Lecture Notes on

Geometry and Topology

Kevin Zhou

kzhou7@gmail.com

These notes cover geometry and topology in physics, as covered in MIT's undergraduate seminar on the subject during the summer of 2016. They focus on how the mathematics is applied, in the context of particle physics and condensed matter, with little emphasis on rigorous proofs. For instance, no point-set topology is developed or assumed. Nothing in these notes is original; they have been compiled from a variety of sources. The primary sources were:

- Nakahara, *Geometry, Topology, and Physics*. The "standard" textbook on the subject for physicists. Covers the usual material (simplicial homology, homotopy groups, de Rham cohomology, bundles), with extra chapters on advanced subjects like characteristic classes and index theorems. There are many errata, and the most comprehensive list is available here.
- Robert Littlejon's Physics 250 notes. These notes primarily follow Nakahara, with some extra physical applications. They are generally more careful and precise.
- Nash and Sen, *Topology and Geometry for Physicists*. A shorter alternative to Nakahara which covers the usual material from a slightly more mathematical perspective. For instance, topological spaces are defined, and excision and long exact sequences are used. However, many propositions are left unproven, and some purported proofs are invalid.
- Schutz, *Geometrical Methods of Mathematical Physics*. A basic introduction to differential geometry that focuses on differential forms.
- Baez, *Gauge Fields, Knots, and Gravity*. Covers differential geometry and fiber bundles as applied in gauge theory. Like Nash and Sen, it has a "math-style" presentation, but not rigorous proofs. Also contains neat applications to Chern-Simons theory and knot theory.
- Jeevanjee, *An Introduction to Tensors and Group Theory for Physicists*. A pedagogical introductory-level book that could serve as a prerequisite for these notes.

The most recent version is here; please report any errors found to kzhou7@gmail.com.

Contents

1 Preliminaries

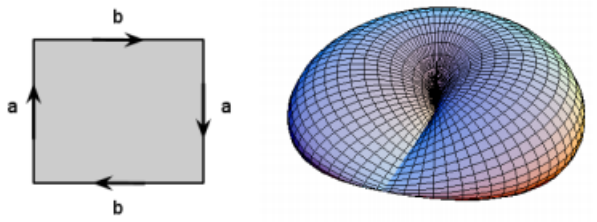
1.1 Constructing Spaces

Before diving into the formal definitions, we'll look at some at examples of spaces with nontrivial topology. Informally, a 'space' X is some set of points, such as the plane. Let $\sim$ be an equivalence relation. Then the quotient space $X/\sim$ is the result of 'gluing together' all points which are equivalent under $\sim$.

Example. Consider the real line $\mathbb{R}$, and let $x \sim y$ if $x - y$ is an integer. Then the quotient space is the unit interval $[0, 1]$ with edges identified, which is topologically the circle S^1.

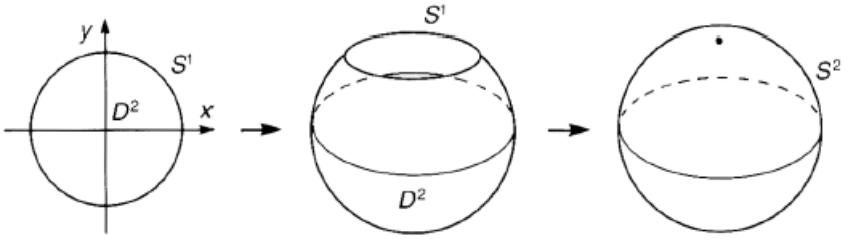
Example. Consider a closed unit square in the plane $\mathbb{R}^2$. We can identify various sides to get different results.

- Identifying two opposite sides in the same direction gives a cylinder, while identifying them in the opposite direction gives a Mobius strip; the latter is an example of a nonorientable surface.
- Identifying both pairs of opposite sides in the same direction gives the torus T^2.
- Identifying one pair of sides in the opposite direction gives the Klein bottle; doing this to both gives the projective plane RP^2.



Both the Klein bottle and the projective plane (shown above) are non-orientable, and cannot be embedded in $\mathbb{R}^3$ without self-interaction. Another construction of the projective plane is to take the two-sphere S^2 and identify opposite points. This is equivalent to taking the closed northern hemisphere of the sphere and identifying opposite edges of the equator; deforming the hemisphere into a square gives our construction.

Example. Identifying all points on the boundary of the closed unit disc D^2 yields the sphere S^2, as shown below.



More generally, identifying the boundary of D^n (i.e. constructing D^n/S^{n-1}) yields S^n.

Example. The n-dimensional real projective space $\mathbb{R}P^n$ is the set of nonzero vectors in $\mathbb{R}^{n+1}$ under the equivalence relation $x \sim y$ if $x = cy$ for real nonzero c. That is, it is the set of two-sided lines in one higher dimension.

This definition differs from our earlier construction of $\mathbb{R}P^2$, but we can link it back. Every two-sided line intersects a unit sphere in two antipodal points, so $\mathbb{R}P^n$ is S^n with antipodes identified. Equivalently, it is the closed northern hemisphere of S^n with opposite points on the equator identified. But this is just D^n with opposite points on the boundary identified. For the case $n = 2$, deforming the disc to a square gives our original construction.

Example. States (kets) in quantum mechanics live in a complex vector space, but they must be normalized and are insensitive to overall phase. Therefore kets are actually equivalence classes,

$$[|\psi\rangle] = \{c|\psi\rangle \mid c \in \mathbb{C}, c \neq 0\}.$$

We call $[|\psi\rangle]$ a ray in the Hilbert space; the set of rays is a complex projective space.

Example. Polarizations of light. The electric field of a monochromatic light wave propagating in the z direction may be written as

$$\mathbf{E} = \mathrm{Re}\left[\mathbf{a}e^{i(kz-\omega t)}\right]$$

where $\mathbf{a}$ is a complex polarization living in $\mathbb{C}^2$. Then the space of polarization states is naively $\mathbb{C}^2 = \mathbb{R}^4$. If we only care about the direction of the electric field, not its magnitude, we identify $\mathbf{a}$ with $c\mathbf{a}$ for any real $c > 0$ and ignore $\mathbf{a} = 0$. The resulting quotient space is S^3.

However, there is a further ambiguity in the phase of the wave. By changing the origin of time, we can move phase factors between $\mathbf{a}$ and the $e^{i(kz-\omega t)}$ factor. We thus identify $\mathbf{a}$ and $e^{i\alpha}\mathbf{a}$. The resulting quotient space $S^3/\sim$ is S^2, and this construction is called the Hopf fibration.

1.2 Topological Invariants

Definition. Let X_1 and X_2 be topological spaces. A continuous map $f\colon X_1 \to X_2$ with a continuous inverse is called a homeomorphism. If there exists a homeomorphism between X_1 and X_2, then the two spaces are homeomorphic.

Homeomorphism is an equivalence relation, and we will use it to identify two topological spaces as "the same". Intuitively, if two spaces are homeomorphic, the spaces may be "deformed into each other", though this comes with some caveats.

Example. A ring in $\mathbb{R}^3$ may be considered a topological space by the subspace topology. However, the topological space consisting of two linked rings is homeomorphic to one containing two unlinked rings, even though there exists no continuous map between the two in $\mathbb{R}^3$. Topologically, the embedding of the rings in $\mathbb{R}^3$ is irrelevant; the only important feature is that there are two of them.

It is possible to detect the linking of the rings topologically, but this takes more work. For example, if S is the set of points in the rings, we may compute the fundamental group of $\mathbb{R}^3 \setminus S$.

In theory, we could try to classify the equivalence classes of homeomorphism, but this is intractable. An easier goal is to define and compute topological invariants, i.e. properties of topological spaces which are invariant under homeomorphism. Then the invariants will give us a coarser, but hopefully still useful, set of equivalence classes. Invariants can be numbers, binary properties, or more generally algebraic objects such as groups. Connectedness, compactness, and the Hausdorff property are all topological invariants.

Example. A closed interval $[a, b]$ is not homeomorphic to an open interval (a, b), since only the former is compact. However, $(-\pi/2, \pi/2)$ is homeomorphic to $(-\infty, \infty)$ by the tangent function, so boundedness is not a topological invariant.

Example. The open disc D^2 is homeomorphic to $\mathbb{R}^2$, by the stereographic projection. Moreover, we know that taking the closed disc and identifying the boundary yields S^2, which implies that $\mathbb{R}^2$ along with a point at infinity is also homeomorphic to S^2. This construction is called the one-point compactification.

Generalizing the reasoning for the open disk to $D^n \cong \mathbb{R}^n$ we have $D^n \times D^m \cong D^{n+m}$. Taking the boundaries of both sides using

$$\partial(M_1 \times M_2) = ((\partial M_1) \times M_2) \cup (M_1 \times (\partial M_2))$$

gives

$$S^{n+m-1} \cong (S^{n-1} \times D^m) \cup (D^n \times S^{m-1}).$$

Definition. Let X and Y be topological spaces, and let $f_1 \colon X \to Y$ and $f_2 \colon X \to Y$ be continuous maps. Then f_1 and f_2 are homotopic if there exists a continuous function

$$g \colon [0,1] \times X \to Y$$

so that $g(0,x) = f_1(x)$ and $g(1,x) = f_2(x)$.

