{02}): -1 at v_0, $+1$ at v_2. And so on.

$\boldsymbol{\partial}_2 : \mathbb{R}^2 \to \mathbb{R}^5$ (rows = edges, cols = triangles t_A, t_B):

$$\boldsymbol{\partial}_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \tag{4.5}$$

Reading column 1 ($t_A = [v_0, v_1, v_2]$): boundary is $e_{12} - e_{02} + e_{01}$, giving $+1$ at e_{01}, $+1$ at e_{12}, -1 at e_{02}, 0 elsewhere.

Verification of $\boldsymbol{\partial}_1\boldsymbol{\partial}_2 = \mathbf{0}$:

$$\boldsymbol{\partial}_1\boldsymbol{\partial}_2 = \begin{pmatrix} -1 & 0 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.6}$$

Every column of $\boldsymbol{\partial}_2$ lies in the null space of $\boldsymbol{\partial}_1$: the boundary of every triangle has no further boundary.

4.6.3 Computing the Cohomology Groups

The coboundary matrices are $\mathbf{D}_0 = \boldsymbol{\partial}_1^\top$ and $\mathbf{D}_1 = \boldsymbol{\partial}_2^\top$.

H^0**:** $\ker \mathbf{D}_0$ consists of vertex-cochains f with $f(v_j) - f(v_i) = 0$ for every edge e_{ij}. This forces all vertex values to be equal: f is constant. Since the square is connected, $\ker \mathbf{D}_0 = \text{span}(1,1,1,1)^\top$, one-dimensional. There are no 0-coboundaries (D_{-1} is the zero map). So $H^0 = \mathbb{R}$, confirming one connected component.

H^1**:** We need $\ker \mathbf{D}_1$ modulo $\text{col}(\mathbf{D}_0)$.

$\text{col}(\mathbf{D}_0)$: all edge-cochains of the form $(f_1 - f_0, f_2 - f_1, f_2 - f_0, f_3 - f_2, f_0 - f_3)$ for some (f_0, f_1, f_2, f_3). This has rank 3 (since $\mathbf{D}_0$ has a 1-dimensional null space, the constants). So the column space is 3-dimensional inside $\mathbb{R}^5$.

$\ker \mathbf{D}_1$: $\mathbf{D}_1\alpha = 0$ means the circulation around both triangles is zero. This gives two equations on the five edge-cochain values. The kernel is $5 - \text{rank}(\mathbf{D}_1) = 5 - 2 = 3$-dimensional (since both triangles give independent constraints).

Since $\dim \ker \mathbf{D}_1 = 3 = \dim \text{col}(\mathbf{D}_0)$, and $\text{col}(\mathbf{D}_0) \subseteq \ker \mathbf{D}_1$ by nilpotency, they must be equal. So $H^1 = 0$: the filled square has no 1-dimensional holes. (Intuitively: every closed path on the filled square bounds some region.)

H^2**:** $\ker \mathbf{D}_2$ (no 3-simplices, so $\ker \mathbf{D}_2 = \mathbb{R}^2$) modulo $\text{col}(\mathbf{D}_1)$ (which is 2-dimensional, since $\mathbf{D}_1$ has rank 2). So $H^2 = 0$.

Euler characteristic check: $\chi = 1 - 0 + 0 = 1 = N_0 - N_1 + N_2 = 4 - 5 + 2 = 1$. ✓

4.7 Homology and Cohomology: Holes Detected from Two Sides

4.7.1 What Cycles and Boundaries Mean

A k**-cycle** is a k-chain with zero boundary: $z \in \ker \boldsymbol{\partial}_k$. Geometrically: a collection of k-simplices whose boundaries cancel. A 1-cycle is a closed loop (every vertex has equal in- and out-flow). A 2-cycle is a closed surface (every edge is shared by an even number of triangles with cancelling orientations).

A k**-boundary** is a k-chain that is the boundary of a $(k+1)$-chain: $b \in \text{im}\, \boldsymbol{\partial}_{k+1}$. A 1-boundary is a loop that bounds a 2-dimensional region in the complex. A 2-boundary is a closed surface that bounds a 3-dimensional region.

By nilpotency, every boundary is a cycle: $\boldsymbol{\partial}_k \circ \boldsymbol{\partial}_{k+1} = 0$ means $\text{im}\, \boldsymbol{\partial}_{k+1} \subseteq \ker \boldsymbol{\partial}_k$.

The k**-th homology group** is

$$H_k(K; \mathbb{R}) = \ker \boldsymbol{\partial}_k \,/\, \text{im}\, \boldsymbol{\partial}_{k+1}. \tag{4.7}$$

Its dimension b_k counts the number of independent closed chains that do not bound: genuinely "hollow" loops, unfilled shells, empty voids. A loop on a sphere bounds (the sphere is hollow but the loop bounds a disk on the surface); a loop around the hole of a torus does not bound.

4.7.2 What Cocycles and Coboundaries Mean

A k**-cocycle** is a k-cochain α with $D_k\alpha = 0$. For a 1-cochain: the circulation around every triangle is zero. For a 2-cochain: the net flux through every tetrahedron is zero. Cocycles are the "conserved" measurement fields.

A k**-coboundary** is a k-cochain of the form $D_{k-1}f$ for some $(k-1)$-cochain f. A 1-coboundary is a gradient (differences of vertex values). The total circulation around any loop in a gradient field is zero (it telescopes).

The k**-th cohomology group** is

$$H^k(K; \mathbb{R}) = \ker \mathbf{D}_k \,/\, \text{im}\, \mathbf{D}_{k-1}. \tag{4.8}$$

Its dimension b_k counts the number of independent conserved fields that cannot be derived from a potential.

4.7.3 The Betti Number Formula via Rank-Nullity

Computing cohomology reduces to linear algebra. By the rank-nullity theorem for the matrix $\mathbf{D}_k$:

$$\dim \ker \mathbf{D}_k = N_k - \operatorname{rank} \mathbf{D}_k. \tag{4.9}$$

Since $\operatorname{im} \mathbf{D}_{k-1}$ has dimension $\operatorname{rank} \mathbf{D}_{k-1}$, and this lies inside $\ker \mathbf{D}_k$ by nilpotency:

$$b_k = \dim H^k = \dim \ker \mathbf{D}_k - \operatorname{rank} \mathbf{D}_{k-1} = (N_k - \operatorname{rank} \mathbf{D}_k) - \operatorname{rank} \mathbf{D}_{k-1}. \tag{4.10}$$

This is the computable formula. No geometry is required beyond computing the ranks of two matrices by Gaussian elimination.

For the circle example ($N_0 = 3$, $N_1 = 3$, no triangles):

$$b_0 = N_0 - \operatorname{rank} \mathbf{D}_0 - 0 = 3 - 2 - 0 = 1, \tag{4.11}$$

$$b_1 = N_1 - \operatorname{rank} \mathbf{D}_1 - \operatorname{rank} \mathbf{D}_0 = 3 - 0 - 2 = 1. \tag{4.12}$$

One connected component, one hole. For the filled square ($N_0 = 4$, $N_1 = 5$, $N_2 = 2$):

$$b_0 = 4 - 3 - 0 = 1, \tag{4.13}$$

$$b_1 = 5 - 2 - 3 = 0, \tag{4.14}$$

$$b_2 = 2 - 0 - 2 = 0. \tag{4.15}$$

One component, no holes. The Euler characteristic $\chi = b_0 - b_1 + b_2 = 1$ matches $N_0 - N_1 + N_2 = 4 - 5 + 2 = 1$.

4.8 The Linear Algebra Picture: Column Space and Null Space

4.8.1 The Four Fundamental Subspaces

Each coboundary matrix $\mathbf{D}_k : \mathbb{R}^{N_k} \to \mathbb{R}^{N_{k+1}}$ has four fundamental subspaces:

Subspace	Lives in	Geometric meaning	Name in cohomology
$\ker \mathbf{D}_k$	$\mathbb{R}^{N_k}$	Closed k-cochains	Cocycles
$\operatorname{col}(\mathbf{D}_{k-1})$	$\mathbb{R}^{N_k}$	Exact k-cochains	Coboundaries
$\ker \mathbf{D}_{k-1}^\top = \ker \boldsymbol{\partial}_k$	$\mathbb{R}^{N_k}$	Co-closed k-cochains	Harmonic
$\operatorname{col}(\mathbf{D}_k^\top) = \operatorname{col}(\boldsymbol{\partial}_{k+1})$	$\mathbb{R}^{N_k}$	Co-exact k-cochains	(see below)

By the fundamental theorem of linear algebra, the full space of k-cochains decomposes orthogonally as:

$$\mathbb{R}^{N_k} = \underbrace{\operatorname{col}(\mathbf{D}_{k-1})}_{\text{exact}} \oplus \underbrace{\ker \mathbf{D}_k \cap \ker \boldsymbol{\partial}_k}_{\text{harmonic}} \oplus \underbrace{\operatorname{col}(\boldsymbol{\partial}_{k+1})}_{\text{co-exact}}. \tag{4.16}$$

This is the discrete Hodge decomposition. Every k-cochain decomposes uniquely into a gradient part, a topological part (harmonic), and a curl part.

4.8.2 Cohomology as the Gap Between Column Space and Null Space

Nilpotency says $\mathrm{col}(\mathbf{D}_{k-1}) \subseteq \ker \mathbf{D}_k$: every exact cochain is closed. The cohomology group is the quotient of the null space by the column space:

$$H^k = \ker \mathbf{D}_k / \mathrm{col}(\mathbf{D}_{k-1}). \tag{4.17}$$

Over $\mathbb{R}$, we can find canonical representatives for each cohomology class: the element of the coset that is orthogonal to $\mathrm{col}(\mathbf{D}_{k-1})$. By the orthogonality relation $\mathbb{R}^{N_k} = \mathrm{col}(\mathbf{D}_{k-1}) \oplus \ker \boldsymbol{\partial}_k$, this is:

$$H^k \cong \ker \mathbf{D}_k \cap \ker \boldsymbol{\partial}_k = \ker \mathbf{D}_k \cap (\mathrm{col}(\mathbf{D}_{k-1}))^{\perp}. \tag{4.18}$$

In words: the cohomology group consists of those closed cochains that are *orthogonal to all exact cochains*. Not the whole null space, but the part of the null space that is invisible to the column space.

Physical reading: a harmonic cochain is one with zero curl (it does not circulate around any triangle) and zero co-curl (it does not accumulate at any vertex). It is in perfect local equilibrium. Its global behaviour is the topology: it circulates only around the "irreducible" loops of the space, the ones that detect the holes. There is one harmonic cochain per hole per degree.

4.8.3 Circle Example: The Column Space and Null Space Computed

For the triangulated circle with three vertices and three edges, no faces:

$\mathbf{D}_0$ has rank 2 (one null vector: the constant cochain). Its column space in $\mathbb{R}^3$ is $\mathrm{col}(\mathbf{D}_0) = \{(a,b,c) : a+b+c=0\}$: the 2-dimensional space of "balanced" edge-cochains (differences between vertex values always sum to zero around any loop).

$\ker \mathbf{D}_1 = \mathbb{R}^3$ (no triangles, so $\mathbf{D}_1 = 0$).

Gap: $\ker \mathbf{D}_1 / \mathrm{col}(\mathbf{D}_0) = \mathbb{R}^3 / \{a+b+c=0\} \cong \mathbb{R}$. The one-dimensional quotient is generated by $(1,1,1)^\top$: the uniform cochain assigning $+1$ to every edge, the discrete $d\theta$.

Harmonic representative: $(1,1,1)^\top$ is orthogonal to $\mathrm{col}(\mathbf{D}_0)$ (since $1\cdot(-1)+1\cdot 1+1\cdot 0 = 0$ and $1\cdot 0+1\cdot(-1)+1\cdot 1 = 0$). Check. It is also in $\ker \boldsymbol{\partial}_1$: at every vertex, exactly one incoming and one outgoing edge with coefficient $+1$, net flow zero. Check.

The harmonic cochain $(1,1,1)^\top$ is the topological invariant: it counts how many times you go around the loop. Its integral over the full circle is $3 \neq 0$: the winding number is one.

4.9 Obstruction in the Simplicial Setting

4.9.1 The Poincaré Lemma for Simplices: Local Exactness

The smooth Poincaré lemma says every closed form is locally exact (exact on any contractible open set). There is a discrete analog: every closed cochain is exact on any tree subcomplex.

A **spanning tree** T of a connected simplicial complex K is a maximal acyclic subgraph: a connected subgraph of the 1-skeleton that includes all vertices but no loops. A spanning tree is contractible (it can be collapsed to a point by successively removing leaves), so there are no topological obstructions on it.

Discrete Poincaré lemma on a tree: if α is a closed 1-cochain ($D_1\alpha = 0$), then on any spanning tree T, the cochain $\alpha|_T$ is exact: there exists a 0-cochain f defined on all vertices such that $D_0 f = \alpha$ on the edges of T.

Proof: Root the tree at any vertex v_0. Set $f(v_0) = 0$. For any other vertex v, there is a unique path in T from v_0 to v; set $f(v) = \sum_{e \text{ on path}} \alpha(e)$ (with appropriate signs). On tree edges, $f(v_j) - f(v_i) = \alpha(e_{ij})$ by construction: the cochain is exactly the gradient of f on T.

4.9.2 The Obstruction Lives on Non-Tree Edges

Now try to extend f from the tree to the whole complex. For each non-tree edge e_{ij} (an edge not in T), we already have values $f(v_i)$ and $f(v_j)$ defined (since T includes all vertices). The question is whether $f(v_j) - f(v_i) = \alpha(e_{ij})$.

If $f(v_j) - f(v_i) \neq \alpha(e_{ij})$, then the edge e_{ij} contributes an **obstruction**. The non-tree edge, together with the unique tree path from v_i to v_j, forms a loop. The total circulation of α around this loop is:

$$\text{loop circulation} = \alpha(e_{ij}) - (f(v_j) - f(v_i)) = \alpha(e_{ij}) - \sum_{e \in T\text{-path}} \alpha(e). \tag{4.19}$$

This is the **period** of α around the loop. If all periods are zero, $\alpha = D_0 f$ globally and the cochain is exact. If any period is nonzero, the cochain is not exact and the nonzero period is the obstruction.

4.9.3 Counting Obstructions: The First Betti Number

For a connected complex with N_0 vertices, N_1 edges, and a spanning tree T:

$$\text{number of non-tree edges} = N_1 - (N_0 - 1) = N_1 - N_0 + 1. \tag{4.20}$$

Each non-tree edge defines one independent loop (adding the edge to the tree creates exactly one cycle). The number of independent loops is $N_1 - N_0 + 1$, which is exactly the first Betti number b_1 for a complex with $H_2 = 0$ (no 2-dimensional holes). This matches the rank-nullity formula: $b_1 = N_1 - \operatorname{rank} \mathbf{D}_1 - \operatorname{rank} \mathbf{D}_0$.

Obstruction data. The vector of all period integrals $(p_1, p_2, \ldots, p_{b_1})$ for the b_1 independent loops is the complete obstruction to exactness. The cochain α is exact if and only if all periods vanish. If α' and α differ by a coboundary ($\alpha' = \alpha + D_0 g$), their periods are the same (exact cochains have zero periods). So the periods depend only on the cohomology class of α. The cohomology group H^1 *is* the space of period vectors: $H^1 \cong \mathbb{R}^{b_1}$, and each independent direction in H^1 corresponds to one independent topological loop.

4.9.4 Worked Obstruction: The Circle

For the triangulated circle with vertices v_0, v_1, v_2 and edges e_{01}, e_{12}, e_{20}:

Spanning tree $T = \{e_{01}, e_{12}\}$ (two edges connecting all three vertices without a loop). The non-tree edge is e_{20}.

Take the closed 1-cochain $\alpha = (1, 1, 1)^\top$ (value 1 on every edge).

On the tree, integrate to find f: $f(v_0) = 0$, then $f(v_1) = f(v_0) + \alpha(e_{01}) = 1$, then $f(v_2) = f(v_1) + \alpha(e_{12}) = 2$.

Now check the non-tree edge e_{20} (going from v_2 to v_0): the gradient says we need $f(v_0) - f(v_2) = 0 - 2 = -2$. But $\alpha(e_{20}) = 1$ (the cochain value on this edge is $+1$ in the direction from v_2 to v_0). The discrepancy is $1 - (-2) = 3$.

This discrepancy is the period of α around the loop $e_{01} + e_{12} + e_{20}$: the total circulation is $1 + 1 + 1 = 3 \neq 0$. The cochain is not exact; the obstruction is the single nonzero period. The first Betti number is $b_1 = 1$ (one independent loop, one period).

If we replace α by $\alpha + D_0 g$ for any 0-cochain g, the period becomes $(1 + g_1 - g_0) + (1 + g_2 - g_1) + (1 + g_0 - g_2) = 3$ regardless of g. The period is cohomology-class invariant.

4.9.5 Obstruction for Homology: When Does a Loop Bound?

The dual question is about cycles, not cocycles. When does a closed chain (loop) bound a higher-dimensional chain (surface)?

A 1-cycle $z \in \ker \boldsymbol{\partial}_1$ bounds if $z \in \operatorname{im} \boldsymbol{\partial}_2$: there exists a 2-chain c with $\partial_2 c = z$. The obstruction is: can we fill in the loop with triangles from the complex?

Example: the circle. The loop $z = e_{01} + e_{12} + e_{20}$ (the entire triangulated circle) is a 1-cycle: its boundary is $\partial_1 z = (v_1 - v_0) + (v_2 - v_1) + (v_0 - v_2) = 0$. Does z bound? Only if there is a 2-chain $c = at_1 + bt_2 + \cdots$ (triangles) with $\partial_2 c = z$. If the complex contains no triangles, there is no such 2-chain. The loop does not bound: $H_1 \neq 0$.

Example: filling in the circle. Now add a single triangle $t = [v_0, v_1, v_2]$ (filling in the disk). Then $\partial_2 t = e_{12} - e_{20} + e_{01}$. For the loop $z = e_{01} + e_{12} + e_{20}$: we need $a\,\partial_2 t = z$, i.e., $a(e_{12} - e_{20} + e_{01}) = e_{01} + e_{12} + e_{20}$. This requires $a = 1$ and $-a = 1$ simultaneously: no solution. The loop still does not bound, because the orientation of the loop is opposite to the boundary of the filled triangle. *(In the correctly oriented version, $z = e_{01} + e_{12} - e_{20}$ does bound.)*

The point is: homological obstruction is also detected by periods, but now via integration of cochains over cycles. The period of a cocycle α around a cycle z is $\langle \alpha, z \rangle$. If α is exact, all periods with any cycle vanish. If z bounds, all periods with any cocycle vanish. This mutual annihilation is Poincaré duality: $H^k \cong H_k^*$.

4.10 The Cochain Complex and the de Rham Correspondence

The chain of coboundary maps

$$0 \to C^0 \xrightarrow{\mathbf{D}_0} C^1 \xrightarrow{\mathbf{D}_1} C^2 \xrightarrow{\mathbf{D}_2} \cdots \tag{4.21}$$

is the **simplicial cochain complex**. Its cohomology groups $H^k = \ker \mathbf{D}_k / \operatorname{im} \mathbf{D}_{k-1}$ are topological invariants of the underlying space, independent of the choice of triangulation.

The connection to de Rham cohomology is through integration. Given a smooth k-form ω on a triangulated manifold, define the integration cochain $I_k(\omega) \in C^k$ by $I_k(\omega)(\sigma^k) = \int_{\sigma^k} \omega$. Stokes' theorem gives $I_{k+1}(d\omega) = D_k \circ I_k(\omega)$: the integration map commutes with the exterior derivative on the smooth side and the coboundary on the discrete side. Integration therefore sends smooth cocycles to discrete cocycles and smooth coboundaries to discrete coboundaries, giving a map on cohomology: $[I] : H^k_{\mathrm{dR}}(M) \to H^k(K; \mathbb{R})$.

The de Rham theorem (Chapter 5) asserts this map is an isomorphism. The smooth theory and the discrete theory detect exactly the same holes, the same obstructions, the same topology.

4.11 The Torus: Reading Off Topology from Matrices

The torus T^2, triangulated as a 3×3 grid with periodic identifications, has $N_0 = 9$, $N_1 = 27$, $N_2 = 18$.

The boundary matrices $\boldsymbol{\partial}_1$ and $\boldsymbol{\partial}_2$ can be written down from the connectivity data. Gaussian elimination (or Smith normal form) gives:

$$\operatorname{rank} \boldsymbol{\partial}_1 = 8, \quad \operatorname{rank} \boldsymbol{\partial}_2 = 17. \tag{4.22}$$

Using the Betti number formula:

$$b_0 = 9 - 8 = 1, \tag{4.23}$$
$$b_1 = 27 - 17 - 8 = 2, \tag{4.24}$$
$$b_2 = 18 - 0 - 17 = 1. \tag{4.25}$$

One connected component, two independent loops, one enclosing "fundamental class". Euler characteristic: $1 - 2 + 1 = 0$.

The two generators of $H^1(T^2)$ are the horizontal and vertical loops. These are the two non-tree edges that survive after removing a spanning tree. The two independent periods — one around the longitude, one around the meridian — are the two independent topological obstructions to global exactness of any 1-cochain on the torus.

Summary: What Simplicial Cohomology Computes

The simplicial theory translates topology into linear algebra. Every step has a matrix:

Object	Linear algebra	Topology
k-cochains	$\mathbb{R}^{N_k}$	Measurements on k-simplices
$\mathbf{D}_k = \boldsymbol{\partial}_{k+1}^\top$	Matrix	Discrete exterior derivative
$\ker \mathbf{D}_k$	Null space	Closed cochains (cocycles)
$\mathrm{col}(\mathbf{D}_{k-1})$	Column space	Exact cochains (coboundaries)
$H^k = \ker / \mathrm{col}$	Quotient	k-dimensional holes
$b_k = \dim H^k$	Rank-nullity	Betti number
Non-tree edges	Free variables	Period integrals (obstructions)
Periods $\neq 0$	Obstruction data	Non-trivial cohomology class

The de Rham theorem (Chapter 5) will show that all of this — the holes, the periods, the obstructions — is exactly the same information that smooth differential forms detect via integration. The continuous and the discrete are, topologically, the same thing.

Chapter 5

De Rham Cohomology

5.1 Collecting the Threads

Three separate stories have been building across the preceding chapters, each circling the same central question from a different direction. It is time to bring them together.

Chapter 2 introduced closed and exact forms through physical intuition. A gradient flow — water running strictly downhill along a height function f — is an exact 1-form $\omega = d_0 f$: it can never circulate, because height is monotone along the flow. A circular canal flow, by contrast, is a closed 1-form ($d_1\omega = 0$, no local vortices) that is not exact (no global height function drives it; the flow goes around a topological loop). The obstruction to exactness was made precise: cover the domain by contractible patches, find local potentials on each, and observe that the potentials *fail to agree on overlaps*. The disagreements accumulate to a nonzero period integral around the loop, and this period is the obstruction.

Chapter 4 restated the same story in discrete language. A closed 1-cochain on a triangulated circle (the uniform cochain $(1, 1, 1)^\top$) is not a gradient cochain because its total circulation is $3 \neq 0$. The spanning-tree argument made the obstruction mechanical: propagate a potential along a spanning tree, observe the discrepancy on the non-tree edge. Each such discrepancy is a period integral, living in a space of dimension equal to the first Betti number b_1.

This chapter develops the smooth, coordinate-free version of the same theory, giving it the name and structure it is known by: **de Rham cohomology**. The groups $H^k_{\mathrm{dR}}(M)$ organize all the obstruction data in degree k into a vector space whose dimension b_k counts the independent topological holes. The de Rham theorem then says that these smooth groups are isomorphic to the simplicial cohomology groups from Chapter 4: the same topology, seen from two different computational perspectives.

5.2 What the Notation Means: The Groups $H^k_{\mathrm{dR}}(M)$

Before computing anything, we should understand what the symbol $H^k_{\mathrm{dR}}(M)$ means and why it takes the form it does.

The letter **H** stands for homology (from the German *Homologie*), a term introduced by Poincaré in 1895 for the algebraic study of shapes via cycles and boundaries. "Cohomology" is the dual version, prefixed by "co" to indicate that it involves functions on chains rather than the chains themselves. The choice of H for both is historical: both theories were discovered together and measure the same topological information from complementary directions.

The **superscript** k is the degree: it tells us which dimension of holes we are counting. H^0 counts

connected components. H^1 counts independent non-contractible loops. H^2 counts independent enclosed voids. In general, H^k counts independent k-dimensional topological holes.

The **subscript dR** stands for "de Rham", indicating that we are using differential forms and the exterior derivative to build the theory, as opposed to singular cochains (algebraic topology) or simplicial cochains (Chapter 4). The de Rham theorem guarantees that all these versions compute the same groups (with real coefficients), so the subscript is often dropped when the context is clear.

The **argument** M is the manifold whose topology we are measuring. Different manifolds have different cohomology groups: the groups are the algebraic fingerprint of the topology of M.

5.3 The de Rham Complex

The starting structure is the **de Rham complex**: the chain of vector spaces and maps

$$0 \longrightarrow \Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \Omega^2(M) \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} \Omega^n(M) \longrightarrow 0, \tag{5.1}$$

where $\Omega^k(M)$ is the (infinite-dimensional) vector space of smooth k-forms on M, and the maps are the successive exterior derivative operators d_k from Chapter 2.

The nilpotency $d_{k+1} \circ d_k = 0$ says that consecutive maps compose to zero. A sequence of vector spaces and linear maps with this property is called a **cochain complex**. The nilpotency is precisely the condition that makes cohomology well-defined: it ensures that $\operatorname{im}(d_{k-1}) \subseteq \ker(d_k)$, so that the exact forms are a subspace of the closed forms.

The complex (5.1) is the smooth analog of the coboundary complex from Chapter 4:

$$\begin{aligned} &\text{Smooth:} && \cdots \xrightarrow{d_{k-1}} \Omega^k(M) \xrightarrow{d_k} \Omega^{k+1}(M) \xrightarrow{d_{k+1}} \cdots \\ &\text{Discrete:} && \cdots \xrightarrow{\mathbf{D}_{k-1}} \mathbb{R}^{N_k} \xrightarrow{\mathbf{D}_k} \mathbb{R}^{N_{k+1}} \xrightarrow{\mathbf{D}_{k+1}} \cdots \end{aligned}$$

The smooth version has infinite-dimensional spaces; the discrete version has finite-dimensional ones. Both have $d^2 = 0$. Both produce the same cohomology groups (by the de Rham theorem, stated below). The discrete version is more computable; the smooth version is more geometric.

5.4 Cohomology as Equivalence Classes

5.4.1 The Quotient Construction

At each degree k, two vector spaces are naturally defined:

$$Z^k(M) = \ker\bigl(d_k : \Omega^k(M) \to \Omega^{k+1}(M)\bigr) \quad \text{(closed } k\text{-forms, or } k\text{-cocycles)}, \tag{5.2}$$

$$B^k(M) = \operatorname{im}\bigl(d_{k-1} : \Omega^{k-1}(M) \to \Omega^k(M)\bigr) \quad \text{(exact } k\text{-forms, or } k\text{-coboundaries)}. \tag{5.3}$$

Nilpotency gives $B^k(M) \subseteq Z^k(M)$: every exact form is closed. The **k-th de Rham cohomology group** is the quotient:

$$H^k_{\mathrm{dR}}(M) = \frac{Z^k(M)}{B^k(M)} = \frac{\ker d_k}{\operatorname{im} d_{k-1}} = \frac{\text{closed } k\text{-forms}}{\text{exact } k\text{-forms}}. \tag{5.4}$$

5.4.2 What an Equivalence Class Is

Two closed k-forms ω and ω' are **cohomologous** if their difference is exact: $\omega - \omega' = d_{k-1}\alpha$ for some $(k-1)$-form α. This is an equivalence relation, and the equivalence class of ω is

$$[\omega] = \{\omega + d_{k-1}\alpha : \alpha \in \Omega^{k-1}(M)\}. \tag{5.5}$$

Think of this as an **orbit**: starting from the closed form ω, we can add any exact form $d\alpha$ without changing the topological information. The orbit is the entire equivalence class, and the cohomology group $H^k_{\mathrm{dR}}(M)$ is the space of all orbits.

Why can we add exact forms freely? Because exact forms have zero periods around every cycle: $\int_\gamma d\alpha = \alpha(\partial\gamma) = 0$ for a closed loop (since $\partial\gamma = \emptyset$). An exact form contributes nothing to the topological measurement — the period integrals that detect holes. So two forms with the same periods are topologically the same, even if they differ pointwise.

Gauge theory interpretation. In electromagnetism, the vector potential A is a 1-form and the field strength $F = d_1 A$ is a 2-form. The gauge transformation $A \mapsto A + d_0\Lambda$ (for any smooth function Λ) does not change F. So A and $A + d_0\Lambda$ are in the same orbit, and the physics (encoded in F) depends only on the cohomology class $[A]$. Every gauge orbit is a single cohomology class.

5.4.3 Why It Is a Group (and a Vector Space)

The set $H^k_{\mathrm{dR}}(M)$ is not merely a set of orbits; it is a vector space, hence an abelian group. Here is the structure.

Addition: $[\alpha] + [\beta] = [\alpha + \beta]$. This is well-defined: if we replace α by $\alpha + d\gamma$ and β by $\beta + d\delta$, the sum changes by $d\gamma + d\delta = d(\gamma + \delta)$, which is exact. The cohomology class of the sum does not depend on which representatives we chose.

Scalar multiplication: $c \cdot [\alpha] = [c\alpha]$ for $c \in \mathbb{R}$.

Zero element: $[0]$, the class of all exact forms. Any two exact forms $d\alpha$ and $d\beta$ differ by $(d\alpha - d\beta) = d(\alpha - \beta)$, which is exact, so they are all in the same class.

Additive inverse: $-[\alpha] = [-\alpha]$.

The axioms are all inherited from the vector space structure of $Z^k(M)$ and are preserved by the quotient, so $H^k_{\mathrm{dR}}(M)$ is a real vector space. Its dimension is the **k-th Betti number** $b_k = \dim H^k_{\mathrm{dR}}(M)$, a non-negative integer (or zero) that counts independent topological holes in dimension k.

Intuition for group elements. Each non-zero element of $H^k_{\mathrm{dR}}(M)$ is one independent topological obstruction: one independent direction in which a closed form can fail to be exact. When $b_k = 2$ (as for $H^1(T^2)$), there are two independent obstructions. A general cohomology class is a linear combination of these, with real coefficients that encode the "how much" of each obstruction.

Parallel with discrete. In Chapter 4, the simplicial cohomology group was the quotient $\ker \mathbf{D}_k / \mathrm{col}(\mathbf{D}_{k-1})$: a finite-dimensional vector space. The de Rham cohomology group is the same construction, replacing finite matrices by differential operators and finite-dimensional vectors by smooth forms. The group structure, the equivalence class picture, and the Betti number all carry over directly.

5.5 The Poincaré Lemma: Locally, Everything Is Exact

The first theorem asserts that the obstructions to exactness are *entirely global*: locally, near any point, a closed form is always exact.

Theorem 5.1 (Poincaré Lemma)**.** *On any contractible open set $U \subseteq M$, every closed k-form is exact: $H^k_{\mathrm{dR}}(U) = 0$ for $k \geq 1$.*

Proof via homotopy operator. Take U to be star-shaped around the origin. Define the **homotopy operator** $K : \Omega^k(U) \to \Omega^{k-1}(U)$ by contracting forms along radial lines:

$$(K\omega)_x(v_1, \ldots, v_{k-1}) = \int_0^1 t^{k-1}\, \omega_{tx}(x, v_1, \ldots, v_{k-1})\, dt. \tag{5.6}$$

A direct computation gives the **homotopy formula**:

$$d_{k-1}(K\omega) + K(d_k\omega) = \omega. \tag{5.7}$$

If ω is closed ($d_k\omega = 0$), then $\omega = d_{k-1}(K\omega)$: the potential is explicitly $\alpha = K\omega$. □

The formula (5.7) is a chain homotopy between the identity map and the zero map on $\Omega^k(U)$. Its existence says that U is "algebraically contractible" — the identity map on its forms is chain-homotopic to zero. This is the smooth analog of the discrete Poincaré lemma (propagating a potential along a spanning tree).

What makes it fail globally? On a contractible space, the homotopy operator K can be constructed globally and gives an explicit potential. On a non-contractible space, there is no global K: the radial contraction does not work because some paths would have to pass through a "hole" (a missing point, a missing region, or a topological obstruction).

5.6 The Central Theme: Local Exactness Does Not Imply Global Exactness

The Poincaré lemma says every closed form is locally exact. This is a complete, unqualified statement about small regions. The de Rham cohomology groups are the systematic measurement of how badly this local exactness fails to extend globally. Every non-trivial element of $H^k_{\mathrm{dR}}(M)$ is precisely one independent mode of failure.

This section builds the intuition for this failure at each degree, working through a progression of examples from the most familiar to the higher-dimensional, and ends by connecting the failure directly to Chern classes.

5.6.1 Degree 1: Flows That Cannot Have a Global Height Function

The most vivid setting is fluid flow in two or three dimensions, where a closed 1-form is a velocity field with no local vorticity, and exactness means the flow is driven by a global potential (height function).

Water Flowing Downhill: The Exact Case

Consider water on a landscape with a height function $f : M \to \mathbb{R}$. The gradient flow $\omega = d_0 f$ is the 1-form whose value on any tangent vector gives the rate of ascent. This form is:

- **Exact** by construction: $\omega = df$.
- **Closed**: $d_1\omega = d_1(d_0 f) = 0$ by nilpotency.

- **Cannot circulate**: $\oint_\gamma \omega = f(\gamma(1)) - f(\gamma(0)) = 0$ for every closed loop γ, because the height returns to its starting value.

Water flows strictly downhill. It can never complete a closed circuit at constant energy. The global potential f is the height function; the flow is globally exact. Here $[\omega] = 0$ in $H^1(M)$: no obstruction.

Flow Around a Cylinder: Locally Exact, Globally Obstructed

Now remove the z-axis from $\mathbb{R}^3$ and consider the domain $M = \mathbb{R}^3 \setminus \{z\text{-axis}\}$. Let r, ϕ, z be cylindrical coordinates. The 1-form

$$\omega = d\phi = \frac{-y\,dx + x\,dy}{x^2 + y^2} \tag{5.8}$$

describes a flow circling around the z-axis. It is:

- **Closed**: $d_1\omega = 0$ everywhere on M (direct computation, or because $\omega = d\phi$ locally).
- **Locally exact**: on any contractible patch $U \subset M$ (any open set not wrapping all the way around the axis), $\omega = d(\text{local angle})$. A local potential always exists.
- **Not globally exact**: $\oint_C \omega = 2\pi$ for any loop C encircling the axis once. No single-valued smooth function $f : M \to \mathbb{R}$ satisfies $df = \omega$ on all of M, because ϕ itself jumps by 2π as you go around.

Why the local potential fails to extend. Cover M by two overlapping half-spaces U_+ (the region $y > 0$ or $x > 0$, say) and U_- (the region $y < 0$ or $x > 0$). On U_+, a local potential is $f_+ = \arctan(y/x)$ (angle measured in standard range). On U_-, a local potential is $f_- = \arctan(y/x) + \text{const}$ (angle on a different branch). On the overlap $U_+ \cap U_-$, the two potentials differ by a constant: $f_+ - f_- = \pm\pi$. This constant discrepancy is the obstruction. Any attempt to patch the two local potentials into one global function fails at the seam: the function jumps by 2π as you cross from one patch to the other.

The domain $M = \mathbb{R}^3 \setminus \{z\text{-axis}\}$ deformation retracts to S^1, and $H^1(S^1) \cong \mathbb{R}$ with generator $[d\phi]$. The obstruction is one-dimensional: a single real number, the period $\oint_C \omega$, determines whether a global potential exists.

Flow Around a Torus: Two Independent Obstructions

The torus $T^2 = S^1 \times S^1$ is parametrized by two angles $(\theta, \phi) \in [0, 2\pi)^2$. Consider the 1-form

$$\omega = a\,d\theta + b\,d\phi, \qquad a, b \in \mathbb{R}. \tag{5.9}$$

This form is closed (both $d\theta$ and $d\phi$ are closed on T^2). It is locally exact: on any small contractible patch, $\omega = d(a\theta + b\phi)$ locally. But it is globally exact if and only if $a = b = 0$: the potential $a\theta + b\phi$ fails to be single-valued on T^2 unless both coefficients vanish.

The two independent obstructions are the two periods:

$$P_1 = \oint_{\theta\text{-circle}} \omega = 2\pi a, \qquad P_2 = \oint_{\phi\text{-circle}} \omega = 2\pi b. \tag{5.10}$$

Both must be zero for a global potential to exist. If either is nonzero, the flow has an irreducible circulation around that cycle of the torus. This is why $H^1(T^2) \cong \mathbb{R}^2$: there are two independent directions in which local exactness can fail to extend globally.