Definition. A path-connected topological space X is simply connected if all loops can be deformed to a point. Alternatively, it means that all continuous functions $f \colon S^1 \to X$ are homotopic to a constant function. Simple-connectedness is a topological invariant; for example, it lets us tell apart $\mathbb{R}^2$ and $\mathbb{R}^2 \setminus \{0\}$.

Definition. Two topological spaces X and Y are of the same homotopy type if there exist continuous functions $f \colon X \to Y$ and $g \colon Y \to X$ so that $f \circ g$ and $g \circ f$ are homotopic to the identity.

Example. Homotopy allows us to "contract" dimensions away, making it much coarser than homeomorphism. However, it is somewhat closer to our idea of "continuous deformation".

- The line $[0,1]$ is of the same homotopy type as a point.
- S^1 is of the same homotopy type as a cylinder, and a Mobius strip.
- The sphere S^n is of the same homotopy type as $\mathbb{R}^{n+1} \setminus \{0\}$.

Homotopy classes of maps $f \colon X \to Y$ may also be used to classify a topological space, as long as either the domain or image are fixed. For example, the homotopy groups of Y are the homotopy classes of maps $f \colon X \to Y$ where $X = S^n$.

1.3 Euler Characteristic

The Euler characteristic is a useful topological invariant for polyhedra. Define a polyhedron as a subset of $\mathbb{R}^3$ bounded by surfaces, called faces. Faces must be simply connected and arranged so that the boundary of two adjacent faces is an edge; two edges can only at a single vertex. The faces, edges, and vertices are called simplexes, and aside from this restriction, they may have any shape.

Definition. Define the Euler characteristic $\chi(K)$ of a polyhedron K as

$$\chi(K) = V - E + F$$

where the quantities on the right are the number of vertices, edges, and faces respectively.

Theorem. If two polyhedra are homeomorphic, they have the same Euler characteristic.

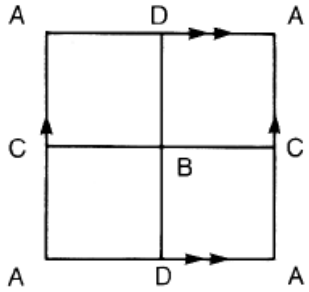
The above theorem allows us to define the Euler characteristic $\chi(X)$ of a general $X \subset \mathbb{R}^3$. To compute it, we form a 'polyhedronization' of X and compute its Euler characteristic.

Example. We give some examples of the Euler characteristic.

- The Euler characteristic of a point, line, or solid disc is 1.
- The Euler characteristic of S^1 is the same as that of the triangle (with no interior), so $\chi(S^1) = 3 - 3 = 0$.
- The Euler characteristic of any polyhedron homeomorphic to S^2 is 2. This is called Euler's theorem; for example, it applies to all Platonic solids.

It's more difficult to compute the Euler characteristics of topologically nontrivial spaces. The embedding of the torus T^2 in $\mathbb{R}^3$ is complicated, and the projective plane can't be embedded in $\mathbb{R}^3$ at all. Instead, we can compute the Euler characteristic by drawing simplexes on squares with edges identified, as we saw in the first section.

Example. The torus T^2. It's tempting to just make the entire square a single face, but this is incorrect because the face is not simply connected once the edges are identified. Instead, we split the square into four faces, shown below.



One must be careful to avoid double-counting edges and vertices. There are only four vertices and eight edges, giving $\chi(T^2) = 4 - 8 + 4 = 0$. Similarly, we have $\chi(\text{Klein bottle}) = 0$ and $\chi(\text{projective plane}) = 1$.

Definition. The connected sum $X \sharp Y$ of two surfaces X and Y is the surface obtained by removing a small disc from both X and Y and connecting the holes with a cylinder.

Theorem. For any two surfaces X and Y,

$$\chi(X \sharp Y) = \chi(X) + \chi(Y) - 2.$$

Proof. Consider polyhedra homeomorphic to X and Y, and suppose the 'small discs' removed are triangle, subsequently connected by a triangular prism. Then the number of vertices is unchanged, the number of edges goes up by 3, and the number of faces goes up by 1.

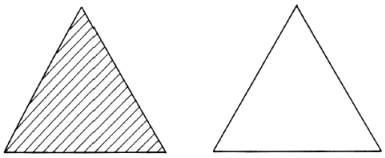
Example. The sum $T^2 \sharp T^2$ is the torus with two holes, Σ_2. Using the above result, the torus with g holes Σ_g has Euler characteristic $2 - 2g$.

Theorem. If two figures X and Y have the same homotopy type, then $\chi(X) = \chi(Y)$.

To get some intuition for this result, consider the fact that a point and line segment have the same homotopy type, because one can shrink a line segment into a point. At the very last stage of this shrinking, a vertex and edge are lost simultaneously, leaving the Euler characteristic the same; in general it does not change upon contractions.

2 Simplicial Homology

Homology groups are a refinement of the Euler characteristic, also computable by a 'polyhedronization' of space. To see the core idea of homology groups, consider the two triangles shown below.



Topologically, the main difference is that the latter has a 'hole', while the former does not. In both cases, the edges form a closed loop, and thus have no boundary. However, in the first case, the edges are themselves the boundary of a face. Thus, the general idea is that to detect holes, we must find regions without boundaries, which are not themselves the boundary of any other region.

2.1 Simplicial Complexes

Homology groups are finitely generated abelian groups, so we begin by reviewing some fundamental facts about abelian groups, writing the group operation as $+$.

Theorem (FIT)**.** Let $f\colon G_1 \to G_2$ be a homomorphism. Then $G_1/\ker f \cong \operatorname{im} f$.

Example. Consider the group $\mathbb{Z}$ and the subgroup of multiples of k, $k\mathbb{Z}$. Quotienting out by $k\mathbb{Z}$ yields $\mathbb{Z}_k$, the cyclic group of order k.

For an Abelian group G with $x \in G$ and $k \in \mathbb{Z}$, let kx denote x added to itself n times. Given elements $x_1, \ldots, x_r$, the most general group element that can be made by them is of the form

$$k_1x_1 + \cdots + k_rx_r.$$

If H is the set of such elements, we say H is generated by the x_i. If there are no nontrivial relations among the elements, i.e. $\sum k_ix_i = 0$ implies $k_i = 0$, then the x_i are said to be linearly independent.

Definition. If G is generated by r linearly independent elements, G is called the free Abelian group of rank r. It is isomorphic to $\mathbb{Z}^r = \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$.

We may also have a nontrivial relation $kx = 0$. Our main claim is that this is essentially the only kind of relation: all Abelian groups look like products of $\mathbb{Z}$'s and $\mathbb{Z}_k$'s.

Lemma. Let G be a free Abelian group of rank r and let H be a nonzero subgroup of G. Then it is possible to choose p generators x_i and p numbers k_i so that $k_1x_1, \ldots, k_px_p$ generate H.

Theorem (Fundamental theorem of finitely generated Abelian groups)**.** Let G be a finitely generated Abelian group with m generators. Then we may write

$$G \cong \mathbb{Z}^r \oplus Z_{k_1} \oplus \cdots \oplus Z_{k_p}$$

where $m = r + p$. We call r the rank of G.

Proof. Let G have m generators $x_1, \ldots, x_m$ and consider the homomorphism

$$f\colon \mathbb{Z}^m \to G, \quad f(n_1, \ldots, n_m) = n_1x_1 + \cdots + n_mx_m.$$

Then the FIT says $\mathbb{Z}^m/\ker f \cong G$. However, since $\ker f$ is a subgroup of $\mathbb{Z}^m$, the lemma implies that we can choose generators so that

$$\ker f \cong k_1\mathbb{Z} \oplus \cdots \oplus k_p\mathbb{Z}.$$

Quotienting $\mathbb{Z}^m$ with this gives the result.

Note. There are several similar-looking sums and products here.

- The Cartesian product of two sets $S \times T$ is the set of elements (s, t) with $s \in S$ and $t \in T$. The direct product $G \times H$ of two groups directly generalizes it.
- The direct sum of two groups $G \oplus H$ is only defined for Abelian groups. For a finite number of summands, it is identical to the direct product, i.e. it contains the set of ordered pairs (g, h). However, for an infinite number of summands, all but a finite number of entries in the tuple must be the identity element, while the direct product has no such restriction. This distinction can be motivated in category theory.
- The tensor product of two groups $G \otimes H$ is more complicated. In the special case of abelian groups, it is the free group generated by elements $g \otimes h$ with relations inherited from G and H, i.e. $(g_1 \otimes h)(g_2 \otimes h) = (g_1 g_2 \otimes h)$.

For example, if $G = H = \mathbb{Z}^3$, then $G \oplus H = \mathbb{Z}^6$, while $G \otimes H = \mathbb{Z}^9$. If $G = H = \mathbb{Z}$, then $G \oplus H = \mathbb{Z}^2$, while $G \otimes H = \mathbb{Z}$ and the tensor product is simply multiplication.

Conceptually, the direct product multiplies the cardinalities of finite objects, while the direct sum adds the number of generators; confusion arises here because we may view finite groups as finite or finitely generated. By contrast, the tensor product multiplies the number of generators. The same confusion arises for finite-dimensional vector spaces, where both the direct product and direct sum add the dimension, while the tensor product multiplies it.

When describing the Euler characteristic, we gave a heuristic definition of the faces, edges, and vertices that made up the polyhedra. We now refine this idea. The standard objects are taken to be triangles and their higher-dimensional analogues, called simplexes.

Definition. A set of points $p_0, \ldots, p_r \in \mathbb{R}^m$ is geometrically independent if there is no $(r-1)$-dimensional plane containing all the points. Equivalently, making $\mathbb{R}^m$ into a vector space with origin at p_0, the vectors $p_1, \ldots, p_r$ are linearly independent.

Definition. Let $p_0, \ldots p_r \in \mathbb{R}^m$ be geometrically independent. The r-simplex $\sigma_r = \langle p_0, \ldots, p_r \rangle$ is the set of points

$$\sigma^r = \left\{ x = \sum c_i p_i \,\middle|\, c_i \geq 0, \sum c_i = 1 \right\}.$$

The c_i are called barycentric coordinates.

Example. A 0-simplex $\langle p_0 \rangle$ is a point/vertex, a 1-simplex $\langle p_0 p_1 \rangle$ is a line/edge, and 2-simple $\langle p_0 p_1 p_2 \rangle$ is a solid triangle and a 3-simplex is a solid tetrahedron. The independence requirement rules out degenerate shapes.

Definition. If we choose $q+1$ points $p_{i_0}, \ldots, p_{i_q}$, then they form a simplex σ_q which is called a q-face of σ_r, and we write $\sigma_q \leq \sigma_r$. If $\sigma_q \neq \sigma_r$, we say σ_q is a proper face of σ_r and write $\sigma_q < \sigma_r$.

Example. There are six proper faces of a 2-simplex, i.e. the three edges and three points. We define a 0-simplex to have no proper faces.

Definition. Let K be a finite set of simplexes in $\mathbb{R}^m$. Then K is a simplicial complex if the simplexes are 'nicely fitted together', meaning that:

1. For all $\sigma \in K$, all faces of σ are in K.
2. If $\sigma, \sigma' \in K$, then their intersection is either empty of a common face of σ and σ', i.e.

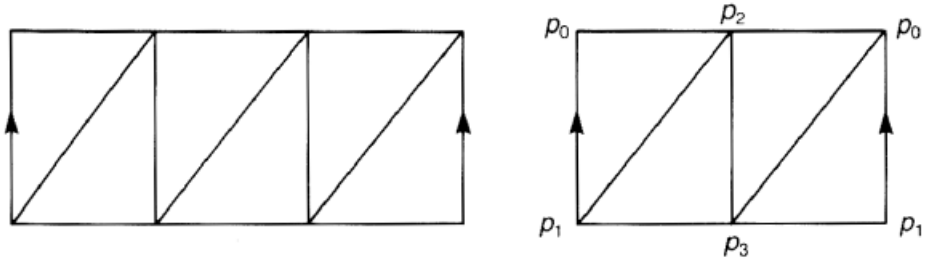
$$\sigma \cap \sigma' = \varnothing \text{ or } \sigma \cap \sigma' \leq \sigma \text{ and } \sigma \cap \sigma' \leq \sigma'.$$

   For example, we can attach triangles together at points or edges, and we can only attach lines together at their endpoints.

Definition. Given a simplicial complex K, define the polyhedron $|K|$ as the union of all elements of K as a subset of $\mathbb{R}^m$.

Definition. A topological space X is said to be triangulable if there exists a simplicial complex K and a homeomorphism $f\colon |K| \to X$. We will only consider triangulable spaces. Note that the triangulation of a space is far from unique.

Example. Consider triangulation of the cylinder $S^1 \times [0,1]$.



We may construct it as a subset of $\mathbb{R}^3$, but for convenience we will draw all our triangulations in $\mathbb{R}^2$ and use arrows to indicate where sides coincide. The simplest construction is at left above. The construction on the right is not a simplicial complex, because $\langle p_0 p_1 p_2 \rangle$ and $\langle p_2 p_3 p_0 \rangle$ intersect in two points, which is not a face.

2.2 Simplicial Homology

We now define oriented simplexes, written as $(\ldots)$ instead of $\langle \ldots \rangle$.

- An oriented 1-simplex $\sigma_1 = (p_0 p_1)$ can be viewed as a directed line segment, traversed from p_0 to p_1. We may assign a group structure to 1-simplexes. We let $(p_0 p_1)$ as a generator, and set $(p_1 p_0) = -(p_0 p_1)$.
- Similarly, for oriented 2-simplexes, we set

$$(p_0 p_1 p_2) = (p_1 p_2 p_0) = (p_2 p_0 p_1) = -(p_2 p_1 p_0) = -(p_1 p_0 p_2) = -(p_0 p_2 p_1).$$

  The two equivalence classes correspond to 'traversing the triangle' clockwise/counterclockwise.
- For r-simplexes, we do the same, determining the sign using the sign of the permutation. For $r = 0$, we formally define $\sigma_0 = p_0$.

Note. One key piece of intuition is that oriented simplexes are "things you can integrate over", i.e. they are directed paths and signed areas/volumes. We'll return to this correspondence when we discuss cohomology, which explicitly deals with integration using differential forms.

We now use oriented simplexes to define groups. Let K be an n-dimensional simplicial complex, and regard the simplexes $\sigma_\alpha \in K$ as oriented simplexes.

Definition. The r-chain group $C_r(K)$ is the free Abelian group generated by the oriented r-simplexes of K. If $r > \dim K$, we define $C_r(K) = 0$. An element of $C_r(K)$ is called an r-chain.

If there are I_r r-simplexes in K, denoted by $\sigma_{r,i}$, then the r-chains are

$$c = \sum c_i \sigma_{r,i}$$

for integers c_i, called the coefficients of c. The r-chain group is $\mathbb{Z}^{I_r}$.

Definition. Let σ_r be an oriented r-simplex. The boundary $\partial_r \sigma_r$ of σ_r is the $(r-1)$-chain

$$\partial_r \sigma_r = \sum_{i=0}^{r} (-1)^i (p_0 p_1 \dots p_{i-1} p_{i+1} \dots p_r).$$

We formally define $\partial_0 \sigma_0 = 0$ for $r = 0$. Also note that $\partial_r(-\sigma_r) = -\partial_r \sigma_r$.

We may extend the domain of ∂_r to all of $C_r(K)$ by letting it act linearly,

$$\partial_r c = \sum_i c_i \partial_r \sigma_{r,i}.$$

We call ∂_r the boundary operator; it is a homomorphism between chain groups. The chain complex $C(K)$ is the sequence of groups and homomorphisms

$$0 \xrightarrow{i} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0$$

where i is the inclusion map.

Note. The minus signs in the definition of ∂_r have an intuitive geometric motivation.

- Consider the 1-simplexes $(p_0 p_1)$ and $(p_1 p_2)$. Geometrically, $(p_0 p_1) + (p_1 p_2)$ can be viewed as equivalent to $(p_0 p_2)$, so they should have the same boundary. This is only possible if a minus sign is present, i.e.

$$\partial_1((p_0 p_1) + (p_1 p_2)) = p_2 - p_1 + p_1 - p_0 = p_2 - p_0.$$

- Consider the sum of the 1-simplexes $(p_0 p_1)$, $(p_1 p_2)$, and $(p_2 p_0)$. These segments form a closed triangle, which has no boundary, so we want the boundary to be zero.
- The boundary of the oriented 2-simplex $(p_0 p_1 p_2)$ should be $(p_0 p_1) + (p_1 p_2) + (p_2 p_0)$ to represent 'going around the triangle'.

Definition. Define the r-cycle group $Z_r(K) = \ker \partial_r$. Elements of $Z_r(K)$ are called r-cycles; they are the r-simplexes with no boundary.

Definition. Define the r-boundary group $B_r(K) = \operatorname{im} \partial_{r+1}$. Elements of $B_r(K)$ are called r-boundaries; they are the r-simplexes that are the boundary of an $(r+1)$-simplex.

Lemma. The composition $\partial_r \circ \partial_{r+1}$ is the zero map, i.e. $B_r(K) \subset Z_r(K)$.