Physically: think of water in a donut-shaped pipe. The flow can circulate around the long way (the θ-direction), or around the short way (the ϕ-direction), or both simultaneously. These two circulations are topologically independent. No height function on the torus can drive either circulation, because the torus has no "downhill" direction that returns to the start.

The General Principle for Degree 1

For any connected manifold M, a closed 1-form ω is globally exact if and only if its period around every closed loop is zero. The space of independent obstructions is $H^1_{\mathrm{dR}}(M) \cong \mathbb{R}^{b_1}$, where b_1 is the number of independent non-contractible loops. Each loop contributes one obstruction; the period around that loop is the obstruction datum. The local potential always exists (Poincaré lemma); the obstruction is entirely in how the local potentials fail to agree across the global topology.

5.6.2 Degree 2: Fluxes That Cannot Have a Global Vector Potential

One dimension up, closed 2-forms are flux densities and exactness means the flux comes from a globally defined vector potential.

Uniform Flux Through a Sphere: The Monopole

Consider $M = S^2$, the 2-sphere, and the area form

$$\omega = \sin\theta \, d\theta \wedge d\phi. \tag{5.11}$$

This is a closed 2-form on S^2 (there are no 3-forms on a 2-dimensional manifold). Its period over S^2 is $\int_{S^2} \omega = 4\pi \neq 0$.

Is ω exact? If $\omega = d_1\alpha$ for some global 1-form α, then by Stokes: $\int_{S^2} \omega = \int_{S^2} d\alpha = \int_{\partial S^2} \alpha = 0$ (since S^2 has no boundary). But $4\pi \neq 0$. Contradiction. No global 1-form α exists with $d\alpha = \omega$.

Local potentials do exist. Cover S^2 by the northern hemisphere U_N (contractible, a disk) and the southern hemisphere U_S (contractible, a disk). On U_N: $\alpha_N = \frac{1-\cos\theta}{2} d\phi$ satisfies $d\alpha_N = \omega$. On U_S: $\alpha_S = -\frac{1+\cos\theta}{2} d\phi$ satisfies $d\alpha_S = \omega$. Both local potentials are valid on their respective patches. On the overlap (the equatorial strip): $\alpha_N - \alpha_S = d\phi$, a closed 1-form with period 2π. The local potentials disagree by a closed-but-not-exact form on the overlap, and this is the obstruction to a global potential.

Physically: this is the Dirac monopole. The area form is the magnetic flux of a magnetic monopole of charge $g = 4\pi$. The local potentials α_N and α_S are the north-gauge and south-gauge vector potentials. The Dirac string (the locus where neither gauge is valid) is the cost of trying to cover the sphere with a single potential. The cohomology class $[\omega] \in H^2(S^2) \cong \mathbb{R}$ is the magnetic charge; its integrality $\int_{S^2} \omega/(2\pi) \in \mathbb{Z}$ is the Dirac quantization condition.

Flux Through the Punctured Space: The Coulomb Field

Take $M = \mathbb{R}^3 \setminus \{0\}$. The 2-form

$$\omega = \frac{x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy}{(x^2+y^2+z^2)^{3/2}} = \frac{\hat{r} \cdot d\mathbf{S}}{r^2} \tag{5.12}$$

is the flux form of the Coulomb (or monopole) field. It is:

- **Closed**: $d\omega = 0$ on $\mathbb{R}^3 \setminus \{0\}$ (the divergence of $\hat{r}/r^2$ is zero away from the origin).

- **Locally exact**: on any contractible patch not enclosing the origin, a local vector potential exists.
- **Not globally exact**: $\int_{S^2} \omega = 4\pi \neq 0$, so no global vector potential exists on $\mathbb{R}^3 \setminus \{0\}$.

The domain $\mathbb{R}^3 \setminus \{0\}$ deformation retracts to S^2, and $H^2(S^2) \cong \mathbb{R}$ is generated by $[\omega]$. The single real number $\int_{S^2} \omega$ is the period that measures the obstruction to a global potential. The origin is the topological hole responsible: it forces the flux lines to converge at a source, creating a non-contractible S^2 surrounding it.

Two Independent 2-Obstructions: The Torus

On T^2, the single 2-form $d\theta \wedge d\phi$ is closed (it is the volume form on the torus) with nonzero period $\int_{T^2} d\theta \wedge d\phi = (2\pi)^2$, so $H^2(T^2) \cong \mathbb{R}$ is one-dimensional.

Now consider $T^3 = S^1 \times S^1 \times S^1$, the 3-torus with coordinates $(\theta_1, \theta_2, \theta_3)$. There are three independent closed 2-forms: $d\theta_1 \wedge d\theta_2$, $d\theta_1 \wedge d\theta_3$, $d\theta_2 \wedge d\theta_3$. Each has a nonzero period over the corresponding 2-torus face. All three are locally exact (on any contractible patch, local potentials exist) but none is globally exact. The three independent periods are three independent obstructions, giving $H^2(T^3) \cong \mathbb{R}^3$.

5.6.3 Degree k: The General Pattern

The pattern generalizes to any degree. A closed k-form ω is:

- **Always locally exact**: on any contractible open set U, the Poincaré lemma gives a $(k-1)$-form α_U with $d\alpha_U = \omega|_U$. This is completely general and requires no topological assumption.
- **Globally exact** if and only if all its period integrals $\int_\Sigma \omega$ vanish for every closed k-cycle Σ in M.
- **Obstructed** when some period is nonzero: the nonzero period is one independent failure mode, living in $H^k_{\mathrm{dR}}(M)$.

The number of independent obstructions is the Betti number $b_k = \dim H^k_{\mathrm{dR}}(M)$. The full set of period integrals $\{\int_{\Sigma_i} \omega\}$ against a basis of k-cycles $\{\Sigma_i\}$ is the coordinate vector of $[\omega]$ in $H^k_{\mathrm{dR}}(M)$.

A table of examples across degrees and manifolds:

Manifold	Degree	Closed form	Local potential?	Global potential?
$\mathbb{R}^n$	any	any closed ω	Yes	Yes ($b_k = 0$)
S^1	1	$d\theta$	Yes (locally)	No ($b_1 = 1$)
$\mathbb{R}^2 \setminus \{0\}$	1	angle form	Yes (locally)	No ($b_1 = 1$)
T^2	1	$a\,d\theta + b\,d\phi$	Yes (locally)	No ($b_1 = 2$)
S^2	2	area form	Yes (locally)	No ($b_2 = 1$)
$\mathbb{R}^3 \setminus \{0\}$	2	Coulomb flux	Yes (locally)	No ($b_2 = 1$)
T^3	2	$d\theta_i \wedge d\theta_j$	Yes (locally)	No ($b_2 = 3$)

5.6.4 Why de Rham Cohomology Is the Right Measure of Failure

The de Rham cohomology group $H^k_{\mathrm{dR}}(M)$ is not merely a collection of examples of the local-to-global failure; it is the *complete, canonical characterization* of all possible failure modes, for the following reasons.

Every failure is detected. If ω is not globally exact, then some period $\int_\Sigma \omega \neq 0$, and this period is nonzero in $H^k_{\mathrm{dR}}(M)$.

Equivalent failures are identified. Two forms ω and $\omega + d\alpha$ fail to be exact in exactly the same ways (they have the same periods over every cycle), so they represent the same element of $H^k_{\mathrm{dR}}(M)$.

The group structure counts independent failures. The dimension b_k counts how many independent directions the failure can take. Having $b_k = 2$ means there are two completely independent reasons a form might fail to be exact, and these reasons cannot be reduced to each other.

Topology, not geometry. The failure is purely topological: it depends only on the global shape of M (which loops are non-contractible, which surfaces are non-bounding), not on the metric, curvature, or local geometry of M. Deforming M without tearing or gluing does not change $H^k_{\mathrm{dR}}(M)$.

5.6.5 Foreshadowing: Chern Classes as Higher-Dimensional Obstructions

The local-to-global failure for scalar potentials (degree 1) and vector potentials (degree 2) is only the beginning. Chern classes generalize this obstruction to *fiber bundles*: instead of asking whether a closed form is the exterior derivative of a globally defined lower-degree form, we ask whether a fiber bundle is globally trivializable.

Consider a complex line bundle $L \to M$. Locally over each contractible patch U_α, the bundle is trivial: $L|_{U_\alpha} \cong U_\alpha \times \mathbb{C}$. A connection on L has local connection 1-forms A_α (local potentials). On overlaps $U_\alpha \cap U_\beta$, the transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to U(1)$ relate the two trivializations. The curvature 2-form $F_\alpha = dA_\alpha$ is locally exact (it is d of the local connection), but globally it may define a nonzero class in $H^2(M;\mathbb{R})$:

$$c_1(L) = \left[\frac{i}{2\pi} F\right] \in H^2(M;\mathbb{R}). \tag{5.13}$$

This is the **first Chern class**, and it measures the obstruction to *globally trivializing the bundle*: the failure of the locally-defined connection to patch together into a globally consistent one. The Chern class is nonzero if and only if the bundle is non-trivial, i.e., if and only if the curvature (local field strength) cannot be written as a globally exact form.

The pattern is identical to the degree-1 and degree-2 stories:

Setting	Local object	Global obstruction
Closed 1-form $\omega = d\phi$ locally	Angle function ϕ	$[\omega] \in H^1(M)$
Closed 2-form $F = dA$ locally	Vector potential A	$[F] \in H^2(M)$
Curvature $F = dA_\alpha$ locally	Global connection A	$c_1(L) \in H^2(M;\mathbb{Z})$

The first Chern class $c_1(L)$ is the complete obstruction to finding a globally consistent connection on L. When $c_1(L) \neq 0$, no global gauge can be chosen; when $c_1(L) = 0$, the bundle is trivial and a global gauge exists. In exactly the same way, a closed 1-form failing to be exact means no global

scalar potential exists, and a closed 2-form failing to be exact means no global vector potential exists.

For higher Chern classes $c_k(L) \in H^{2k}(M;\mathbb{Z})$, the story continues: each is an obstruction to a successively more refined global consistency condition, living in higher even-degree cohomology. De Rham cohomology is the home of all these obstructions; the Chern-Weil theorem (Chapter 9) gives their explicit construction from the curvature of any connection. The key insight carried from this chapter into the next: *every characteristic class is a cohomology class measuring how a local object fails to be global.*

5.7 Periods: The Coordinates on the Cohomology Group

The most concrete way to understand a cohomology class is through its **period integrals**. Given a closed k-form ω and a closed k-cycle γ (a closed k-dimensional submanifold or chain with no boundary), the period of ω around γ is

$$P(\omega, \gamma) = \int_\gamma \omega. \tag{5.14}$$

This integral depends only on the cohomology class $[\omega]$ and the homology class $[\gamma]$, not on the specific representatives:

- If we replace ω by $\omega + d\alpha$, the period changes by $\int_\gamma d\alpha = \int_{\partial\gamma} \alpha = 0$ (since γ is closed: $\partial\gamma = \emptyset$).
- If we replace γ by $\gamma + \partial\sigma$ (add a boundary), the period changes by $\int_{\partial\sigma} \omega = \int_\sigma d\omega = 0$ (since ω is closed).

So periods are well-defined pairings between cohomology classes and homology classes:

$$H^k_{\mathrm{dR}}(M) \times H_k(M;\mathbb{R}) \to \mathbb{R}, \qquad ([\omega], [\gamma]) \mapsto \int_\gamma \omega. \tag{5.15}$$

This is the **de Rham pairing**. Over $\mathbb{R}$, it is a perfect non-degenerate pairing: every element of H^k is completely determined by its periods against all k-cycles. This is one half of the de Rham theorem.

The periods are the coordinates. If $b_k = \dim H^k = r$, choose r independent cycles $\gamma_1, \ldots, \gamma_r$. The period vector $(P(\omega, \gamma_1), \ldots, P(\omega, \gamma_r))$ is the coordinate of $[\omega]$ in H^k. Two closed forms are cohomologous if and only if all their periods agree.

A closed form is exact if and only if all its periods vanish: $\omega = d\alpha$ if and only if $\int_\gamma \omega = 0$ for every closed cycle γ. This is the intrinsic meaning of the cohomology group: H^k is the vector space of period vectors.

Connection to the discrete spanning-tree argument. In Chapter 4, we found potential discrepancies on non-tree edges, and these discrepancies were exactly the period integrals of the cochain around the independent loops defined by those edges. The smooth version is the same: for each independent loop γ_i on the manifold, the period $\int_{\gamma_i} \omega$ is the obstruction datum for that loop. The Betti number b_k is the number of independent such loops (in homology).

5.8 Canonical Examples

5.8.1 Euclidean Space $\mathbb{R}^n$: No Holes

$\mathbb{R}^n$ is contractible, so the Poincaré lemma applies globally: every closed form is exact. All cohomology groups vanish except $H^0_{\mathrm{dR}}(\mathbb{R}^n) \cong \mathbb{R}$ (one connected component, spanned by the constant function $f = 1$).

No topological holes, no obstructions, all periods zero. The canonical example of trivial cohomology.

5.8.2 The Circle S^1: One Hole

The circle is the simplest space with non-trivial cohomology. It has one connected component and one independent non-contractible loop (going all the way around).

$H^0_{\mathrm{dR}}(S^1) \cong \mathbb{R}$, generated by the constant function.

$H^1_{\mathrm{dR}}(S^1) \cong \mathbb{R}$, generated by the **angle form** $d\theta$ (formally $(-y\,dx + x\,dy)/(x^2+y^2)$ in Cartesian coordinates). This form is closed ($d(d\theta) = 0$) but not exact (the angle θ is not a single-valued function on S^1). Its period around the circle is $\oint_{S^1} d\theta = 2\pi \neq 0$, confirming it represents a non-trivial class.

Any other closed 1-form ω on S^1 is cohomologous to $c\,d\theta$ for some $c \in \mathbb{R}$: the coefficient $c = \frac{1}{2\pi}\oint_{S^1} \omega$ is the period, and it completely determines the cohomology class. The group $H^1(S^1) \cong \mathbb{R}$ is parametrized by this one real number.

Betti numbers: $b_0 = 1$, $b_1 = 1$, $b_k = 0$ for $k \geq 2$. Euler characteristic: $\chi = 1 - 1 = 0$.

5.8.3 The Punctured Plane $\mathbb{R}^2 \setminus \{0\}$: Same as S^1

The punctured plane $\mathbb{R}^2 \setminus \{0\}$ deformation retracts onto the unit circle S^1 via the map $\mathbf{x} \mapsto \mathbf{x}/|\mathbf{x}|$. Since homotopy-equivalent spaces have isomorphic cohomology (by homotopy invariance, proved below), $H^k_{\mathrm{dR}}(\mathbb{R}^2 \setminus \{0\}) \cong H^k_{\mathrm{dR}}(S^1)$.

The generator of H^1 is the angle form $\eta = (-y\,dx + x\,dy)/(x^2+y^2)$. This is closed everywhere on the punctured plane, but not exact: $\oint_{S^1} \eta = 2\pi \neq 0$. The hole at the origin is the obstruction. The period around the unit circle is 2π, and the cohomology class $[\eta]$ is determined by this single period.

Visualizing the obstruction. Cover $\mathbb{R}^2 \setminus \{0\}$ with two patches: $U_+ = \{y > -\epsilon\}$ (upper region) and $U_- = \{y < \epsilon\}$ (lower region). On each patch, η is exact: $\eta = d(\arctan(y/x))$ on an appropriate branch. But the two branches disagree on the overlap $U_+ \cap U_-$ (which has two connected components, left and right of the origin): the angle function jumps by $\pm\pi$ as we pass through the left component. This jump, accumulated around the full circle, gives the period 2π.

5.8.4 The Two-Sphere S^2: A Hollow Ball

S^2 has one connected component, no non-contractible loops, and one independent enclosed void.

$H^0_{\mathrm{dR}}(S^2) \cong \mathbb{R}$ (one component).

$H^1_{\mathrm{dR}}(S^2) = 0$ (no holes in dimension 1: every loop on S^2 bounds a disk). This can be seen via the Mayer-Vietoris sequence: covering S^2 by two contractible patches (sphere minus a point in each hemisphere), the overlapping ring is homotopy equivalent to S^1, but the long exact sequence forces $H^1(S^2) = 0$.

$H^2_{\mathrm{dR}}(S^2) \cong \mathbb{R}$, generated by the **area form** $\omega_{\mathrm{area}} = \sin\theta\, d\theta \wedge d\phi$. This 2-form is closed (there is no $d_2\omega_{\mathrm{area}}$ that is nonzero since S^2 is 2-dimensional) and its integral over S^2 is $\int_{S^2} \omega_{\mathrm{area}} = 4\pi \neq 0$. Any closed 2-form ω on S^2 is cohomologous to $c\,\omega_{\mathrm{area}}$ where $c = (4\pi)^{-1} \int_{S^2} \omega$ is its period over S^2.

Betti numbers: $b_0 = 1$, $b_1 = 0$, $b_2 = 1$. Euler characteristic: $\chi = 1 - 0 + 1 = 2$.

5.8.5 The Torus T^2: Two Independent Loops

$T^2 = S^1 \times S^1$ has one connected component, two independent non-contractible loops (the longitude and the meridian), and one fundamental class.

$H^0_{\mathrm{dR}}(T^2) \cong \mathbb{R}$ (one component).

$H^1_{\mathrm{dR}}(T^2) \cong \mathbb{R}^2$, generated by $d\theta$ and $d\phi$ (the angle forms in the two circular directions). These two forms are closed but not cohomologous (their periods around the two independent loops are different), and together they span the 2-dimensional cohomology group. A closed 1-form ω on T^2 is cohomologous to $a\,d\theta + b\,d\phi$ where $a = \frac{1}{2\pi} \oint_{\theta\text{-circle}} \omega$ and $b = \frac{1}{2\pi} \oint_{\phi\text{-circle}} \omega$ are its two periods.

$H^2_{\mathrm{dR}}(T^2) \cong \mathbb{R}$, generated by $d\theta \wedge d\phi$.

Betti numbers: $b_0 = 1$, $b_1 = 2$, $b_2 = 1$. Euler characteristic: $\chi = 1 - 2 + 1 = 0$.

Matching the discrete computation. In Chapter 4, the triangulated torus had $N_0 = 9$, $N_1 = 27$, $N_2 = 18$, with ranks $\operatorname{rank} \mathbf{D}_0 = 8$ and $\operatorname{rank} \mathbf{D}_1 = 17$, giving $b_1 = 27 - 17 - 8 = 2$. The two non-tree edges in the spanning tree argument correspond exactly to the two generators $d\theta$ and $d\phi$ of H^1, and their periods are the two independent obstruction numbers.

Space	b_0	b_1	b_2	χ	Generators of H^1
$\mathbb{R}^n$	1	0	0	1	none
S^1	1	1	0	0	$d\theta$
$\mathbb{R}^2 \setminus \{0\}$	1	1	0	–	$(-y\,dx + x\,dy)/(x^2+y^2)$
S^2	1	0	1	2	none
T^2	1	2	1	0	$d\theta$, $d\phi$

5.9 Homotopy Invariance

Two smooth maps $f_0, f_1 : M \to N$ are **smoothly homotopic** if there exists a smooth map $F : M \times [0,1] \to N$ with $F(\cdot, 0) = f_0$ and $F(\cdot, 1) = f_1$. A smooth homotopy F induces a chain homotopy between the cochain maps f_0^* and f_1^* (pulling back forms), which implies $f_0^* = f_1^*$ on cohomology: homotopic maps induce the same map on H^k_{dR}.

Corollary 5.2 (Homotopy invariance)**.** *If M and N are smoothly homotopy equivalent (each is a deformation retract of the other), then $H^k_{\mathrm{dR}}(M) \cong H^k_{\mathrm{dR}}(N)$ for all k.*

This is why $\mathbb{R}^2 \setminus \{0\}$ has the same cohomology as S^1: the retraction $\mathbf{x} \mapsto \mathbf{x}/|\mathbf{x}|$ is a homotopy equivalence. Cohomology is insensitive to stretching, squashing, or any deformation that does not tear or glue the space: it is a topological invariant.

5.10 The Čech Perspective: Gluing and Disagreements

The period-integral picture makes cohomology feel like a measurement of global integrals. The Čech picture makes it feel like a measurement of gluing failures. Both describe the same objects; each has computational advantages.

Cover M by contractible open sets $\{U_\alpha\}$. Given a closed k-form ω, the Poincaré lemma gives a local potential A_α on each U_α with $dA_\alpha = \omega$. On each overlap $U_\alpha \cap U_\beta$, both potentials are valid, so:

$$d(A_\alpha - A_\beta) = \omega - \omega = 0 \quad \text{on } U_\alpha \cap U_\beta. \tag{5.16}$$

The difference $A_\alpha - A_\beta$ is a closed $(k-1)$-form on the overlap. If the overlap is contractible, this is exact: $A_\alpha - A_\beta = d\Lambda_{\alpha\beta}$ for some $(k-2)$-form $\Lambda_{\alpha\beta}$.

A global potential A (with $dA = \omega$) exists if and only if we can choose the local potentials A_α so that all differences $A_\alpha - A_\beta$ vanish. Adjusting A_α by a closed form on U_α changes the differences; the question is whether we can adjust them all to zero simultaneously. The obstruction to doing so is the cohomology class $[\omega]$: it vanishes if and only if all the differences can be made to vanish.

For 1-forms in the circle example: Cover S^1 with $U_+ = \{y > -\epsilon\}$ and $U_- = \{y < \epsilon\}$. On each, $d\theta = d(\text{local angle})$. On the overlap (two arcs), the two angle branches differ by a constant: $\theta_+ - \theta_- = 0$ on the right arc and $\theta_+ - \theta_- = -2\pi$ on the left arc. These constant differences encode the period 2π. The cohomology class lives in these constants: changing the local potentials (by constants) can eliminate one disagreement but not both simultaneously.

5.11 The de Rham Theorem

The de Rham theorem unifies the smooth and discrete pictures into a single statement.

Theorem 5.3 (de Rham, 1931)**.** *Let M be a smooth manifold and K any triangulation of it. The integration map*

$$I_k : \Omega^k(M) \to C^k(K;\mathbb{R}), \qquad (I_k\omega)(\sigma) = \int_\sigma \omega \tag{5.17}$$

(integrating ω over each k-simplex σ) induces an isomorphism on cohomology:

$$H^k_{\mathrm{dR}}(M) \xrightarrow{\sim} H^k(K;\mathbb{R}). \tag{5.18}$$

Every topological hole detectable by smooth differential forms is detected by the discrete simplicial theory, and vice versa. The Betti numbers b_k are the same from both perspectives.

The bridge is Stokes' theorem: $I_{k+1}(d\omega)(\tau) = \int_\tau d\omega = \int_{\partial\tau} \omega = (D_k I_k \omega)(\tau)$. This says the integration map commutes with the exterior derivatives on both sides — it is a **cochain map**. Such a map induces a map on cohomology. The de Rham theorem asserts this induced map is an isomorphism.

What this means in practice: to compute de Rham cohomology groups of a compact manifold, triangulate it and compute the ranks of the coboundary matrices. No analysis, only linear algebra. Conversely, the smooth theory gives an explicit representative (a differential form) for each cohomology class, and integration computes the period data.

5.12 The Mayer-Vietoris Sequence

The Mayer-Vietoris sequence is the main computational tool for computing cohomology by cutting a space into simpler pieces.

If $M = U \cup V$ with U, V open, the restriction maps and the connecting homomorphism fit into the exact sequence:

$$\cdots \to H^k(M) \xrightarrow{(r_U, r_V)} H^k(U) \oplus H^k(V) \xrightarrow{r_U - r_V} H^k(U \cap V) \xrightarrow{\delta} H^{k+1}(M) \to \cdots \tag{5.19}$$

Exactness means: image of each map equals kernel of the next. This constrains the cohomology of M given the cohomologies of U, V, and $U \cap V$.

Application: $H^2(S^2)$. Cover S^2 by $U = S^2 \setminus \{\text{south pole}\}$ and $V = S^2 \setminus \{\text{north pole}\}$, both contractible. Their intersection $U \cap V$ is homotopy equivalent to S^1. The relevant piece of the sequence is:

$$H^1(U) \oplus H^1(V) \to H^1(U \cap V) \xrightarrow{\delta} H^2(S^2) \to H^2(U) \oplus H^2(V). \tag{5.20}$$

Since U and V are contractible, $H^1(U) = H^1(V) = H^2(U) = H^2(V) = 0$. So $\delta : H^1(S^1) \xrightarrow{\sim} H^2(S^2)$ is an isomorphism. The generator of $H^2(S^2)$ comes from the generator $[d\theta]$ of $H^1(S^1)$ via the connecting homomorphism, which "lifts" the loop to a global 2-class. Explicitly, the generator of $H^2(S^2)$ is the area form $\sin\theta \, d\theta \wedge d\phi$.

5.13 Physics Examples: Cohomology in Action

5.13.1 Aharonov-Bohm Effect: H^1 of a Punctured Region

In the Aharonov-Bohm experiment, an electron travels around a solenoid. Outside the solenoid, the magnetic field $\mathbf{B} = 0$, but the vector potential A is non-zero and satisfies $d_1 A = F = 0$ (closed 1-form). The question is whether A is exact: can we find a global function Λ with $A = d_0 \Lambda$?

The domain is $\mathbb{R}^3 \setminus \{\text{solenoid axis}\}$, which deformation retracts to S^1 (a circle around the solenoid). Since $H^1(S^1) \cong \mathbb{R}$, closed 1-forms on this domain need not be exact: they can have a non-zero period around the solenoid. The period is $\oint_{S^1} A = \Phi/\hbar$ where Φ is the magnetic flux through the solenoid.

When the electron goes around the solenoid, it accumulates a phase $e^{i\Phi/\hbar}$. This is the holonomy of the connection A around the non-contractible loop. The cohomology class $[A] \in H^1(\mathbb{R}^3 \setminus \text{axis})$ determines the phase; if $[A] = 0$ (exact), the phase is trivial and no effect is observed. The observable effect arises because $H^1 \neq 0$ and $[A] \neq 0$.

The Aharonov-Bohm effect is a *direct physical measurement of a cohomology class*: the interference pattern is shifted by $e^{i\Phi/\hbar}$, which is the period of $[A]$ around the fundamental loop, and it changes by 2π when Φ changes by $h = 2\pi\hbar$ (one unit of flux quantum), reflecting the periodicity of $H^1(S^1; \mathbb{Z}) \cong \mathbb{Z}$.

5.13.2 Magnetic Monopoles: H^2 of Punctured Space

A magnetic monopole at the origin creates a magnetic 2-form $B = \frac{g}{4\pi r^2}(x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy)$ on $\mathbb{R}^3 \setminus \{0\}$.

This satisfies $d_2 B = 0$ (the 3-form divergence of $\mathbf{B}$ vanishes away from the origin), so B is a closed 2-form. The question is whether B is exact: does a global vector potential A exist with $d_1 A = B$?

The domain $\mathbb{R}^3 \setminus \{0\}$ deformation retracts to S^2, and $H^2(S^2) \cong \mathbb{R}$. The period of B over S^2 is $\int_{S^2} B = g$. If $g \neq 0$, then $[B] \neq 0$ in H^2, so B is not exact: no global vector potential exists.

This is the Dirac string: to define A globally, we must remove a line (the Dirac string) from $\mathbb{R}^3$, making the domain simply connected. On $\mathbb{R}^3 \setminus \{\text{string}\}$, B becomes exact (the domain is contractible). The price is a singular gauge potential along the string.

The Dirac quantization condition $qg \in 2\pi\mathbb{Z}$ comes from requiring that $[B]$ lie in the integer lattice of $H^2(S^2; \mathbb{R})$: the cohomology class must be integral for the wave function to be single-valued. This is a restriction to $H^2(S^2; \mathbb{Z})$, the integer cohomology. Magnetic charge is quantized because cohomology classes that arise from bundles must be integral.

5.13.3 Berry Phases: H^1 of Parameter Space

A quantum system with Hamiltonian $H(\lambda)$ depending on parameters $\lambda \in \mathcal{P}$ defines a vector bundle of ground states over $\mathcal{P}$. The Berry connection

$$\mathcal{A} = i\langle \psi(\lambda) | d\psi(\lambda) \rangle \tag{5.21}$$

is a 1-form on $\mathcal{P}$ taking values in $i\mathbb{R}$ (since the wavefunction $|\psi\rangle$ is defined up to phase). Under a phase redefinition $|\psi\rangle \mapsto e^{i\chi(\lambda)}|\psi\rangle$, the Berry connection changes by $\mathcal{A} \mapsto \mathcal{A} - d\chi$: a gauge transformation. The Berry curvature $\mathcal{F} = d\mathcal{A}$ is gauge-invariant and represents a cohomology class in $H^2(\mathcal{P})$.

When the parameters traverse a closed loop $\gamma \subset \mathcal{P}$, the accumulated phase is the Berry phase $\exp\left(i \oint_\gamma \mathcal{A}\right)$, the holonomy of the Berry connection. This is nonzero if and only if $[\mathcal{A}] \in H^1(\mathcal{P})$ is non-trivial.

For a spin-$\frac{1}{2}$ particle in a magnetic field $\mathbf{B}$ of varying direction, the parameter space is S^2 (the sphere of field directions). The Berry curvature is the area form of S^2 (up to a factor), and the first Chern number is $C_1 = \frac{1}{2\pi}\int_{S^2} \mathcal{F} = \pm\frac{1}{2}$. A loop enclosing a solid angle Ω gives Berry phase $\exp(i\Omega/2)$. When the parameters traverse the full sphere, the phase is $\exp\left(i \cdot 2\pi \cdot \frac{1}{2}\right) = e^{i\pi} = -1$: the spinor flips sign under a full rotation of the field, the familiar half-integer statistics.

5.13.4 The Quantum Hall Effect: H^2 of the Brillouin Zone

In the integer quantum Hall effect, electrons in a 2D crystal with periodic potential fill Bloch bands $|u_n(\mathbf{k})\rangle$ over the Brillouin zone torus T^2 (parametrized by crystal momentum $\mathbf{k}$). The Berry connection of the occupied bands is

$$\mathcal{A}_n = i\langle u_n(\mathbf{k}) | d | u_n(\mathbf{k}) \rangle, \qquad \mathcal{F}_n = d\mathcal{A}_n. \tag{5.22}$$

The **first Chern number** of the n-th band is

$$C_n = \frac{1}{2\pi}\int_{T^2} \mathcal{F}_n \in \mathbb{Z}. \tag{5.23}$$

This is an integer because the Chern number is the period of the closed 2-form $\mathcal{F}_n$ over the closed 2-manifold T^2, and the underlying bundle structure forces the period to be integral (the class lives in $H^2(T^2; \mathbb{Z})$, the integer cohomology).

The TKNN formula says the Hall conductivity is $\sigma_{xy} = \frac{e^2}{h}\sum_n C_n$: a topological invariant that cannot change continuously. It changes only when the gap closes (a topological phase transition), and it is robust against disorder, interactions, and all local perturbations that preserve the gap — because it is a cohomological invariant.

5.14 Summary: What De Rham Cohomology Computes

De Rham cohomology is the systematic, coordinate-free theory of topological obstruction for smooth manifolds. Each piece of the structure has a precise meaning:

Object	Meaning	Example
Closed k-form	Conserved measurement, zero source	$dB = 0$ (Gauss)
Exact k-form	Has a $(k-1)$-potential globally	$B = dA$ (no monopole)
Cohomology class	Orbit under adding exact forms	Gauge equivalence
$H^k(M)$	Vector space of independent obstructions	$H^2(S^2) = \mathbb{R}$
Betti number b_k	Number of independent k-holes	$b_1(T^2) = 2$
Period integral	Coordinate of the cohomology class	$\oint d\theta = 2\pi$
Chern class	Integer cohomology class from bundle	$c_1 \in H^2(M;\mathbb{Z})$

The unifying theme is the same obstruction principle traced through Chapters 2, 3, and 4: a closed form is locally exact everywhere, but the local potentials can fail to glue globally. The failure is measured by period integrals around non-contractible cycles. The de Rham cohomology group $H^k(M)$ is the space of all independent obstructions in degree k, organized as a real vector space whose dimension is the Betti number b_k and whose elements are equivalence classes of closed forms under the gauge action of adding exact forms.

Chapter 6

Hodge Theory

6.1 The Problem Hodge Theory Solves

De Rham cohomology gave us a way to classify topological holes: the group $H^k_{\mathrm{dR}}(M)$ is the quotient of closed k-forms by exact k-forms, and its dimension b_k counts the independent k-dimensional holes. But this quotient leaves us with a puzzle.

Each cohomology class is an entire infinite-dimensional orbit: starting from one closed form ω, the orbit is all forms $\omega + d\alpha$ for arbitrary $(k-1)$-forms α. These forms all represent the same topological class, but they can look very different pointwise. When we want to *compute* something concrete — an integral, an inner product, a spectral invariant — we must pick one representative from the orbit. But which one?

This is the problem Hodge theory solves: on a compact Riemannian manifold, there is a **canonical** representative in every cohomology class — the one with the smallest L^2 norm. This **harmonic form** is the "most spread out", "most symmetric", "quietest" representative of its class. It is the form that fits the geometry of the manifold as smoothly as possible while still encoding the correct topological information.

6.1.1 Why the Problem Matters

The existence of canonical representatives has concrete consequences:

Betti numbers from analysis. Each Betti number b_k is a topological integer, but Hodge theory computes it analytically: $b_k = \dim \ker \Delta_k$, the dimension of the kernel of the Hodge Laplacian on k-forms. Topology becomes a question about solutions to a differential equation.

Poincaré duality. The harmonic representative of a k-class on a compact n-manifold maps, under the Hodge star $\star$, to a harmonic $(n-k)$-form, which represents a class in H^{n-k}. This immediately gives $H^k \cong H^{n-k}$: the symmetry of Betti numbers is a consequence of the metric duality between k-forms and $(n-k)$-forms.

Connection to physics. In gauge theory, the canonical representative of a curvature class corresponds to the unique "best" gauge field in its gauge orbit. In TQFT, path integrals localize on harmonic forms, reducing infinite-dimensional integrals to finite-dimensional ones over $\mathcal{H}^k$.

Bridge to the index theorem. The Hodge theorem says $\mathcal{H}^k \cong H^k_{\mathrm{dR}}$: the analytic object (harmonic forms, kernel of the Laplacian) equals the topological object (cohomology classes). This identification is the prototype for the Atiyah-Singer index theorem, which equates the analytic index of any elliptic operator with a topological invariant expressed in characteristic classes.

6.1.2 The Position in the Story

Looking back at the preceding chapters, each one answered a question the previous one raised:

Chapter 2 asked: when is a closed form exact? Chapter 5 (de Rham cohomology) organized the obstruction into a group H^k. Hodge theory (this chapter) asks: among all forms in a cohomology class, which is the best representative? The answer requires the metric, which is why Hodge theory is the first chapter in this book that genuinely uses geometric (not just topological) structure.

6.2 The Hodge Star: Measuring the Complement

6.2.1 The Geometric Idea

In an n-dimensional oriented inner-product space, every k-dimensional subspace has a unique $(n-k)$-dimensional orthogonal complement. A k-form measures densities on k-dimensional subspaces; its Hodge dual measures densities on the complementary $(n-k)$-dimensional subspaces.

The most familiar case is $\mathbb{R}^3$ ($n=3$). A 1-form ω measures something along a line (1-dimensional). Its Hodge dual $\star\omega$ is a 2-form measuring something through the plane perpendicular to that line (2-dimensional). This is the duality between a vector and the plane perpendicular to it: the electric field $\mathbf{E}$ (a 1-form along field lines) is dual to the flux form (a 2-form through surfaces perpendicular to the field lines).

Concretely in $\mathbb{R}^3$: the x-direction is a 1-dimensional subspace. Its orthogonal complement is the yz-plane, a 2-dimensional subspace. The Hodge star converts dx (measuring flow along x) into $dy \wedge dz$ (measuring area in the yz-plane):

$$\star dx = dy \wedge dz, \quad \star dy = dz \wedge dx, \quad \star dz = dx \wedge dy. \tag{6.1}$$

And the volume form converts back:

$$\star(dy \wedge dz) = dx, \quad \star(dz \wedge dx) = dy, \quad \star(dx \wedge dy) = dz, \quad \star(dx \wedge dy \wedge dz) = 1. \tag{6.2}$$

The full table for $\mathbb{R}^3$ (standard orientation):

ω	$\star\omega$	ω	$\star\omega$
1	$dx \wedge dy \wedge dz$	$dx \wedge dy \wedge dz$	1
dx	$dy \wedge dz$	$dy \wedge dz$	dx
dy	$dz \wedge dx$	$dz \wedge dx$	dy
dz	$dx \wedge dy$	$dx \wedge dy$	dz

The pattern: $\star$ rotates each basis form into the "orthogonal complement" basis form, with a sign given by whether the resulting permutation is even or odd.