Proof. Geometrically, the boundary of a boundary is zero. Algebraically, since the boundary operators are linear, it is sufficient to show that all oriented $(r+1)$-simplexes $\sigma = (p_0 \dots p_{r+1})$ are sent to zero. Let σ_i be the r-simplex with p_i removed and let σ_{ij} be the $(r-1)$-simplex with p_i and p_j removed. Then

$$\partial_r \partial_{r+1} \sigma = \partial_r \sum_{i=0}^{r+1} (-1)^i \sigma_i = \sum_{i=0}^{r+1} (-1)^i \left(\sum_{j=0}^{i-1} (-1)^j \sigma_{ij} + \sum_{j=i+1}^{r+1} (-1)^{j-1} \sigma_{ij} \right)$$

There are two cases, depending on whether p_j comes before or after p_i, giving contributions

$$\sum_{j<i} (-1)^{i+j} \sigma_{ij} - \sum_{j>i} (-1)^{i+j} \sigma_{ij} = 0.$$

None of the groups we have defined are topological invariants. For example, the 1-chain group of a triangle is $\mathbb{Z}^3$, and the 1-chain group of a square is $\mathbb{Z}^4$, but the two are homeomorphic. We now define a group that is.

Definition. Define the r^{th} homology group $H_r(K) = Z_r(K)/B_r(K)$. Let $H_r(K) = 0$ for $r > n$ or $r < 0$. The elements of $H_r(K)$ are equivalence classes of r-cycles $[z]$, called homology classes; if $[z] = [z']$ we say z and z' are homologous.

Note. Geometrically, two r-cycles are homologous if they differ by an r-boundary. The general intuition is that each generator of $H_r(K)$ represents an "r-dimensional hole", where an r-dimensional hole is an empty region bounded by an r-sphere. For example, the circle has one 1-dimensional hole, while the torus has *two*. (This differs from our earlier nomenclature, where we called Σ_n the "n-holed torus".) Another way to phrase this intuition is that $H_r(K)$ counts the number of distinct ways to embed an r-simplex nontrivially into K.

Theorem. Homology groups are homotopy invariants, which implies they are topological invariants. As with Euler characteristic, this lets us extend them to apply to all triangulable topological spaces.

2.3 Computation of Homology Groups

Example. A single point, $K = \{p_0\}$. Then $C_0(K) = \mathbb{Z}$ and $Z_0(K) = C_0(K)$. However, the boundary group $B_0(K)$ is trivial, so $H_0(K) = Z_0(K)/B_0(K) = \mathbb{Z}$. By similar reasoning, n points gives $H_0(K) = \mathbb{Z}^n$.

Example. A single line, $K = \{p_0, p_1, (p_0 p_1)\}$. Then $C_0(K) = \{ip_0 + jp_1\}$ and $C_1(K) = \{k(p_0 p_1)\}$. As before, the highest boundary group $B_1(K)$ is trivial, and we can compute to show that $Z_1(K)$ is trivial too. Then $H_1(K) = 0$.

Now, the boundary group $B_0(K)$ contains simplexes of the form $k(p_1 - p_0)$, while the cycle group $Z_0(K) = C_0(K)$. Then the zeroth homology group is $H_0(K) = \mathbb{Z}^2/\mathbb{Z} = \mathbb{Z}$. We can show this more formally by defining the homomorphism

$$f \colon Z_0(K) \to \mathbb{Z}, \quad f(ip_0 + jp_1) = i + j.$$

The kernel of this homomorphism is $B_0(K)$, so the quotient $H_0(K)$ is isomorphic to the image $\mathbb{Z}$.

Example. The triangle $K = \{p_0, p_1, p_2, (p_0p_1), (p_1p_2), (p_2p_0)\}$. This is a triangulation of S^1. We have $B_1(K) = 0$, so $H_1(K) = Z_1(K)$. To compute this group, let

$$z = i(p_0p_1) + j(p_1p_2) + k(p_2p_0) \in Z_1(K).$$

Then we have

$$\partial_1 z = i(p_1 - p_0) + j(p_2 - p_1) + k(p_0 - p_2) = 0$$

which implies $i = j = k$. Thus $H_1(K) = Z_1(K) = \mathbb{Z}$, identifying a "1-dimensional hole" in the space. Next, we compute $H_0(K)$. We have $Z_0(K) = C_0(K) = \mathbb{Z}^3$, and the 0-boundaries are

$$\partial_1 \left(l(p_0p_1) + m(p_1p_2) + n(p_2p_0)\right) = (n - l)p_0 + (l - m)p_1 + (m - n)p_2.$$

We now repeat the trick from the last example, defining the homomorphism

$$f\colon Z_0(K) \to \mathbb{Z}, \quad f(ip_0 + jp_1 + kp_2) = i + j + k.$$

The kernel is $B_0(K)$, so the quotient $H_0(K)$ is isomorphic to the image $\mathbb{Z}$.

Example. The solid triangle; add the simplex $(p_0p_1p_2)$ to the triangle. The 0-simplexes and 1-simplexes remain the same. However, $B_1(K)$ is no longer trivial; its elements are

$$\partial_2(m(p_0p_1p_2)) = m\left((p_1p_2) - (p_0p_2) + (p_0p_1)\right) = m\left((p_0p_1) + (p_1p_2) + (p_2p_0)\right).$$

Then $B_1(K) = Z_1(K)$, so $H_1(K) = 0$. That is, the hole has been removed.

Next, $B_2(K) = 0$ and we must compute $Z_2(K)$. However, we've just shown above that $m(p_0p_1p_2)$ has nonzero boundary unless $m = 0$, so $Z_2(K) = 0$ and thus $H_2(K) = 0$.

Example. Spheres and discs. In general, the simplicial complex containing all the proper faces of $(p_0p_1 \dots p_n)$ is homeomorphic to S^{n-1}. Including the central face gives D^n.

Through similar computations, we find that the nontrivial homology groups of S^n are $H_0(S^n) = H_n(S^n) = \mathbb{Z}$. The only nontrivial homology group of the disc is $H_0(D^n) = \mathbb{Z}$.

Prop. If K is connected, then $H_0(K) \cong \mathbb{Z}$.

Proof. If K is connected, then for any two 0-simplexes p_i and p_j, there exists a sequence of 1-simplexes $(p_ip_k), \dots, (p_mp_j)$ connecting them. The boundary of this set is $p_j - p_i$, which implies that p_i and p_j are homologous. Therefore, for $z = \sum n_ip_i$, we have

$$[z] = \sum n_i[p_i] = \sum n_i[p_1].$$

Then $[z] = 0$ (i.e. $z \in B_0(K)$) if $\sum n_i = 0$.

Next, we compute $B_0(K)$ directly. All elements of this group have the form

$$\sum n_i\partial_1(p_{i,1}p_{i,2}) = \sum n_i(p_{i,1} - p_{i,2}).$$

Then if $\sum n_jp_j \in B_0(K)$, we must have $\sum n_j = 0$. Combined with the previous fact, we have completely characterized the group $B_0(K)$.

Finally, to get $H_0(K)$, we use the usual trick. Define the homomorphism

$$f\colon Z_0(K) \to \mathbb{Z}, \quad f\left(\sum n_ip_i\right) = \sum n_i.$$

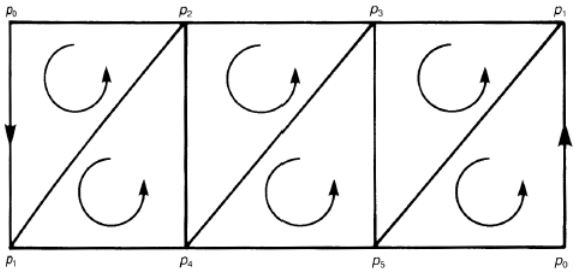
Then the kernel is $B_0(K)$, and the image is $\mathbb{Z}$. Thus $H_0(K) \cong \mathbb{Z}$.

Note. There are some links between homology and homotopy.

- If two 1-cycles are homotopic, then they are homologous. This is because they differ by a boundary, i.e. the area swept out by the homotopy.
- The converse is not true. Consider the sum $T^2 \sharp T^2$. A 1-cycle through the connecting tube is a boundary, so it is homologous to the zero cycle. But it can't be deformed continuously into the zero cycle: trying to pull it through either torus 'snaps it in half'.
- In general, the first homology group is the abelianization of the fundamental group.

We now explore some non-orientable spaces. If our space K is an n-dimensional manifold, then the highest possible nontrivial homology group is $H_n(K) = Z_n(K)$. If the manifold is orientable and closed, then the entire manifold is an n-cycle, so $H_n(K) \cong \mathbb{Z}$. If the manifold is not closed, this fails since we pick up the overall boundary (as seen in the disc example), so $H_n(K)$ is trivial. More subtly, it also fails if the manifold is not orientable. If one tried to form an n-cycle out of all the n-simplexes in the manifold, it would be impossible to do it coherently: one would 'wrap back around' in the opposite orientation.

Example. The Mobius strip. The triangulation is almost the same as the cylinder's.