6.2.2 Definition

Let (M, g) be a compact oriented n-dimensional Riemannian manifold. The metric g provides a pointwise inner product on each $\Lambda^k T_p^* M$, and defines the Riemannian volume form $\mathrm{vol}_g = \sqrt{\det g_{ij}}\, dx^1 \wedge \cdots \wedge dx^n$.

The **Hodge star** $\star : \Omega^k(M) \to \Omega^{n-k}(M)$ is the unique operator satisfying

$$\alpha \wedge \star\beta = \langle \alpha, \beta \rangle_g \, \mathrm{vol}_g \qquad \text{for all } \alpha, \beta \in \Omega^k(M). \tag{6.3}$$

In words: $\star\beta$ is the $(n-k)$-form that, when wedged with β itself, gives $|\beta|^2$ times the volume form. The star converts a k-form into the $(n-k)$-form that fills out the rest of the volume.

In orthonormal coordinates $\{dx^1, \ldots, dx^n\}$:

$$\star(dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = \epsilon_{i_1 \cdots i_k j_1 \cdots j_{n-k}}\, dx^{j_1} \wedge \cdots \wedge dx^{j_{n-k}}, \tag{6.4}$$

where $\{j_1, \ldots, j_{n-k}\}$ are the complementary indices and ϵ is the sign of the permutation $(i_1, \ldots, i_k, j_1, \ldots, j_{n-k})$.

6.2.3 Double Application: $\star\star = \pm 1$

Applying the Hodge star twice returns the original form up to a sign:

$$\star \star \omega = (-1)^{k(n-k)}\omega \qquad \text{for } \omega \in \Omega^k(M). \tag{6.5}$$

In $\mathbb{R}^3$ ($n = 3$): on 1-forms, $\star\star = (-1)^{1\cdot 2} = +1$ (so $\star(\star dx) = dx$, as confirmed by the table). On 2-forms, $\star\star = (-1)^{2\cdot 1} = -1$.

In $\mathbb{R}^4$ ($n = 4$, Euclidean): on 2-forms, $\star\star = (-1)^{2\cdot 2} = +1$. So $\star$ is an involution on 2-forms, with eigenvalues ± 1. The $+1$ eigenspace consists of **self-dual** 2-forms ($\star\omega = +\omega$); the -1 eigenspace consists of **anti-self-dual** forms ($\star\omega = -\omega$). This splitting is central to instanton theory.

6.2.4 The L^2 Inner Product

From equation (6.3), the L^2 inner product on forms is:

$$(\alpha, \beta)_{L^2} = \int_M \alpha \wedge \star\beta. \tag{6.6}$$

This integral makes sense because $\alpha \wedge \star\beta$ is an n-form (integrable over an n-manifold). For $\alpha = \beta$, it gives the squared L^2 norm $\|\omega\|^2_{L^2} = \int_M \omega \wedge \star\omega \geq 0$, with equality only if $\omega = 0$.

Why the metric is necessary. The Hodge star uses the metric twice: to define the inner product $\langle \alpha, \beta \rangle_g$ and to define the volume form vol_g. Without a metric, there is no canonical way to pair k-forms with $(n-k)$-forms and no canonical way to define a norm on forms. This is why the Hodge theorem is not a purely topological result: it depends on the choice of Riemannian metric g. The cohomology groups H^k are topological; the harmonic representatives within each class are geometric.

6.3 Electromagnetism as Hodge Duality Made Explicit

Maxwell's equations in vacuum, written in the language of differential forms, become strikingly symmetric:

$$dF = 0, \qquad d \star F = 0, \tag{6.7}$$

where the 2-form F encodes both electric and magnetic fields. These two equations are related by the Hodge star: the second is the first applied to $\star F$ in place of F.

The symmetry $F \leftrightarrow \star F$ is **electromagnetic duality**: it exchanges electric fields $\mathbf{E}$ with magnetic fields $\mathbf{B}$ (up to signs and factors). In the standard formulation of Maxwell's theory, electric and magnetic fields are treated asymmetrically (the electric field is polar, the magnetic is axial). The Hodge formulation reveals the symmetry that exists in vacuum.

Gauge invariance is also transparent. The equation $dF = 0$ means F is a closed 2-form. On any contractible region, by the Poincaré lemma, $F = d_1 A$ for a local 1-form potential A. Changing $A \to A + d_0\Lambda$ leaves F unchanged (by $d^2 = 0$). The cohomology class $[F] \in H^2_{\mathrm{dR}}(M)$ is the gauge-invariant content; different gauge choices A are different representatives of the same class.

Connection to de Rham cohomology. If M has non-trivial H^2 (e.g., if $M = \mathbb{R}^3 \setminus \{\text{origin}\}$ with $H^2 \cong \mathbb{R}$), then F can be closed but not exact, corresponding to a magnetic monopole. The cohomology class $[F]$ is the charge. If $H^2(M) = 0$ (e.g., $M = \mathbb{R}^3$), then every closed F is exact and a global potential always exists. Hodge theory refines this: among all potentials A with $dA = F$, the one satisfying the **Lorenz gauge condition** $d^*A = 0$ is the harmonic representative in the gauge orbit, minimizing the L^2 norm $\|A\|_{L^2}$ within its gauge class.

6.4 The Codifferential: d Running Backwards

Given the L^2 inner product on forms, the exterior derivative d has a formal adjoint d^*, which runs in the opposite direction: $d^* : \Omega^k \to \Omega^{k-1}$. It is defined by the adjointness relation

$$(d\alpha, \beta)_{L^2} = (\alpha, d^*\beta)_{L^2} \tag{6.8}$$

for all compactly supported $\alpha \in \Omega^{k-1}$, $\beta \in \Omega^k$. Integrating by parts (using Stokes' theorem on a compact manifold without boundary) gives the explicit formula:

$$d^* = (-1)^{n(k+1)+1} \star d \star \quad \text{(on } k\text{-forms in dimension } n\text{)}. \tag{6.9}$$

In $\mathbb{R}^3$:

- On 1-forms: $d^*(A\,dx + B\,dy + C\,dz) = -\left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right)$, the negative divergence.
- On 2-forms: $d^*(A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy)$ gives the curl of (A, B, C), translated to a 1-form.

Just as d measures the exterior derivative in *increasing* dimension (boundary-effect to bulk-source), d^* measures the "interior derivative" in *decreasing* dimension (divergence of a flux back to the source it came from).

Nilpotency of d^*. Since $d^2 = 0$, we have $(d^*)^2 = 0$ as well:

$$(d^*)^2 = (\star d \star)(\star d \star) = \pm \star\, d(\star\star) d \star = \pm \star\, d^2 \star = 0. \tag{6.10}$$

The codifferential defines a second cochain complex, running backwards:

$$0 \longleftarrow \Omega^0 \xleftarrow{d^*} \Omega^1 \xleftarrow{d^*} \Omega^2 \xleftarrow{d^*} \cdots \xleftarrow{d^*} \Omega^n \longleftarrow 0. \tag{6.11}$$

Its "cohomology" is isomorphic to the de Rham cohomology via $\star$ (by Poincaré duality). The two complexes (d going up, d^* going down) are dual to each other, and their combination in the Hodge Laplacian captures all topological information.

Coclosed forms. A k-form ω is **coclosed** if $d^*\omega = 0$. For a 1-form in $\mathbb{R}^3$: coclosed means divergence-free. For a 2-form: coclosed means curl-free (in the appropriate sense). Being simultaneously closed ($d\omega = 0$) and coclosed ($d^*\omega = 0$) is the harmonic condition.

6.5 The Hodge Laplacian: Adding and Subtracting Dimensions

6.5.1 Definition and Basic Properties

The **Hodge Laplacian** on k-forms is:

$$\Delta_k = d_{k-1}d_k^* + d_{k+1}^* d_k : \Omega^k(M) \to \Omega^k(M). \tag{6.12}$$

It combines two operations: $d_k^* d_k$ (take the exterior derivative, then take the co-derivative back down) and $d_{k-1} d_k^*$ (take the co-derivative down, then take the exterior derivative back up). Together, they measure all ways that ω can vary in either dimension direction.

On functions ($k = 0$): $\Delta_0 f = d^* df = -\mathrm{div}(\nabla f) = -\nabla^2 f$. This is the classical Laplacian (up to sign convention), confirming the intuition: the Hodge Laplacian on forms is the correct generalization of the ordinary Laplacian to differential forms of all degrees.

Non-negativity: for any compact k-form ω,

$$(\Delta_k \omega, \omega)_{L^2} = \|d_k \omega\|_{L^2}^2 + \|d_k^* \omega\|_{L^2}^2 \geq 0. \tag{6.13}$$

The Hodge Laplacian is non-negative. Its kernel consists of forms for which both $\|d\omega\| = 0$ and $\|d^*\omega\| = 0$: forms that are both closed and coclosed.

6.5.2 Harmonic Forms and Why They Are Special

A k-form is **harmonic** if $\Delta_k \omega = 0$, equivalently if $d\omega = 0$ and $d^*\omega = 0$ simultaneously. The space of harmonic k-forms is denoted $\mathcal{H}^k(M)$.

Why is this special? A closed form ω (with $d\omega = 0$) represents a cohomology class in H^k. But there are infinitely many forms in the same class: $\omega + d\alpha$ for any $(k-1)$-form α. Among all these representatives, the harmonic one is the unique element of the class satisfying $d^*\omega = 0$ as well.

The minimum-norm property. For any exact form $d\alpha$:

$$\|\omega + d\alpha\|_{L^2}^2 = \|\omega\|_{L^2}^2 + 2(\omega, d\alpha)_{L^2} + \|d\alpha\|_{L^2}^2. \tag{6.14}$$

If ω is harmonic ($d^*\omega = 0$), then $(\omega, d\alpha)_{L^2} = (d^*\omega, \alpha)_{L^2} = 0$. So $\|\omega + d\alpha\|_{L^2}^2 = \|\omega\|_{L^2}^2 + \|d\alpha\|_{L^2}^2 \geq \|\omega\|_{L^2}^2$. The harmonic representative has the minimum L^2 norm in its cohomology class.

Uniqueness. If ω and ω' are both harmonic and cohomologous ($\omega - \omega' = d\alpha$), then $d^*(\omega - \omega') = 0$ and $d(\omega - \omega') = 0$, so $\omega - \omega' = d\alpha$ is both exact and harmonic. But $\|d\alpha\|_{L^2}^2 = (d\alpha, d\alpha)_{L^2} = (\alpha, d^* d\alpha)_{L^2}$; combined with $\Delta(d\alpha) = 0$ implies $d\alpha = 0$. So $\omega = \omega'$: the harmonic representative is unique.

Intuition: the least-oscillatory representative. Among all forms in a cohomology class, the harmonic form is the one that "does the least work". Adding $d\alpha$ to a form introduces oscillations (since $d\alpha$ vanishes on all cohomology classes but has nonzero variation). The harmonic form has none of these extra oscillations: its variation is the minimum required by the topology.

For a 1-form on S^1, the cohomology class in $H^1(S^1)$ is determined by its period $\oint d\theta$. The harmonic representative is the constant form $c\, d\theta$ with $c = \frac{1}{2\pi} \oint \omega$. It is the "smoothest" form with the correct period: uniform, no bumps, no oscillations.

6.6 The Hodge Decomposition Theorem

Theorem 6.1 (Hodge Decomposition)**.** *Let (M, g) be a compact oriented Riemannian manifold without boundary. Every smooth k-form $\omega \in \Omega^k(M)$ decomposes uniquely and orthogonally as*

$$\omega = \underbrace{d\alpha}_{\text{exact part}} + \underbrace{d^*\beta}_{\text{co-exact part}} + \underbrace{\gamma}_{\text{harmonic part}}, \tag{6.15}$$

where $\alpha \in \Omega^{k-1}$, $\beta \in \Omega^{k+1}$, and $\gamma \in \mathcal{H}^k$. The three pieces are L^2-orthogonal, and the harmonic part γ is the unique harmonic representative of the cohomology class $[\omega] \in H^k_{\mathrm{dR}}(M)$.

6.6.1 What the Three Parts Mean

The decomposition splits every k-form into three independent, orthogonal components, each with a distinct geometric character.

The exact part $d\alpha$ is the component that is a gradient: it came from a $(k-1)$-form α. Integrating $d\alpha$ around any closed k-cycle gives zero (it is exact, so has zero periods). It is the "locally generated" part, the part with no topological content.

The co-exact part $d^*\beta$ is the component that is a divergence: it came from a $(k+1)$-form β via the co-differential. Integrating $d^*\beta$ against any closed cycle also gives a computable answer, but this part contributes nothing to cohomology either (it is in the image of d^*, which is L^2-orthogonal to closed forms).

The harmonic part γ is the remainder. It is closed ($d\gamma = 0$) and coclosed ($d^*\gamma = 0$). It cannot be further decomposed: it is the irreducible topological residue of ω after removing all the "generated" parts. Its cohomology class $[\gamma] = [\omega]$ is the topological content of ω.

In classical vector calculus on $\mathbb{R}^3$, for a vector field $\mathbf{F}$ (1-form):

- Exact part: ∇f (gradient of a scalar potential).
- Co-exact part: $\nabla \times \mathbf{A}$ (curl of a vector potential).
- Harmonic part: constant fields, or fields encoding topological structure.

This is the Helmholtz decomposition: any vector field decomposes into a gradient, a curl, and a harmonic remainder. Hodge theory is the generalization to forms of any degree on any Riemannian manifold.

6.6.2 Connection to de Rham Cohomology

The Hodge decomposition (6.15) sharpens the de Rham picture considerably. In de Rham theory:

$$\Omega^k / B^k \cong H^k_{\mathrm{dR}}, \tag{6.16}$$

where $B^k = \mathrm{im}(d_{k-1})$ is the space of exact forms. The Hodge theorem provides a canonical splitting of this quotient:

$$\Omega^k = B^k \oplus (d^*\Omega^{k+1}) \oplus \mathcal{H}^k, \tag{6.17}$$

an orthogonal direct sum (no quotient needed). The harmonic subspace $\mathcal{H}^k$ is a *canonical linear subspace* of Ω^k that represents the quotient H^k_{dR} without any choice:

$$\mathcal{H}^k(M) \cong H^k_{\mathrm{dR}}(M). \tag{6.18}$$

Every cohomology class has exactly one harmonic representative. The abstract quotient becomes a concrete subspace.

Dimension count. Since $\mathcal{H}^k \cong H^k_{\mathrm{dR}}$:

$$\dim \mathcal{H}^k = b_k = \dim H^k_{\mathrm{dR}}(M). \tag{6.19}$$

On a compact connected manifold: $\mathcal{H}^0$ consists of constant functions (one-dimensional, confirming $b_0 = 1$). On S^1: $\mathcal{H}^1 = \mathrm{span}(d\theta)$ (one-dimensional, confirming $b_1 = 1$). On T^2: $\mathcal{H}^1 = \mathrm{span}(d\theta, d\phi)$ (two-dimensional, confirming $b_1 = 2$).

6.6.3 Proof Sketch via Elliptic Operator Theory

The key analytic input is that Δ_k is an **elliptic operator**: its highest-order symbol is positive-definite. For elliptic operators on compact manifolds, the Fredholm alternative guarantees:

- The kernel $\ker \Delta_k = \mathcal{H}^k$ is finite-dimensional.
- The range $\operatorname{im} \Delta_k$ is closed with finite codimension.
- $\Omega^k = \ker \Delta_k \oplus \operatorname{im} \Delta_k$.

The identity $\Delta = dd^* + d^*d$ then shows that $\operatorname{im} \Delta \subseteq \operatorname{im}(d) + \operatorname{im}(d^*)$, and the L^2-orthogonality of $\operatorname{im}(d)$, $\operatorname{im}(d^*)$, and $\ker \Delta$ completes the decomposition.

The finiteness of $\ker \Delta_k$ is what makes the Betti numbers finite: even though Ω^k is infinite-dimensional, the space of harmonic forms (the topological part) is finite-dimensional.

6.7 Hodge Theory and Poincaré Duality

Theorem 6.2 (Poincaré Duality)**.** *For a compact oriented n-manifold without boundary, $H^k_{\mathrm{dR}}(M) \cong H^{n-k}_{\mathrm{dR}}(M)$ for all k.*

Proof via harmonic forms. The Hodge star $\star : \Omega^k \to \Omega^{n-k}$ maps harmonic forms to harmonic forms: if $d\omega = 0$ and $d^*\omega = 0$, then $d^*(\star\omega) = \pm \star (d\omega) = 0$ and $d(\star\omega) = \pm \star (d^*\omega) = 0$. So $\star : \mathcal{H}^k \to \mathcal{H}^{n-k}$ is an isomorphism (since $\star\star = \pm\mathrm{id}$). Via $\mathcal{H}^k \cong H^k_{\mathrm{dR}}$ and $\mathcal{H}^{n-k} \cong H^{n-k}_{\mathrm{dR}}$, we obtain $H^k \cong H^{n-k}$. $\square$

What this means. Poincaré duality says the number of independent k-dimensional holes in a compact manifold equals the number of independent $(n-k)$-dimensional holes. For $n = 2$: $b_0 = b_2$ (components equal void count), and $b_1 = b_1$ (trivial, self-dual). For $n = 3$: $b_0 = b_3$, $b_1 = b_2$.

The Betti numbers are symmetric around the middle degree, a constraint visible in the table:

Space	b_0	b_1	b_2	b_3
S^1	1	1	–	–
S^2	1	0	1	–
T^2	1	2	1	–
S^3	1	0	0	1
T^3	1	3	3	1

Each row is palindromic: the sequence reads the same forwards and backwards. This is Poincaré duality, enforced by the Hodge star.

6.8 The Spectrum of Δ_k and Topology vs. Geometry

On a compact Riemannian manifold, the Hodge Laplacian Δ_k has a discrete, non-negative spectrum:

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \nearrow +\infty. \tag{6.20}$$

The zero eigenvalue has multiplicity $b_k = \dim \mathcal{H}^k$ (the Betti number). All other eigenvalues are strictly positive.

Topology is in the zero modes. The zero eigenspace $\ker \Delta_k = \mathcal{H}^k$ captures the topological information: the Betti numbers. This part is unchanged by any smooth deformation of the metric (the Betti numbers are topological invariants).

Geometry is in the positive spectrum. The positive eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots$ depend on the metric: they encode how the manifold is curved, how large it is, how tightly its topology is packed. Changing the metric (while keeping the topology fixed) changes these eigenvalues but not the number of zero modes.

Example: the flat torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$. The Laplacian on functions ($k = 0$) has eigenfunctions $e^{2\pi i(mx+ny)}$ with eigenvalues $4\pi^2(m^2 + n^2)$ for $(m, n) \in \mathbb{Z}^2$. The zero eigenvalue has multiplicity 1 (constant functions, $b_0 = 1$). The positive eigenvalues are $4\pi^2, 8\pi^2, 16\pi^2, \ldots$, depending on the lattice geometry. If we stretch the torus (change the metric), these eigenvalues change, but $b_0 = 1$ remains fixed. The torus has one connected component regardless of how we stretch it.

The TQFT connection. In topological quantum field theories, path integrals are computed by restricting to harmonic forms (the zero modes of Δ_k). This is what makes TQFT invariants finite: even though the space of all forms is infinite-dimensional, the zero-mode space is finite-dimensional. The separation between topology (zero modes) and geometry (positive modes) is the mathematical mechanism behind topological protection.

6.9 Instantons and Self-Dual Curvature

The decomposition of 2-forms in dimension $n = 4$ into self-dual and anti-self-dual parts has a profound application to gauge theory.

On an oriented Riemannian 4-manifold, $\star\star = +1$ on 2-forms, so $\star$ is an involution with eigenvalues ± 1. Every 2-form splits as $\omega = \omega^+ + \omega^-$ with $\star\omega^\pm = \pm\omega^\pm$. In terms of the Yang-Mills functional:

$$\|F\|_{L^2}^2 = \|F^+\|_{L^2}^2 + \|F^-\|_{L^2}^2, \tag{6.21}$$

and the second Chern number (topological charge) is:

$$k = \frac{1}{8\pi^2}\int_M \mathrm{Tr}(F \wedge F) = \frac{1}{8\pi^2}(\|F^+\|_{L^2}^2 - \|F^-\|_{L^2}^2). \tag{6.22}$$

For $k > 0$, the Yang-Mills action satisfies $\|F\|^2 \geq 8\pi^2 k$, with equality if and only if $F^- = 0$, i.e., $F = F^+$: the curvature is **self-dual**. These energy-minimizing connections within a topological class are called **instantons**.

Instantons satisfy $d^*F = 0$ (the Yang-Mills equation) automatically via the Bianchi identity $dF = 0$: if $F = \star F$, then $d^*F = \pm \star d \star F = \pm \star dF = 0$. So every instanton is automatically a critical point of the Yang-Mills functional.

Connection to de Rham cohomology. The instanton number $k = c_2(E)$ is the second Chern class of the gauge bundle: an integer in $H^4(M; \mathbb{Z}) \cong \mathbb{Z}$ (for $M = S^4$). The Hodge decomposition of the curvature $F = F^+ + F^-$ is the Hodge decomposition of the representative of the second Chern class into self-dual and anti-self-dual components. The instanton condition $F = F^+$ is the condition that the curvature is its own harmonic representative relative to the self-dual-form complex.

6.10 The Discrete Hodge Decomposition

Everything in this chapter has a discrete version, providing the computational backbone for all numerical implementations.

For a simplicial complex with coboundary matrices $\mathbf{D}_k$ and an inner product on cochains (given by mass matrices $\mathbf{M}_k$), the discrete codifferential is $\mathbf{D}_k^* = \mathbf{M}_{k-1}^{-1}\mathbf{D}_{k-1}^\top \mathbf{M}_k$ and the discrete Hodge Laplacian is $\mathbf{\Delta}_k = \mathbf{D}_{k-1}\mathbf{D}_k^* + \mathbf{D}_{k+1}^*\mathbf{D}_k$.

The discrete Hodge decomposition of any k-cochain ω:

$$\omega = \underbrace{\mathbf{D}_{k-1}\alpha}_{\text{gradient part}} + \underbrace{\mathbf{D}_{k+1}^{*}\beta}_{\text{curl part}} + \underbrace{\gamma}_{\text{harmonic part}}, \tag{6.23}$$

is computed by solving $\mathbf{\Delta}_k\gamma = 0$ (the harmonic part spans the null space of $\mathbf{\Delta}_k$). This is a finite-dimensional linear algebra problem, solvable by Gaussian elimination or iterative methods.

Connecting to Chapter 4. In the simplicial cohomology chapter, the harmonic cochains were identified as the intersection $\ker \mathbf{D}_k \cap \ker \mathbf{D}_{k-1}^{\top}$: simultaneously cocycles (closed) and cycles (coclosed relative to the inner product). This is exactly the discrete analog of $\ker \Delta_k = \ker d_k \cap \ker d_k^*$. The Betti numbers computed by rank-nullity in Chapter 4 are the dimensions of these harmonic spaces.

6.11 Summary: Hodge Theory as the Bridge

Hodge theory sits at the crossroads of topology (de Rham cohomology), analysis (elliptic operators, L^2 theory), and geometry (the Riemannian metric). Its central achievements are:

Theorem	Statement	Significance
Hodge decomposition	$\Omega^k = \mathrm{im}(d) \oplus \mathcal{H}^k \oplus \mathrm{im}(d^*)$	Canonical splitting
Hodge isomorphism	$\mathcal{H}^k \cong H^k_{\mathrm{dR}}$	Topology from analysis
Poincaré duality	$H^k \cong H^{n-k}$ via $\star$	Palindromic Betti numbers
Minimum norm	Harmonic = smallest L^2 norm in class	Canonical representative
Discrete version	$\ker \mathbf{\Delta}_k$ computable by linear algebra	Computational tool

The deep connection to de Rham cohomology is this: the abstract quotient $H^k = Z^k/B^k$ becomes, via Hodge theory, the concrete subspace $\mathcal{H}^k \subset \Omega^k$. Topology (which classes exist) is determined by the global structure of M; geometry (which representative is harmonic) depends on the metric. The two can be cleanly separated — by the Hodge decomposition — and each studied independently.

In the chapters ahead, this separation is used repeatedly. Characteristic classes (Chapter 9) are computed by integrating curvature forms, which are representatives of the relevant cohomology classes; changing the connection changes the representative but not the class. TQFT invariants (Chapter 12) are computed by restricting to the zero modes of Laplacians on spaces of fields; the finite-dimensionality of the harmonic space is what makes the invariants well-defined.

Chapter 7

Homotopy Theory: Loops, Winding, and Obstruction

7.1 The Central Question: When Is a Deformation Possible?

Imagine holding a rubber band in your hands. You can stretch it, compress it, twist it, or bend it into any shape you like. In topology, all these deformed shapes are considered the same. But if the rubber band is looped around a lamp post, no amount of stretching can free it without cutting it. The lamp post creates an **obstruction** to the deformation.

This chapter makes obstructions precise. The key tool is the concept of **homotopy**: a continuous interpolation between two maps, parameterized by a time variable $t \in [0, 1]$. Two maps are homotopic if one can be continuously deformed into the other. Homotopy is an equivalence relation, and the equivalence classes — the **homotopy classes** — are the topological invariants that detect whether a deformation is possible.

The central insight is that homotopy classes are often **integers**: a loop around a circle winds an integer number of times, a map between spheres covers the target an integer number of times, and a gauge field configuration wraps around its target space an integer number of times. These integers are topological charges: they cannot change under continuous deformation, and they govern quantized phenomena from superconducting flux to magnetic monopole charges to QCD instanton numbers.

The chapter is organized around escalating examples. We begin with the simplest: loops in the plane around a missing point ($\pi_1(S^1) = \mathbb{Z}$, the winding number). We then generalize to maps between spheres ($\pi_n(S^n) = \mathbb{Z}$, the degree). Finally we reach the remarkable $\pi_3(S^2) = \mathbb{Z}$, generated by the Hopf fibration — a case where topology surprises even experts. Throughout, the physical applications drive the mathematics.

7.2 Homotopy of Maps

Definition 7.1 (Homotopy)**.** Two continuous maps $f_0, f_1 : X \to Y$ are **homotopic**, written $f_0 \simeq f_1$, if there exists a continuous map $H : X \times [0, 1] \to Y$ with $H(x, 0) = f_0(x)$ and $H(x, 1) = f_1(x)$ for all $x \in X$.

Think of t as time: the homotopy H is a film in which f_0 morphs continuously into f_1. At each time t, the map $H(\cdot, t)$ is a snapshot. The film must be continuous: no instantaneous jumps, no

tearing, no gluing. Two maps are in the same homotopy class if there exists such a film between them.

Homotopy is an equivalence relation (reflexive: use constant film; symmetric: reverse time; transitive: concatenate films). The equivalence classes are **homotopy classes**, and the set of all homotopy classes of maps from X to Y is written $[X, Y]$.

Homotopy equivalence of spaces. Two spaces X and Y are homotopy equivalent if there are maps $f : X \to Y$ and $g : Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$. Homotopy-equivalent spaces have the same homotopy groups and cohomology. Key examples: $\mathbb{R}^n \simeq \{\mathrm{pt}\}$ (contractible: shrink to origin); $\mathbb{R}^n \setminus \{0\} \simeq S^{n-1}$ (retract radially onto the unit sphere).

7.3 The Fundamental Group: Loops and Holes

7.3.1 Definition and Group Structure

Fix a basepoint $x_0 \in X$. A **loop based at** x_0 is a continuous map $\gamma : [0, 1] \to X$ with $\gamma(0) = \gamma(1) = x_0$. Two loops are homotopic (as elements of π_1) if there is a homotopy $H : [0, 1] \times [0, 1] \to X$ with $H(s, 0) = \gamma_0(s)$, $H(s, 1) = \gamma_1(s)$, and $H(0, t) = H(1, t) = x_0$ for all t (the endpoints stay at the basepoint throughout).

Definition 7.2 (Fundamental group)**.** The **fundamental group** $\pi_1(X, x_0)$ is the set of homotopy classes of loops based at x_0, with the group operation of concatenation: $[\gamma_1] \cdot [\gamma_2]$ traverses γ_1 then γ_2.

The identity is the constant loop at x_0. The inverse of $[\gamma]$ is the same loop traversed backwards. Concatenation is well-defined on homotopy classes: if $\gamma_1 \simeq \gamma_1'$ and $\gamma_2 \simeq \gamma_2'$ then $\gamma_1\gamma_2 \simeq \gamma_1'\gamma_2'$.

Why a group? The set of homotopy classes of loops, with concatenation, satisfies all group axioms. The associativity $([\alpha][\beta])[\gamma] = [\alpha]([\beta][\gamma])$ holds up to homotopy (re-parametrize the time spent on each sub-loop). The identity and inverse properties are clear. This makes π_1 a genuine group, encoding the algebraic structure of loop deformations.

For $n \geq 2$, the higher homotopy groups $\pi_n(X, x_0)$ (defined in Section 7.6) are always abelian, because spheres of dimension ≥ 2 have enough room to slide past each other. But π_1 can be non-abelian: the fundamental group of a figure-eight is the free group $\mathbb{Z} * \mathbb{Z}$, in which the order of traversal of the two loops matters.

7.3.2 The Pole Obstruction: $\pi_1(\mathbb{R}^2 \setminus \{0\})$

The canonical obstruction example: the plane minus the origin. A loop that does not wind around the origin can be pulled away from it and contracted to the basepoint. But a loop that winds once around the origin cannot be contracted without crossing the origin. The missing point is the obstruction.

The set of homotopy classes of loops is classified by the integer winding number: how many net times the loop goes around the origin, with counterclockwise positive. This gives $\pi_1(\mathbb{R}^2 \setminus \{0\}) \cong \mathbb{Z}$.

Concretely: the loop $\gamma_n(\theta) = e^{in\theta}$ for $\theta \in [0, 2\pi]$ has winding number n. For $n \neq m$, the loops γ_n and γ_m are not homotopic: no continuous deformation converts one winding number into another. The integer n is an obstruction to the existence of a homotopy.

7.3.3 Physical Example: Vortices in Superfluids

In a superfluid, the condensate is described by a complex order parameter $\psi(\mathbf{r}) = |\psi|e^{i\theta(\mathbf{r})}$. The phase θ takes values in $S^1 = U(1) \cong \mathbb{R}/2\pi\mathbb{Z}$. A **vortex** is a point in 2D where $|\psi| = 0$ (the core)

and the phase θ winds around the core:

$$\oint_C d\theta = 2\pi n, \qquad n \in \mathbb{Z}. \tag{7.1}$$

The integer n is the **vortex charge**, an element of $\pi_1(S^1) \cong \mathbb{Z}$. It counts how many times the phase winds around the core as we traverse the loop C encircling the vortex. In a superconductor, each vortex traps one quantum of magnetic flux $\Phi_0 = h/2e$: the topological charge n is directly related to the flux quantum, and the quantization of flux is a direct consequence of $\pi_1(S^1) = \mathbb{Z}$.

A single vortex (winding number $n = 1$) cannot be continuously removed from the condensate without passing the phase through a discontinuity or collapsing $|\psi|$ to zero across a macroscopic region, which costs energy proportional to system size. This is topological stability: the vortex is protected by $\pi_1 \neq 0$.

7.3.4 Key Examples of π_1

Space	π_1	Physical meaning
$\mathbb{R}^n$ (any n)	0	No loops, no defects
S^1 (circle)	$\mathbb{Z}$	Winding number
S^n ($n \geq 2$)	0	Simply connected
$\mathbb{R}^2 \setminus \{0\}$, S^1	$\mathbb{Z}$	Vortex charge
T^2 (torus)	$\mathbb{Z} \times \mathbb{Z}$	Two independent loops
$S^1 \vee S^1$ (figure-eight)	$\mathbb{Z} * \mathbb{Z}$	Non-abelian!
$\mathbb{R}^3 \setminus \{\text{line}\}$	$\mathbb{Z}$	Aharonov-Bohm
$SO(3)$	$\mathbb{Z}/2\mathbb{Z}$	Spinor double cover
$U(1)$	$\mathbb{Z}$	Electromagnetic gauge

The non-abelian example. On the figure-eight, the loop a around the first circle followed by loop b around the second is not homotopic to ba (first b, then a). To deform ab into ba, you would need to "slide a past b", which requires crossing the junction point, which would disconnect the figure-eight. The fundamental group captures this: $\pi_1(S^1 \vee S^1) = \langle a, b \,|\rangle$ is the free group, with no relation between a and b.

$SO(3)$ **and spinors.** The rotation group $SO(3)$ has $\pi_1(SO(3)) = \mathbb{Z}/2\mathbb{Z}$: a 2π rotation is not contractible (not homotopic to the identity), but a 4π rotation is. This is the Dirac belt trick: a belt with one 2π twist cannot be untwisted without moving the endpoints, but with two 2π twists it can. This $\mathbb{Z}/2\mathbb{Z}$ fundamental group is the mathematical source of spinors: the universal cover of $SO(3)$ is $SU(2) \cong S^3$, and spin-$\frac{1}{2}$ particles transform under $SU(2)$ rather than $SO(3)$, picking up a factor of -1 under 2π rotations.

7.4 Winding Numbers: An Integer from an Integral

7.4.1 The Formula

The winding number is the prototype for all topological charges. For a loop $\gamma : S^1 \to S^1$ (a map from the circle to itself, or equivalently a loop in $\mathbb{R}^2 \setminus \{0\}$ parametrized as $\gamma(\theta) = e^{i\phi(\theta)}$ where $\phi : [0, 2\pi] \to \mathbb{R}$ is a continuous lift of the angle), the winding number is

$$w(\gamma) = \frac{\phi(2\pi) - \phi(0)}{2\pi} = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\phi}{d\theta}\, d\theta. \tag{7.2}$$

Since γ is a loop on S^1, we need $e^{i\phi(2\pi)} = e^{i\phi(0)}$, forcing $\phi(2\pi) - \phi(0) \in 2\pi\mathbb{Z}$. The winding number is always an integer.

In de Rham language, using the angle form $d\theta = (-y\, dx + x\, dy)/(x^2 + y^2)$ on S^1 (the generator of $H^1_{\mathrm{dR}}(S^1)$):

$$w(\gamma) = \frac{1}{2\pi} \oint_\gamma \gamma^*(d\theta) = \frac{1}{2\pi} \oint_\gamma \frac{-y\, dx + x\, dy}{x^2 + y^2}. \tag{7.3}$$

This formula is profound: it expresses a homotopy invariant (the winding number, an element of $\pi_1(S^1) = \mathbb{Z}$) as an integral of a differential form. This is the bridge between homotopy theory and de Rham cohomology. The angle form generates $H^1_{\mathrm{dR}}(S^1) \cong \mathbb{R}$, and its period integral over γ computes the homotopy class.

7.4.2 Worked Calculations

Example 1. $\gamma_n(\theta) = e^{in\theta}$, wrapping n times around. The lift is $\phi(\theta) = n\theta$, so $\phi(2\pi) - \phi(0) = 2\pi n$ and $w = n$. Directly via the integral: $\gamma_n^*(d\theta) = n\, d\theta$, so $\frac{1}{2\pi} \oint n\, d\theta = n$.

Example 2. The loop $\gamma(\theta) = (\cos 2\theta, \sin 2\theta)$ in $\mathbb{R}^2 \setminus \{0\}$. This traverses the unit circle twice. Applying the formula:

$$w(\gamma) = \frac{1}{2\pi} \int_0^{2\pi} \frac{-\sin 2\theta \cdot (-2\sin 2\theta) + \cos 2\theta \cdot 2\cos 2\theta}{\cos^2 2\theta + \sin^2 2\theta}\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} 2\, d\theta = 2. \tag{7.4}$$

Example 3. A figure-eight loop: starts at $(1, 0)$, goes counterclockwise around the circle of radius 1, then clockwise around the same circle. The net winding is $1 + (-1) = 0$: this loop is null-homotopic in $\mathbb{R}^2 \setminus \{0\}$ if the loop doesn't pass through the origin.

Example 4: Fluid circulation. In fluid mechanics, the circulation of a velocity field $\mathbf{v} = (v_x, v_y)$ around a closed curve C is $\Gamma = \oint_C (v_x\, dx + v_y\, dy)$. For an irrotational vortex with velocity field $\mathbf{v} = \frac{\Gamma}{2\pi} \frac{(-y,x)}{x^2+y^2}$, the circulation around any loop encircling the vortex core is Γ. This is $\frac{\Gamma}{2\pi} \times 2\pi = \Gamma$ times the winding number of the loop. The circulation is quantized (if we require $\mathbf{v}$ to come from a single-valued velocity potential) in multiples of the quantum of circulation.

7.4.3 Why the Winding Number Is Stable

If γ_t is a smooth family of loops (a homotopy), then $w(\gamma_t) = \frac{1}{2\pi} \oint_{\gamma_t}$ is a continuous function of t taking only integer values. A continuous integer-valued function must be constant. Therefore: homotopic loops have the same winding number. The winding number is a topological invariant.