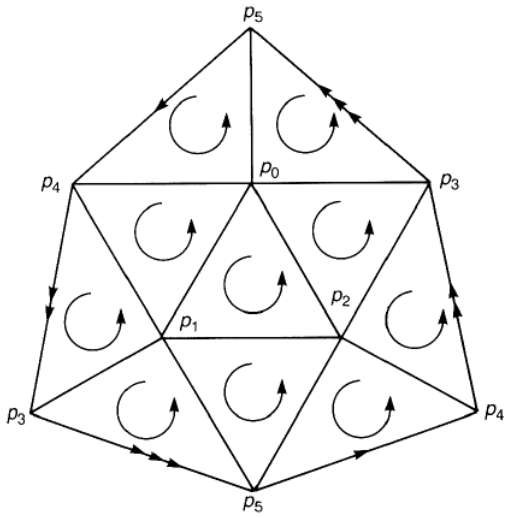


First, let's consider $H_2(K) = Z_2(K)$. A 2-cycle must have no boundary, yet each triangle has a unique edge (along the single edge of the Mobius strip) not contained in any other triangle. Therefore, there are no 2-cycles and $H_2(K)$ is trivial.

We can get additional insight for considering our 'best guess' for a 2-cycle, the set of all triangles with the orientation shown above. The boundary of this 2-chain contains the edge of the Mobius strip, but it also contains $2(p_0p_1)$. If we had glued the sides in the same direction, these would have canceled; the extra factor appears because the Mobius strip is not orientable. This presents a totally independent obstacle to having a 2-cycle.

We already know $H_0(K) = \mathbb{Z}$, so we turn to $H_1(K)$. We work intuitively. The Mobius strip is somewhat like the circle, so we should get a homology class from a chain that runs around the loop. We consider $(p_0p_2) + (p_2p_3) + (p_3p_1) + (p_1p_0)$ as a candidate; we want to show it is not the boundary of a 2-chain. Any such 2-chain would have to include the top three triangles, but then to cancel the internal edges, we would have to include the bottom three triangles. But then we cannot cancel the bottom edges, so the construction fails. Any other path around the strip differs from this one by a boundary, so there's only one kind of nontrivial homology class. We conclude $H_1(K) = \mathbb{Z}$.

Example. The projective plane, i.e. the disc with opposite points identified. The simplest candidate triangulation is a hexagon, but it's an illegal simplicial complex. As with the cylinder and Mobius strip, we need to put at least three triangles between each edge identification. We thus arrive at a correct triangulation by adding an internal triangle.



As before, the nonorientability forbids 2-cycles. We must include all the triangles, and when we do, the total boundary doesn't vanish. An explicit computation shows that it is

$$2\left((p_3p_5) + (p_5p_4) + (p_4p_3)\right).$$

Thus $H_2(K)$ is trivial.

Next, we compute $H_1(K)$. The fundamental group is $\mathbb{Z}_2$ and is generated by a loop that goes across the plane and wraps back; this translates to the chain $z = (p_3p_5) + (p_5p_4) + (p_4p_3)$. By the same logic as with the Mobius strip, this is not the boundary of any 2-chain, so it is in a nontrivial homology class. However, $2z$ is exactly the boundary of the entire projective plane, so $H_1(K) = \mathbb{Z}_2$.

This is our first encounter with a homology group that is not free. The non-free part of the group is called the torsion subgroup, and they measure in some sense the 'twisting' of the space.

Example. The torus T^2. We work entirely by intuition. The torus is a closed orientable manifold (or, alternatively, it's hollow), so $H_2(K) = \mathbb{Z}$. There are two loops we can draw, so $H_1(K) = \mathbb{Z}^2$.

2.4 Properties of Homology Groups

We now step back and consider general properties of homology groups.

Prop. If K is the disjoint union of connected components K_i, then $H_r(K) = \bigoplus H_r(K_i)$.

Proof. This direct sum decomposition clearly holds for the r-chain groups, and similarly for the subgroups $Z_r(K)$ and $B_r(K)$. Then we have

$$H_r(K) = Z_r(K)/B_r(K) = \bigoplus Z_r(K_i)/\bigoplus B_r(K_i) = \bigoplus(Z_r(K_i)/B_r(K_i)) = \bigoplus H_r(K_i).$$

Corollary. If K has n connected components, then $H_0(K) = \mathbb{Z}^n$.

Note. We may also define homology groups over the real numbers, so that all groups become real vector spaces. The free parts of homology groups don't change, as $\mathbb{Z}^n$ just becomes $\mathbb{R}^n$. However, all torsion subgroups vanish, as the quotient $H_r(K) = Z_r(K)/B_r(K)$ is always of the form $\mathbb{R}^n/\mathbb{R}^m = \mathbb{R}^{n-m}$. More concretely, a torsion subgroup $\mathbb{Z}_n$ arises from a cycle z with $nz = \partial(z')$. But if we allow real coefficients, then z is a boundary, since $z = \partial(z'/n)$. Similarly, if we define homology groups over $\mathbb{Z}_2$, torsion subgroups cannot appear because $\mathbb{Z}_2$ has no nontrivial subgroups.

Theorem (Kunneth formula)**.** For homology groups over $\mathbb{R}$, we have

$$H_r(X \times Y) = \bigoplus_{p+q=r} H_p(X) \otimes H_q(Y).$$

A more general version of this theorem also accounts for torsion. We will prove this later using de Rham cohomology.

Example. The torus T^2 is $S^1 \times S^1$. Hence the Kunneth formula gives

$$H_2(T^2) = \mathbb{R}, \quad H_1(T^2) = \mathbb{R}^2, \quad H_0(T^2) = \mathbb{R}.$$

This is intuitive, and follows because the torus is "hollow", forming a two-dimensional hole, has two independent nontrivial loops, and one connected component. More generally, we see that

$$H_k(T^n) = \mathbb{R}^{\binom{n}{k}}$$

which would be quite difficult to show directly.

Definition. The dimension of the free part of $H_r(K)$ is called the r^{th} Betti number of K, and denoted $b_r(K)$.

Theorem (Euler–Poincare)**.** Let K be an n-dimensional simplicial complex with I_r r-simplexes. Then the Euler characteristic is

$$\chi(K) = \sum_{r=0}^{N} (-1)^r I_r = \sum_{r=0}^{n} (-1)^r b_r(K)$$

where the first equality is the definition of the Euler characteristic generalized to arbitrary dimension, i.e. the Euler characteristic is a topological invariant.

Proof. Since we are only looking at dimensions of free groups, we can work over $\mathbb{R}$. Then we may apply the rank-nullity theorem to find

$$I_r = \dim C_r = \dim(\ker \partial_r) + \dim(\operatorname{im} \partial_r) = \dim Z_r + \dim B_{r-1}$$

where we are abbreviating notation. We also have

$$b_r = \dim H_r = \dim(Z_r/B_r) = \dim Z_r - \dim B_r.$$

Comparing the two sums and using $\dim B_{-1} = \dim B_n = 0$ gives the result.

We may compute homology groups efficiently using relative homology.

Definition. For a simplicial complex K and subcomplex L, the relative chain group is

$$C_r(K; L) = C_r(K)/C_r(L).$$

The relative boundary operator

$$\overline{\partial}_p \colon C_r(K; L) \to C_{r-1}(K; L)$$

is defined by mapping the coset of $c_r \in C_r(K)$ to the coset of $\partial_r c_r$. As before we define

$$Z_r(K; L) = \ker \overline{\partial}_r, \quad B_r(K) = \operatorname{im} \overline{\partial}_{r+1}, \quad H_r(K; L) = Z_r(K; L)/B_r(K; L).$$

Note that elements of $Z_r(K; L)$ need not be elements of $Z_r(K)$.

The basic intuition for relative homology is that we simply "shrink L to a point", though it's slightly more complex than that.

Example. Let $K = \{p_0, p_1, p_2, (p_0p_1), (p_1p_2), (p_2p_0)\}$ and $L = \{p_0, p_1, (p_0p_1)\}$. Then

$$C_0(K;L) = \langle p_2 \rangle, \quad C_1(K;L) = \langle (p_1p_2), (p_2, p_0) \rangle.$$

By direct computation, we have

$$B_0(K;L) = \langle p_2 \rangle, \quad Z_0(K;L) = \langle p_2 \rangle, \quad H_0(K;L) = \{0\}.$$

This is one sense in which relative homology differs from collapsing L. The homology class of the connected component is "eaten up" by taking relative homology. Next,

$$B_1(K;L) = \{0\}, \quad Z_1(K;L) = \langle (p_0p_2) + (p_2p_1) \rangle, \quad H_1(K;L) = \mathbb{Z}.$$

All higher homology groups are trivial.

Theorem (Excision)**.** Let K be a simplicial complex containing a closed subcomplex L. If L_0 is an open subcomplex of L so that the closure $\overline{L_0}$ is contained in the interior of L. Then

$$H_r(K;L) = H_r(K - L_0, L - L_0)$$

where the $-$ denotes set subtraction.