Conversely, any two loops with the same winding number are homotopic. This makes the winding number a *complete* invariant for maps $S^1 \to S^1$: two such maps are homotopic if and only if their winding numbers agree. The set $[S^1, S^1] = \pi_1(S^1) \cong \mathbb{Z}$ is classified by a single integer.

7.4.4 The Aharonov-Bohm Effect: Winding as Phase

Consider an electron circling around a solenoid carrying magnetic flux Φ. The vector potential outside the solenoid satisfies $\nabla \times \mathbf{A} = 0$ but $\mathbf{A} \neq 0$, so the 1-form $A = A_\phi \, d\phi = \frac{\Phi}{2\pi r} r \, d\phi$ is a closed but non-exact 1-form on $\mathbb{R}^2 \setminus \{\text{axis}\}$. The electron's wavefunction accumulates a phase:

$$\Delta\phi_{\text{AB}} = \frac{e}{\hbar} \oint_C A = \frac{e}{\hbar} \cdot \Phi. \tag{7.5}$$

This is the **Aharonov-Bohm phase**: a quantum phase that depends on the winding number of the electron's path around the solenoid. When $\Phi = nh/e$ (an integer number of flux quanta), the phase is $2\pi n$ and gives no observable effect. For other values of Φ, the phase shifts the electron interference pattern even though $\mathbf{B} = 0$ along the electron's path. The winding number of the path, combined with the topological content of A as a class in $H^1(\mathbb{R}^2 \setminus \{0\})$, determines the physical observable.

7.5 Degree of Maps: Covering the Target

7.5.1 Definition and Preimage Formula

The winding number generalizes from loops on circles to maps between manifolds of the same dimension. Let $f : M \to N$ be a smooth map between compact oriented n-manifolds. The **degree** of f is

$$\deg(f) = \int_M f^* \omega_N, \tag{7.6}$$

where ω_N is the volume form on N normalized so $\int_N \omega_N = 1$. Equivalently, choose any regular value $y \in N$ (by Sard's theorem, almost all points are regular: those where df is surjective at every preimage). The preimage $f^{-1}(y) = \{x_1, \ldots, x_k\}$ is finite, and

$$\deg(f) = \sum_{i=1}^{k} \operatorname{sgn}(\det df_{x_i}), \tag{7.7}$$

where the sign is $+1$ if f is orientation-preserving at x_i and -1 if orientation-reversing. The degree counts how many times the domain covers the target, with orientation.

7.5.2 Worked Calculations

The map $z \mapsto z^n$ on S^1. Writing $z = e^{i\theta}$, the map $f(\theta) = e^{in\theta}$ has n preimages of any regular value (the n values $\theta_k = \theta_0 + 2\pi k/n$ for $k = 0, \ldots, n-1$), each with $df = n > 0$, contributing $+1$. The degree is n.

The magnetic hedgehog $\hat{B} : S^2 \to S^2$. A magnetic monopole of unit charge has field $\mathbf{B} = g/(4\pi r^2)\hat{r}$. The normalized field $\hat{B}(\hat{r}) = \hat{r}$ is the identity map $S^2 \to S^2$, which has degree $+1$. Every direction on the target sphere is covered exactly once, orientation-preserving. Using the integral formula:

$$\deg(\hat{B}) = \frac{1}{4\pi} \int_{S^2} \hat{B}^* \, \omega_{S^2} = \frac{1}{4\pi} \int_{S^2} \sin\theta \, d\theta \, d\phi = 1. \tag{7.8}$$

For a monopole of charge n, the field is n times stronger and the normalized map $\hat{B} : S^2 \to S^2$ wraps n times, giving $\deg(\hat{B}) = n$. The degree is the magnetic charge.

A degree-2 map $S^2 \to S^2$. Take the map that wraps the ϕ-coordinate twice: $(\theta, \phi) \mapsto (\theta, 2\phi)$. Except at the two poles (where the mapping is singular), every point in the target has two preimages with the same orientation, contributing $+1$ each. The degree is 2.

Degree-(-1): reflection. The map $f(x_0, x_1, \ldots, x_n) = (-x_0, x_1, \ldots, x_n)$ on S^n reflects one coordinate. Every regular value has one preimage, but the map reverses orientation. Degree $= -1$.

7.5.3 Degree Is a Complete Invariant for $S^n \to S^n$

Two smooth maps $f, g : S^n \to S^n$ are homotopic if and only if $\deg(f) = \deg(g)$. This is the Hopf theorem (1926 for $n = 2$, Hopf for general n). Combined with $\pi_n(S^n) \cong \mathbb{Z}$, the classification of maps $S^n \to S^n$ up to homotopy is complete: they are classified by a single integer, the degree.

Homotopy groups of spheres. The degree gives us $\pi_n(S^n) \cong \mathbb{Z}$ for all $n \geq 1$. But there are also non-trivial maps from higher-dimensional spheres to lower-dimensional ones, which we address next.

7.6 Higher Homotopy Groups

7.6.1 Definition

Definition 7.3 (Higher homotopy groups)**.** The **n-th homotopy group** $\pi_n(X, x_0)$ consists of homotopy classes of continuous maps $(S^n, s_0) \to (X, x_0)$ that send a fixed basepoint s_0 to x_0. The group operation is defined by pinching the equator of S^n to a point and mapping each hemisphere separately.

For $n \geq 2$, π_n is abelian: two maps from S^n can be "slid past each other" in the extra dimensions, making the order of composition irrelevant up to homotopy.

7.6.2 The Physical Relevance Table

The following table collects the most important homotopy groups and their physical applications:

Group	Value	Physical system	Topological charge
$\pi_1(S^1)$	$\mathbb{Z}$	Vortex in superfluid	Vortex number
$\pi_1(S^1)$	$\mathbb{Z}$	Superconductor	Quantized flux
$\pi_1(\mathbb{R}^2 \setminus \{0\})$	$\mathbb{Z}$	Aharonov-Bohm	Winding
$\pi_1(SO(3))$	$\mathbb{Z}/2$	Spinor	Double cover
$\pi_2(S^2)$	$\mathbb{Z}$	Magnetic monopole	Magnetic charge
$\pi_2(S^2)$	$\mathbb{Z}$	2D Skyrmion	Skyrmion number
$\pi_3(S^2)$	$\mathbb{Z}$	Hopfion	Hopf charge
$\pi_3(S^3)$	$\mathbb{Z}$	Yang-Mills instanton	Instanton number
$\pi_3(SU(2))$	$\mathbb{Z}$	Sphaleron / baryon	Winding number
$\pi_4(S^3)$	$\mathbb{Z}/2$	Spin structures	(more subtle)

7.6.3 $\pi_2(S^2) = \mathbb{Z}$: Wrapping a Sphere Around a Sphere

A map $f : S^2 \to S^2$ is classified by its degree (the number of times it wraps the domain sphere around the target). The magnetic monopole of charge g has the normalized field $\hat{B} : S^2_\infty \to S^2_{\text{directions}}$ with $\deg(\hat{B}) = g$, and this is an element of $\pi_2(S^2) \cong \mathbb{Z}$.

Explicit calculation for the hedgehog (unit monopole). The field $\mathbf{B} = \hat{r}/(4\pi r^2)$ satisfies $\hat{B}(\hat{r}) = \hat{r}$ (the identity map). Using the area form $\omega_{S^2} = \sin\theta\, d\theta \wedge d\phi$:

$$\deg(\hat{B}) = \frac{1}{4\pi}\int_{S^2} \hat{B}^*(\sin\theta\, d\theta \wedge d\phi) = \frac{1}{4\pi}\int_{S^2} \sin\theta\, d\theta \wedge d\phi = \frac{1}{4\pi}\cdot 4\pi = 1. \tag{7.9}$$

The charge-g monopole has $\hat{B}(\hat{r}) =$ the degree-g map, giving $\pi_2(S^2) \ni [\hat{B}] = g \in \mathbb{Z}$.

7.6.4 $\pi_3(S^3) = \mathbb{Z}$: Yang-Mills Instantons

Since $SU(2) \cong S^3$ as a topological space, maps $S^3 \to SU(2)$ are classified by $\pi_3(S^3) = \mathbb{Z}$. In Yang-Mills theory on $\mathbb{R}^4$, a gauge field A with finite action must approach a pure gauge at spatial infinity: $A_\mu \to g^{-1}\partial_\mu g$ on the boundary S^3_∞. The map $g : S^3_\infty \to SU(2) \cong S^3$ has a degree, the **instanton number**:

$$k = \deg(g) = \frac{1}{24\pi^2}\int_{S^3_\infty} \mathrm{Tr}[(g^{-1}dg)^{\wedge 3}] \in \mathbb{Z}. \tag{7.10}$$

Equivalently, k is computed in the bulk as the second Chern class:

$$k = \frac{1}{8\pi^2}\int_{\mathbb{R}^4} \mathrm{Tr}(F \wedge F) \in \mathbb{Z}. \tag{7.11}$$

For the BPST one-instanton solution, $k = 1$. The instanton cannot be continuously deformed to the trivial connection ($k = 0$) while keeping the action finite, because k is a topological invariant. The QCD θ-vacuum $|\theta\rangle = \sum_n e^{in\theta}|n\rangle$ is the quantum superposition of all instanton sectors, in exact analogy with a Bloch wave being the superposition of all crystal cells.

7.6.5 $\pi_3(S^2) = \mathbb{Z}$: The Hopf Invariant and Hopfions

The group $\pi_3(S^2)$ is generated by the Hopf fibration $h : S^3 \to S^2$, a remarkable map that classifies genuine 3D linking topology.

Realize $S^3 \subset \mathbb{C}^2$ as $\{(z_1, z_2) : |z_1|^2 + |z_2|^2 = 1\}$ and $S^2 \cong \mathbb{CP}^1$ as the space of complex lines. The Hopf map is:

$$h(z_1, z_2) = [z_1 : z_2] \in \mathbb{CP}^1 \cong S^2. \tag{7.12}$$

The preimage $h^{-1}(p)$ of any point $p \in S^2$ is a great circle in S^3. Any two such circles (preimages of distinct points $p, q \in S^2$) are **linked** with linking number 1: they cannot be separated by any continuous deformation.

The Hopf invariant as an integral. For a map $f : S^3 \to S^2$, choose a normalized area form ω on S^2 ($\int_{S^2} \omega = 1$). Since $H^2(S^3) = 0$, we can write $f^*\omega = d\alpha$ for some 1-form α on S^3. The Hopf invariant is:

$$H(f) = \int_{S^3} \alpha \wedge d\alpha = \int_{S^3} \alpha \wedge f^*\omega. \tag{7.13}$$

This is manifestly a Chern-Simons integral: $\alpha \wedge d\alpha$ is the Chern-Simons 3-form for the "gauge field" α on S^3. For the standard Hopf map, $H(h) = 1$.

Physical realization: Hopfions. A **Hopfion** is a field configuration $\hat{n} : \mathbb{R}^3 \to S^2$ (a unit vector field in 3D space) with nonzero Hopf charge. It exists in 3D ferromagnets and chiral liquid crystals. The Hopf charge measures the linking of the preimage curves: for two different target directions $p, q \in S^2$, the sets $\{\mathbf{r} : \hat{n}(\mathbf{r}) = p\}$ and $\{\mathbf{r} : \hat{n}(\mathbf{r}) = q\}$ are closed curves in $\mathbb{R}^3$ whose linking number is the Hopf charge.

The surprise of $\pi_3(S^2) = \mathbb{Z}$. One might expect maps from a higher-dimensional sphere to a lower-dimensional one to be trivial (deformable to a constant map). But $\pi_3(S^2) \cong \mathbb{Z}$ is nontrivial, generated by the Hopf map. Intuitively: S^3 is very tightly "wrapped around" S^2 by the Hopf fibration, and no continuous deformation can undo this linking. The Hopf invariant (7.13) detects it.

7.7 Covering Spaces and the Universality of Integers

7.7.1 The Universal Cover of S^1

The **universal covering space** of S^1 is the real line $\mathbb{R}$, with covering map $p : \mathbb{R} \to S^1$, $p(t) = e^{2\pi i t}$. Every integer on $\mathbb{R}$ maps to the basepoint $1 \in S^1$; the fiber $p^{-1}(1) = \mathbb{Z}$ is the integer lattice.

A loop $\gamma : [0, 1] \to S^1$ (with $\gamma(0) = \gamma(1) = 1$) lifts uniquely to a path $\tilde{\gamma} : [0, 1] \to \mathbb{R}$ starting at $\tilde{\gamma}(0) = 0$. The endpoint $\tilde{\gamma}(1)$ must lie in the fiber $p^{-1}(1) = \mathbb{Z}$: it is an integer. That integer is the winding number.

This is the **deepest reason winding numbers are integers**: lifting a loop to the universal cover produces a path in $\mathbb{R}$ whose endpoint is forced by the fiber structure to be an integer. The covering space "unrolls" the periodic identification of $S^1 = \mathbb{R}/\mathbb{Z}$ and reveals the integer lattice underneath.

7.7.2 Gauge Theory and Covering Spaces

In electromagnetism, the gauge group is $U(1) \cong S^1$. The electromagnetic vector potential A is a connection on a principal $U(1)$-bundle. Gauge transformations $A \to A + d\Lambda$ correspond to choosing

a different "branch" of the multivalued function Λ. The fundamental group $\pi_1(U(1)) = \mathbb{Z}$ governs the "large gauge transformations" that cannot be continuously deformed to the identity, and the quantization of magnetic charge is the statement that the winding number of the transition function around the equator of S^2 must be an integer.

7.8 The Hopf Fibration: Architecture of the 3-Sphere

The Hopf fibration $h : S^3 \to S^2$ is not just a map; it gives S^3 the structure of a principal $U(1)$-bundle over S^2.

The fiber structure. Each fiber $h^{-1}(p) \cong S^1$ is a great circle in S^3. These circles foliate S^3: every point of S^3 lies on exactly one fiber circle, and the base parametrizes which circle. The total space is S^3 (compact, simply connected), the base is S^2 (compact), and the fiber is S^1 (compact). The long exact sequence of homotopy groups for a fibration (Section 7.10) gives:

$$\cdots \to \pi_2(S^1) \to \pi_2(S^3) \to \pi_2(S^2) \xrightarrow{\delta} \pi_1(S^1) \to \pi_1(S^3) \to \cdots \tag{7.14}$$

Since $\pi_2(S^1) = 0$, $\pi_2(S^3) = 0$, and $\pi_1(S^3) = 0$, this gives $\pi_2(S^2) \cong \pi_1(S^1) \cong \mathbb{Z}$, recovering the classification of monopole charges.

First Chern class $c_1 = 1$. As a $U(1)$-bundle over S^2, the Hopf bundle is classified by its first Chern class $c_1 \in H^2(S^2; \mathbb{Z}) \cong \mathbb{Z}$. The connection 1-form on S^3 has curvature $F = d\alpha$ equal to the pullback of the area form of S^2, so

$$c_1 = \frac{1}{2\pi} \int_{S^2} F = \frac{1}{2\pi} \int_{S^2} \omega_{S^2} = \frac{4\pi}{2\pi} = 2. \tag{7.15}$$

Wait — with the normalization $\int_{S^2} \omega = 4\pi$ (area of unit sphere), the first Chern number is 1. The Hopf bundle is the minimal nontrivial $U(1)$-bundle over S^2, the one with unit magnetic charge.

7.9 Obstruction Theory: Why Maps Can Fail to Exist

7.9.1 The General Principle

Homotopy theory is the systematic study of **obstructions**: topological features that prevent certain constructions from succeeding globally, even when they work locally.

The general setup: we want to build a continuous map $f : X \to Y$ (or to extend a partial map, or to deform one map to another). We proceed by building f over successive "skeleta" of X — first the vertices (X^0), then the edges (X^1), then the triangles (X^2), and so on. At each stage k, we have defined f on X^{k-1} and want to extend it over each k-cell. Each k-cell e^k has its boundary $\partial e^k \cong S^{k-1}$ already mapped to Y (since $\partial e^k \subset X^{k-1}$). The question is whether we can fill in: extend $f|_{\partial e^k}$ to a map $f : e^k \to Y$.

The **obstruction** to this extension is the class $[f|_{\partial e^k}] \in \pi_{k-1}(Y)$. If this class is trivial, $f|_{\partial e^k} : S^{k-1} \to Y$ is null-homotopic and the extension exists. If the class is nontrivial, the extension is impossible. The total obstruction lives in the cohomology group $H^k(X; \pi_{k-1}(Y))$ (cohomology with coefficients in the homotopy group).

7.9.2 The Hairy Ball Theorem as an Obstruction

The hairy ball theorem says: there is no continuous non-vanishing vector field on S^2. Equivalently, the tangent bundle TS^2 has no non-vanishing global section.

The obstruction argument. A non-vanishing vector field on S^2 is a section of the frame bundle, which is a map $v : S^2 \to TS^2 \setminus \{0\} \simeq S^2 \times S^1$ (locally). Trying to build such a section cell by cell over the CW structure of S^2 (one 0-cell, one 1-cell, one 2-cell), we can always extend over the 0-skeleton and 1-skeleton. But on the 2-cell (the top hemisphere), the boundary map goes around the equator and the vector must rotate by 2π (due to the Gauss-Bonnet theorem: total curvature 4π spread over S^2, requiring the vector to rotate). The resulting map $S^1 \to S^1$ has winding number $\pm 1 \neq 0$, preventing extension over the 2-cell. The obstruction is in $H^2(S^2; \pi_1(S^1)) = H^2(S^2; \mathbb{Z}) \cong \mathbb{Z}$.

The obstruction is the Euler class $e(TS^2) = 2 \in H^2(S^2; \mathbb{Z})$, measuring the total rotation of the tangent frame as we traverse the boundary of the 2-cell.

Physical consequence. The hairy ball theorem implies that any continuous tangent vector field on a sphere (like the Earth's surface) must vanish somewhere. Applied to wind fields, this predicts that there is always at least one point on Earth with zero horizontal wind speed (a hurricane eye, so to speak). In mathematical physics, it implies that any electromagnetic field configuration on S^2 must have at least one singular point (a Dirac string) unless the field itself vanishes.

7.9.3 The Dirac Quantization as an Obstruction

The Dirac quantization condition $qg \in 2\pi\mathbb{Z}$ for the product of electric charge q and magnetic charge g is an obstruction argument.

We want to define a wave function for a charge-q particle in the field of a monopole of charge g. Locally, the wave function is a complex scalar $\psi_\alpha : U_\alpha \to \mathbb{C}$. On overlaps, $\psi_\alpha = e^{iq\Lambda_{\alpha\beta}}\psi_\beta$ for transition functions $\Lambda_{\alpha\beta}$. These transition functions must be consistent on triple overlaps: $e^{iq(\Lambda_{\alpha\beta}+\Lambda_{\beta\gamma}+\Lambda_{\gamma\alpha})} = 1$. This forces q times the integral of the transition function to be an integer multiple of 2π, which gives $qg \in 2\pi\mathbb{Z}$.

This is the obstruction to defining a globally consistent quantum state: the obstruction class lives in $H^2(S^2; \pi_1(U(1))) = H^2(S^2; \mathbb{Z})$, and it must be the trivial class for a consistent wave function to exist. The condition $qg \in 2\pi\mathbb{Z}$ is exactly the condition that the Chern class $c_1 = qg/2\pi \in \mathbb{Z}$ is an integer.

7.9.4 Characteristic Classes as Systematic Obstructions

The most systematic form of obstruction theory produces the characteristic classes studied in Chapter 9. Given a vector bundle $E \to M$, the question of whether E can be globally trivialized (i.e., whether $E \cong M \times \mathbb{R}^n$) is an obstruction problem. The successive obstructions to trivialization live in the cohomology groups $H^k(M; \pi_{k-1}(GL(n)))$, and they are exactly the Stiefel-Whitney classes, Pontryagin classes, and Euler class of E.

For a complex vector bundle, the obstructions to finding a complex trivialization are the Chern classes $c_k(E) \in H^{2k}(M; \mathbb{Z})$. Each nonzero Chern class is one independent obstruction, at one stage of the skeleton, to building a global frame. The Chern classes are zero if and only if the bundle is trivializable.

7.10 The Long Exact Sequence of Homotopy Groups

A fibration $F \to E \to B$ (a fiber bundle in which local triviality holds in a strong sense) gives a **long exact sequence** of homotopy groups:

$$\cdots \to \pi_n(F) \xrightarrow{i_*} \pi_n(E) \xrightarrow{p_*} \pi_n(B) \xrightarrow{\delta} \pi_{n-1}(F) \to \cdots \to \pi_0(B). \tag{7.16}$$

Exactness means: image of each map equals kernel of the next. This is one of the most powerful tools for computing homotopy groups.

Example: path-loop fibration. The based path space $P(X) = \{\gamma : [0,1] \to X, \gamma(0) = x_0\}$ is contractible (every path can be retracted to the constant path). The fiber over x_0 of the endpoint map $P(X) \to X$ is the loop space ΩX. The long exact sequence gives $\pi_n(X) \cong \pi_{n-1}(\Omega X)$: the n-th homotopy group of X is the $(n-1)$-th homotopy group of the loop space.

Example: Hopf fibration $S^1 \to S^3 \to S^2$. The relevant piece of the long exact sequence:

$$\pi_3(S^1) \to \pi_3(S^3) \xrightarrow{h_*} \pi_3(S^2) \xrightarrow{\delta} \pi_2(S^1) \to \pi_2(S^3) \to \pi_2(S^2) \xrightarrow{\delta} \pi_1(S^1) \to \pi_1(S^3). \tag{7.17}$$

Using $\pi_3(S^1) = 0$, $\pi_3(S^3) = \mathbb{Z}$, $\pi_2(S^1) = 0$, $\pi_2(S^3) = 0$, $\pi_1(S^1) = \mathbb{Z}$, $\pi_1(S^3) = 0$:

$$0 \to \mathbb{Z} \xrightarrow{h_*} \pi_3(S^2) \to 0, \qquad 0 \to \pi_2(S^2) \xrightarrow{\delta} \mathbb{Z} \to 0. \tag{7.18}$$

So $h_* : \pi_3(S^3) \xrightarrow{\sim} \pi_3(S^2)$ is an isomorphism: $\pi_3(S^2) \cong \mathbb{Z}$, generated by the Hopf map. And $\pi_2(S^2) \cong \mathbb{Z}$.

7.11 Homotopy and Cohomology: The Hurewicz Bridge

Homotopy groups and cohomology groups both classify topological spaces and maps, but from complementary directions. Homotopy probes a space using spheres as inputs; cohomology probes it using closed forms as measurements.

The Hurewicz homomorphism. For a simply connected space with $\pi_k(X) = 0$ for $k < n$, there is a canonical isomorphism $\pi_n(X) \cong H_n(X;\mathbb{Z})$ (the Hurewicz theorem). This relates the homotopy class of a map $f : S^n \to X$ to a homology class, and via the de Rham pairing, to an integral of a closed form.

From homotopy class to period integral. A map $f : S^n \to M$ in a homotopy class $[f] \in \pi_n(M)$ gives a map on cohomology $f^* : H^n_{\mathrm{dR}}(M) \to H^n_{\mathrm{dR}}(S^n) \cong \mathbb{R}$. For a closed n-form ω on M:

$$\int_{S^n} f^*\omega \in \mathbb{Z} \qquad (\text{if } [\omega] \in H^n(M;\mathbb{Z})). \tag{7.19}$$

This integral is an integer because it is simultaneously a period integral (of the cohomology class $[\omega]$) and the degree of the map f (which is an integer by the preimage-counting argument). The identity of these two computations is one half of the de Rham theorem: the period integrals of integral cohomology classes over spherical cycles must be integers.

Physical example: quantization from homotopy. The magnetic charge g is simultaneously: (1) the degree of the map $\hat{B} : S^2 \to S^2$ (element of $\pi_2(S^2) = \mathbb{Z}$), and (2) the period integral $\frac{1}{4\pi}\int_{S^2} F$ of the magnetic flux 2-form (element of $H^2(S^2;\mathbb{Z}) = \mathbb{Z}$). These are two descriptions of the same integer. The fact that it is an integer from the homotopy side forces the integral to be an integer on the cohomology side, giving Dirac quantization.

7.12 The Obstruction Hierarchy

Every topological invariant in this book fits into a hierarchy of obstructions, organized by dimension:

Dim	Homotopy group	Physical system	Invariant	Consequence
1	$\pi_1(S^1) = \mathbb{Z}$	Vortex, A-B effect	Winding	Flux quantum
2	$\pi_2(S^2) = \mathbb{Z}$	Monopole, Skyrmion	Degree	Charge quantization
3	$\pi_3(S^3) = \mathbb{Z}$	YM instanton	Pontryagin	θ-vacuum
3	$\pi_3(S^2) = \mathbb{Z}$	Hopfion	Hopf invariant	Linking
all	$\pi_{2k}(U(n))$	Bundles over S^{2k}	Chern class	Hall conductance

Each row is one stage of the obstruction tower. At dimension 1, the obstruction is to contracting a loop (living in π_1). At dimension 2, the obstruction is to contracting a sphere map (living in π_2). And so on. Each higher stage detects subtler topology that the lower stages cannot see.

The Chern classes are the most systematic: the k-th Chern class $c_k(E) \in H^{2k}(M;\mathbb{Z})$ is the obstruction, at dimension $2k$, to trivializing the bundle E over the $2k$-skeleton of M. The complete set of Chern classes is the complete obstruction to trivialization.

7.13 Summary: Obstruction as the Organizing Principle

Homotopy theory can be summarized as the systematic mathematics of what cannot be done continuously. Every construction that seems locally possible but globally fails has a topological obstruction, which is an integer living in a homotopy group.

The three themes that will recur throughout this book are:

Winding numbers are integers. Whether it is π_1 (loops), π_n (spheres), or π_3 (the Hopf invariant), the homotopy class of a map is an integer. This integer is computed either by preimage counting (the degree formula) or by a period integral (the de Rham formula). Both methods give the same answer, for the same underlying reason: both are counting how many times one space wraps around another.

Obstructions live in homotopy groups. The failure of a local construction to extend globally is measured by an element of a homotopy group $\pi_k(Y)$. If the element is trivial, the extension exists. If not, it is impossible. The systematic organization of all these obstructions produces the characteristic classes.

Cohomology classes are periods of homotopy classes. Via the Hurewicz homomorphism and the de Rham theorem, every homotopy class gives a period integral that is an integer. The cohomology group $H^k(M;\mathbb{Z})$ records exactly these integers. The de Rham cohomology $H^k_{\mathrm{dR}}(M)$ (over $\mathbb{R}$) contains the integral cohomology as an integer lattice, and the Chern-Weil theorem (Chapter 9) shows how to compute these integers from the curvature of a connection.

Chapter 8

Fiber Bundles: Combing the Basketball

8.1 The Basketball and the Compass

Imagine holding a basketball. At every point on its surface, you attach a small **compass needle** that can point in any direction tangent to the surface. You now have a *field of needles* on the sphere: at each of the infinitely many surface points, a chosen direction. This is a **section** of a fiber bundle.

The basketball is the **base space** $M = S^2$: the arena where physics happens. The compass needle at each point lives in the **fiber** F: the space of possible directions (here, the circle S^1 of tangent directions). The collection of all needle-choices at all points is a **section**. And the question of whether a consistent global choice is possible is the first topological question about this bundle.

Can you comb the basketball? Can you choose a needle direction at every point of the sphere, varying smoothly, without any needle ever pointing radially inward (vanishing) anywhere? Equivalently: can you give the basketball a consistent "hairdo" with no cowlick?

The answer is **no**: the hairy ball theorem. No matter how cleverly you arrange the compass needles, at least one must vanish or point radially. This is not a failure of imagination; it is a genuine topological obstruction. The sphere has no non-vanishing tangent vector field. In contrast, a torus *can* be combed globally — the torus has no such obstruction, because its Euler characteristic is zero.

Now upgrade from a 2D compass needle to a full **3D compass**: at each point on the basketball, attach a tiny coordinate frame — three mutually perpendicular arrows. This is a section of the **frame bundle**: a choice of oriented orthonormal basis for the tangent space at every point. The group $SO(3)$ acts on each fiber by rotating the frame, and a gauge transformation at a point is a rotation of its frame. Whether a global frame can be chosen consistently — the question of whether the sphere is *parallelizable* — is a deeper obstruction question, answered by the existence and properties of the characteristic classes we develop in Chapter 9.

Fiber bundles are the mathematical language for all of this: fields that live over a space, with values in some fixed "internal" space. The global topology of the base space imposes constraints — obstructions — on what kinds of global fields can exist.

8.2 Local Product, Global Twist

8.2.1 Paper Strips: The Cylinder and the Möbius Band

Take a strip of paper. Bring the two short ends together. If you glue them *straight*, you get a **cylinder**: the strip closes into a tube. If you give one end a **half-twist before gluing**, you get the **Möbius band**: the strip closes into the famous one-sided surface.

Both objects have the same local appearance: any short segment of either strip is just a flat rectangle. A small ant on either surface cannot tell which one it lives on by examining only its immediate neighborhood. The difference is entirely in the *global assembly*: how the two ends are glued.

This is the prototype for every fiber bundle. The circle S^1 is the base space (think of the central circle running along the midline of the strip). The fiber is the real line $\mathbb{R}$ (the cross-sections perpendicular to the central circle). The bundle is the entire strip. Local triviality holds: over any small arc of the circle, the strip looks like (arc)$\times\mathbb{R}$. The global topology — cylinder vs. Möbius band — depends on how the local pieces are glued together.

A **section** of the Möbius band is a function that picks one point in each fiber: a height assignment on every cross-section of the strip, varying smoothly as you walk around. For the *cylinder*, such a function can be everywhere positive: just set the height to $+1$ everywhere. For the *Möbius band*, no everywhere-positive (or everywhere-negative) section exists. After going halfway around the strip, "up" and "down" have been swapped by the half-twist. For a continuous section to return to its starting value after a full circuit, it must have crossed zero somewhere. The Möbius band has no non-vanishing global section.

The half-twist is the obstruction. It lives in the transition function.

8.2.2 Constructing the Möbius Band Explicitly

Cover S^1 with two open arcs:

$$U_+ = \{\theta \in (-\epsilon, \pi + \epsilon)\} \quad \text{(the "right half")},$$
$$U_- = \{\theta \in (\pi - \epsilon, 2\pi + \epsilon)\} \quad \text{(the "left half")},$$

for small $\epsilon > 0$. The overlap $U_+ \cap U_-$ has two connected components: the **top arc** near $\theta = \pi$ and the **bottom arc** near $\theta = 0 \equiv 2\pi$.

On each patch define a trivial bundle $U_\pm \times \mathbb{R}$. To assemble a global bundle, specify how to glue: the transition function $g_{+-} : U_+ \cap U_- \to GL(1, \mathbb{R}) = \mathbb{R}^*$ says: "the fiber coordinate t in the U_+ description equals $g_{+-}(\theta) \cdot t$ in the U_- description."

Cylinder: $g_{+-}(\theta) = +1$ everywhere. Both patches use the same orientation. Gluing is parallel; the strip closes without twisting.

Möbius band: $g_{+-}(\theta) = +1$ on the top arc and $g_{+-}(\theta) = -1$ on the bottom arc. The transition function changes sign between the two overlap components. When you cross from the top arc to the bottom arc — walking around the back of the circle — "up" in the U_+ description becomes "down" in the U_- description. This is the half-twist.

Why the section must vanish. Suppose a section $s : S^1 \to E$ is everywhere positive (say $s > 0$). In the U_+ patch, $s_+(\theta) > 0$. In the U_- patch, $s_- = g_{+-} \cdot s_+$. On the top arc, $g_{+-} = +1$, so $s_- = s_+ > 0$. On the bottom arc, $g_{+-} = -1$, so $s_- = -s_+ < 0$. But s_- must be positive everywhere. Contradiction. No such section exists.

The caterpillar metaphor: imagine a tiny caterpillar on the Möbius band, always walking along the center circle with its back pointing "up" (in the local fiber direction). After one full circuit, it

returns to the starting point — but the half-twist means it is now upside-down relative to where it started. No consistent notion of "up" survives the journey around the band.

8.2.3 What Local Triviality Means

A **fiber bundle** $\pi : E \to M$ with fiber F is locally a product: every point $x \in M$ has a neighborhood U over which

$$\pi^{-1}(U) \cong U \times F. \tag{8.1}$$

The homeomorphism in (8.1) is a **local trivialization**, and the requirement that it exists for every small U is **local triviality**.

Local triviality says: look at a small enough patch, and every bundle looks like a boring product. The global twist, if any, only shows up when you try to extend the local description around the entire base. This is the same principle we met in cohomology: every closed form is locally exact, but the global obstruction to exactness encodes topology. For bundles: every bundle is locally trivial, but the global obstruction to triviality encodes topology.

8.3 The Formal Definition

Definition 8.1 (Fiber Bundle)**.** A **fiber bundle** (E, M, π, F) consists of smooth manifolds E (total space), M (base space), F (fiber), a smooth surjective projection $\pi : E \to M$, and an open cover $\{U_\alpha\}$ of M with homeomorphisms (local trivializations)

$$\phi_\alpha : \pi^{-1}(U_\alpha) \xrightarrow{\sim} U_\alpha \times F \tag{8.2}$$

such that $\text{proj}_1 \circ \phi_\alpha = \pi|_{\pi^{-1}(U_\alpha)}$ (the trivialization respects the projection).

Unpacking each piece with the basketball.

The **total space** E is the basketball together with all the compass needles: the disjoint union $E = \bigsqcup_{x \in M} F_x$ of all fiber spaces. Every point of E is a pair: a location on the basketball, plus a needle direction at that location.

The **projection** $\pi : E \to M$ forgets the needle and remembers only the location: $\pi(\text{needle at } x) = x$.

The **fiber over** x is $F_x = \pi^{-1}(x)$: all possible needle directions at the single point x. Each $F_x \cong F$ as a topological space, but there is no canonical identification between F_x and F_y for $x \neq y$. The fibers are parallel sheets, not identified sheets.

The **local trivialization** ϕ_α says: over a small patch U_α of the basketball, we can simultaneously describe all the needles in that patch using a product $(x, \text{direction})$ with a consistent orientation convention. Outside this patch, a different convention may be needed.

The **structure group** G is the group of self-homeomorphisms of F that relate different local trivializations: for the compass bundle, $G = SO(2)$ (rotations of the needle). For the 3D compass frame bundle, $G = SO(3)$ (rotations of the frame). For a complex line bundle, $G = U(1)$ (phases). The structure group encodes what kinds of changes-of-convention are allowed.

Examples.

- **Cylinder** $S^1 \times \mathbb{R}$: trivial line bundle. No twist. A section is any smooth real-valued function on S^1: can be non-vanishing.
- **Möbius band**: non-trivial line bundle over S^1. Half-twist. A section must vanish at least once.

- **Tangent bundle** TS^2: the compass-needle bundle on the basketball. Non-trivial. No non-vanishing section (hairy ball theorem).
- **Tangent bundle** TS^1: line bundle over the circle. Trivial! The unit tangent vector field ("always go counterclockwise") is a non-vanishing global section.
- **Hopf bundle** $S^1 \hookrightarrow S^3 \to S^2$: the simplest non-trivial principal $U(1)$-bundle. Total space S^3, base S^2, fiber S^1. Every two fibers (circles in S^3) are linked; no global section exists.

8.4 Sections and Transition Functions: A Critical Distinction

The most important conceptual divide in bundle theory separates two completely different objects. Mixing them up is the most common source of confusion.

8.4.1 Sections Are Fields

A **section** $s : M \to E$ is a smooth map satisfying $\pi \circ s = \mathrm{id}_M$: it picks one element of each fiber, smoothly over the base. A section is a *physical field*: it is data you choose after the bundle is defined.

The compass picture. A section of the tangent bundle TS^2 is a tangent vector field on the basketball: at each point, one specific compass needle direction. The hairy ball theorem says no section of TS^2 is non-vanishing: every hairdo has a cowlick.

The 3D compass picture. A section of the frame bundle is a *gauge choice*: at each point, a specific orthonormal frame (three perpendicular arrows). Changing the frame at each point independently is a gauge transformation. Two sections of the same principal bundle are related by a gauge transformation; they describe the same bundle, just with different local conventions.

Examples of sections across physics.

- A scalar field $\phi : M \to \mathbb{R}$ is a section of the trivial bundle $M \times \mathbb{R}$.
- A vector field X is a section of TM (one tangent vector per point).
- A wavefunction ψ is a section of a complex line bundle over spacetime (one complex number per spacetime point, but the complex phase requires a gauge choice to be meaningful).
- A gauge potential A is a section of the adjoint bundle (not of the principal bundle: connections are more subtle and will be treated in the next chapter).

Sections are data on top of the bundle. Two different sections live on the same bundle. The bundle is fixed; sections are choices made within it.

8.4.2 Transition Functions Are Gluing Data

A **transition function** $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ is part of the definition of the bundle itself. It is determined before any section is chosen. It encodes how the local trivializations over overlapping patches are related.

The compass picture. Over the patch U_α (say, the northern hemisphere of the basketball), we have a local compass convention: at each point in U_α, we label the needle directions using a fixed reference frame e_α. Over the patch U_β (the southern hemisphere), we use a different reference frame e_β. On the overlap (the equatorial strip), both conventions are valid. The transition function

$g_{\alpha\beta}$ tells us: "if the needle points in direction v according to the northern convention e_α, it points in direction $g_{\alpha\beta}(x) \cdot v$ according to the southern convention e_β."

For the Hopf bundle, $g_{\alpha\beta} : S^1 \to U(1)$ is the map $e^{i\phi} \mapsto e^{i\phi}$ on the equatorial circle. Its winding number is 1. That winding number is the first Chern class $c_1 = 1$: the topological classification of the bundle.

Transition functions encode global topology. They are not arbitrary: they must satisfy the cocycle condition on triple overlaps. Two collections of transition functions that differ by a gauge transformation define the same bundle. The isomorphism class of the bundle is the equivalence class of its transition functions.

8.4.3 The Key Analogy

Concept	Manifold	Fiber Bundle
Local description	Chart $\phi_\alpha : U_\alpha \to \mathbb{R}^n$	Trivialization $\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times F$
Change of description	Chart overlap map	Transition function $g_{\alpha\beta}$
Global topology	Encoded in chart overlaps	Encoded in transition functions
A field	Smooth function $f : M \to \mathbb{R}$	Section $s : M \to E$

A coordinate chart tells us how to measure positions locally; the chart overlaps encode the manifold's global structure. Similarly, a local trivialization tells us how to label fiber elements locally; the transition functions encode the bundle's global structure. A smooth function on a manifold and a section of a bundle are both "data on top of the structure"; neither is part of the structure itself.

8.5 Transition Functions Classify Bundles

8.5.1 The Cocycle Condition

On each triple overlap $U_\alpha \cap U_\beta \cap U_\gamma$, the transition functions must satisfy the **cocycle condition**:

$$g_{\alpha\gamma}(x) = g_{\alpha\beta}(x) \cdot g_{\beta\gamma}(x). \tag{8.3}$$

Think of this as a consistency requirement: converting directly from patch γ to patch α must give the same answer as going through patch β as an intermediary.

A collection $\{g_{\alpha\beta}\}$ satisfying (8.3) is a **Čech 1-cocycle**. Given such a cocycle, we can construct the bundle by taking $\bigsqcup_\alpha (U_\alpha \times F)$ and gluing: (x, f) in patch α is identified with $(x, g_{\alpha\beta}(x) \cdot f)$ in patch β.

8.5.2 Gauge Equivalence of Cocycles

Two cocycles $\{g_{\alpha\beta}\}$ and $\{g'_{\alpha\beta}\}$ define **isomorphic bundles** if there exist maps $h_\alpha : U_\alpha \to G$ (one per patch) such that:

$$g'_{\alpha\beta}(x) = h_\alpha(x)^{-1} \cdot g_{\alpha\beta}(x) \cdot h_\beta(x) \quad \text{on every overlap.} \tag{8.4}$$

The transformation (8.4) is a **gauge transformation**: changing local trivializations by h_α on each patch changes the transition functions by (8.4).

The orbit picture. The set of all cocycles is a large space. Two cocycles in the same orbit under gauge transformations define the same bundle. The space of orbits classifies all bundles:

$$\{\text{isomorphism classes of } G\text{-bundles over } M\} \cong \check{H}^1(M;G) = \frac{\{\text{cocycles}\}}{\{\text{gauge transformations}\}}. \tag{8.5}$$

This is **Čech cohomology**: the set of gauge-equivalence classes of transition function collections. Each orbit is one topological type of bundle.

8.5.3 Example: Line Bundles over S^1

For real line bundles ($G = \mathbb{R}^*$) over S^1, the two-patch construction gives transition functions on two overlap components. The only topological invariant is the product of the two signs: $+1$ gives the cylinder, -1 gives the Möbius band. There are exactly two isomorphism classes: $\check{H}^1(S^1;\mathbb{R}^*) \cong \mathbb{Z}/2\mathbb{Z}$.

8.5.4 Example: $U(1)$ Bundles over S^2 and the Monopole

Cover S^2 by north and south hemispheres U_N, U_S overlapping in the equatorial strip $\cong S^1 \times (-\epsilon, \epsilon)$. The transition function is a map $g_{NS} : S^1 \to U(1)$.

Picture. Imagine two "caps" (northern and southern hemispheres), each carrying a local compass convention for the $U(1)$-fiber. Where the caps overlap — the equatorial band — both conventions are valid, and the transition function says how to convert between them. The transition function is a map from the equatorial circle to $U(1)$, and it can wind around $U(1)$ any integer number of times.

The winding number of g_{NS} is the **first Chern class** $c_1 \in \mathbb{Z}$. For $g_{NS}(\phi) = e^{in\phi}$:

- $n = 0$: trivial bundle. The conventions agree; no winding.
- $n = 1$: the Hopf bundle. The equatorial convention winds once as you go around the equator. This is the magnetic monopole of unit charge.
- $n = -1$: the anti-Hopf bundle, winding once in the opposite direction. Monopole of charge -1.

The Dirac quantization condition $g \in \mathbb{Z}$ is the statement that the transition function must have integer winding number: only then is g_{NS} a well-defined map $S^1 \to U(1)$.

8.6 Sections: Existence and Obstruction

8.6.1 Combing the Sphere: A Topological Impossibility

Return to the basketball. We want to comb it: assign a non-zero compass needle direction to every point, smoothly. Try to build this comb inductively.

Start at the north pole: point the needle south. Now extend outward from the north pole. In a small disk around the north pole, everything is fine: just rotate the south-pointing needle smoothly as you move away from the pole.

Keep extending. As you approach the equator, the needles all point approximately southward. Cross the equator; continue extending. As you approach the south pole, all needles pointing from all directions converge inward. But a continuous field cannot converge to a single point from all directions without vanishing: the field would have to "choose" a direction at the south pole

consistent with all the needles arriving from different directions, but by symmetry there is no preferred direction. The field must vanish at the south pole.

This argument shows that *any* attempt to comb S^2 produces at least one cowlick. The cowlick is topologically mandatory.

The Poincaré-Hopf theorem makes this precise: for any vector field on a compact manifold M with only isolated zeros, the sum of the **indices** of the zeros (each index measures how many times the field rotates as you circle the zero) equals the Euler characteristic $\chi(M)$. For S^2, $\chi = 2$, so the total index sum is 2. A non-vanishing field has no zeros, so its index sum is $0 \neq 2$. Contradiction.

For the torus: $\chi(T^2) = 0$. A constant vector field on the flat torus is non-vanishing: total index sum $= 0 = \chi(T^2)$. The torus can be combed. Geometrically: the torus has two holes whose contributions to the total curvature cancel, leaving no mandatory cowlick.

8.6.2 The Obstruction Theory Argument

More generally, the obstruction to finding a non-vanishing section of a rank-n vector bundle $E \to M$ is the **Euler class** $e(E) \in H^n(M; \mathbb{Z})$. When $e(E) \neq 0$, every section must vanish somewhere.

The argument goes cell-by-cell over the CW structure of M. On each k-cell, the section can be extended from the boundary $\partial e^k \cong S^{k-1}$ to the interior iff the class $[s|_{\partial e^k}] \in \pi_{k-1}(F \setminus \{0\})$ is trivial. For the tangent bundle of S^2 (fiber $\mathbb{R}^2 \setminus \{0\} \simeq S^1$), the obstruction on the single 2-cell is:

$$e(TS^2) = \chi(S^2) = 2 \in H^2(S^2; \mathbb{Z}) \cong \mathbb{Z}. \tag{8.6}$$

This is nonzero: the combing is impossible, and the size of the obstruction (2) predicts that the unavoidable zeros have total index sum 2.

8.6.3 Sections vs. Trivial Bundles

For a *principal* G-bundle, a global section is equivalent to a global trivialization (the section selects one element per fiber, identifying each fiber with G via that element's position). So: a principal bundle is trivializable iff it has a global section.

For a general vector bundle, having a *non-vanishing* section is weaker than being trivial: the tangent bundle of S^{2n-1} has a non-vanishing section for all $n \geq 1$ (odd-dimensional spheres can be combed) but is non-trivial for $n \geq 2$.

8.7 Principal Bundles: The Master Template

8.7.1 The 3D Compass Frame

Upgrade the compass from a single needle to a full **3D compass frame**: at each point of the basketball, attach a tiny coordinate frame — three mutually perpendicular arrows $\{e_1, e_2, e_3\}$ spanning the tangent space. Now the "fiber" at each point is not a single direction but the entire group of possible frames. That group is $SO(3)$ (for a 3D frame bundle): the 6-dimensional group of orthogonal transformations.

Choosing a specific frame at each point is a **gauge choice**. Different gauge choices (different global sections of the frame bundle) are related by $SO(3)$-valued functions: gauge transformations. Physical quantities must not depend on the gauge choice: they must be gauge-invariant.

This is the structure of a **principal G-bundle**: the fiber is the group G itself, the total space carries a free transitive right G-action on each fiber, and the base is the orbit space $M = P/G$.

Definition 8.2 (Principal G-Bundle)**.** A **principal G-bundle** is a fiber bundle $\pi : P \to M$ where the fiber is G, G acts on P from the right freely and transitively on each fiber, and $M = P/G$.

Freely: $p \cdot g = p \implies g = e$. No frame is fixed by any non-trivial rotation. This ensures the G-action has no redundancy.

Transitively on fibers: any two frames at the same base point are related by a unique group element. The fiber is a principal homogeneous space (torsor) for G: it looks like G, but without a preferred identity element.

No preferred zero. Unlike a vector bundle, no element of the fiber of a principal bundle is "the zero". All frames are gauge-equivalent, and physical content lives in the base M, not in the fiber. The gauge freedom is the fiber; the physics is the base.

8.7.2 Associated Bundles: All Fibers from One Principal Bundle

Given the 3D frame bundle $P \to M$ (principal $SO(3)$-bundle), every representation $\rho : SO(3) \to GL(V)$ of $SO(3)$ on a vector space V gives an associated vector bundle:

$$E_V = P \times_{SO(3)} V = (P \times V)/\sim, \quad (p, v) \sim (p \cdot g, \rho(g^{-1})v). \tag{8.7}$$

The equivalence identifies gauge-related frames: (p, v) and $(p \cdot g, \rho(g^{-1})v)$ describe the same physical field value v in different gauges. The fiber of E_V over x is a copy of V, but the copy depends on the gauge.

The physics dictionary. In electromagnetism:

- Principal bundle: the $U(1)$ gauge bundle $P \to M$ with fiber $U(1)$ = all possible phase conventions at each spacetime point.
- Associated bundle (fundamental representation $U(1)$ on $\mathbb{C}$): the line bundle $L = P \times_{U(1)} \mathbb{C}$ where charged scalar fields live.
- A section of L: a charged scalar field ϕ. Under gauge transformation $h : M \to U(1)$, it transforms as $\phi \to h\phi$.
- A connection on P: the electromagnetic vector potential A. It tells us how to parallel-transport the phase as we move in spacetime.

In the Standard Model, one principal bundle with gauge group $G = SU(3) \times SU(2) \times U(1)$ generates all particle content: quarks as sections of the associated bundle for the fundamental $SU(3)$ representation, gluons in the adjoint, electrons in the $U(1)$ representation, and so on.

8.8 Classifying Spaces: The Universal Catalog

8.8.1 The Catalog Idea

Every fiber bundle over every space is drawn from a single **universal catalog**: a space BG (the classifying space) that contains, in a precise sense, one entry per isomorphism class of principal G-bundles. To specify which bundle you have over your manifold M, you need only specify a map $f : M \to BG$ (a "catalog entry selection"), up to homotopy. Two homotopic maps select the same bundle.

Theorem 8.1 (Classifying Space)**.** *For any Lie group G, there exists a contractible space EG with a free G-action, with quotient $BG = EG/G$. For any CW complex M:*

$$\{\textit{isomorphism classes of principal } G\textit{-bundles over } M\} \cong [M, BG]. \tag{8.8}$$

Every bundle arises as the pullback $f^*(EG \to BG)$ for some classifying map $f : M \to BG$.

8.8.2 BG as a Space of Orbits

Think of EG as a huge space of all possible **gauges**: at each point, a gauge is a choice of frame, connection, or reference. The group G acts on EG by changing the gauge (rotating the compass frame). Two points of EG in the same G-orbit are the same physical configuration described in two different gauges.

The quotient $BG = EG/G$ is the space of **gauge-equivalence classes**: each point of BG is one orbit, one gauge-invariant configuration. BG is "what remains after modding out the gauge redundancy."

A map $f : M \to BG$ picks, for each point $x \in M$, one orbit in EG (one gauge-equivalence class). The pullback bundle assigns to x the actual pre-orbit element in EG — but now varying over M, so the total space of the pullback bundle is itself a principal G-bundle over M.

8.8.3 Why EG Must Be Contractible

The key property of EG is contractibility: $\pi_n(EG) = 0$ for all $n \geq 0$. Why is this necessary?

If EG had nontrivial topology (say $\pi_n(EG) \neq 0$), then the pullback bundle $f^*(EG) \to M$ would carry "extra" topological information from EG itself, contaminating the classification. Bundles would not be cleanly classified by the homotopy class of $f : M \to BG$; the topology of EG would interfere.

With EG contractible, all topological complexity of the bundle $f^*(EG)$ comes entirely from two sources: the map f and the quotient structure EG/G. The map f is classified up to homotopy by $[M, BG]$. The quotient $BG = EG/G$ encodes the group G. The bundle is fully classified by $[f] \in [M, BG]$, with no contamination.

The contractibility of EG also makes the long exact sequence of homotopy groups for the fibration $G \to EG \to BG$ produce:

$$\cdots \to \pi_n(EG) \to \pi_n(BG) \xrightarrow{\sim} \pi_{n-1}(G) \to \pi_{n-1}(EG) \to \cdots \tag{8.9}$$

Since $\pi_n(EG) = 0$: $\pi_n(BG) \cong \pi_{n-1}(G)$. The classifying space BG has homotopy groups that are one step up from those of G. For $G = U(1)$: $\pi_2(BU(1)) = \pi_1(U(1)) = \mathbb{Z}$ and all other $\pi_n(BU(1)) = 0$, making $BU(1) = K(\mathbb{Z}, 2)$, an Eilenberg-MacLane space.

8.8.4 Why Infinite-Dimensional Spaces Are Needed

For $G = U(1)$, we need a contractible space with a free $U(1)$ action. The sequence of spheres provides the answer:

Try S^1: $U(1)$ acts on S^1 freely (rotation). But $\pi_1(S^1) = \mathbb{Z} \neq 0$: not contractible. The quotient $S^1/U(1) = \{\text{pt}\}$ is too small.

Try $S^3 \subset \mathbb{C}^2$: $U(1)$ acts on S^3 freely via $e^{i\theta} \cdot (z_1, z_2) = (e^{i\theta} z_1, e^{i\theta} z_2)$. The quotient is $S^3/U(1) = \mathbb{CP}^1 = S^2$. But S^3 still has $\pi_3(S^3) = \mathbb{Z} \neq 0$: not contractible. The quotient $\mathbb{CP}^1 = S^2$ can only classify $U(1)$-bundles over spaces of dimension ≤ 2.

Try $S^{2n+1} \subset \mathbb{C}^{n+1}$: $U(1)$ acts freely. The quotient is $\mathbb{CP}^n$. The sphere S^{2n+1} has $\pi_k(S^{2n+1}) = 0$ for $k \leq 2n$, so $\mathbb{CP}^n$ classifies $U(1)$-bundles over manifolds of dimension $\leq 2n$.

The limit:

$$S^1 \subset S^3 \subset S^5 \subset \cdots \subset S^\infty = \bigcup_{n \geq 1} S^{2n-1}. \tag{8.10}$$

The infinite-dimensional sphere S^∞ is contractible ($\pi_k(S^\infty) = 0$ for all k): it "absorbs" all homotopy groups, one finite-dimensional approximation at a time. The free $U(1)$ action on S^∞ gives:

$$EU(1) = S^\infty, \qquad BU(1) = S^\infty/U(1) = \mathbb{CP}^\infty. \tag{8.11}$$

The infinite-dimensional projective space $\mathbb{CP}^\infty$ is the universal catalog for $U(1)$-bundles: any $U(1)$-bundle over any space is a pullback of the Hopf bundle $S^\infty \to \mathbb{CP}^\infty$.

The key point: we do not need infinite dimensions for their own sake. We need them because every finite-dimensional sphere has some nonzero homotopy group that limits its classifying ability. Only the infinite limit kills all homotopy groups and achieves true universality.

8.8.5 BG Is Not a Product: The Universal Twist

It is tempting to write $EG = BG \times G$ (since $BG = EG/G$, why not?). This is wrong, and the reason illuminates the whole theory.

The contradiction. If $EG = BG \times G$, then $EG \to BG$ is the trivial bundle. Every pullback of a trivial bundle is trivial. So every principal G-bundle over every M would be trivial. But the Möbius band and the Hopf bundle are non-trivial. Contradiction.

What is true locally. Over any contractible open set $U \subset BG$, the bundle $EU(1) \to BU(1)$ is trivializable: $\pi^{-1}(U) \cong U \times G$. The global non-triviality is encoded in how these local trivializations are glued, i.e., in the transition functions of the universal bundle.

The universal bundle contains all possible twists. Every possible transition function collection, up to gauge equivalence, appears as a sub-bundle of the universal bundle. This is what makes the classification work: any bundle over M is obtained by "selecting" a particular twisted sub-family of the universal bundle via the classifying map $f : M \to BG$.

8.8.6 The Cohomology of BG: Characteristic Classes

The cohomology ring $H^*(BG; \mathbb{Z})$ is the ring of **characteristic classes**: all natural invariants that can be assigned to any principal G-bundle.

Given a classifying map $f : M \to BG$, any cohomology class $c \in H^k(BG; \mathbb{Z})$ pulls back to $f^*(c) \in H^k(M; \mathbb{Z})$, which depends only on the isomorphism class of the bundle (since homotopic maps have the same pullback on cohomology). This is a characteristic class of the bundle.

For $G = U(n)$:

$$H^*(BU(n); \mathbb{Z}) = \mathbb{Z}[c_1, c_2, \ldots, c_n], \tag{8.12}$$

a polynomial ring. The generators $c_k \in H^{2k}(BU(n))$ are the **universal Chern classes**. For any rank-n complex vector bundle $E \to M$ classified by $f : M \to BU(n)$:

$$c_k(E) = f^*(c_k) \in H^{2k}(M; \mathbb{Z}). \tag{8.13}$$

These are the Chern classes of E: the complete set of topological obstructions to trivializing E. The bundle is trivializable iff all $c_k(E) = 0$.

8.9 Three Levels and a Summary Dictionary

A fiber bundle carries three levels of structure, each independent of the others:

Level 1: The bundle topology. Encoded in transition functions $\{g_{\alpha\beta}\}$, classified by $\check{H}^1(M;G) \cong [M, BG]$. This is the topological type: is the bundle the Möbius band or the cylinder? The Hopf bundle or the trivial bundle over S^2?

Level 2: A connection. How to compare neighboring fibers (parallel transport). Multiple connections can live on the same bundle. The curvature of a connection is a measure of how parallel transport fails to commute. This is the subject of Chapter 10.

Level 3: A section. A specific field configuration. May or may not exist; if it exists, there are many of them. For principal bundles, a section is a gauge choice.

Object	Mathematical meaning	Physical meaning
Base M	Spacetime/parameter space	Where physics lives
Fiber F	Space of internal states	Gauge degrees of freedom
Projection π	"Forget the fiber"	Which spacetime point
Local trivialization	$\pi^{-1}(U) \cong U \times F$	Local gauge convention
Transition function $g_{\alpha\beta}$	Gluing data on overlaps	Gauge transformation
Cocycle condition	$g_{\alpha\gamma} = g_{\alpha\beta} g_{\beta\gamma}$	Consistency
Gauge equivalence	$g' = h_\alpha^{-1} g\, h_\beta$	Relabeling conventions
$\check{H}^1(M;G)$	Space of orbits of cocycles	Topological types of bundles
Section $s : M \to E$	One fiber element per base point	A field configuration
Non-vanishing section	$s(x) \neq 0$ everywhere	Globally non-zero field
Trivial bundle	$E \cong M \times F$	No global twist
Euler class $e(E)$	Obstruction in $H^n(M;\mathbb{Z})$	Mandatory zeros of sections
Classifying map $f : M \to BG$	Selects bundle from catalog	Topological type
Chern class $c_k(E) = f^*(c_k)$	Obstruction to triviality	Topological charge

The basketball and compass return here as the ultimate summary: the basketball is M, the compass needle at each point is a section of TS^2, the combing question is whether a non-vanishing section exists, the answer (no) is the hairy ball theorem, and the reason (Euler class $= 2 \neq 0$) is the first characteristic class of the tangent bundle. The 3D compass frame is a section of the frame bundle, gauge transformations rotate the frame, and the topology of the frame bundle is classified by the classifying map $M \to BSO(n)$, whose cohomological pullbacks are the Pontryagin and Euler classes.

Chapter 9

Connections: How Fibers Twist

9.1 The Fundamental Idea: Fibers Twist as You Move

In Chapter 8 we built the Möbius band by introducing a half-twist into the transition function: as you cross from one patch to the other, the fiber flips sign. That discrete twist — encoded in a single ± 1 — was the entire difference between the trivial cylinder and the non-trivial Möbius band.

A **connection** is the continuous, infinitesimal, differential version of exactly that twist. Instead of specifying a discrete flip at patch boundaries, a connection specifies, at every point and in every direction, *how much the fiber twists as you take an infinitesimal step.* The transition functions of the bundle are the integrated version of the connection; the connection is the derivative of the transition function.

The three key quantities in this chapter are all aspects of the same twist:

- The **connection** A is the instantaneous twist rate: how much the fiber rotates per unit displacement in each direction.
- The **curvature** $F = dA + A \wedge A$ is the twist rate per unit area: how much twist accumulates around an infinitesimally small loop.
- The **holonomy** $\mathrm{Hol}(\gamma)$ is the total accumulated twist after going all the way around a finite loop γ.

If the fiber never twists regardless of path, the connection is **flat** ($F = 0$) and the holonomy is trivial. If the fiber twists as you move, the curvature is non-zero and the holonomy records the total twist.

The bridge from Chapter 8 to Chapter 11.

- Chapter 8 (bundles): global topological twist, encoded discretely in transition functions, classified by $\check{H}^1(M; G)$.
- This chapter (connections): continuous infinitesimal twist at each point, encoded in the connection 1-form A, with curvature F measuring the local twist rate.
- Chapter 11 (characteristic classes): total integrated twist over the whole manifold, encoded in integers like $\frac{1}{2\pi}\int F$ (first Chern class), topologically protected because they are the integrals of closed forms.

Why "parallel transport" is a misleading name. The standard term for carrying a fiber element along a path is *parallel transport*, which suggests "no change, kept constant". This is accurate locally — you are doing the most parallel thing the curved geometry allows — but misses the global point entirely. The whole point is that parallel transport *does* change things: after a round trip, the fiber element is twisted relative to where it started. A Foucault pendulum is "parallel transported" for twenty-four hours, and it returns twisted by $2\pi \sin(\text{latitude})$. The electron in the Aharonov-Bohm experiment is "parallel transported" around the solenoid and returns twisted by the phase $e^{-ie\Phi/\hbar}$. We use the word "transport" but the reader should always hear "twist accumulation".

An analogy. Think of a long flexible rod being pushed along a curved surface. Locally, at each instant, the rod is being held as straight as possible: no deliberate twisting is applied. But when you carry the rod around a loop and return to the start, the rod has been twisted relative to its original orientation, because the surface curved underneath it. The curvature of the surface is the twist-per-unit-area; the total twist after the circuit is the holonomy. The act of carrying the rod — always keeping it "as straight as possible" — is the connection.

This is not a deficiency we can ignore. In physics, we constantly need to compare: is the electron's phase at x consistent with its phase at y? Does the tangent vector at x "point the same way" as the tangent vector at y? Is the color of a quark field the same at two nearby points? None of these questions have answers without a rule for how the fiber twists from one point to the next. That rule is the connection.

The key structural point. A connection is an *additional choice* made on top of the bundle. It is not determined by the bundle topology alone; many different connections can live on the same bundle. The bundle tells us what fibers are attached at each point and how they are globally twisted (transition functions); the connection tells us how the fiber twists infinitesimally as we move. Different connections on the same bundle represent different rules for twisting, all consistent with the same global topology.

9.2 Three Ways to Specify the Twist Rate

A connection specifies the infinitesimal twist of the fiber at each point. There are three equivalent ways to make this precise, each revealing a different facet of the same geometry.

9.2.1 Picture 1: Following the Twist Along Paths

A connection on a principal G-bundle $\pi : P \to M$ specifies, for each smooth path $\gamma : [0,1] \to M$ and each starting point $p_0 \in \pi^{-1}(\gamma(0))$, a unique **horizontal lift**: a path $\tilde{\gamma} : [0,1] \to P$ with $\pi \circ \tilde{\gamma} = \gamma$ (it projects down to γ), $\tilde{\gamma}(0) = p_0$ (it starts at p_0), and $\dot{\tilde{\gamma}}(t)$ is *horizontal* at each t (it never moves along the fiber direction).

The horizontal lift is uniquely determined by the path and the starting point: given where you start, the rule "no fiber motion at each instant" determines the entire trajectory. This is the "as straight as possible" rule from our rod analogy. The endpoint $\tilde{\gamma}(1) \in \pi^{-1}(\gamma(1))$ is where the fiber element ends up after being carried along γ without deliberate twisting. The total twist is the difference between where it ends up and where it would have been if the bundle were trivial.

9.2.2 Picture 2: The Twist Direction at Each Point

At each point $p \in P$, the tangent space T_pP decomposes:

$$T_pP = V_pP \oplus H_pP, \tag{9.1}$$

where $V_pP = \ker d\pi_p$ is the **vertical subspace** (the fiber twist directions: moving in V_pP is pure fiber motion with no base displacement) and H_pP is the **horizontal subspace** chosen by the connection (the "no twist" directions: moving in H_pP carries you sideways in the base without twisting the fiber).

The connection declares, at each point, which directions count as "no fiber twist" (horizontal) and which are "pure fiber twist" (vertical). A path is horizontal — "carries the fiber element along without twisting it" — precisely when its velocity is always horizontal.

Different connections declare different directions to be "horizontal", giving different rules for what counts as "no twist in this direction". On a sphere, the horizontal subspace at each tangent frame point is the set of directions "along the sphere surface" with the fiber not rotating; but there is a family of choices for exactly what "not rotating" means, and different choices give different connections.

9.2.3 Picture 3: The Twist Rate as a Lie-Algebra-Valued Form

The most computationally useful picture: a connection is captured by a Lie-algebra-valued 1-form

$$A \in \Omega^1(M; \mathfrak{g}), \tag{9.2}$$

the **connection 1-form** or **gauge potential**. At each point $x \in M$, A takes a tangent vector $v \in T_xM$ and returns a Lie algebra element $A(v) \in \mathfrak{g}$: the infinitesimal group element that the fiber twists by when you step in direction v.

For $G = U(1)$ (electromagnetic gauge theory), $\mathfrak{g} = i\mathbb{R}$, and $A(v) = iA_\mu v^\mu$ tells us how much the $U(1)$ phase twists (per unit step in direction v). For $G = SU(2)$ (Yang-Mills), $\mathfrak{g} = \mathfrak{su}(2)$, and A describes how the color frame twists.

The connection encodes the twist rate of the transition functions. If the transition functions of the bundle are $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$, then the connection's local form A satisfies (near the boundary of U_α) a relation analogous to the derivative of $g_{\alpha\beta}$. The discrete boundary twist of Chapter 8 becomes the differential instantaneous twist of this chapter.

Under a gauge transformation $h : U \to G$ (change of local trivialization), the connection 1-form transforms as:

$$A \mapsto hAh^{-1} + h\,dh^{-1}. \tag{9.3}$$

For $G = U(1)$: $A \mapsto A + d\log h = A - id\Lambda$, the electromagnetic gauge transformation. The extra term $h\,dh^{-1}$ is the derivative of the change-of-gauge map: it is the infinitesimal twist introduced by choosing a new local convention.

The accumulated twist along a path. The twist of the fiber element f_0 as it is carried along γ is governed by the ODE "no net twist per step":

$$\frac{Df}{dt} = \dot{f}(t) + A(\dot{\gamma}(t)) \cdot f(t) = 0, \tag{9.4}$$

with solution — the total accumulated twist map:

$$f(t) = \mathcal{P}\exp\left(-\int_0^t A(\dot{\gamma}(s))\,ds\right) \cdot f_0. \tag{9.5}$$

The path-ordered exponential $\mathcal{P}\exp(-\int A)$ is the total finite twist accumulated along γ. For abelian $G = U(1)$, path-ordering is unnecessary and the twist is simply the phase $e^{-i\int_0^t A}$.

9.3 Holonomy: The Total Accumulated Twist

9.3.1 Definition: Twist That Does Not Return

For a closed loop $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1) = x_0$ and a starting fiber element p_0 over x_0, the horizontal lift $\tilde{\gamma}$ (the "no deliberate twist" path in the total space) starts at p_0 and returns to the fiber over x_0, but not necessarily to p_0 itself. Instead:

$$\tilde{\gamma}(1) = p_0 \cdot g \quad \text{for some } g \in G. \tag{9.6}$$

This g is the **holonomy** of the connection around γ:

$$\mathrm{Hol}(\gamma) = g = \mathcal{P} \exp\left(- \oint_\gamma A \right) \in G. \tag{9.7}$$

The holonomy is the total twist accumulated around the loop. You started with a fiber element p_0, carried it around without deliberately twisting it (staying horizontal at each step), and yet when you returned the fiber had been twisted by g relative to where it started. The twist came from the geometry of the base space and the fiber bundle, not from any applied force or deliberate rotation.

> *Zero local twist rate, nonzero total twist: this is the Aharonov-Bohm effect, the Foucault pendulum, the Berry phase, and geodetic precession — all in one sentence.*

If $g = e$ (identity) for every loop, the connection is **flat**: no twist is ever accumulated, regardless of the path. If $g \neq e$ for some loop, the connection has **curvature**: there are loops around which the fiber twists non-trivially. The curvature F is the infinitesimal twist per unit area; the holonomy is the finite total twist around a finite loop.

Global twist without local twist. The most striking case: $F = 0$ everywhere (zero local twist rate) yet $\mathrm{Hol}(\gamma) \neq e$ for some loop. This is possible when the domain is *not simply connected*: the loop cannot be contracted to a point, so the zero local twist cannot be "unrolled" into a global trivialization. The hole in the domain stores a global twist that has no local signature. This is the Aharonov-Bohm effect.

9.3.2 Twist on the Sphere: An Explicit Calculation

The most vivid example of accumulated twist is parallel transport on the unit sphere S^2. We carry a tangent vector around a triangular loop, doing our best to "not twist it" at each step. The result: it comes back twisted.

Setup. Start at the point $P = (1,0,0)$ on the equator with a tangent vector $\mathbf{v}_0$ pointing north.

Leg 1: Along the equator from $P = (1,0,0)$ to $Q = (0,1,0)$ (a quarter of the equator, a geodesic). The "no twist" rule along a geodesic keeps the vector at a constant angle to the geodesic. Since $\mathbf{v}_0$ points north and the equator runs east, the vector continues to point north. No twist on this leg. Arriving at Q: $\mathbf{v}_1$ still points north.

Leg 2: Along the meridian from $Q = (0,1,0)$ to the north pole $N = (0,0,1)$ (another geodesic). Again no twist along the geodesic. The vector $\mathbf{v}_1$ (pointing north, aligned with the meridian direction) stays aligned with the meridian. Arriving at N: the vector points toward Q, i.e., in the $-y$ direction.

Leg 3: Down the prime meridian from $N = (0,0,1)$ back to $P = (1,0,0)$. No twist along the geodesic. The vector (pointing in the $-y$ direction at N) is carried down maintaining its angle with the meridian. Arriving at P: the vector points west, in the $-y$ direction.

The twist. We started pointing north ($+z$ direction at P) and returned pointing west ($-y$ direction at P): a 90ř twist. We never deliberately twisted the vector at any step. The twist came entirely from carrying it around a curved surface.

Why 90ř**?** The twist angle equals the **solid angle** enclosed by the loop. Our triangle encloses one-eighth of S^2: solid angle $\frac{1}{8} \cdot 4\pi = \frac{\pi}{2}$ steradians. Twist $= \frac{\pi}{2} = 90$ř. By Stokes:

$$\mathrm{Hol}(\gamma) = \exp\Big(-\int_\Sigma K \, dA\Big), \tag{9.8}$$

where $K = 1$ is the Gaussian curvature and Σ is any surface bounded by γ. The curvature is the twist-per-unit-area; integrating it gives the total twist.

Spherical excess = total twist. For a geodesic triangle with angles α, β, γ:

$$\text{total twist} = \alpha + \beta + \gamma - \pi = \text{area}. \tag{9.9}$$

On a flat plane, $\alpha + \beta + \gamma = \pi$ exactly: no twist. On a positively curved surface, the angle sum exceeds π: the excess measures the twist. Curvature = angle excess = holonomy = total accumulated twist.

9.3.3 The Foucault Pendulum: Twist from Living on a Sphere

A Foucault pendulum swings in a plane. In Paris (latitude 48.9řN), the plane of oscillation rotates approximately 11.3ř per hour — one full rotation in about 31.8 hours. No horizontal force acts on the pendulum. Why does the plane twist?

The answer: the Earth's daily rotation carries each point on the surface around a latitude circle, and the pendulum's oscillation plane is the fiber being carried. The fiber (the plane of oscillation) undergoes the "no deliberate twist" transport — the Levi-Civita connection on the sphere — as the latitude circle closes after one day. After a full circuit, the plane has been twisted by the holonomy.

The solid angle subtended by a latitude circle at colatitude θ is $\Omega = 2\pi(1 - \cos\theta)$. At Paris, $\theta \approx 41.1$ř, giving $\Omega \approx 1.974$ steradians and a twist of ≈ 113ř per day (≈ 4.7ř per hour). The precise formula is twist per day $= 2\pi \sin(\text{latitude})$.

The twist came from nowhere and everywhere. There is no force twisting the pendulum. The twist came from going around a loop on a curved surface. The curvature of the Earth — its Gaussian curvature — is the local twist rate of the Levi-Civita connection; the daily accumulated twist is the holonomy of that connection around the latitude circle.

9.3.4 The Aharonov-Bohm Effect: Phase Twist Without Local Field

An electron moves outside a solenoid carrying magnetic flux Φ. Outside the solenoid: $\mathbf{B} = 0$, so the curvature $F = dA = 0$. The local twist rate of the $U(1)$ connection is zero everywhere on the electron's path. Yet the connection is not flat globally — it is a closed-but-not-exact 1-form, a familiar obstruction from Chapter 5.

When the electron completes a loop γ around the solenoid, it accumulates a total phase twist:

$$\mathrm{Hol}(\gamma) = e^{-i \oint_\gamma A} = e^{-i\Phi/\hbar}. \tag{9.10}$$

The local twist rate is zero everywhere; the total twist after one circuit is $\Phi/\hbar$. The twist did not accumulate from any local source on the path: it was stored globally in the topology of the domain. The solenoid creates a hole in $\mathbb{R}^2 \setminus \{\text{solenoid}\}$, and the hole stores a non-trivial holonomy that has no local curvature signature.

Zero local twist, nonzero global twist. This is the clearest possible example of the distinction between the local twist rate (curvature, F) and the total accumulated twist (holonomy, $\oint A$). On a simply connected domain, these are related by Stokes: $\oint A = \int F$. On a non-simply connected domain, the Stokes relation fails because there is no spanning surface: the total twist $\oint A$ is not determined by any integral of the local twist rate F. The topology of the domain allows a nonzero total twist even with zero local twist rate.

9.3.5 The Berry Phase: Quantum State Twist in Parameter Space

When a quantum system's Hamiltonian $H(\lambda)$ depends on parameters λ that are adiabatically varied around a closed loop Γ in parameter space $\mathcal{P}$, the ground state $|\psi(\lambda)\rangle$ is the fiber being carried. The Berry connection

$$\mathcal{A} = i\langle\psi(\lambda)|d|\psi(\lambda)\rangle \tag{9.11}$$

is the connection 1-form specifying the instantaneous twist rate of the quantum state as parameters change. The "no deliberate twist" rule (the horizontal lift condition) is precisely the adiabatic condition: the state follows the instantaneous ground state without absorbing excitations, which is the "as straight as possible" transport in parameter space.