Theorem. There is a long exact sequence of homology groups

$$\ldots \xrightarrow{\partial^*} H_r(L) \xrightarrow{i^*} H_r(K) \xrightarrow{j^*} H_r(K;L) \xrightarrow{\partial^*} H_{r-1}(L) \xrightarrow{i^*} \ldots.$$

Proof. The proof has many pieces which we will only sketch. First we define the maps i^*, j^*, and ∂^*. **(finish)**

3 Homotopy Groups

Homotopy is an equivalence relation between maps f between two topological spaces X and Y. Homotopy groups are constructed from homotopy classes of such maps, where Y is the space under investigation and $X = S^n$. The resulting groups, heuristically, tell us about "n-dimensional holes" in the space, but in a more powerful way than homology groups.

3.1 The Fundamental Group

Definition. Let X be a topological space and let $I = [0, 1]$. A loop is a continuous map $\alpha\colon I \to X$ with $\alpha(0) = \alpha(1) = x_0$. We call x_0 the base point.

Prop. The set of homotopy classes of loops with in X with base point x has the structure of a group, called the fundamental group $\pi_1(X, x)$, under the operation

$$\alpha * \beta(s) = \begin{cases} \alpha(2s) & 0 \le s \le 1/2 \\ \beta(2s-1) & 1/2 \le s \le 1 \end{cases}, \quad \alpha^{-1}(s) = \alpha(1-s).$$

The unit element is the homotopy class of the constant map, $[c_x]$.

Proof. There are many things to manually check. For example, we must show that $[\alpha * \alpha^{-1}] = [c_x]$. This is verified by the homotopy

$$F(s,t) = \begin{cases} \alpha(2s(1-t)) & 0 \le s \le 1/2 \\ \alpha(2(1-s)(1-t)) & 1/2 \le s \le 1 \end{cases}.$$

The confirmation of the other parts is similar.

Definition. A topological space X is arcwise connected if, for any $x_0, x_1 \in X$, there exists a continuous map $\alpha\colon I \to X$ with $\alpha(0) = x_0$ and $\alpha(1) = x_1$.

Prop. Let X be arcwise connected. Then for any $x_0, x_1 \in X$, $\pi_1(X, x_0)$ is isomorphic to $\pi_1(X, x_1)$. Therefore we may write the fundamental group as simply $\pi_1(X)$.

Proof. Let η be a path from x_0 to x_1. Then the isomorphism is

$$P_\eta([\alpha]) = [\eta^{-1} * \alpha * \eta].$$

It is clearly a homomorphism, and it is an isomorphism because it has an inverse map,

$$P_\eta^{-1}([\alpha']) = [\eta * \alpha' * \eta^{-1}].$$

This concludes the proof. However, note that different choices of η yield different isomorphisms: prepending a loop to η affects the isomorphism by conjugation by that loop. Hence the isomorphism is not 'natural'.

Note. Arcwise connectedness is a stronger property than connectedness. For example, consider the subset of $\mathbb{R}^2$ given by

$$\{(0, y) \,|\, -1 < y < 1\} \cup \{(x, \sin \pi/x) \,|\, 0 < x < 1\}.$$

It is connected, but one cannot travel between the two pieces by a continuous path. However, for reasonable spaces, which include all spaces we study, arcwise connectedness is equivalent to connectedness. From this point onward we assume all spaces are arcwise connected.

Prop. Let X and Y be homotopic with homotopy equivalence $f\colon X \to Y$. Then $\pi_1(X, x_0)$ is isomorphic to $\pi_1(Y, f(x_0))$, so the fundamental group is homotopy invariant.

Proof. The isomorphism is to send a loop $\alpha\colon I \to X$ to $f \circ \alpha\colon I \to Y$. This is a well-defined operation on homotopy classes: if α and β are homotopic, then so are $f \circ \alpha$ and $f \circ \beta$. It is also a group homomorphism, since $f \circ (\alpha * \beta) = (f \circ \alpha) * (f \circ \beta)$.

To show that it is an isomorphism, note that it has an inverse, namely composition with the homotopy inverse g. The path $g \circ f \circ \alpha$ is homotopic to α, since $g \circ f$ is homotopic to the identity.

The most convenient homotopies for finding the fundamental group are deformation retractions.

Definition. Let X and Y be topological spaces with $Y \subset X$. A deformation refraction is a continuous map $F\colon X \times [0,1] \to X$ such that

$$F(x,0) = x, \quad F(x,1) \in Y, \quad F(y,t) = y.$$

That is, F 'shrinks' the spaces down from X to Y. We say Y is a deformation retract of X.

Prop. If Y is a deformation retract of X, then X and Y are homotopic.

Proof. Define $f\colon X \to Y$ by $f(x) = F(x,1)$ and $g\colon Y \to X$ by inclusion. Then the composition $f \circ g$ is the identity, and $g \circ f$ is homotopic to the identity by the existence of the deformation refraction.

Definition. If $\pi_1(X)$ is trivial, then X is simply connected.

Definition. If X may be deformation retracted to a single point, then X is contractible. This implies X is simply connected.

Prop. For topological spaces X and Y, $\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y)$. This can be proven by playing around with projection operators.

3.2 Examples of Fundamental Groups

We now use our results to find fundamental groups.

- $\pi_1(\mathbb{R}^n)$ is trivial because $\mathbb{R}^n$ is contractible.
- $\pi_1(S^1) = \mathbb{Z}$, where maps are indexed by their 'winding number' around the sphere. (We'll justify this more carefully below.) Then the fundamental group of the punctured disk is also $\mathbb{Z}$ by deformation retraction.
- $\pi_1(S^n)$ is trivial for $n > 1$, as we can shrink the loops to a point. More rigorously, we can always deform the path so it doesn't hit some point. Performing stereographic projection with this point as the North pole gives a path in $\mathbb{R}^n$, which must be contractible.
- $\pi_1(T^n) = \mathbb{Z}^n$, because $T^n = S^1 \times \cdots \times S^1$.
- $\pi_1(\mathbb{R}P^2) = \mathbb{Z}_2$. We found the generator when computing $H_1(\mathbb{R}P^2)$. Similarly, $\pi_1(\mathbb{R}P^n) = \mathbb{Z}_2$ for $n > 2$. The case $\mathbb{R}P^1$ is distinct since $\mathbb{R}P^1 \cong S^1$.

Covering spaces provide another mathematical tool to compute fundamental groups. We motivate them using the physical example of spin 1/2.

- Classical rotations live in $SO(3)$, a three-dimensional manifold. Consider a 3×3 rotation matrix R with unit determinant. Then

$$|R-I| = |(R-I)^T| = |R^{-1}-I| = |R^{-1}||I-R| = -|R-I|.$$

  Therefore, R has an eigenvector with eigenvalue 1, and hence it fixes an axis.

- Restricting to the orthogonal subspace, R is just a 2D rotation. Therefore, all rotations can be parametrized as $R(\hat{n},\theta)$, in terms of an axis $\hat{n}$ and an angle θ. Writing $\boldsymbol{\theta} = \theta\hat{n}$, shows the possible values of $\boldsymbol{\theta}$ fill the ball D^3. However,

$$R(\hat{n},\pi) = R(-\hat{n},\pi)$$

  so opposite points on the ball's surface are identified.

- In quantum mechanics, spin rotations are described by $SU(2)$. Elements of $SU(2)$ may be written in terms of the Cayley–Klein parameters,

$$U = x_0 I - i\mathbf{x}\cdot\boldsymbol{\sigma}, \quad x_0^2 + \mathbf{x}^2 = 1.$$

  This shows that, topologically, $SU(2) = S^3$ and thus $SO(3) = SU(2)/\{\pm1\}$. That is, $SO(3)$ is a sphere with antipodal points identified, i.e. it is isomorphic to $\mathbb{R}P^3$.

- To see these two descriptions are equivalent, it is useful to go to one lower dimension. An S^2 with antipodal points identified is the same as a hemisphere with opposite points on its boundary identified, but a hemisphere is homeomorphic to a disc D^2.

- Consider the evolution of a spin 1/2 particle in a magnetic field $\mathbf{B}(t)$. Define

$$\boldsymbol{\omega}(t) = g\frac{e}{2mc}\mathbf{B}(t).$$

  We describe the spin state of the particle with a spinor χ, and the Schrodinger equation reads

$$i\hbar\frac{\partial\chi}{\partial t} = \boldsymbol{\omega}(t)\cdot\left(\frac{\hbar}{2}\boldsymbol{\sigma}\right)\chi$$

  Here, $\boldsymbol{\sigma}$ is a vector operator containing the Pauli matrices.