After completing the loop Γ, the state returns to itself up to a phase twist — the Berry phase:

$$e^{i\gamma_B} = e^{i\oint_\Gamma \mathcal{A}} = e^{i\int_\Sigma \mathcal{F}}, \tag{9.12}$$

where $\mathcal{F} = d\mathcal{A}$ is the Berry curvature (local twist rate in parameter space). For a spin-$\frac{1}{2}$ particle with parameter space $\cong S^2$ (directions of the magnetic field), the Berry phase for a loop enclosing solid angle Ω is $\gamma_B = -\Omega/2$.

The spinor sign flip. A spin-$\frac{1}{2}$ particle acquires a phase of $e^{i\pi} = -1$ when the magnetic field is rotated by 2π (enclosing solid angle 4π, giving $\gamma_B = -2\pi$). This is the topological origin of spinor sign flip under 2π rotation: the quantum state is twisted by -1 as the field completes a full rotation, a direct consequence of the Berry connection's holonomy around the parameter-space sphere.

9.3.6 Twist Accumulation: A Physical Dictionary

Every observable in the following table is an instance of the same mathematical object — the holonomy of a connection — seen in five different physical settings. In each case a fiber element is carried around a loop without deliberate twisting, and the total accumulated twist is the observable.

Physical system	What twists	Connection	Loop	Total twist (holonomy)
Transport on S^2	Tangent vector	Levi-Civita	Geodesic triangle	Solid angle
Foucault pendulum	Oscillation plane	Levi-Civita	Latitude circle/day	$2\pi\sin(\text{lat})$
Aharonov-Bohm	$U(1)$ phase	$A = \Phi/2\pi r\, d\phi$	Loop around solenoid	$e^{-i\Phi/\hbar}$
Berry phase	Quantum state	$\mathcal{A} = i\langle\psi\|d\psi\rangle$	Loop in $\mathcal{P}$	$e^{i\int_\Sigma \mathcal{F}}$
Geodetic precession	Gyroscope axis	Levi-Civita	Satellite orbit	$6.6''$/year

9.4 Curvature: The Local Twist Rate per Unit Area

9.4.1 Definition

The **curvature** of a connection A is the measure of how fast the fiber twists as you enclose an infinitesimally small area. Formally, it is the $\mathfrak{g}$-valued 2-form:

$$F = dA + A \wedge A. \tag{9.13}$$

The dA term is the exterior derivative of the twist rate (how the twist rate changes from point to point). The $A \wedge A$ term is the self-coupling of the twist: for non-abelian groups, a rotation in one fiber direction followed by a rotation in another is not the same as doing them in the reverse order, and the commutator $[A, A]$ captures this non-commutativity.

For $G = U(1)$ (abelian): $A \wedge A = 0$ and $F = dA$. The curvature is simply the exterior derivative of the connection: the local rate of change of the twist rate.

Infinitesimal twist around a tiny loop. For an infinitesimal parallelogram in the (μ, ν)-plane with sides ϵ and δ, the total accumulated twist (holonomy) around the loop is:

$$\text{Hol}(\text{tiny loop}) = e^{-F_{\mu\nu}\epsilon\delta} + O(\epsilon^2\delta^2), \tag{9.14}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$. The curvature 2-form F is exactly the twist per unit area. Integrating F over a finite region gives the holonomy around its boundary (Stokes):

$$\text{Hol}(\partial\Sigma) = \mathcal{P}\exp\left(-\int_\Sigma F\right). \tag{9.15}$$

Flat = no local twist. The connection is **flat** if $F = 0$ everywhere: no region accumulates any twist. A connection that twists everywhere by the same amount in all directions has $F = 0$: it is flat (the twists cancel each other when integrated around any contractible loop). Only twists that fail to cancel when you go around a loop contribute to holonomy.

9.4.2 Gauge Transformation of the Twist Rate

Under a gauge transformation $h : M \to G$, the curvature transforms *covariantly*:

$$F \mapsto hFh^{-1}. \tag{9.16}$$

For abelian $G = U(1)$: $F \mapsto F$ (invariant). The curvature is gauge-invariant in electromagnetism. For non-abelian G: the curvature transforms by conjugation. The trace $\text{Tr}(F^k)$ is invariant under conjugation, so it is gauge-invariant and gives physical observables (the characteristic classes).

9.4.3 The Bianchi Identity: The Twist Rate Has No Divergence

The curvature (twist rate) satisfies:

$$DF \coloneqq dF + [A, F] = 0, \tag{9.17}$$

the **Bianchi identity**. This is the bundle-theoretic analog of $d^2 = 0$: the covariant derivative of the curvature vanishes. In physical terms: *the twist rate has no further source.* Just as $d^2 = 0$ says the exterior derivative of an exterior derivative is zero (bulk of a bulk is zero), $DF = 0$ says the covariant derivative of the curvature is zero: the curvature 2-form is "already the full twist rate" with no further derivative-level structure.

For $G = U(1)$: $DF = dF = d(dA) = 0$ by nilpotency. This encodes Faraday's law and $\nabla \cdot \mathbf{B} = 0$: the magnetic field (twist rate) has no magnetic charges as sources. For non-abelian G: the extra term $[A, F]$ accounts for the self-coupling of the twist rate in the non-abelian group structure.

9.5 Form Notation vs. Tensor Notation

Both the form (exterior calculus) notation and the tensor index notation describe the same geometry, but they emphasize different aspects and have complementary strengths. Knowing when to use each is essential for reading modern physics and geometry literature.

9.5.1 What Each Notation Is Independent Of

Form notation $F = dA + A \wedge A$ is:

- **Coordinate-independent**: No indices μ, ν; no choice of chart.
- **Frame-independent**: No choice of local basis for the fiber; no vierbein needed.
- **Metric-independent** (mostly): Exterior derivatives d, wedge products $\wedge$, and curvature $F = dA + A \wedge A$ require no metric. Only the Hodge star $\star$ (needed for $d \star F$) requires a metric.
- Naturally adapted to **integration**: forms integrate over chains without any additional structure.
- Makes **gauge invariance manifest**: $F \to hFh^{-1}$ is clearly covariant; $\mathrm{Tr}(F \wedge F)$ is manifestly gauge-invariant.

Tensor index notation $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ is:

- **Coordinate-dependent**: indices μ, ν refer to a specific chart.
- **Metric-dependent for contractions**: raising/lowering indices requires $g^{\mu\nu}$.
- Makes **components explicit**: $E_x, B_y, \ldots$ have direct physical meaning.
- Essential for **explicit computation**: wave equations, perturbation theory, coupling to matter.
- More natural for **variational calculations**: deriving Euler-Lagrange equations from $\mathcal{L}(A_\mu)$.

9.5.2 Side-by-Side Comparison

Concept	Form notation	Tensor notation
Connection	$A \in \Omega^1(M; \mathfrak{g})$	A^a_μ, gauge potential
Curvature	$F = dA + A \wedge A$	$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\nu$
Gauge transform	$A \to hAh^{-1} + hdh^{-1}$	$A_\mu \to hA_\mu h^{-1} + h\partial_\mu h^{-1}$
Bianchi identity	$DF = 0$	$D_{[\mu} F_{\nu\rho]} = 0$
Field equation	$D \star F = J$	$D^\mu F^a_{\mu\nu} = J^a_\nu$
Gauge invariant	$\mathrm{Tr}(F \wedge F)$	$\mathrm{Tr}(F_{\mu\nu} F^{\mu\nu})$
Stokes theorem	$\oint_\gamma A = \int_\Sigma F$	$\oint A_\mu dx^\mu = \int F_{\mu\nu} dS^{\mu\nu}$

The form notation is the language of mathematical structure; the tensor notation is the language of physical computation. Modern physics uses both, switching between them depending on the task.

9.6 Electromagnetism: A Complete Treatment

9.6.1 The Geometric Setting

Electromagnetism is a gauge theory with structure group $G = U(1)$. The gauge bundle is a principal $U(1)$-bundle $\pi : P \to M$ over spacetime M (Minkowski space $\mathbb{R}^{1,3}$ or a general 4-manifold). The electromagnetic vector potential A is the connection 1-form; the field strength F is the curvature 2-form.

9.6.2 The Electromagnetic Field in Form Notation

The **vector potential** is a real-valued 1-form $A = A_\mu\, dx^\mu$ on M (the $\mathfrak{u}(1) = i\mathbb{R}$ component is factored out). The **field strength** (curvature) is:

$$F = dA = \sum_{\mu<\nu} F_{\mu\nu}\, dx^\mu \wedge dx^\nu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{9.18}$$

($A \wedge A = 0$ for $U(1)$.) In terms of electric and magnetic fields (with $x^0 = t$, x^i spatial):

$$F = E_i\, dt \wedge dx^i + \tfrac{1}{2}\epsilon_{ijk} B_k\, dx^i \wedge dx^j. \tag{9.19}$$

Explicitly: $F_{0i} = E_i$ (electric field components), $F_{ij} = \epsilon_{ijk} B_k$ (magnetic field components).

Maxwell's equations in form notation:

$$dF = 0, \tag{9.20}$$

$$d \star F = J. \tag{9.21}$$

Here J is the 3-form current and $\star$ is the Hodge star with respect to the Minkowski metric.

Decoding (9.20): $dF = d(dA) = 0$ is automatic by nilpotency. Expanding in components, $dF = 0$ gives

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0 \quad \text{(6 independent equations)}. \tag{9.22}$$

Setting $\mu\nu\rho = 0ij$: $\partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i} = 0$, which is Faraday's law $\partial_t \mathbf{B} + \nabla \times \mathbf{E} = 0$. Setting $\mu\nu\rho = ijk$: $\nabla \cdot \mathbf{B} = 0$.

Decoding (9.21): The Hodge star converts $F_{\mu\nu}$ to its dual $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}$ (exchanging $\mathbf{E}$ and $\mathbf{B}$). Then $d \star F = J$ gives:

$$\partial_\mu \tilde{F}^{\mu\nu} = J^\nu \quad \Leftrightarrow \quad \nabla \cdot \mathbf{E} = \rho, \quad \nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{j}. \tag{9.23}$$

All four Maxwell equations in two lines: $dF = 0$ and $d \star F = J$. The first pair (Faraday + Gauss for B) are the Bianchi identity, automatic from $F = dA$. The second pair (Ampère + Gauss for E) are the dynamical equations of motion.

In tensor notation (for comparison):

$$\partial_{[\mu} F_{\nu\rho]} = 0, \qquad \partial^\mu F_{\mu\nu} = J_\nu \quad (= 0 \text{ in vacuum}), \tag{9.24}$$

where the first uses antisymmetrization and the second requires the metric to raise the index on ∂.

9.6.3 Gauge Invariance

The gauge transformation $A \mapsto A + d\Lambda$ (for any smooth function $\Lambda : M \to \mathbb{R}$) leaves the field strength invariant:

$$F \mapsto d(A + d\Lambda) = dA + d(d\Lambda) = dA = F. \tag{9.25}$$

Physical observables are gauge-invariant: they depend on F, not on the specific choice of A.

The gauge orbit. All potentials $A' = A + d\Lambda$ for any Λ lie in the same gauge orbit and produce the same physics. The orbit is a coset of the space of all 1-forms modulo the subspace of exact 1-forms — precisely the de Rham cohomology perspective: two potentials describe the same physics iff they are cohomologous.

Global obstructions. If M has non-trivial $H^2_{\mathrm{dR}}(M)$ (e.g., $M = \mathbb{R}^3 \setminus \{0\}$ for a monopole), then F may be a closed 2-form that is not globally exact: no global potential A exists. The obstruction is the cohomology class $[F] \in H^2(M; \mathbb{Z})$, the first Chern class c_1 of the $U(1)$ bundle.

9.6.4 Holonomy in Electromagnetism: The Wilson Loop

For a closed curve γ in spacetime, the holonomy of the electromagnetic connection is:

$$W(\gamma) = e^{-i \oint_\gamma A}. \tag{9.26}$$

This is the **Wilson loop**: a gauge-invariant observable (the trace of the holonomy). By Stokes: $W(\gamma) = e^{-i \int_\Sigma F}$ for any surface Σ bounded by γ. The Wilson loop is physical: it is what the electron's wavefunction accumulates in phase as it traverses γ. If $F = 0$ inside Σ (like outside a solenoid) but $\int_\Sigma F = \Phi \neq 0$ (because A is non-exact), the holonomy is still non-trivial: the Aharonov-Bohm effect.

9.7 General Relativity: A Complete Treatment

9.7.1 Gravity as a Connection on the Frame Bundle

General relativity is a gauge theory with structure group $G = SO(1,3)$ (the Lorentz group). The bundle is the **frame bundle** P of spacetime: the principal $SO(1,3)$-bundle of oriented orthonormal frames (local Lorentz frames, or vierbeins). The connection on this bundle is the **spin connection** ω; its curvature is the **Riemann curvature tensor**.

9.7.2 The Vierbein: A Local Frame at Each Point

A **vierbein** (or tetrad) is a local section of the frame bundle: a choice of orthonormal frame $\{e_a\}_{a=0}^3$ at each spacetime point, where a is a **local Lorentz index** (frame index) and the spacetime coordinates give a separate **coordinate index** μ. The vierbein 1-forms are:

$$e^a = e^a{}_\mu \, dx^\mu, \qquad g_{\mu\nu} = \eta_{ab} \, e^a{}_\mu e^b{}_\nu, \tag{9.27}$$

where $\eta_{ab} = \mathrm{diag}(-1, +1, +1, +1)$ is the flat Minkowski metric. The vierbein encodes the metric; the metric encodes the vierbein up to a local Lorentz transformation.

The **spin connection** $\omega^a{}_b = \omega^a{}_{b\mu} \, dx^\mu$ is the connection 1-form on the frame bundle. It tells us how the local frame rotates as we move from one spacetime point to another.

9.7.3 Cartan's Structure Equations: Form Notation for GR

Cartan's structure equations give the torsion and curvature in form notation:

$$T^a = de^a + \omega^a{}_b \wedge e^b \qquad \text{(torsion 2-form)}, \tag{9.28}$$

$$R^a{}_b = d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b \qquad \text{(curvature 2-form)}. \tag{9.29}$$

Compare (9.29) with the Yang-Mills curvature $F = dA + A \wedge A$: they are the same formula. The spin connection ω plays the role of the gauge potential A, and the curvature 2-form $R^a{}_b$ plays the role of the field strength F.

The **torsion-free condition** $T^a = 0$ says that the vierbein (encoding the metric) and the spin connection are consistent: the connection is determined by the metric. This determines ω uniquely from e, giving the **Levi-Civita connection**.

9.7.4 The Riemann Tensor in Both Notations

In form notation: The Riemann curvature is $R^a{}_b \in \Omega^2(M)$ with values in $\mathfrak{so}(1,3)$. Its components give the Riemann tensor:

$$R^a{}_b = \tfrac{1}{2} R^a{}_{b\mu\nu}\, dx^\mu \wedge dx^\nu. \tag{9.30}$$

In tensor notation: The Riemann tensor $R^\rho{}_{\sigma\mu\nu}$ is defined by:

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu) V^\rho = R^\rho{}_{\sigma\mu\nu} V^\sigma, \tag{9.31}$$

the failure of covariant derivatives to commute. In terms of the Christoffel symbols $\Gamma^\rho_{\mu\nu}$ (the tensor-notation version of the connection):

$$R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}. \tag{9.32}$$

This is the same formula as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$, with Γ playing the role of A and the commutator term $[\Gamma, \Gamma]$ reflecting the non-abelian structure of $SO(1,3)$.

Holonomy interpretation. The equation (9.31) says: commuting covariant derivatives is not free. Transport a vector V around an infinitesimal parallelogram in the (μ,ν)-plane; it returns rotated by $R^\rho{}_{\sigma\mu\nu}\epsilon\delta$ (to leading order). This is the gravitational version of the infinitesimal holonomy formula: curvature is infinitesimal holonomy.

9.7.5 Einstein's Equations in Form Notation

From the curvature 2-form $R^a{}_b$, define the Ricci scalar and Einstein tensor by contraction:

$$R_{ab} = R^c{}_{acb} \quad \text{(Ricci tensor)}, \tag{9.33}$$

$$G_{ab} = R_{ab} - \tfrac{1}{2}\eta_{ab} R \quad \text{(Einstein tensor)}. \tag{9.34}$$

Einstein's field equations, in Cartan's form language, are:

$$\epsilon_{abcd}\, R^{ab} \wedge e^c \wedge e^d = 16\pi G \star T, \tag{9.35}$$

where T is the stress-energy 3-form and ϵ_{abcd} is the totally antisymmetric symbol. This is a single equation of 2-forms, equivalent to the ten component equations:

$$G_{\mu\nu} = 8\pi G\, T_{\mu\nu}. \tag{9.36}$$

Comparison with Yang-Mills. The Yang-Mills equation is $D \star F = J$; Einstein's equation (in first-order form) is $D(R \wedge e) = T$. Both equate the covariant divergence of the curvature to the source. The structural parallel is exact:

	Electromagnetism	**General Relativity**
Bundle	$U(1)$-bundle over M	Frame bundle $P \to M$
Group	$U(1)$	$SO(1,3)$
Connection	$A = A_\mu\, dx^\mu$	$\omega^a{}_b = \omega^a{}_{b\mu}\, dx^\mu$
Curvature	$F = dA$	$R^a{}_b = d\omega^a{}_b + \omega \wedge \omega$
Bianchi	$dF = 0$	$DR^a{}_b = 0$
Field equation	$d \star F = J$	$\epsilon_{abcd} R^{ab} \wedge e^c \wedge e^d = T$
Gauge transform	$A \to A + d\Lambda$	$e \to \Lambda e$ (diffeomorphism)
Physical field	$F_{\mu\nu}$ ($\mathbf{E}, \mathbf{B}$)	$R^\rho{}_{\sigma\mu\nu}$ (tidal forces)

9.7.6 The Levi-Civita Connection and Geodesics

In tensor notation, the Levi-Civita connection is the Christoffel symbol:

$$\Gamma^\rho_{\mu\nu} = \tfrac{1}{2} g^{\rho\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}). \tag{9.37}$$

This is the unique torsion-free metric-compatible connection.

A **geodesic** is a curve that parallel transports its own tangent vector:

$$\frac{D\dot\gamma^\mu}{dt} = \ddot\gamma^\mu + \Gamma^\mu_{\rho\nu}\dot\gamma^\rho\dot\gamma^\nu = 0. \tag{9.38}$$

In flat space, $\Gamma = 0$ and geodesics are straight lines. In curved spacetime, geodesics are the worldlines of freely falling observers: the paths of "no acceleration" in the gravitational field. The curvature of the connection causes geodesics to converge or diverge: geodesic deviation is measured by the Riemann tensor.

Holonomy and tidal forces. If you carry a gyroscope along a geodesic (freely falling orbit) in curved spacetime, its axis precesses relative to the distant stars. This **geodetic precession** is the holonomy of the Levi-Civita connection along the orbit. For a low-orbit satellite, it amounts to about 6.6 arcseconds per year (measured by Gravity Probe B). The precession is:

$$\Delta\phi = \oint_\gamma \omega \sim \int_\Sigma R, \tag{9.39}$$

a direct macroscopic measurement of spacetime curvature via holonomy.

9.8 Yang-Mills Theory: Non-Abelian Gauge Fields

For a non-abelian gauge group G (e.g., $SU(2)$ for electroweak, $SU(3)$ for QCD), the curvature is:

$$F^a = dA^a + \tfrac{1}{2} f^{abc} A^b \wedge A^c, \tag{9.40}$$

where f^{abc} are the structure constants of $\mathfrak{g}$. The $A \wedge A$ term represents the self-interaction of the non-abelian gauge field: gluons interact with other gluons (unlike photons, which do not interact with each other).

The Yang-Mills equations:

$$D \star F = J \qquad \text{(field equation)}, \tag{9.41}$$

$$DF = 0 \qquad \text{(Bianchi identity)}. \tag{9.42}$$

These generalize Maxwell's equations to non-abelian gauge groups. For $G = U(1)$: $D = d$ and we recover Maxwell. For $G = SU(3)$: these are the equations of QCD (quantum chromodynamics) governing gluon fields and strong interactions.

Instantons. On an oriented Euclidean 4-manifold, $\star$ acts on 2-forms and satisfies $\star^2 = 1$. Decompose $F = F^+ + F^-$ into self-dual ($\star F^+ = F^+$) and anti-self-dual ($\star F^- = -F^-$) parts. An **instanton** satisfies $F = \star F$ (self-dual curvature). Instantons:

- Automatically satisfy the Yang-Mills equations ($DF = 0$ by Bianchi, and $D \star F = DF = 0$).
- Minimize the Yang-Mills action in their topological class (the Bogomolny bound $\|F\|^2 \geq 8\pi^2 |c_2|$, with equality iff $F = \pm \star F$).
- Are classified by the second Chern class $c_2 \in \mathbb{Z}$ (the instanton number), a homotopy invariant.

9.9 Flat Connections and Holonomy Representations

Theorem 9.1. *A connection on a principal bundle over a simply connected base is flat ($F = 0$) if and only if all holonomies are trivial.*

On a non-simply connected base, a flat connection can still have non-trivial holonomy around non-contractible loops. The holonomy group of a flat connection depends only on the loop's homotopy class: parallel transport around homotopic loops gives the same group element. This defines a **holonomy representation**:

$$\rho : \pi_1(M) \to G, \qquad [\gamma] \mapsto \mathrm{Hol}(\gamma). \tag{9.43}$$

Theorem 9.2 (Holonomy Theorem)**.** *Isomorphism classes of flat G-bundles over M with basepoint correspond bijectively to representations $\rho : \pi_1(M) \to G$, up to conjugation in G.*

Examples.

- The Aharonov-Bohm effect: flat $U(1)$ connection on $\mathbb{R}^2 \setminus \{0\}$ with $\pi_1 = \mathbb{Z}$. The representation $\rho : \mathbb{Z} \to U(1)$ sends the generator to $e^{-i\Phi/\hbar}$. Each value of Φ gives a distinct (inequivalent) flat connection and distinct Aharonov-Bohm phase.
- On a torus T^2: $\pi_1(T^2) = \mathbb{Z}^2$. Flat $U(1)$ connections are parametrized by two angles (θ_1, θ_2), the holonomies around the two fundamental cycles. These are the **Wilson lines** wrapping the two cycles.
- In Chern-Simons theory on a 3-manifold M_3: the partition function is a sum over flat connections on M_3, which by the holonomy theorem is a sum over representations $\rho : \pi_1(M_3) \to G$. Each representation contributes a quantum amplitude, giving topological invariants of M_3.

9.10 Summary: The Three Faces of Twist

Everything in this chapter is a facet of one idea: fibers twist as you move through a bundle. The connection specifies the instantaneous twist rate; the curvature is the twist rate per unit area; the holonomy is the total accumulated twist around a loop.

Concept	Twist language	Mathematical form	Physical example
Connection A	Instantaneous twist rate	$A \in \Omega^1(M;\mathfrak{g})$	**E**, **B** potential
Gauge transform	Relabeling twist convention	$A \to hAh^{-1} + hdh^{-1}$	Change of gauge
Curvature F	Twist rate per unit area	$F = dA + A \wedge A$	**E**, **B** field
Flat connection	Zero local twist rate	$F = 0$ everywhere	Vacuum EM
Holonomy $\mathrm{Hol}(\gamma)$	Total accumulated twist	$\mathcal{P}\exp(-\oint A)$	AB phase, precession
Global twist without local	Hole stores twist	$F = 0$ but $\oint A \neq 0$	Aharonov-Bohm
Bianchi $DF = 0$	Twist rate has no source	$d\mathrm{Tr}(F^k) = 0$	Faraday / $\nabla \cdot B = 0$
Field equation $D \star F = J$	Matter sources twist	$\partial^\mu F_{\mu\nu} = J_\nu$	Maxwell / Einstein
Wilson loop	Observable twist	$e^{-i\oint A}$	Interference pattern
Characteristic class	Total twist = integer	$\frac{1}{2\pi}\int F \in \mathbb{Z}$	Quantized charge

The narrative arc from Chapter 8 to Chapter 11. In Chapter 8, a bundle was built from discrete twist data: transition functions specifying how the fiber flips or rotates at patch boundaries. The global topology of the bundle is completely encoded in these discrete twists. This chapter adds the continuous version: a connection specifies the infinitesimal twist at every point in every direction. Integrating this twist around a loop gives the holonomy; integrating the curvature (twist rate per area) over the whole manifold gives a characteristic class.

The three quantities — connection (twist rate), curvature (twist rate per area), holonomy (total twist) — are related by:

$$\text{connection } A \xrightarrow{d+A\wedge A} \text{curvature } F \xrightarrow{\int_\Sigma} \text{holonomy } \mathrm{Hol}(\partial\Sigma). \tag{9.44}$$

The Bianchi identity $DF = 0$ says the curvature is "already complete": it needs no further differentiation to be a closed form. The Chern-Weil theorem (Chapter 11) says the integral of the curvature over the whole manifold is a topological integer: the total twist, accumulated by integrating the local twist rate over all of M, is quantized. The topology of the bundle is the ultimate origin of the quantization.

In every physical example — the Foucault pendulum, the Aharonov-Bohm effect, the Berry phase, the geodetic precession of a gyroscope — the observable effect is accumulated twist: a fiber element is carried around a loop without deliberate twisting, and returns twisted. The twist is real, measurable, and quantized. That is the content of this chapter.

Chapter 10

Characteristic Classes: Topology from Curvature

10.1 The Question Behind Everything

Every chapter since Chapter 5 has been building toward a single question: given a vector bundle $E \to M$, what topological invariants can we extract from it?

We know the answer in principle. The classifying map $f : M \to BG$ (Chapter 7) encodes the bundle's topology, and cohomology classes in $H^*(BG; \mathbb{Z})$ pull back to give invariants of E. These are the characteristic classes. But this answer is abstract: working with $BG = \mathbb{CP}^\infty$ and classifying maps is computationally unwieldy.

The Chern-Weil theorem gives us a completely different route to the same invariants: *compute them analytically from the curvature of any connection on E*. Pick any gauge potential A, compute its curvature $F = dA + A \wedge A$, form the invariant polynomial $\mathrm{Tr}(F^k)$, and you have a closed differential form whose cohomology class is a characteristic class of E.

The argument, once you have the right tools, is almost a one-liner. Building the right tools is what this chapter is about.

10.2 From Curvature to Cohomology: The One-Line Argument

Let $E \to M$ be a rank-n complex vector bundle with connection A and curvature $F = dA + A \wedge A \in \Omega^2(M; \mathfrak{gl}(n))$ (a matrix-valued 2-form). We want to construct closed differential forms from F that are natural invariants of E.

10.2.1 Step 1: Kill the gauge dependence with the trace

The curvature F is not gauge-invariant: under $A \mapsto gAg^{-1} + gdg^{-1}$, the curvature transforms as $F \mapsto gFg^{-1}$. It is only covariant, not invariant.

But the **trace** of any power of F is gauge-invariant:

$$\mathrm{Tr}((gFg^{-1})^k) = \mathrm{Tr}(gF^kg^{-1}) = \mathrm{Tr}(F^k). \tag{10.1}$$

The cyclic property of the trace absorbs the gauge transformation. More generally, any **symmetric invariant polynomial** P on $\mathfrak{gl}(n)$ — a polynomial satisfying $P(gXg^{-1}) = P(X)$ for all g — gives a gauge-invariant form $P(F)$.

The invariant polynomials are generated by $\mathrm{Tr}(F^k)$ for $k = 1, \ldots, n$ (or equivalently by the elementary symmetric polynomials σ_k of the eigenvalues of F). So we have a recipe: start with the curvature, take symmetric invariant polynomials, and get gauge-invariant differential forms.

10.2.2 Step 2: The Bianchi identity makes them closed

The curvature satisfies the Bianchi identity $DF = dF + [A, F] = 0$. From this:

$$\begin{aligned} d\,\mathrm{Tr}(F^k) &= k\,\mathrm{Tr}(dF \cdot F^{k-1}) = k\,\mathrm{Tr}((DF - [A,F]) \cdot F^{k-1}) \\ &= k\,\mathrm{Tr}(0 \cdot F^{k-1}) - k\,\mathrm{Tr}([A,F] \cdot F^{k-1}) \\ &= -k\,\mathrm{Tr}(AF^k - F^k A) = -k\,\mathrm{Tr}([A, F^k]) = 0, \end{aligned} \tag{10.2}$$

using the cyclic property of the trace in the last step. So $\mathrm{Tr}(F^k)$ is **closed**: $d\,\mathrm{Tr}(F^k) = 0$.

10.2.3 The result so far

From these two steps: any symmetric invariant polynomial applied to the curvature gives a **gauge-invariant closed differential form**. Such a form represents a class in $H^*_{\mathrm{dR}}(M; \mathbb{R})$.

This is the essence of Chern-Weil: curvature, passed through an invariant polynomial, produces a de Rham cohomology class. Steps 1 and 2 are literally two lines.

10.2.4 Step 3: The class is independent of the connection

The non-trivial content is that different connections on E give cohomologous forms. This is what makes the class a topological invariant of the bundle, not just of the connection.

Suppose A_0 and A_1 are two connections on E. Connect them by a linear path $A_t = A_0 + t(A_1 - A_0) = A_0 + ta$ where $a = A_1 - A_0 \in \Omega^1(M; \mathfrak{gl}(n))$. The curvature along the path is $F_t = dA_t + A_t \wedge A_t$.

Computing the derivative:

$$\frac{d}{dt}\mathrm{Tr}(F_t^k) = k\,\mathrm{Tr}\left(\frac{dF_t}{dt} \cdot F_t^{k-1}\right) = k\,\mathrm{Tr}(D_t a \cdot F_t^{k-1}) = d\Big(k\,\mathrm{Tr}(a \cdot F_t^{k-1})\Big), \tag{10.3}$$

where $D_t a = da + [A_t, a]$ is the covariant derivative of a, and the last equality uses the Bianchi identity $D_t F_t = 0$ combined with the cyclic trace. Integrating:

$$\mathrm{Tr}(F_1^k) - \mathrm{Tr}(F_0^k) = d \underbrace{\int_0^1 k\,\mathrm{Tr}(a \cdot F_t^{k-1})\,dt}_{\text{the transgression form}}. \tag{10.4}$$

The difference is **exact**: $\mathrm{Tr}(F_1^k)$ and $\mathrm{Tr}(F_0^k)$ represent the same de Rham cohomology class.

Summary of Chern-Weil: gauge invariance of Tr (step 1) + Bianchi identity (step 2) = closed gauge-invariant form. The path argument (step 3) shows the cohomology class is independent of the connection. All three steps are direct consequences of the structure we built in the preceding chapters.

10.3 What It Means for a Class to Live in Cohomology

10.3.1 The orbit picture

A cohomology class $[c] \in H^{2k}_{\mathrm{dR}}(M;\mathbb{R})$ is not a single differential form; it is an **equivalence class** (orbit) of closed forms under the equivalence relation:

$$\omega \sim \omega' \quad \text{iff} \quad \omega - \omega' = d\alpha \text{ for some } (2k-1)\text{-form } \alpha. \tag{10.5}$$

Saying "the characteristic class $c_k(E)$ lives in $H^{2k}(M;\mathbb{R})$" means: $c_k(E)$ *is* such an equivalence class. The curvature form $\sigma_k(F/(2\pi i))$ is one representative of this class; another connection gives another representative of the same class.

The orbit structure mirrors the gauge orbit of the connection itself. Under a gauge transformation, $A \to A^g$ and $F \to gFg^{-1}$, so $\sigma_k(F)$ changes by a gauge transformation. But the gauge orbit of $\sigma_k(F)$ is a single cohomology class.

Under a change of connection (without gauge transformation), $\sigma_k(F) \to \sigma_k(F')$ changes by an exact form (equation (10.4)). So the cohomology class $[\sigma_k(F)]$ is the same for all connections.

The characteristic class $c_k(E)$ is the orbit of any representative $\sigma_k(F/(2\pi i))$ under the action of adding exact forms. It is an intrinsic property of the bundle E, not of any particular connection on E.

10.3.2 What "lives in cohomology" does NOT mean

It does not mean there is a preferred differential form called $c_k(E)$. There are infinitely many closed $2k$-forms representing $c_k(E)$, one for each connection on E. What is unique is the cohomology class: the set of all these forms, quotiented by the exact form ambiguity.

It also does not mean the class is determined by local data. The cohomology class is a global invariant: it encodes information about the bundle over the entire manifold M. Locally, over any contractible patch, every connection is flat and every characteristic class vanishes (Poincaré lemma). The class is detected only by global integration.

10.3.3 Integer classes and real classes

The Chern-Weil construction gives classes in $H^{2k}(M;\mathbb{R})$ (real coefficients). But a deeper fact says these classes are *integral*: they lie in the image of $H^{2k}(M;\mathbb{Z}) \to H^{2k}(M;\mathbb{R})$. The Chern numbers $\int_\Sigma c_k(E)$ (integrals over appropriate closed submanifolds) are always integers.

The integrality comes from the topological classification: bundles are classified by $[M, BU(n)]$, and the cohomology classes in $H^*(BU(n);\mathbb{Z})$ are integral. The Chern-Weil construction provides real representatives (differential forms) for these integer classes; the integrality is a separate fact about the homotopy theory.

10.4 The Chern-Weil Theorem, Stated Precisely

Theorem 10.1 (Chern-Weil)**.** *Let $E \to M$ be a complex rank-n vector bundle with connection A and curvature F. Let P be any symmetric* $\mathrm{GL}(n)$*-invariant polynomial of degree k on $\mathfrak{gl}(n,\mathbb{C})$. Then:*

1. *$P(F)$ is a closed $2k$-form on M.*

2. *The de Rham cohomology class* $[P(F)] \in H^{2k}_{\mathrm{dR}}(M;\mathbb{R})$ *is independent of the choice of connection on* E.

3. *The map* $P \mapsto [P(F)]$ *is a ring homomorphism (**Chern-Weil homomorphism**) from the ring* $\mathrm{Inv}(\mathfrak{g})$ *of invariant polynomials to the cohomology ring* $H^*(M;\mathbb{R})$.

The proof is exactly the three-step argument above. Property (1) is the Bianchi identity. Property (2) is the transgression calculation. Property (3) follows from the multiplicativity of the trace and the cup product.

The **Chern-Weil homomorphism** is:

$$\mathrm{CW} : \mathrm{Inv}(\mathfrak{g}) \to H^*(M;\mathbb{R}), \qquad P \mapsto [P(F)]. \tag{10.6}$$

Every invariant polynomial produces a cohomology class; combining them gives a ring map from invariant polynomials to cohomology.

10.5 Chern Classes: The Most Important Output

10.5.1 The generating formula

The Chern classes are defined by the determinant formula:

$$\det\left(I + \frac{i}{2\pi}F\right) = \sum_{k=0}^{n} c_k(E), \qquad c_k(E) \in H^{2k}(M;\mathbb{R}). \tag{10.7}$$

The left side is a formal expansion: since F is a 2-form (not a number), powers of F are 4-forms, 6-forms, etc. The k-th term in the expansion is a $2k$-form:

$$c_0(E) = 1, \tag{10.8}$$

$$c_1(E) = \frac{i}{2\pi}\mathrm{Tr}(F), \tag{10.9}$$

$$c_2(E) = \frac{1}{2}\left(\frac{i}{2\pi}\right)^2 \left(\mathrm{Tr}(F)^2 - \mathrm{Tr}(F^2)\right), \tag{10.10}$$

$$c_k(E) = \left(\frac{i}{2\pi}\right)^k \sigma_k(F), \tag{10.11}$$

where σ_k is the k-th elementary symmetric polynomial of the eigenvalues of F (as a matrix of 2-forms).

The factor $i/(2\pi)$ is a normalization ensuring the classes are integral: $\int c_k(E) \in \mathbb{Z}$ over appropriate cycles.

10.5.2 Why this particular polynomial?

The determinant formula (10.7) is not arbitrary. The elementary symmetric polynomials σ_k are exactly the generators of the ring of $\mathrm{GL}(n)$-invariant polynomials (by the fundamental theorem of symmetric functions). By the Whitney formula, the total Chern class $c(E) = \sum c_k(E)$ satisfies:

$$c(E \oplus F) = c(E) \cup c(F), \tag{10.12}$$

which determines the Chern classes on all bundles from line bundles. On a line bundle (rank 1), the only Chern class is $c_1(E) = \frac{i}{2\pi}\mathrm{Tr}(F) = \frac{i}{2\pi}F$ (since $\mathfrak{u}(1) = i\mathbb{R}$ and F is a scalar). The determinant formula then gives the correct classes for all ranks by the splitting principle.

10.6 The First Chern Class: The Main Character

10.6.1 What it measures

The first Chern class $c_1(E) = \frac{i}{2\pi}\mathrm{Tr}(F) \in H^2(M;\mathbb{Z})$ is the primary obstruction to trivializing a line bundle. For a complex line bundle ($n = 1$, fiber $\mathbb{C}$, group $U(1)$):

- $c_1(E) = 0$ if and only if the bundle is trivial (admits a non-vanishing global section).
- $c_1(E) = n \in H^2(M;\mathbb{Z}) \cong \mathbb{Z}$ (for $M = S^2$) classifies the bundle completely.
- Changing the connection changes F but not $[F]/(2\pi)$.