- Now, let $\mathbf{S}(t)$ be the expectation value of the spin operator,

$$\mathbf{S}(t) = \langle\chi(t)|\frac{\hbar}{2}\boldsymbol{\sigma}|\chi(t)\rangle.$$

  This quantity is just a vector in $\mathbb{R}^3$. With some work, we can show that

$$\frac{d\mathbf{S}}{dt} = \boldsymbol{\omega}(t)\times\mathbf{S}.$$

  This is the "classical" equation of motion.

- For every time evolution of the spinor $\chi(t)$, we have a corresponding time evolution of the classical spin $\mathbf{S}(t)$. Mathematically, we have a correspondence from paths $\overline{\alpha}(t)$ on $SU(2)$ to paths $\alpha(t)$ on $SO(3)$, and we say $\overline{\alpha}$ is a lift of α.

- Now consider the closed loops on $SO(3)$. Since $\pi_1(SO(3)) = \mathbb{Z}_2$, there are two homotopy classes; these correspond to closed loops on $SU(2)$, and paths that connect antipodal points, respectively. That is, $SU(2)$ is big enough to keep track of homotopy in $SO(3)$. As a result, a spinor has to turn twice to return to itself.
- Mathematically, we are consider a projection map $p\colon SU(2) \to SO(3)$ defined by

$$R_{ij} = \frac{1}{2} \operatorname{tr} U^\dagger \sigma_i U \sigma_j.$$

  This is a two-to-one projection, since $p(U) = p(-U)$. We say $SU(2)$ is a double cover of $SO(3)$, and since it is simply connected, it is a universal cover.
- In general, the double cover of $SO(n)$ is called the spin group $\mathrm{Spin}(n)$. Other examples are

$$\mathrm{Spin}(4) = SU(2) \times SU(2), \quad \mathrm{Spin}(5) = Sp(2), \quad \mathrm{Spin}(6) = SU(4).$$

Intuitively, a covering space is just an "unrolling" of a space that is not simply connected into a larger space. When the unrolling is complete, we arrive at the universal cover, which is simply connected. We now use the universal cover to find a fundamental group; along the way, we will motivate the formal definition of a covering space.

Example. The circle S^1 has universal cover $\mathbb{R}$. The projection map $p\colon \mathbb{R} \to S^1$ is $p(x) = e^{ix}$. We can picture the cover as a helix sitting above S^1. The inverse function p^{-1} is multivalued, with

$$p^{-1}(0) = \{2\pi n \mid n \in \mathbb{Z}\}.$$

If we consider an open interval about $1 \in S^1$ that is small enough, it will be simply connected. The set of its preimages in $\mathbb{R}$ will each be homeomorphic to the original interval in S^1.

Consider a continuous path $\alpha\colon [0,1] \to S^1$. A lift of α is a map $\overline{\alpha}\colon [0,1] \to \mathbb{R}$ satisfying $\alpha = p \circ \overline{\alpha}$, and we claim that up to the choice of starting point, $\overline{\alpha}$ is unique. Intuitively, this is because the set of preimages of a small interval is discrete; therefore we always have 'only one possible choice' as $\overline{\alpha}$ must be continuous.

Now consider loops on S^1 with

$$\alpha(0) = \alpha(1) = 1, \quad \overline{\alpha}(0) = 0, \quad \overline{\alpha}(1) = 2\pi n.$$

We claim that the winding number n indexes the homotopy classes. It is invariant under homotopy by continuity; conversely, if two loops have the same winding number, then their lifts are homotopic, so projecting shows that the loops are homotopic. Therefore, $\pi_1(S^1) = \mathbb{Z}$.

Example. The circle S^1 can also cover itself by wrapping around n times, though these covers wouldn't work for the proof above. The cylinder is a double cover of the Mobius strip; this is easiest to see by looking at gluing diagrams. Similarly, the torus T^2 is a cover of the Klein bottle. Generally, we may use double covers to make nonorientable manifolds orientable.

We now formalize the definitions and theorems we used above.

Definition. Let M and $\overline{M}$ be connected topological spaces with a surjective map $p\colon \overline{M} \to M$. Suppose that for every $x \in M$, there is a connected open neighborhood U so that $p^{-1}(U)$ is a disjoint union of open sets $\{U_\alpha\}$ in $\overline{M}$, each mapped homeomorphically onto U by p. Then $\overline{M}$ is a cover of M, and if $\overline{M}$ is simply connected, it is the universal cover of M.

Lemma. Given a continuous path α in M with $\alpha(0) = x_0$, and a choice of point $\overline{x}_0$ in the preimage $p^{-1}(x_0)$, there is a unique continuous path $\overline{\alpha}$ in $\overline{M}$ so that $\overline{\alpha}(0) = \overline{x}_0$ and $\alpha = p \circ \overline{\alpha}$.

Theorem. Every connected space M has a universal cover $\overline{M}$.

Proof. We explicitly construct $\overline{M}$. Choose an arbitrary $x_0 \in M$, and let (x, γ) denote a path γ in M from x_0 to x. The points of $\overline{M}$ are equivalence classes $[(x, \gamma)]$ of these tuples, under the relation

$$(x, \gamma) \sim (x', \gamma') \text{ if } x = x', \quad \gamma \text{ homotopic to } \gamma'.$$

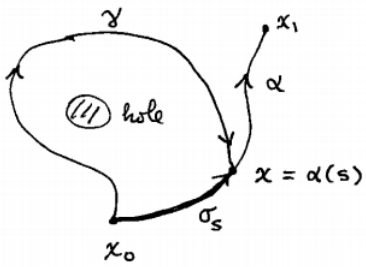
We define the projection map

$$p \colon \overline{M} \to M \colon [(x, \gamma)] \to x.$$

Now, if M is already simply connected, this construction clearly works. To see what this construction does in general, consider $x = x_0$. Then the corresponding points in $\overline{M}$ are loops based at x_0 up to homotopy, i.e. they correspond to the fundamental group $G = \pi_1(M, x_0)$. The branches of $p^{-1}(x_0)$ are labeled by $g \in G$. So far, this corresponds well with our idea of 'unraveling loops'.

Now consider a simply connected region U containing x_0. We want to show that $p^{-1}(U)$ consists of regions homeomorphic to U and indexed by G. To do this, note that for any $x \in U$, we may draw a conventional path τ_x from x_0 to x lying entirely in U. Given a curve γ from x_0 to x, we may label it by the homotopy class of the loop $\gamma \circ \tau_x^{-1}$. This gives the desired labeling of $p^{-1}(U)$. (We are ignoring some formal issues, like how to define a topology on $\overline{M}$ or to construct U.)

Now we show that $\overline{M}$ is simply connected. Consider a path α in M starting from x_0. We may continuously identify a group element with $[(x, \gamma)]$ for $x = \alpha(s)$ by the same method as before: let σ_s be the restriction of α to $[0, s]$, and let $g = [\gamma \circ \sigma_s^{-1}]$.



If the path α crosses itself, so that $\alpha(s_1) = \alpha(s_2)$, the group assignments for the two points may disagree. In particular, suppose α is a loop based at x_0. Choose a branch of p^{-1} at x_0 labeled by $g_0 \in G$, and follow it continuously along the loop. This is the lift $\overline{\alpha}(s) = [(\alpha(s), \gamma_s)]$ of $\alpha(s)$.

By construction, our assignment tells us that

$$[\gamma_s \circ \sigma_s^{-1}] = [\gamma_0]$$

In particular, setting $s = 1$, this gives

$$[\gamma_1 \circ \alpha^{-1}] = [\gamma_0]$$

which implies that $g_1 = g_0 * [\alpha]$. That is, if $[\alpha]$ is not the identity, the lift is not closed in $\overline{M}$. If the lift is closed (i.e. a loop in $\overline{M}$), then $[\alpha]$ is the identity, so the loop is contractible in M. By continuity, the lifted loop is contractible as well, so $\overline{M}$ is simply connected.

3.3 Fundamental Groups of Polyhedra

To compute fundamental groups of polyhedra, we need a little more group theory.

- Let G be a group and $x = \{a, b, c, \ldots\}$ be a finite subset of G. If every element of G may be written in terms of a finite product of elements of x, then G is finitely generated by x.
- We call such a product of generators a word. A word is reduced if all zero powers (e.g. a^0) are removed and all elements are canceled with their inverses (e.g. $a^3a^{-1} \to a^2$).
- If every element of G can be written uniquely as a reduced word, G is freely generated. If not, there are relations connecting the generators, i.e. specific words that are equal to zero.
- More formally, let F be the free group generated by $(x_1, \ldots, x_n)$, i.e. the set of reduced words of the x_i. The group operation is concatenation and subsequent reduction of reduced words.
- Suppose G is generated by $\{x_1, \ldots, x_n\}$ but not freely. We may define a map $f\colon F \to G$ that simply maps words in F to identical words in G, so $G = F/\ker f$. The members of $\ker f$ tell us about the relations in G. (More precisely, $\ker f$ is generated by elements of the form grg^{-1} where $g \in G$ and r is a relation in G.)