Physical meaning: $c_1(E) = \frac{1}{2\pi}[F]$ is the curvature class of the $U(1)$ gauge field (magnetic flux per unit area, integrated). For a magnetic monopole of charge g:

$$c_1(E) = \frac{1}{2\pi}\int_{S^2} F = \frac{g}{2\pi}\int_{S^2} \sin\theta\, d\theta \wedge d\phi = g \in \mathbb{Z}. \tag{10.13}$$

The integer g is the magnetic charge. The Dirac quantization condition $eg \in 2\pi\mathbb{Z}$ is the statement that c_1 must be integral: without integrality, the transition functions of the bundle are not well-defined.

10.6.2 Changing the connection: the invariance is not obvious

An explicit example. On the $U(1)$ bundle over S^2 with charge $g = 1$, consider two connections:

Connection 1 (monopole gauge north): $A_N = \frac{1-\cos\theta}{2}d\phi$ on the northern hemisphere U_N. Curvature: $F_N = dA_N = \frac{1}{2}\sin\theta\, d\theta \wedge d\phi$.

Connection 2 (monopole gauge south): $A_S = -\frac{1+\cos\theta}{2}d\phi$ on the southern hemisphere U_S. Curvature: $F_S = dA_S = \frac{1}{2}\sin\theta\, d\theta \wedge d\phi$.

Both give the same curvature form! So in this case, the representatives are literally equal, not just cohomologous. (This is because both are the unique $U(1)$-invariant connection with charge 1.) The transition function on the equator is $g_{NS}(\phi) = e^{i\phi}$, which has winding number 1. And indeed:

$$c_1 = \frac{1}{2\pi}\int_{S^2} F = \frac{1}{2\pi}\cdot 2\pi = 1.\checkmark \tag{10.14}$$

For a different connection (say one with a concentrated flux near the north pole), the integrand F would look different — peaked near the north pole instead of uniform — but the integral over S^2 is still 1. The class is the integral; the specific form of the integrand is a representative, not the class itself.

10.6.3 The quantum Hall effect: c_1 of the Bloch band

In the integer quantum Hall effect, electrons in a periodic potential fill energy bands. Each filled band defines a complex line bundle over the Brillouin zone torus T^2 (the momentum space). The connection is the Berry connection $\mathcal{A}_n = i\langle u_n(\mathbf{k})|d|u_n(\mathbf{k})\rangle$ and the curvature is the Berry curvature $\mathcal{F}_n = d\mathcal{A}_n$.

The first Chern number of the n-th band:

$$C_n = \frac{1}{2\pi}\int_{T^2} \mathcal{F}_n \in \mathbb{Z}, \tag{10.15}$$

is the TKNN invariant. The Hall conductivity is $\sigma_{xy} = \frac{e^2}{h}\sum_n C_n$. The integrality of C_n (guaranteed by Chern-Weil) is the mathematical reason for the remarkable precision of the quantum Hall effect: the Hall resistance is quantized to parts in 10^{10} because it equals an integer times h/e^2, and that integer cannot change continuously.

Independence of connection. We could compute the Hall conductance using any gauge (any Berry connection), and the answer is the same integer. This is Chern-Weil: the cohomology class $[\mathcal{F}_n/(2\pi)]$ is independent of the gauge choice.

10.7 Curvature Cannot Be Spread Too Thin: The Gauss-Bonnet Theorem

10.7.1 The topology constrains the geometry

Here is a striking physical consequence of Chern-Weil. The Euler characteristic of a compact surface M is a topological integer: $\chi(S^2) = 2$, $\chi(T^2) = 0$, $\chi(\Sigma_g) = 2 - 2g$ (for a genus-g surface). The Gauss-Bonnet-Chern theorem says:

$$\chi(M) = \frac{1}{2\pi}\int_M K\, dA, \tag{10.16}$$

where K is the Gaussian curvature. This integral is a characteristic class: $\frac{K\,dA}{2\pi}$ is the Euler class $e(TM)$, computed by the Chern-Weil formula from the Levi-Civita curvature.

What this means geometrically. You can reshape S^2 any way you want — inflate it, squash it, make it bumpy, flatten most of it — and the total Gaussian curvature is always 4π. The topology ($\chi = 2$) forces the integral to be $2\pi \cdot 2 = 4\pi$.

Imagine trying to make the sphere "almost flat": $K \approx 0$ everywhere. You can make K very small in most places, but then you must compensate by making K very large somewhere else (like at sharp corners or poles). The total is fixed at 4π no matter what. **Topology is the floor on the total curvature.** The curvature cannot be spread out to nothing; some minimum total curvature is required.

Contrast with the torus. The torus has $\chi = 0$, so $\int K\, dA = 0$. You can make a perfectly flat torus (with $K = 0$ everywhere) because the topological constraint allows zero total curvature. The standard embedding of the torus in $\mathbb{R}^3$ has positive curvature on the outside and negative curvature on the inside, which cancel exactly.

Connection to holonomy. By Stokes (in the form of the holonomy formula from Chapter 10), the holonomy of the Levi-Civita connection around a loop equals the integral of K over the enclosed region. The total curvature 4π is the holonomy of a loop that encloses the entire sphere: bringing a vector all the way around the sphere gives a full rotation by 4π (or 2π — the sign depends on convention). The Euler characteristic "counts" how many full rotations of holonomy the sphere accumulates.

10.7.2 Gauss-Bonnet in higher dimensions

For an even-dimensional compact oriented Riemannian manifold M^{2m}:

$$\chi(M) = \int_M e(TM) = \int_M \mathrm{Pf}\left(\frac{R}{2\pi}\right), \tag{10.17}$$

where Pf is the Pfaffian of the curvature matrix R (a polynomial in R whose square is $\det R$). For $m = 1$ this is the classical Gauss-Bonnet formula. For $m = 2$ (a 4-manifold):

$$\chi(M^4) = \int_{M^4} \frac{1}{8\pi^2} \left(\mathrm{Tr}(R \wedge R) - \tfrac{1}{4}(\mathrm{Tr}\, R)^2 \right). \tag{10.18}$$

10.8 The Second Chern Class: Instanton Number

The second Chern class is a 4-form:

$$c_2(E) = \frac{-1}{8\pi^2} \left(\mathrm{Tr}(F \wedge F) - \mathrm{Tr}(F)^2/2 \right). \tag{10.19}$$

For an $SU(n)$ bundle ($\mathrm{Tr}\, F = 0$):

$$c_2(E) = \frac{-1}{8\pi^2} \mathrm{Tr}(F \wedge F). \tag{10.20}$$

The second Chern number $\int_{M^4} c_2(E)$ is an integer for any compact 4-manifold M^4.

The instanton number. For a Yang-Mills field on $\mathbb{R}^4$ with finite action (which forces the field to be a pure gauge at infinity), the second Chern number is:

$$k = \int_{\mathbb{R}^4} c_2(E) = \frac{-1}{8\pi^2} \int_{\mathbb{R}^4} \mathrm{Tr}(F \wedge F) \in \mathbb{Z}. \tag{10.21}$$

This integer k classifies the gauge field topologically. $k = 0$: topologically trivial (can be continuously deformed to $A = 0$). $k = 1$: the one-instanton sector. $k = -1$: one anti-instanton.

Physical meaning. Each instanton sector represents a tunneling process in the path integral: the gauge field tunnels between topologically inequivalent vacua, separated by a topological barrier measured by k. The QCD θ-vacuum is a superposition of all instanton sectors with weight $e^{ik\theta}$: $|\theta\rangle = \sum_{k \in \mathbb{Z}} e^{ik\theta} |k\rangle$.

Chern-Weil tells us two things. First, $\mathrm{Tr}(F \wedge F)$ is a closed 4-form (Bianchi identity). Second, its integral is independent of the connection (in its topological class): you cannot change the instanton number k by a small deformation of the gauge field. The integer k is rigid, topologically protected.

The action bound. In Euclidean Yang-Mills theory, the action $S = \frac{1}{2g^2} \int \mathrm{Tr}(F \wedge \star F)$ satisfies:

$$S = \frac{1}{2g^2} \int \mathrm{Tr}(F \wedge \star F) \geq \frac{1}{2g^2} \left| \int \mathrm{Tr}(F \wedge F) \right| = \frac{8\pi^2 |k|}{2g^2}. \tag{10.22}$$

(Using $\|F^{\pm}\|^2 \geq 0$ and $\int F \wedge \star F = \|F^+\|^2 + \|F^-\|^2$, $\int F \wedge F = \|F^+\|^2 - \|F^-\|^2$.) Equality holds iff $F = \star F$ (self-dual for $k > 0$) or $F = -\star F$ (anti-self-dual for $k < 0$). These are the instantons: minimum action configurations in their topological class.

10.9 Pontryagin Classes and the Signature

For a real vector bundle $E \to M$ of rank n, complexify to get $E_{\mathbb{C}} = E \otimes_{\mathbb{R}} \mathbb{C}$. The Chern classes of $E_{\mathbb{C}}$ satisfy $c_{2k+1}(E_{\mathbb{C}}) = 0$, and the **Pontryagin classes** are:

$$p_k(E) = (-1)^k c_{2k}(E_{\mathbb{C}}) \in H^{4k}(M; \mathbb{Z}). \tag{10.23}$$

They live in degrees divisible by 4 (unlike Chern classes which live in degrees divisible by 2). Via Chern-Weil, with the Levi-Civita curvature R:

$$p_1(TM) = \frac{-1}{8\pi^2}\mathrm{Tr}(R \wedge R) \in H^4(M;\mathbb{R}). \tag{10.24}$$

The signature theorem. For a compact oriented 4-manifold:

$$\mathrm{sign}(M) = \frac{1}{3}\int_M p_1(TM) = \frac{1}{24\pi^2}\int_M \mathrm{Tr}(R \wedge R), \tag{10.25}$$

where $\mathrm{sign}(M) = b_2^+ - b_2^-$ is the signature (the difference between positive and negative eigenvalues of the intersection form on $H^2(M;\mathbb{R})$). The signature is an integer; the right side is a geometric integral — the same Chern-Weil miracle as Gauss-Bonnet.

10.10 Stiefel-Whitney Classes: Obstructions over $\mathbb{Z}/2$

For real vector bundles, the **Stiefel-Whitney classes** $w_k(E) \in H^k(M;\mathbb{Z}/2\mathbb{Z})$ are characteristic classes with mod-2 coefficients. They are *not* produced by the Chern-Weil construction (which requires a Lie algebra and trace, and gives real classes). They require a separate construction using the classifying space $BO(n)$.

The key applications:

$w_1(E) = 0$ iff E is orientable. For the Möbius band over S^1: $w_1 \neq 0$ (non-orientable).

$w_2(TM) = 0$ iff M admits a **spin structure**: a consistent choice of spinors over M. Without a spin structure, no globally consistent Dirac fermion exists on M. Since the Standard Model includes spin-$\frac{1}{2}$ particles (electrons, quarks), any spacetime manifold supporting the Standard Model must have $w_2(TM) = 0$. This is a topological constraint on spacetime.

The complex projective plane $\mathbb{CP}^2$ has $w_2 \neq 0$: it cannot support global spinor fields. By contrast, all orientable 3-manifolds have $w_2 = 0$ (they all admit spin structures).

10.11 The Chern-Simons Form: The Primitive of c_2

Recall from the transgression formula (10.4): the difference $\mathrm{Tr}(F_1^k) - \mathrm{Tr}(F_0^k)$ is exact, with the explicit primitive being the integral along the path. For $k = 2$ and comparing to the zero connection $A_0 = 0$:

$$\mathrm{Tr}(F \wedge F) = d\,\mathrm{CS}(A), \qquad \mathrm{CS}(A) = \mathrm{Tr}\Big(A \wedge dA + \tfrac{2}{3}A \wedge A \wedge A\Big). \tag{10.26}$$

The **Chern-Simons 3-form** $\mathrm{CS}(A)$ is the "primitive" of the second Chern class: its exterior derivative gives back $\mathrm{Tr}(F \wedge F)$. On a 3-manifold, $\mathrm{CS}(A)$ is a 3-form that can be integrated.

Why it exists. On a 3-manifold M^3, the 4-form $\mathrm{Tr}(F \wedge F)$ vanishes (there are no 4-forms). So $\mathrm{Tr}(F \wedge F)$ is automatically closed (and zero). By Poincaré's lemma, it is locally exact: $\mathrm{Tr}(F \wedge F) = d\,\mathrm{CS}(A)$ locally. But $\mathrm{CS}(A)$ is a globally defined 3-form because M^3 is 3-dimensional.

Gauge non-invariance. Under a gauge transformation $A \to A^g$, the Chern-Simons form changes:

$$\mathrm{CS}(A^g) = \mathrm{CS}(A) + d(\ldots) + \mathrm{WZ}(g), \tag{10.27}$$

where $\mathrm{WZ}(g) = \frac{1}{3}\mathrm{Tr}((g^{-1}dg)^3)$ is the Wess-Zumino term. For a "large" gauge transformation ($g : M^3 \to G$ with winding number n), $\int_{M^3} \mathrm{WZ}(g) = 8\pi^2 n$. So the Chern-Simons action $S_{\mathrm{CS}}(A) =$

$\frac{k}{4\pi}\int_{M^3}\mathrm{CS}(A)$ shifts by $2\pi kn$ under a large gauge transformation. For quantum gauge invariance (e^{iS} invariant), we need $k \in \mathbb{Z}$: the **level quantization** of Chern-Simons theory.

Chern-Simons as a TQFT. The Chern-Simons theory is a topological field theory (its action has no metric dependence). Its equations of motion are $F = 0$ (flat connections). Its partition function on a closed 3-manifold is a topological invariant, and its Wilson loop observables compute knot invariants (the Jones polynomial for $G = SU(2)$).

10.12 The Chern Character and the Index Theorem

The **Chern character** is a ring homomorphism:

$$\mathrm{ch}(E) = \mathrm{Tr}\left(\exp\frac{iF}{2\pi}\right) = \mathrm{rank}(E) + c_1(E) + \frac{c_1(E)^2 - 2c_2(E)}{2} + \cdots . \tag{10.28}$$

Unlike the total Chern class $c(E)$, which is multiplicative under tensor product with respect to the cup product, $\mathrm{ch}(E)$ converts both direct sums and tensor products into simple additions and multiplications: $\mathrm{ch}(E \oplus F) = \mathrm{ch}(E) + \mathrm{ch}(F)$ and $\mathrm{ch}(E \otimes F) = \mathrm{ch}(E) \cdot \mathrm{ch}(F)$. This makes it a ring homomorphism from K-theory to cohomology.

The **Atiyah-Singer index theorem** is the deepest application of characteristic classes:

$$\mathrm{index}(D) = \int_M \mathrm{ch}(E - F) \cdot \mathrm{td}(TM_{\mathbb{C}}), \tag{10.29}$$

where $D : \Gamma(E) \to \Gamma(F)$ is any elliptic operator, ch is the Chern character, and td is the Todd class (another characteristic class computed from the curvature). The left side is an analytic integer (kernel minus cokernel dimension). The right side is a topological integral. The theorem says they always agree.

Special cases: Gauss-Bonnet ($D = d + d^*$, index $= \chi(M)$), Hirzebruch signature ($D =$ signature operator, index $= \mathrm{sign}(M)$), Riemann-Roch (complex manifolds), Dirac index (counting zero modes of the Dirac operator, relevant to anomaly cancellation).

10.13 The Full Obstruction Hierarchy

All characteristic classes fit into a unified hierarchy of topological obstructions. Every obstruction to trivializing a bundle, extending a section, or equipping a manifold with extra structure is measured by a characteristic class:

Class	Lives in	Bundle type	Obstruction	Physics
$c_1(E)$	$H^2(M;\mathbb{Z})$	Complex line	Triviality	Monopole charge
$c_2(E)$	$H^4(M;\mathbb{Z})$	Complex rank 2	Triviality	Instanton number
$p_1(TM)$	$H^4(M;\mathbb{Z})$	Real rank 4	Triviality	Gravitational anomaly
$e(TM)$	$H^n(M;\mathbb{Z})$	Real rank n	Non-van. section	Hairy ball
$w_1(TM)$	$H^1(M;\mathbb{Z}/2)$	Real	Orientability	Möbius band
$w_2(TM)$	$H^2(M;\mathbb{Z}/2)$	Real	Spin structure	Dirac fermions

10.14 Summary: The Three Levels of Understanding

Level 1: Geometric. Curvature is the local infinitesimal holonomy. Characteristic classes are global integrals of curvature. The Gauss-Bonnet theorem says: the total Gaussian curvature of a surface is $2\pi\chi$, a topological integer. Topology constrains geometry: you cannot spread the curvature to zero if $\chi \neq 0$.

Level 2: Analytic. Given any connection on a bundle, applying any invariant polynomial to the curvature gives a closed gauge-invariant differential form. This form represents a de Rham cohomology class. Different connections give cohomologous forms: the class is a topological invariant of the bundle, not the connection. This is the Chern-Weil theorem.

Level 3: Cohomological. A characteristic class is an element of $H^{2k}(M;\mathbb{Z})$: an equivalence class of closed forms, where two forms are equivalent if they differ by an exact form. The class is the orbit of $\sigma_k(F/(2\pi i))$ under the action of adding exact forms. Different connections give different representatives of the same orbit. The class is fixed by the bundle topology. Its integral over appropriate cycles is always an integer.

These three levels are three perspectives on the same object. The Chern-Weil theorem is the bridge between them: it identifies the geometric object (curvature integral) with the analytic object (closed form) with the cohomological object (cohomology class). All three are the same topological integer, computed in three different ways.

Chapter 11

Physics of Topology: Applications

11.1 Three Levels of Topological Physics

The topological structures developed in the preceding chapters — homotopy groups, fiber bundles, connections, curvature, characteristic classes — appear throughout modern physics in three physically distinct ways that share the same mathematical language but differ fundamentally in their Hilbert space structure and Hamiltonian. Conflating these three levels generates persistent confusion; the first task of this chapter is to separate them clearly, and to use that separation as the organizing principle for all the examples that follow.

At the **first level** are topological collective modes: solitons, vortices, Skyrmions, and Hopfions. These inhabit an ordinary infinite-dimensional Fock space with a non-trivial Hamiltonian ($H \neq 0$). The order parameter field takes values in a target space with non-trivial homotopy groups, and as a result the Fock space decomposes into topological sectors,

$$\mathcal{F} = \bigoplus_{Q \in \mathbb{Z}} \mathcal{F}^{(Q)}, \tag{11.1}$$

labelled by a conserved integer topological charge Q. The Hamiltonian is block-diagonal in this decomposition: it generates rich dynamics within each sector but has no matrix elements between sectors of different Q. No smooth Schrödinger evolution can move the Fock hidden state between sectors.

At the **second level** is topological quantum field theory (TQFT): Chern-Simons theory, BF theory, the G/G model, and their relatives. Here $H = 0$ identically. There are no propagating particles, no energy spectrum, no local dynamics. The Hilbert space is finite-dimensional, consisting of topological invariants of the spatial manifold. The theory computes knot polynomials, 3-manifold invariants, and ground-state degeneracies that depend only on topology.

At the **third level** are topological phases of matter: the quantum Hall effect, topological insulators, topological superconductors, spin liquids. These are ordinary many-body systems ($H \neq 0$, infinite-dimensional Fock space), but the ground state is separated from all excitations by an energy gap $\Delta > 0$. At energies below Δ, the effective theory is a TQFT. The topological invariant of the ground state bundle is protected jointly by the gap and by the integrality of characteristic classes.

The organizing question is: *what is the Hilbert space, and what is the Hamiltonian?*

Level	Hilbert space	Hamiltonian	Protection	Example
1: Topological collective modes	Infinite Fock space $\bigoplus_Q \mathcal{F}^{(Q)}$	$H \neq 0$, block-diagonal	Geometric superselection	Soliton, Skyrmion
2: TQFT	Finite-dimensional; not a Fock space	$H = 0$; no dynamics	Topological invariant of manifold	Chern-Simons
3: Topological phase	Infinite Fock space; gapped ground state	$H \neq 0$, gap $\Delta > 0$	Gap plus integrality of Chern number	Quantum Hall

11.2 Level 1: Topological Collective Modes

11.2.1 The Superselection Principle

A topological charge $Q[\phi]$ is an integer-valued functional of the field configuration $\phi(\mathbf{x})$ that is invariant under any continuous deformation. To change Q, the field must pass through a singularity or a configuration of infinite energy. This makes the superselection rule

$$\langle \psi_m | H | \psi_n \rangle = 0, \qquad |\psi_m\rangle \in \mathcal{F}^{(Q_1)},\ |\psi_n\rangle \in \mathcal{F}^{(Q_2)},\ Q_1 \neq Q_2, \tag{11.2}$$

exact and geometrical, requiring no symmetry assumption on H. The Schrödinger equation preserves the sector; the topological charge observable gives an integer with probability one regardless of all local details of the state within the sector.

The key contrast with Noether charges is this: a Noether charge is protected by a symmetry of H; break the symmetry and the charge is no longer conserved. A topological charge is protected by the geometry of configuration space; change H arbitrarily and the charge is still conserved, because no smooth time evolution connects configurations of different Q at finite energy.

11.2.2 Kinks in $(1+1)$ Dimensions: π_0 Topology

The ϕ^4 theory in $(1+1)$ dimensions with potential $V(\phi) = \lambda(\phi^2 - v^2)^2$ is the canonical example. The two classical vacua $\phi = \pm v$ are disconnected components of the vacuum manifold; a field configuration is characterized by its values at spatial infinity, classified by $\pi_0(\{-v, +v\}) = \mathbb{Z}_2$. The topological charge

$$Q = \frac{\phi(+\infty) - \phi(-\infty)}{2v} \in \{-1, 0, +1\} \tag{11.3}$$

distinguishes the vacuum sector ($Q = 0$), the kink ($Q = +1$), and the antikink ($Q = -1$). The kink solution

$$\phi_{\text{kink}}(x,t) = v \tanh\left(\frac{x - x_0 - ut}{\xi\sqrt{1 - u^2/c^2}}\right), \qquad \xi = \frac{1}{v\sqrt{2\lambda}}, \tag{11.4}$$

is a relativistic soliton of rest mass $M_{\text{kink}} = \frac{4v^3\sqrt{\lambda}}{3}$. It has its own Fock space of quantum fluctuations around the classical profile: small oscillations of the field about ϕ_{kink} give phonon-like modes (including a zero mode from translation invariance) quantized into a new particle spectrum. A kink and antikink can annihilate, since $Q_{\text{kink}} + Q_{\text{antikink}} = 0$, but a single kink cannot decay.

The **sine-Gordon model** is a closely related theory in $(1+1)$d with potential $V(\phi) = \frac{m^2}{\beta^2}(1-\cos\beta\phi)$, which has an infinite family of degenerate vacua at $\phi = 2\pi n/\beta$. Its kink solution

$$\phi_{\mathrm{SG}}(x,t) = \frac{4}{\beta}\,\mathrm{arctanexp}\!\left(\frac{m(x-ut)}{\sqrt{1-u^2}}\right) \tag{11.5}$$

interpolates from 0 to $2\pi/\beta$. The sine-Gordon model is *exactly integrable* via the inverse scattering method, and its quantum kinks are exact bound states with an S-matrix computable in closed form. Moreover, it is dual (via bosonization) to the massive Thirring model of interacting fermions, a remarkable correspondence between a bosonic soliton and a fundamental fermion.

Physical realizations: dislocations in crystals (Frenkel-Kontorova model), magnetic domain walls, Josephson junctions.

11.2.3 Vortices in $(2+1)$ Dimensions: π_1 Topology

In a superfluid with order parameter $\psi_0 = \sqrt{n_0}\,e^{i\theta}$, the phase $\theta \in S^1$ can wind around a singular point (the vortex core where $n_0 = 0$):

$$\oint_C d\theta = 2\pi Q, \qquad Q \in \mathbb{Z} = \pi_1(S^1). \tag{11.6}$$

The vortex solution in polar coordinates has $n_0(r) \to 0$ as $r \to 0$ and $n_0(r) \to n_{\mathrm{bulk}}$ as $r \to \infty$, with $\theta(\varphi) = Q\varphi$. The energy of a single vortex is $E_v \sim \pi n_s(\hbar/m)^2 \ln(R/\xi)$, where R is the system size and ξ is the coherence length; vortices are therefore logarithmically confined, and the transition between the vortex-pair (low-T) and vortex-plasma (high-T) phases is the Berezinskii-Kosterlitz-Thouless (BKT) transition.

In a **superconductor**, the order parameter is $\psi_s = |\psi_s|e^{i\theta}$ and the magnetic flux through the vortex is quantized:

$$\Phi = Q\,\frac{h}{2e} \equiv Q\,\Phi_0, \tag{11.7}$$

where $\Phi_0 = h/2e \approx 2.07 \times 10^{-15}$ Wb is the flux quantum. This is the Abrikosov vortex. The factor of $2e$ (not e) comes from the Cooper pair charge. A type-II superconductor in an external field $H_{c1} < H < H_{c2}$ admits a lattice of Abrikosov vortices, each carrying one flux quantum Φ_0. The Fock hidden state carrying a single vortex lives in $\mathcal{F}^{(Q=1)}$ permanently; the quantized flux is a direct Born rendering of the topological charge.

In a **spinor Bose-Einstein condensate** ($F = 1$ atoms), the order parameter lives on a more complex manifold and vortices can carry half-integer winding numbers, with the vortex core filled by a different spin component. These non-abelian vortices have richer dynamics and can reconnect in ways that simple $U(1)$ vortices cannot.

11.2.4 Monopoles in $(3+1)$ Dimensions: π_2 Topology

Dirac monopole

A point magnetic charge at the origin creates a field $\mathbf{B} = \frac{g}{4\pi r^2}\hat{\mathbf{r}}$ with total flux g through any enclosing S^2. Stokes' theorem obstructs a global vector potential: $\int_{S^2} dA = g \neq 0$ but $\int_{S^2} dA = \int_{\partial S^2} A = 0$. The bundle solution covers S^2 with two patches:

$$A_N = \frac{g}{4\pi}\frac{1-\cos\theta}{\sin\theta}\,d\phi, \qquad A_S = -\frac{g}{4\pi}\frac{1+\cos\theta}{\sin\theta}\,d\phi. \tag{11.8}$$

The transition function $e^{iqg\phi/2\pi}$ must be single-valued, giving the **Dirac quantization condition** $qg = 2\pi n$, $n \in \mathbb{Z}$. The integer $n = c_1 \in H^2(S^2;\mathbb{Z})$ is the magnetic charge. The quantization of all electric charges in nature ($q = ne$, $n \in \mathbb{Z}$) is explained: if even one magnetic monopole exists, Dirac quantization forces all charges to be integer multiples of $2\pi/g$.

't Hooft-Polyakov monopole

In $SU(2)$ Yang-Mills-Higgs theory with the Higgs in the adjoint representation and vacuum $|\phi| = v$, the boundary condition at spatial infinity defines a map $\hat{\phi} : S^2_\infty \to SU(2)/U(1) \cong S^2$, classified by $\pi_2(S^2) = \mathbb{Z}$. The hedgehog ansatz

$$\phi^a = v\hat{r}^a H(r), \qquad A^a_i = \epsilon_{aij}\hat{r}^j \frac{1 - K(r)}{er}, \tag{11.9}$$

with $H(0) = 0$, $H(\infty) = 1$, $K(0) = 1$, $K(\infty) = 0$, gives a smooth finite-energy monopole of magnetic charge $g = 4\pi/e$ and mass $M = 4\pi v/e \cdot f(\lambda/e^2)$ where f is a slowly varying function. In the BPS limit $\lambda \to 0$, the mass saturates the Bogomolny bound $M = 4\pi v/e$, the equations of motion reduce to the first-order BPS equations $\mathbf{B}^a = (D_i\phi)^a$, and the multi-monopole moduli space has dimension $4k$ for a k-monopole solution.

Grand unified theories (GUTs) in which the Standard Model gauge group $SU(3) \times SU(2) \times U(1)$ arises from breaking a simple group like $SU(5)$ necessarily produce monopoles via the Kibble mechanism during cosmological phase transitions. The overproduction of such monopoles (each carrying $M \sim 10^{16}$ GeV) is the magnetic monopole problem, one of the primary motivations for inflationary cosmology.

11.2.5 Skyrmions: Baryons as Topological Solitons

The Skyrme model

The Skyrme model has field $U(\mathbf{x}) \in SU(2)$, a map from compactified space S^3 to $SU(2) \cong S^3$, classified by $\pi_3(SU(2)) = \mathbb{Z}$. The Lagrangian

$$\mathcal{L} = \frac{f_\pi^2}{16}\mathrm{Tr}(\partial_\mu U^\dagger \partial^\mu U) + \frac{1}{32e^2}\mathrm{Tr}([U^\dagger\partial_\mu U, U^\dagger \partial_\nu U]^2) \tag{11.10}$$

involves two parameters: the pion decay constant f_π and the Skyrme parameter e. The topological charge

$$B = \frac{1}{24\pi^2}\int d^3x\, \epsilon^{ijk}\mathrm{Tr}[(U^\dagger\partial_i U)(U^\dagger\partial_j U)(U^\dagger\partial_k U)] \in \mathbb{Z} \tag{11.11}$$

is identified with baryon number. The minimal-energy configuration in each sector $B = Q$ is called the *Skyrmion of charge* Q, and with the Skyrme parameters fitted to f_π and the proton mass, the model gives a surprisingly accurate description of nuclear physics: binding energies, spin-isospin spectra, magnetic moments. Nuclear physics emerges from the topology of maps $S^3 \to SU(2)$.

Magnetic Skyrmions in two dimensions

In 2d ferromagnets, the magnetization $\hat{n}(\mathbf{x}) \in S^2$ defines a map from the compactified plane S^2 to S^2:

$$Q = \frac{1}{4\pi}\int d^2\mathbf{x}\, \hat{n} \cdot (\partial_x \hat{n} \times \partial_y \hat{n}) \in \mathbb{Z} = \pi_2(S^2). \tag{11.12}$$

In chiral magnets (MnSi, FeGe, Co/Pt multilayers), the competition between ferromagnetic exchange J, Dzyaloshinskii-Moriya (DM) interaction $D\,\hat{n}\cdot(\nabla\times\hat{n})$, and Zeeman coupling $-B\,n_z$ stabilizes a Skyrmion lattice phase at intermediate fields and temperatures. Individual Skyrmions can be nucleated, moved by spin-transfer torque, and detected by the topological Hall effect. Their diameter ranges from ~ 1 nm (in multilayers) to ~ 100 nm (in bulk MnSi). They are being actively studied for topological magnetic memory: a bit is stored as $Q=0$ (uniform) or $Q=1$ (Skyrmion), protected against thermal noise by the topological barrier.

The Skyrmion number is also relevant in quantum Hall ferromagnets (at Landau level filling $\nu=1$), where Skyrmion quasiparticles are energetically favored over spin-flip excitations when the Zeeman energy is small compared to the exchange energy.

11.2.6 Hopfions: Linking in Three Dimensions

A Hopfion is a field configuration $\hat{n}(\mathbf{x}):S^3\to S^2$ classified by $\pi_3(S^2)=\mathbb{Z}$. The Hopf invariant

$$H=\frac{1}{16\pi^2}\int d^3\mathbf{x}\;\epsilon^{\mu\nu\rho}\,\mathcal{A}_\mu\,\partial_\nu\mathcal{A}_\rho, \tag{11.13}$$

where $d\mathcal{A}=\hat{n}^*\omega_{S^2}$ is the pullback of the area form, counts how many times the preimage circles of any two generic points on S^2 link with each other in S^3. For the canonical Hopf fibration, $H=1$ and every pair of preimage circles links exactly once.

The **Faddeev-Niemi** (or Faddeev-Skyrme) model has Lagrangian

$$\mathcal{L}=\frac{1}{2}(\partial_\mu\hat{n})^2-\frac{e^2}{4}(\hat{n}\cdot\partial_\mu\hat{n}\times\partial_\nu\hat{n})^2, \tag{11.14}$$

and is proposed as an effective model for knotted gluon field configurations in the infrared limit of QCD. Hopfions appear as stable classical solutions in three-dimensional chiral magnets and frustrated nematics. The Hopf charge is a superselection label protected geometrically: $\mathcal{F}=\bigoplus_{H\in\mathbb{Z}}\mathcal{F}^{(H)}$ with H conserved by any smooth evolution.

11.2.7 Instantons: Tunnelling Between Vacua

An instanton is a Euclidean saddle point that computes a tunnelling amplitude between topological sectors, not a state in any sector.

Double-well instanton

For $V(\phi)=\lambda(\phi^2-v^2)^2$, the tunnelling amplitude $\langle+v|e^{-HT}|-v\rangle$ is dominated by the instanton

$$\phi_{\rm inst}(\tau)=v\tanh\!\big(v\sqrt{2\lambda}(\tau-\tau_0)\big),\qquad S_{\rm inst}=\frac{4v^3\sqrt{\lambda}}{3}. \tag{11.15}$$

The true ground states are superpositions $|\Omega_\pm\rangle=\frac{1}{\sqrt{2}}(|+v\rangle\pm|-v\rangle)$ split by $\Delta E\sim e^{-S_{\rm inst}}$, entirely non-perturbative.

BPST Yang-Mills instanton

On Euclidean $\mathbb{R}^4\cong S^4$, the one-instanton solution of Belavin-Polyakov-Schwarz-Tyupkin is

$$A_\mu^{\rm inst}=\frac{2\rho^2\,\bar{\eta}^a_{\mu\nu}(x-x_0)_\nu}{(x-x_0)^2\big[(x-x_0)^2+\rho^2\big]}\,\frac{\tau^a}{2}, \tag{11.16}$$

where $\bar{\eta}^a_{\mu\nu}$ are the anti-self-dual 't Hooft symbols and τ^a are Pauli matrices. The instanton number

$$k = \frac{1}{8\pi^2}\int \mathrm{Tr}(F \wedge F) = 1 \tag{11.17}$$

counts the winding of the gauge field at spatial infinity: $A_\mu \to g^{-1}\partial_\mu g$ with $g : S^3_\infty \to SU(2)$ of winding number 1, consistent with $\pi_3(SU(2)) = \mathbb{Z}$. The moduli space of k-instantons in $SU(2)$ is $8k$-dimensional (positions, sizes, and orientations).

The QCD θ-vacuum

The Yang-Mills vacua $|n\rangle$ ($n \in \mathbb{Z}$) are connected by instantons. The true QCD vacuum is the θ-vacuum $|\theta\rangle = \sum_n e^{in\theta}|n\rangle$, analogous to the energy eigenstates of a Bloch electron in a crystal. The topological θ-term

$$\mathcal{L}_\theta = \frac{\theta g_s^2}{32\pi^2}\mathrm{Tr}(F \wedge F) \tag{11.18}$$

violates CP symmetry. The experimental bound $|\theta| < 10^{-10}$ (from the neutron electric dipole moment) is the strong CP problem. Instanton-mediated effects include chiral symmetry breaking, the η' mass via the 't Hooft determinant interaction, and the θ-dependence of hadron masses.

11.3 Level 2: Topological Quantum Field Theory ($H = 0$)

11.3.1 The Atiyah-Segal Axioms

When $H = 0$ identically, there is no Schrödinger evolution: the Fock hidden state does not move. All physics is topological. The Atiyah-Segal axioms capture this structure as a functor:

$$Z : \mathbf{Cob}_d \to \mathbf{Vect}_{\mathbb{C}}, \tag{11.19}$$

assigning to each closed oriented $(d-1)$-manifold Σ a finite-dimensional Hilbert space $Z(\Sigma)$, and to each d-dimensional cobordism $M : \Sigma_1 \to \Sigma_2$ a linear map $Z(M) : Z(\Sigma_1) \to Z(\Sigma_2)$. The axioms encode:

Topological invariance: Diffeomorphic cobordisms give the same linear map. No metric data enters.

Composition: $Z(M_1 \cup_\Sigma M_2) = Z(M_2) \circ Z(M_1)$ (cutting and regluing composes amplitudes).

Monoidal: $Z(\Sigma_1 \sqcup \Sigma_2) = Z(\Sigma_1) \otimes Z(\Sigma_2)$ (spatial locality without local dynamics).

Duality: $Z(\bar{\Sigma}) = Z(\Sigma)^*$ (reversing orientation gives the dual space).

Partition function: For a closed d-manifold M (a cobordism $\emptyset \to \emptyset$), $Z(M) \in \mathbb{C}$ is a diffeomorphism invariant.

In 2 dimensions, TQFTs are completely classified: they correspond to commutative Frobenius algebras, with $Z(S^1)$ as the algebra. In 3 dimensions, the classification is richer and involves modular tensor categories. The theory becomes increasingly sophisticated in higher dimensions, where the Baez-Dolan-Lurie cobordism hypothesis characterizes fully extended TQFTs via (∞, n)-categories.

11.3.2 Chern-Simons Theory

Action and equations of motion

Chern-Simons theory in three dimensions has action

$$S_{\mathrm{CS}}(A) = \frac{k}{4\pi}\int_{M^3} \mathrm{Tr}\Big(A \wedge dA + \frac{2}{3}A \wedge A \wedge A\Big), \tag{11.20}$$

where A is a connection on a principal G-bundle over M^3 and $k \in \mathbb{Z}$ is the **level**, quantized by requiring gauge invariance under large gauge transformations: under $A \to A^g = g^{-1}Ag + g^{-1}dg$, the action shifts by $S_{\rm CS}(A^g) = S_{\rm CS}(A) + 2\pi k\, w_3(g)$ where $w_3(g) \in \mathbb{Z}$ is the winding number of $g : M^3 \to G$. For $e^{iS_{\rm CS}}$ to be gauge-invariant, $k \in \mathbb{Z}$.

The equations of motion are

$$F = dA + A \wedge A = 0, \tag{11.21}$$

flat connections. There are no propagating degrees of freedom; all physics is encoded in the holonomies of A around non-contractible loops in M^3.

Hilbert space and ground state degeneracy

For $M^3 = \Sigma_g \times \mathbb{R}$ with Σ_g a genus-g surface, the Hilbert space $\mathcal{H}_{\Sigma_g} = Z(\Sigma_g)$ is the space of holomorphic sections of the k-th power of the determinant line bundle over the moduli space of flat G-connections on Σ_g. For $G = U(1)$ at level k:

$$\dim \mathcal{H}_{\Sigma_g} = k^g. \tag{11.22}$$

For $G = SU(2)$ at level k, the dimension is given by the Verlinde formula:

$$\dim \mathcal{H}_{\Sigma_g} = \left(\frac{k+2}{2}\right)^{g-1} \sum_{j=0}^{k/2} \left(\sin\frac{(2j+1)\pi}{k+2}\right)^{2-2g}, \tag{11.23}$$

a remarkable formula counting conformal blocks of the $SU(2)_k$ Wess-Zumino-Witten (WZW) model on Σ_g.

Wilson loops and knot invariants

Physical observables are Wilson loop operators

$$W_R(\mathcal{C}) = \mathrm{Tr}_R\, \mathcal{P} \exp\!\left(i \oint_{\mathcal{C}} A\right), \tag{11.24}$$

for a representation R of G and a loop $\mathcal{C} \subset M^3$. Their expectation values

$$\langle W_{R_1}(\mathcal{C}_1) \cdots W_{R_n}(\mathcal{C}_n) \rangle_{M^3} = \frac{\int \mathcal{D}A\, e^{iS_{\rm CS}(A)} \prod_i W_{R_i}(\mathcal{C}_i)}{\int \mathcal{D}A\, e^{iS_{\rm CS}(A)}} \tag{11.25}$$

depend only on the isotopy class of the colored link $\{(\mathcal{C}_i, R_i)\}$ in M^3.

For $G = SU(2)$, $M^3 = S^3$, $R =$ fundamental (spin-$\frac{1}{2}$) representation, and $\mathcal{C} =$ the unknot, one obtains

$$\langle W(\text{unknot}) \rangle = \frac{\sin(\pi/(k+2))}{\sin(\pi/(k+2))} = 1. \tag{11.26}$$

For a general link L, the normalized expectation value is the **Jones polynomial** $V_L(q)$ evaluated at $q = e^{2\pi i/(k+2)}$. The Jones polynomial is a topological invariant of oriented knots and links, discovered by Vaughan Jones in 1984 and previously unknown to mathematicians; Witten's Chern-Simons computation (1989) gave the first field-theoretic derivation and earned a Fields Medal.

The **HOMFLY-PT polynomial**, which depends on two variables and specializes to the Jones polynomial (and also to the Alexander polynomial), arises from $SU(N)$ Chern-Simons theory, with $q = e^{2\pi i/(k+N)}$.

Surgery, Reshetikhin-Turaev invariants, and 3-manifolds

The Chern-Simons partition function $Z(M^3)$ on a closed 3-manifold is a 3-manifold invariant. It can be computed via surgery: any closed orientable 3-manifold is obtained from S^3 by surgery on a link L (Dehn surgery, adding solid tori with prescribed gluing). This gives

$$Z(M^3) = \frac{1}{\mathcal{D}} \sum_{\text{colorings}} \langle W_{\text{colored } L} \rangle_{S^3}, \tag{11.27}$$

where $\mathcal{D}$ is a normalization and the sum is over colorings of the surgery link by representations of G weighted by quantum dimensions. This is the **Reshetikhin-Turaev (RT) invariant**, a rigorous mathematical construction that preceded and motivated the Witten-Reshetikhin-Turaev (WRT) quantum invariants of 3-manifolds.

11.3.3 BF Theory

Action and structure

BF theory in d dimensions has action

$$S_{\text{BF}} = \frac{1}{2\pi} \int_{M^d} B \wedge F, \tag{11.28}$$

where A is a connection, $F = dA + A \wedge A$ is its curvature, and B is a $(d-2)$-form. The equations of motion are $F = 0$ (flat connection) and $DB = 0$ (covariantly closed B). BF theory is a TQFT in any dimension and is topological because the action contains no metric.

2d BF theory: Yang-Mills in the limit of zero area

In 2 dimensions, B is a 0-form (a function $\phi \in \mathfrak{g}^*$) and F is a 2-form. The action $\int_\Sigma \phi F$ is the two-dimensional Yang-Mills action in the limit of infinite coupling (zero area): $S_{2\text{d YM}} = \frac{1}{2e^2} \int \text{Tr}(F^2) + \int \phi F \to \int \phi F$ as $e \to \infty$. The 2d BF partition function on a genus-g surface gives $Z(\Sigma_g) = \sum_R (\dim R)^{2-2g}$, a sum over representations of G.

3d and 4d BF theory: discrete gauge theory

In 3 dimensions, B is a 1-form taking values in $\mathfrak{g}$. Abelian 3d BF theory with gauge group $\mathbb{Z}_N$ (obtained by $B \wedge dA + \frac{N}{4\pi} A \wedge A$) is the effective theory of topological order with $\mathbb{Z}_N$ anyons: it describes $\mathbb{Z}_N$ discrete gauge theories and their anyonic excitations with statistics $\theta = 2\pi/N$.

In 4 dimensions, B is a 2-form. The 4d BF term $\int B \wedge F$ is the topological sector of various gravitational and gauge theories. Adding a mass term $\frac{m^2}{2} \int B \wedge \star B$ gives the Cremmer-Scherk-Kalb-Ramond massive 2-form theory.

Dijkgraaf-Witten theory

For a finite group G, there is a TQFT called Dijkgraaf-Witten theory, which counts (with weights) principal G-bundles over the spacetime manifold M. The action is

$$S_{\text{DW}}(P) = \langle \alpha, [M] \rangle, \tag{11.29}$$

where $\alpha \in H^d(BG; \mathbb{R}/\mathbb{Z})$ is a cohomology class of the classifying space (the "twist") and $[M]$ is the fundamental class. For $G = \mathbb{Z}_N$ in $d = 3$, this produces the $\mathbb{Z}_N$ BF theory above. Dijkgraaf-Witten theory is a fully explicit, finite (no renormalization needed) TQFT that serves as a testing ground for the general theory.

11.3.4 The G/G Gauged WZW Model and 2d Yang-Mills

WZW model

The Wess-Zumino-Witten model describes a map $g : \Sigma \to G$ from a Riemann surface to a Lie group, with action

$$S_{\mathrm{WZW}}(g) = \frac{k}{4\pi}\int_\Sigma \mathrm{Tr}(g^{-1}\partial g \wedge g^{-1}\bar\partial g) + \frac{k}{12\pi}\int_B \mathrm{Tr}(g^{-1}dg)^3, \tag{11.30}$$

where $\partial\Sigma = \emptyset$ (closed surface) and $\partial B = \Sigma$. The WZW term (the last integral over the 3-ball B bounded by Σ) is topological and quantized in integer multiples k (the level) by the same argument as the Chern-Simons level. The WZW model is a conformal field theory with current algebra symmetry $\hat{\mathfrak{g}}_k$ (an affine Kac-Moody algebra), and the Verlinde formula for the dimension of the Chern-Simons Hilbert space is derived from its partition function.

The G/G model

The G/G gauged WZW model is obtained by gauging the diagonal G-symmetry $g \to hgh^{-1}$, $h : \Sigma \to G$, which makes the theory topological. Its partition function on Σ_g gives exactly the Verlinde dimension formula, establishing a deep link between 2d conformal field theory and 3d Chern-Simons TQFT: the Chern-Simons Hilbert space on Σ_g is the space of conformal blocks of the WZW model on Σ_g.

11.3.5 Donaldson-Thomas Theory and 4-Manifold Invariants

In four dimensions, the Donaldson invariants of smooth 4-manifolds arise from a topological twist of $\mathcal{N} = 2$ super Yang-Mills theory. The partition function receives contributions only from instantons (self-dual connections, $F^+ = 0$), and counts instantons in each topological sector:

$$Z = \sum_{k=0}^{\infty} q^k \,\#\{\text{instantons with } c_2 = k\}. \tag{11.31}$$

These Donaldson invariants distinguish smooth structures on 4-manifolds: two 4-manifolds can be homeomorphic but not diffeomorphic (Freedman-Donaldson), and the Donaldson invariants detect the difference. This is purely a topological fact about smooth 4-manifolds, uncovered by physics-inspired mathematics.

The simpler **Seiberg-Witten invariants** replace instanton moduli spaces by the moduli of solutions to the Seiberg-Witten equations (the monopole equations from the electromagnetic dual of $\mathcal{N} = 2$ Yang-Mills). They are easier to compute and contain the same diffeomorphism-type information for most 4-manifolds.

11.3.6 Gravitational TQFTs

3d gravity as Chern-Simons theory

Pure gravity in three dimensions with cosmological constant Λ can be formulated as a Chern-Simons theory. The gravitational connection is $\omega^a{}_b$ (the spin connection) and the triad is e^a. Together they form a gauge field for the Lie algebra of $ISO(2,1)$ (Poincaré group, $\Lambda = 0$), $SO(2,2)$ ($\Lambda < 0$, anti-de Sitter), or $SO(3,1)$ ($\Lambda > 0$, de Sitter). The action is the Einstein-Hilbert action:

$$S_{\text{3d grav}} = \frac{1}{16\pi G}\int_{M^3}\left(e^a \wedge R_a + \frac{\Lambda}{6}\epsilon_{abc}e^a \wedge e^b \wedge e^c\right), \tag{11.32}$$

and with the identifications $A = \omega + e/\ell$ and $\tilde{A} = \omega - e/\ell$ (where $\ell^{-2} = -\Lambda$), this becomes two copies of Chern-Simons theory with gauge group $SL(2,\mathbb{R})$ (for $\Lambda < 0$). Three-dimensional gravity therefore has no local degrees of freedom (consistent with $H = 0$ in the TQFT sense) and is exactly solvable, with the physical content residing in the global holonomies of the gauge field.

Topological gravity and Witten-type TQFTs

Topological gravity (Witten 1988) is a TQFT on 2-dimensional moduli spaces of Riemann surfaces. It computes intersection numbers on the Deligne-Mumford compactification $\overline{\mathcal{M}}_{g,n}$ of the moduli space of genus-g Riemann surfaces with n marked points. Its partition function is related to the KdV hierarchy of integrable PDEs by Witten's conjecture (proved by Kontsevich), a non-trivial result connecting topology of moduli spaces to mathematical physics.

11.4 Level 3: Topological Phases of Matter

A topological phase is an ordinary gapped many-body system whose low-energy effective theory is a TQFT. The gap protects the ground state from local perturbations; the characteristic class of the ground state bundle provides the topological invariant. The same topological integer — Chern number, $\mathbb{Z}_2$ index, Pontryagin number — that classifies the fiber bundle also determines measurable transport coefficients.

11.4.1 The Integer Quantum Hall Effect

Landau levels and band topology

In a 2d electron system with perpendicular field B, the spectrum forms Landau levels $E_n = \hbar\omega_c(n+\frac{1}{2})$ with degeneracy $N_\phi = BA/2\pi\ell_B^2$ per level. With a periodic crystal potential, each Landau level broadens into a Bloch band $E_n(\mathbf{k})$, and the filled bands form a vector bundle over the Brillouin zone torus T^2.

The Berry connection on the n-th filled band is $\mathcal{A}_n(\mathbf{k}) = i\langle u_n(\mathbf{k})|\nabla_{\mathbf{k}}|u_n(\mathbf{k})\rangle$ and the Berry curvature is $\mathcal{F}_n = \partial_{k_x}\mathcal{A}_{n,y} - \partial_{k_y}\mathcal{A}_{n,x}$.

TKNN formula

The Thouless-Kohmoto-Nightingale-den Nijs formula (1982) gives the Hall conductivity of N filled bands as

$$\sigma_{xy} = \frac{e^2}{h}\sum_{n=1}^{N} C_n, \qquad C_n = \frac{1}{2\pi}\int_{T^2} \mathcal{F}_n\, d^2k \in \mathbb{Z}. \tag{11.33}$$

Each C_n is the first Chern number of the n-th band bundle. The Hall conductivity is quantized to parts in 10^{10} and defines the SI ohm since 2019. The protection: C_n is an integer that cannot change without closing the spectral gap, and the gap suppresses all local excitations.

The bulk Chern number $C = \sum_n C_n$ equals the number of chiral edge modes by the bulk-boundary correspondence (Hatsugai 1993): one chiral propagating mode per unit of Hall conductance.

Laughlin's argument

The quantization can be understood by Laughlin's gauge argument (1981): thread a flux Φ through a cylinder. Adiabatically increasing Φ by one flux quantum $\Phi_0 = h/e$ implements a large gauge

transformation, which maps the Hamiltonian back to itself. The current pumped around the cylinder in this process is

$$I = \frac{\partial E}{\partial \Phi} = \sigma_{xy}\, V, \tag{11.34}$$

and the quantization $\sigma_{xy} = n\, e^2/h$ follows because n electrons must be transferred across the cylinder (by the spectral flow of the Landau levels) during one flux quantum cycle. This argument is topological: it does not require any specific model, only the existence of a gap.

Hofstadter butterfly

When the ratio of the magnetic flux to a flux quantum per unit cell is rational, $\Phi/\Phi_0 = p/q$, the Bloch bands split into q subbands. The resulting fractal energy spectrum — the **Hofstadter butterfly** — has been directly observed in graphene-boron nitride moiré superlattices. The Hall conductivity in each gap is given by the Chern number of the occupied subbands, which changes as the Fermi energy crosses each subband. The quantum Hall effect of the Hofstadter butterfly has been measured in moiré systems.

11.4.2 The Fractional Quantum Hall Effect and Anyons

Laughlin wavefunctions

At filling $\nu = 1/m$ (odd m), the Laughlin wavefunction

$$\Psi_m(z_1, \ldots, z_N) = \prod_{i<j} (z_i - z_j)^m \exp\!\left(-\frac{1}{4\ell_B^2} \sum_k |z_k|^2\right) \tag{11.35}$$

is an exact ground state of the model Hamiltonian with contact interactions, and an excellent approximation for Coulomb interactions. It has a spectral gap $\Delta \sim e^2/\ell_B$, quasihole excitations of charge $e^* = e/m$, and the Hall conductivity $\sigma_{xy} = \frac{1}{m}\frac{e^2}{h}$.

Composite fermion theory

Jain's composite fermion picture (1989) maps the FQHE at $\nu = p/(2ps+1)$ to the IQHE of composite fermions — electrons bound to $2s$ flux quanta. The composite fermion fills p effective Landau levels in the residual magnetic field $B^* = B - 2s\, n\, \Phi_0$, giving a hierarchy of fractions

$$\nu = \frac{p}{2ps+1}, \qquad p, s \in \mathbb{Z}. \tag{11.36}$$

This explains the observed sequence $\nu = 1/3, 2/5, 3/7, \ldots$ and $1/5, 2/9, \ldots$ from a unified framework.

Abelian anyons and Chern-Simons effective theory

The effective field theory for the Laughlin state at $\nu = 1/m$ is

$$\mathcal{L}_{\text{eff}} = \frac{m}{4\pi} a \wedge da + \frac{1}{2\pi} A \wedge da, \tag{11.37}$$

where a is an emergent statistical gauge field and A is the electromagnetic potential. This is a $U(1)$ Chern-Simons theory at level m. Its observables: quasihole charge e/m (from the $A \wedge da$ coupling), exchange statistics phase $e^{i\pi/m}$ (from the Chern-Simons braiding of Wilson loops), and ground-state degeneracy m^g on genus-g surfaces. The K-matrix formalism generalizes this to arbitrary Abelian states: $\mathcal{L}_{\text{eff}} = \frac{1}{4\pi} K_{IJ} a^I \wedge da^J + \frac{t_I}{2\pi} A \wedge da^I$, with K an integer symmetric matrix and t_I a charge vector.

The $\nu = 5/2$ state and non-abelian anyons

The $\nu = 5/2$ fractional quantum Hall state (observed 1987) is believed to be the Moore-Read Pfaffian state, with wavefunction

$$\Psi_{5/2} = \mathrm{Pf}\left(\frac{1}{z_i - z_j}\right) \prod_{i<j} (z_i - z_j)^2 \exp\left(-\frac{1}{4\ell_B^2} \sum_k |z_k|^2\right), \tag{11.38}$$

where Pf is the Pfaffian of the matrix $1/(z_i - z_j)$. The quasiparticles are **Ising anyons** with non-abelian fusion rules $\sigma \times \sigma = 1 + \psi$, where σ is the fundamental quasihole and ψ is a neutral fermion. Moving one σ around another implements a unitary transformation on the degenerate ground state manifold. This is the platform for non-abelian topological quantum computation (see Section 11.5).

11.4.3 Topological Insulators and Superconductors

$\mathbb{Z}_2$ topological insulators in 2d: QSHE

The quantum spin Hall effect (Kane-Mele 2005) arises in 2d systems with time-reversal symmetry T ($T^2 = -1$ for spin-$\frac{1}{2}$). The first Chern number vanishes (time-reversal maps $\mathcal{F}_n \to -\mathcal{F}_n$), but a $\mathbb{Z}_2$ invariant

$$\nu = \frac{1}{2\pi} \int_{\text{half-BZ}} \mathcal{F} - \oint_{\partial\,\text{half-BZ}} \mathcal{A} \pmod 2 \in \{0, 1\} \tag{11.39}$$

is well-defined. When $\nu = 1$, the system is a 2d topological insulator with one Kramers pair of edge modes per boundary. These edge modes are helical: electrons with opposite spins propagate in opposite directions and are protected against any time-reversal-preserving backscattering. First realized in HgTe/CdTe quantum wells (König et al. 2007) and inverted InAs/GaSb bilayers.

3d strong topological insulators

In 3d, the topological classification gives a $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ invariant $(\nu_0; \nu_1\nu_2\nu_3)$. A material with $\nu_0 = 1$ is a **strong topological insulator**: it has an odd number of surface Dirac cones on any surface, protected against any perturbation that preserves T. The $\mathbb{Z}_2$ invariant is the second Stiefel-Whitney class w_2 of the occupied Bloch bundle.

Bulk-to-surface correspondence: the surface hosts $2\nu_0 + 1$ Dirac cones per time-reversal-invariant momentum, an odd total number. Since an even number of Dirac cones can always be gapped pairwise, only the odd case is topologically protected. Bi_2Se_3 and Bi_2Te_3 (both $\nu_0 = 1$, one Dirac cone) are paradigmatic strong topological insulators observed by ARPES.

Topological crystalline insulators

Beyond time-reversal, crystal symmetries (rotations, reflections, inversion) can protect topological phases not captured by the AZ classification. **Topological crystalline insulators** (TCIs) are protected by crystal symmetry rather than time-reversal. SnTe is the canonical example: the (001) surface hosts two Dirac cones protected by the mirror symmetry of the rock-salt crystal. The topological invariant is a mirror Chern number $n_{\mathcal{M}} \in \mathbb{Z}$ defined on the mirror-invariant planes in the Brillouin zone.

The periodic table of topological insulators and superconductors

The Altland-Zirnbauer (AZ) symmetry classes classify free-fermion Hamiltonians according to the presence or absence of three symmetries: time-reversal T ($T^2 = \pm 1$), particle-hole C ($C^2 = \pm 1$), and chiral symmetry $S = TC$. This gives ten classes (the Bott clock). In each dimension d and each symmetry class, the topological classification is a group:

$$\begin{array}{rcccccccccc} & \text{A} & \text{AIII} & \text{AI} & \text{BDI} & \text{D} & \text{DIII} & \text{AII} & \text{CII} & \text{C} & \text{CI} \\ d=1: & 0 & \mathbb{Z} & 0 & \mathbb{Z} & \mathbb{Z}_2 & \mathbb{Z}_2 & 0 & 2\mathbb{Z} & 0 & 0 \\ d=2: & \mathbb{Z} & 0 & 0 & 0 & \mathbb{Z}_2 & \mathbb{Z}_2 & \mathbb{Z} & 0 & 2\mathbb{Z} & 0 \\ d=3: & 0 & \mathbb{Z} & 0 & 0 & 0 & \mathbb{Z} & \mathbb{Z}_2 & 0 & 0 & 2\mathbb{Z} \end{array} \tag{11.40}$$

This periodic table (Schnyder-Ryu-Furusaki-Ludwig 2008, Kitaev 2009) classifies all free-fermion topological phases. Class A in $d = 2$ is the IQHE ($\mathbb{Z}$: Chern number); class AIII in $d = 1$ is the SSH chain ($\mathbb{Z}$: winding number); class D in $d = 2$ is the $p + ip$ superconductor ($\mathbb{Z}_2$); class DIII in $d = 3$ is the topological superconductor with time-reversal ($\mathbb{Z}$: $\mathbb{Z}$-valued winding number).

Topological superconductors and Majorana modes

A topological superconductor has a superconducting gap in the bulk and gapless Majorana modes on the boundary. For the 1d Kitaev chain (class BDI),

$$H = -\mu \sum_j c_j^\dagger c_j - t \sum_j (c_{j+1}^\dagger c_j + \text{h.c.}) + \Delta \sum_j (c_{j+1}^\dagger c_j^\dagger + \text{h.c.}), \tag{11.41}$$

the topological phase ($|\mu| < 2t$) has a Majorana zero mode at each end: $\gamma_L = c_1 + c_1^\dagger$ and $\gamma_R = i(c_N^\dagger - c_N)$. These operators satisfy $\gamma^2 = 1$ and $\{\gamma_L, \gamma_R\} = 0$: they are Majorana fermions localized at the boundaries. The non-local fermionic operator $f = (\gamma_L + i\gamma_R)/2$ defines a zero-energy degenerate ground state manifold, protected by the bulk gap. In 2d, the $p + ip$ topological superconductor (class D) has chiral Majorana edge modes and Majorana zero modes bound to vortices.

3d topological superconductors

The class DIII 3d topological superconductor has a $\mathbb{Z}$ invariant (the winding number of the topological term in the boundary theory). The surface hosts a single 2d Majorana cone: a massless 2d Dirac fermion of Majorana type (with reality condition). He-3 B-phase is the physical realization: a neutral superfluid whose surface supports a Majorana cone, confirmed by specific heat measurements.

11.4.4 Topological Semimetals

Weyl semimetals

A Weyl semimetal has point-like band crossings (Weyl nodes) in the Brillouin zone. Near each node, the Hamiltonian is

$$H = \hbar v_F \boldsymbol{\sigma} \cdot \mathbf{k} \qquad \text{(chirality } +1\text{)} \quad \text{or} \quad H = -\hbar v_F \boldsymbol{\sigma} \cdot \mathbf{k} \qquad \text{(chirality } -1\text{)}. \tag{11.42}$$

Each Weyl node is a monopole of Berry curvature with charge ± 1:

$$\frac{1}{2\pi} \oint_{S^2} \mathcal{F}\, d^2k = \pm 1, \tag{11.43}$$

where S^2 encloses the node. Nodes of opposite chirality appear in pairs (Nielsen-Ninomiya theorem, proven by topological arguments). The surface Brillouin zone hosts **Fermi arcs**: open segments of Fermi surface connecting the surface projections of nodes of opposite chirality. These have been observed by ARPES in TaAs, NbAs, and many other materials.

Transport properties: the **chiral magnetic effect** ($\mathbf{j} = \frac{e^2}{4\pi^2\hbar}\Delta\mu_5\,\mathbf{B}$, a current along $\mathbf{B}$ proportional to the chiral chemical potential $\Delta\mu_5$), **anomalous Hall effect** with $\sigma_{xy} = \frac{e^2}{h}\frac{\Delta k}{2\pi}$ (where Δk is the Weyl node separation in $\mathbf{k}$-space), and **negative magnetoresistance** (from the chiral anomaly pumping charge between nodes of opposite chirality).

Dirac semimetals

A Dirac semimetal has fourfold-degenerate Dirac nodes (a pair of Weyl nodes of opposite chirality at the same momentum), protected by additional crystal symmetry. Na_3Bi and Cd_3As_2 are examples. They can be driven to topological insulator or superconductor phases by breaking the protecting symmetry.

Nodal-line semimetals

Some materials have band crossings along one-dimensional lines (nodal lines) in the Brillouin zone, protected by a combination of time-reversal and inversion symmetry. The topological invariant is a $\mathbb{Z}_2$ Berry phase around the line, and the surfaces host nearly flat bands ("drumhead surface states").

11.4.5 Topological Order and Spin Liquids

Toric code

Kitaev's toric code (2003) is a spin-$\frac{1}{2}$ model on a square lattice with Hamiltonian

$$H_{\mathrm{TC}} = -J_e \sum_v A_v - J_m \sum_p B_p, \tag{11.44}$$

where $A_v = \prod_{e\ni v}\sigma_e^z$ is the product of σ^z on all edges touching vertex v, and $B_p = \prod_{e\in p}\sigma_e^x$ is the product of σ^x around the plaquette p. All A_v and B_p commute ($[A_v, B_p] = 0$), and the ground state satisfies $A_v|\Psi_0\rangle = |\Psi_0\rangle$ and $B_p|\Psi_0\rangle = |\Psi_0\rangle$ for all v, p.

On a torus, the ground state is 4-fold degenerate (corresponding to the four ways to thread $\mathbb{Z}_2$ flux through the two cycles). Excitations are anyons: violations of $A_v = 1$ are electric charges e, violations of $B_p = 1$ are magnetic vortices m, and their bound state $\epsilon = e \times m$ is a fermion. The braiding phase of e around m is -1: these are $\mathbb{Z}_2$ anyons described by the $\mathbb{Z}_2$ BF theory $\mathcal{L} = \frac{2}{2\pi} b \wedge da$.

The toric code is the prototypical example of topological order (Wen 1990): its ground state degeneracy on genus-g surfaces is 4^g, its excitations are anyonic, and no local operator can distinguish the degenerate ground states. Topological order is long-range entanglement: the ground state has a finite **topological entanglement entropy** $S_{\mathrm{topo}} = \gamma = \log 2$ that subtracts from the area law.

Quantum spin liquids

A quantum spin liquid (QSL) is a frustrated magnet that fails to order at $T = 0$, instead forming a coherent quantum superposition of many spin configurations. The effective description is often a deconfined gauge theory. The $\mathbb{Z}_2$ spin liquid has the toric code as its fixed point; the $U(1)$ spin liquid (Dirac spin liquid) has an emergent photon and Dirac fermion excitations.

Candidate materials include herbertsmithite ($ZnCu_3(OH)_6Cl_2$), a 2d kagome antiferromagnet, and various triangular-lattice compounds. The key experimental signatures of topological order in spin liquids are: a T-linear specific heat (from gapless excitations), no long-range magnetic order, and fractionalized excitations (spinons carrying $S = 1/2$ without a full spin flip).

Non-abelian topological order: the Kitaev honeycomb model

Kitaev's honeycomb model (2006) has spin-$\frac{1}{2}$ on a honeycomb lattice with bond-dependent interactions:

$$H = -J_x \sum_{x\text{-bonds}} \sigma_i^x \sigma_j^x - J_y \sum_{y\text{-bonds}} \sigma_i^y \sigma_j^y - J_z \sum_{z\text{-bonds}} \sigma_i^z \sigma_j^z - K \sum_i \sigma_i^x \sigma_i^y \sigma_i^z. \tag{11.45}$$

In the phase $J_x = J_y \ll J_z$, the model reduces to the toric code ($\mathbb{Z}_2$ topological order). In the phase where all J's are comparable, with a non-zero three-spin term K (a time-reversal-breaking perturbation), the model is exactly solvable and realizes non-abelian topological order: the vortex excitations (fluxes of the $\mathbb{Z}_2$ gauge field) carry Majorana zero modes and have non-abelian Ising anyon statistics. The effective theory is $SU(2)$ Chern-Simons at level $k = 1$, describing the same universality class as the $\nu = 5/2$ FQH state. Candidate materials include α-$RuCl_3$ and Na_2IrO_3.

11.4.6 Symmetry-Protected Topological Phases

Beyond topologically ordered phases (which have intrinsic topological order), there exist **symmetry-protected topological (SPT) phases**: gapped phases whose non-trivial properties depend on a symmetry G. If the symmetry is broken, SPT phases become trivial.

In 1d, the Haldane phase of the spin-1 chain is an SPT protected by a combination of time-reversal, inversion, and π-rotation symmetries. Its edge has spin-$\frac{1}{2}$ degrees of freedom (protected by the symmetry). The bulk gap and the $\mathbb{Z}_2$ invariant (the quantized Berry phase $\theta = 0$ or π) classify it.

The general classification of SPT phases uses the cohomology of the symmetry group: $H^{d+1}(G, U(1))$ classifies d-dimensional bosonic SPT phases with symmetry G. For $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ (time-reversal and an internal $\mathbb{Z}_2$) in 1d, this gives $\mathbb{Z}_2$: two phases, trivial and Haldane. For fermionic SPT phases, the classification uses group supercohomology or, more generally, spin cobordism.

11.5 Topological Quantum Computing

Topological quantum computing (TQC) exploits non-abelian anyons to implement fault-tolerant quantum gates. The key idea: quantum information is stored not in the local state of individual qubits but in the global, topologically protected degeneracy of the ground state of a collection of non-abelian anyons. Gates are implemented by braiding anyons around each other.

11.5.1 Non-Abelian Anyons and Braiding

For a system of $2n$ Ising anyons σ (all fusing to the vacuum in pairs), the ground state is 2^{n-1}-fold degenerate. The braiding of anyons i and j (exchanging them counterclockwise) is described by the **braid group** B_{2n}, acting as a unitary matrix $\rho(\sigma_{ij})$ on the degenerate ground state.

For Ising anyons, the fusion rules are $\sigma \times \sigma = 1 + \psi$, and the braiding matrices (in the basis of fusion channels) are

$$\rho(\sigma_{12}) = e^{i\pi/8} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{pmatrix}, \tag{11.46}$$

generating a representation of the braid group. Moving anyon 1 around anyon 2 enacts a controlled phase gate on the qubit encoded in the fusion channel of the pair.

For Fibonacci anyons ($\tau \times \tau = 1 + \tau$), the braiding matrices generate a dense subset of $SU(2)$ (for the appropriate sector), making the computation **universal**: any quantum gate can be approximated to arbitrary precision by a sequence of braidings. This is the TQC scheme of Freedman-Kitaev-Larsen-Wang (2002).

11.5.2 Physical Platforms

Fractional quantum Hall $\nu = 5/2$: Non-abelian Ising anyons. Evidence from interference experiments in quantum point contacts and thermal Hall measurements.

Semiconductor-superconductor nanowires: A spin-orbit-coupled semiconductor wire (InAs, InSb) proximitized by an s-wave superconductor (Al) and placed in a magnetic field realizes the Kitaev chain in the topological phase. Zero-bias conductance peaks observed in many experiments; debate about whether they arise from true Majorana modes or from trivial Andreev states is ongoing.

Kitaev honeycomb materials: α-$RuCl_3$ and Na_2IrO_3 may host Kitaev spin liquid ground states with non-abelian anyon excitations. Current experiments probe the half-integer thermal Hall conductance as a signature.

Topological superconductors in 2d: Chiral $p + ip$ superconductors (proposed in Sr_2RuO_4, and various Majorana platforms) host chiral Majorana edge modes and Majorana zero modes in vortex cores.

11.5.3 Error Protection from Topology

The key advantage of TQC is that errors require local operators to change the topological sector. In a system with anyons separated by distance L, an error that moves anyon i to the position of anyon j requires a string of local operators of length $\sim L$, giving an error rate $\sim e^{-L/\xi}$. For macroscopic systems, this error rate is negligible. By contrast, conventional quantum error correction requires active syndrome measurement and correction, adding overhead.

In the toric code, the logical qubit is encoded in the ground state degeneracy on the torus; a logical error requires an operator string wrapping around the torus, exponentially suppressed in L.

11.6 The Mathematical Unity Across Three Levels

Despite their physical differences, the three levels share a single mathematical framework: the Chern-Weil homomorphism, applied to bundles over different base spaces.

The Yang-Mills instanton has Pontryagin index $k = \frac{1}{8\pi^2}\int_{M_4} \mathrm{Tr}(F \wedge F)$. By Stokes,

$$\int_{M_4} \mathrm{Tr}(F \wedge F) = \int_{\partial M_4} \mathrm{Tr}\Big(A \wedge dA + \tfrac{2}{3} A \wedge A \wedge A\Big), \tag{11.47}$$

so the Chern-Simons Lagrangian on ∂M_4 is the boundary term of the instanton number. The Skyrmion charge is the same $\pi_3(SU(2))$ invariant as the instanton, but in Minkowski spacetime.

The Hopf invariant is the Chern-Simons invariant on S^3. The quantum Hall Chern number is the $\mathrm{Tr}(F \wedge F)$ integral over the ground state bundle on the Brillouin zone torus. The toric code topological degeneracy is the $\mathbb{Z}_2$ BF theory partition function on the torus.

What differs is the bundle:

System	**Bundle**	**Base space**	**Invariant**
Soliton (kink)	Order parameter	Physical space	π_0 winding
Vortex	Phase bundle	Space minus core	π_1 winding
Monopole	$U(1)$ gauge bundle	S^2 at infinity	$c_1 \in \mathbb{Z}$
Skyrmion	$SU(2)$ field	S^3 (compactified space)	$\pi_3 \in \mathbb{Z}$
Instanton	$SU(2)$ gauge bundle	S^4 (Euclidean spacetime)	$c_2 \in \mathbb{Z}$
QHE	Bloch band bundle	Brillouin zone T^2	$c_1 \in \mathbb{Z}$
3d TI	Bloch band bundle	Brillouin zone T^3	$w_2 \in \mathbb{Z}_2$
Toric code	$\mathbb{Z}_2$ flat bundle	Torus (spatial)	Holonomy
Chern-Simons	Flat G-bundle	3-manifold	Holonomy, Jones poly.

The same mathematical machine — characteristic classes via curvature integrals — produces topological integers in every case. The physical interpretation (stable particle, tunnelling amplitude, Hall conductance, manifold invariant, ground state degeneracy) is completely different in each, but the mathematical structure is identical.

11.7 Conclusion: Obstruction at Three Scales

Every phenomenon in this chapter is an instance of the same geometric principle: a topological obstruction prevents a field configuration from being continuously deformed to the trivial one, and this obstruction is an integer (or a $\mathbb{Z}_2$ element), robust and protected.

At **Level 1**, the obstruction is a superselection rule: the Fock space decomposes into sectors, the Hamiltonian is block-diagonal, and the Schrödinger equation cannot cross sector boundaries. The topological charge gives an integer with certainty. Rich dynamics occurs within each sector; the game engine runs at full power, just not across boundaries.

At **Level 2**, there is no game engine. The Hamiltonian vanishes; the Hilbert space is finite-dimensional; the only observables are topological invariants of the manifold or embedded links, giving discrete values with certainty. The theory is exactly soluble and its partition functions are manifold invariants, knot polynomials, or 3-manifold invariants.

At **Level 3**, the game engine runs in the full many-body Fock space, but the spectral gap suppresses all local dynamics in the ground state. The effective low-energy theory is Level 2. The topological invariant (Chern number, $\mathbb{Z}_2$ index, winding number) is protected jointly by the gap and the integrality of the characteristic class. Measurements of transport coefficients or entanglement entropy give rational multiples of fundamental constants with extraordinary precision.

These three voices speak the same mathematical language — characteristic classes, homotopy groups, the Chern-Weil homomorphism — but in fundamentally different physical registers. The

preceding chapters have built the mathematical vocabulary; this chapter has shown where that vocabulary is spoken, and what it reveals about the universe when it is.