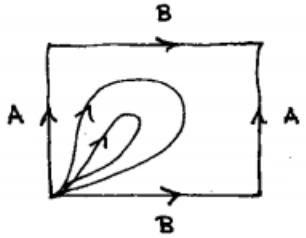
Example. Let $G = \mathbb{Z}^2$ be an Abelian group generated by $\{x, y\}$. The relation is $xyx^{-1}y^{-1} = 1$, and we can present the group as

$$G = \{x, y; xyx^{-1}y^{-1}\}.$$

As another example, we have

$$\mathbb{Z}_k = \{x; x^k\}.$$

Example. Consider the torus as a square with opposite sides identified.



We guess that the fundamental group is generated by A and B. However, we have the relation $ABA^{-1}B^{-1}$ by the contraction shown above. Therefore, $\pi_1(T^2) = \mathbb{Z}^2$. If we instead consider the Klein bottle, the relation would be $ABAB^{-1}$, so the fundamental group is not abelian.

Note. For a general connected simplicial complex K, the fundamental group can be computed as follows. We associate every oriented 1-simplex (ij) with a generator g_{ij} so that $g_{ij} = g_{ji}^{-1}$. To allow for deformations of paths through triangles, we set $g_{ij}g_{jk} = g_{ik}$ if $(ijk) \in K$. To associate homotopically trivial paths with the identity, it suffices to choose a one-dimensional subpolyhedron L of K which is contractible and contains all of the vertices of K. If L contains (ij), we set $g_{ij} = 1$, which intuitively has the effect of contracting L to a point.

This procedure clearly only depends on the 1-simplexes and 2-simplexes of K, formalizing the notion that the fundamental group only sees "one-dimensional holes". This procedure does not generalize to higher homotopy groups; in general computing them is quite difficult.

We now connect the fundamental group and the first homology group.

Definition. Let G be a group. The subgroup generated by all commutators $xyx^{-1}y^{-1}$ in G is called the commutator subgroup C.

Prop. The commutator subgroup $C \subset G$ is a normal subgroup of G, and G/C is abelian.

Proof. All generators of C are mapped to C by conjugation, because

$$gxyx^{-1}y^{-1}g^{-1} = (gxg^{-1})(gyg^{-1})(gx^{-1}g^{-1})(gy^{-1}g^{-1}) = x'y'x'^{-1}y'^{-1}.$$

Therefore C is a normal subgroup. Now, consider the cosets $[g_1]$ and $[g_2]$. We have

$$g_1g_2(g_2^{-1}g_1^{-1}g_2g_1) = g_2g_1$$

so $[g_1g_2] = [g_2g_1]$. Using coset multiplication, this gives $[g_1][g_2] = [g_2][g_1]$ so G/C is abelian.

Note. A perfect group has $G = C$. Then if the fundamental group is perfect, the first homology group is trivial. Example of perfect groups are quite rare, with the simplest being A_5, so it is sometimes claimed that a trivial first homology group implies simple connectedness. This claim is true for all examples considered in these notes.

Theorem. Let K be a simplicial complex, let $G = \pi_1(K)$, and let C be its commutator subgroup. Then $H_1(K, \mathbb{Z}) = \pi_1(K)/C$. This is a special case of the Hurewicz theorem, which relates homology and homotopy groups.

Example. The Klein bottle has

$$\pi_1(M) = \{x, y; xyxy^{-1}\}.$$

Quotienting by commutators gives the extra relation $xyx^{-1}y^{-1}$. Combining these relations, we find $x^2 = 1$, giving $H_1(M) = \mathbb{Z} \times \mathbb{Z}_2$.

Corollary. If X and Y are of the same homotopy type, their first homology groups are the same.

3.4 Higher Homotopy Groups

Higher homotopy groups are defined as homotopy classes of maps $S^n \to M$. However, this formulation is slightly inconvenient. Recall that for the fundamental group, we mapped from $I = [0, 1]$ and demanded the map was equal at the endpoints. Similarly, to study maps $S^2 \to M$ it is sufficient to consider homotopy classes of maps

$$\alpha \colon I \times I \to M$$

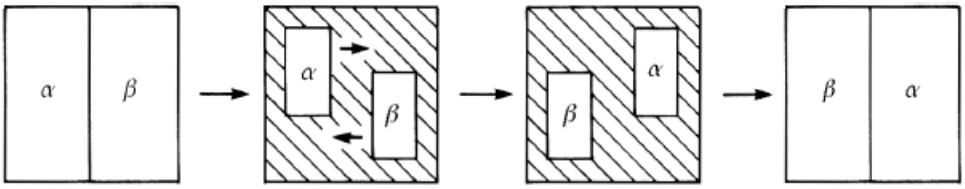
where the boundary of the square is mapped to a single point x_0. This works since $I \times I/\sim \cong S^2$. We may now easily define a group operation by giving each map half of the square,

$$(\alpha * \beta)(s_1, s_2) = \begin{cases} \alpha(2s_1, s_2) & 0 \leq 1/2 \leq s, \\ \beta(2s_1 - 1, s_2) & 1/2 \leq s_1 \leq 1. \end{cases}$$

As with the fundamental group, it is straightforward to check this is a group, which is homotopy invariant and does not depend on the base point x_0. The definitions of $\pi_n(M)$ are similar.

Prop. All higher homotopy groups $\pi_n(M)$ for $n > 1$ are abelian.

Proof. This is apparent from the following diagram.



Above, all shaded regions are filled with the base point x_0. There is 'enough room' for $n > 1$ to move any two maps past each other.

Prop. Let $\overline{M}$ be the universal cover of M. Then $\pi_n(\overline{M}) = \pi_n(M)$ for $n > 1$.

Proof. The essential difference is that the sphere S^n is simply connected for $n > 1$. Specifically, consider an n-loop in M with $n > 1$. Unlike the $n = 1$ case, this loop always lifts to a loop in $\overline{M}$. To see this, consider the equation $[\gamma_1 \circ \alpha^{-1}] = [\gamma_0]$ from the universal cover proof. Since the sphere is simply connected, $[\alpha]$ is always trivial, so $[\gamma_1] = [\gamma_0]$ and the group index g is uniquely defined on the lifted loop.

Prop. All higher homotopy groups are homotopy invariants.

Prop. Just as for the fundamental group, we have $\pi_n(X \times Y) = \pi_n(X) \times \pi_n(Y)$.

Next, we list facts about higher homotopy groups.

- We first assert that
$$\pi_n(S^n) = \mathbb{Z}.$$
This can be proven by defining a higher-dimensional analogue of the winding number, called the wrapping number. For $n = 2$, we work in spherical coordinates, mapping (θ, ϕ) to (α, β). Then the wrapping number is
$$N = \frac{1}{4\pi} \int d\alpha d\beta \, \sin\alpha = \frac{1}{4\pi} \int d\theta d\phi \, \sin\alpha \left(\frac{d\alpha}{d\theta}\frac{d\beta}{d\phi} - \frac{d\beta}{d\theta}\frac{d\alpha}{d\phi} \right)$$
where the integral is over all (θ, ϕ), and we simply used a Jacobian to change variables.

- We must then show that configurations with different N cannot be deformed into each other, while those with the same N can. The first part simply holds because N is an integer, and we omit the proof of the second.

- One particular application of this result is
$$\pi_3(SO(3)) = \pi_3(SU(2)) = \mathbb{Z}.$$
All of these results for S^n also hold for $\mathbb{R}P^n$, since its universal cover is S^n for $n > 1$.

- For $k < n$, we know $\pi_k(S^n)$ is trivial, because there is 'enough' space to contract all loops. For $k > n$, the homotopy groups are surprisingly not trivial. For example,
$$\pi_3(S^2) = \mathbb{Z}$$
and the generator of the group is the Hopf fibration.

- There exists a map J called the J-homomorphism

$$J\colon \pi_k(SO(n)) \to \pi_{k+n}(S^n)$$

which is an isomorphism for $k = 1$, giving

$$\pi_1(SO(n)) = \pi_{n+1}(S^n).$$

In particular, this tells us that $\pi_1(SO(2)) = \pi_3(S^2)$, giving an alternate proof that $\pi_3(S^2) = \mathbb{Z}$. We also have $\pi_1(SO(3)) = \pi_4(S^3) = \pi_4(SO(3))$, which shows $\pi_4(SO(3)) = \mathbb{Z}_2$.

- There are strong constraints on the homotopy groups of Lie groups. It can be shown that

$$\pi_2(G) = 0 \text{ for } G \text{ compact, connected,} \quad \pi_3(G) = \mathbb{Z} \text{ for } G \text{ compact, connected, simple.}$$

The latter result means that instantons in $SU(2)$ are representative of all instantons.

- Finally, many homotopy groups can be computed using the long exact sequence

$$\ldots \to \pi_2(Y) \to \pi_2(X) \to \pi_2(X/Y) \to \pi_1(Y) \to \pi_1(X) \to \pi_1(X/Y) \to \pi_0(Y) \to \ldots.$$

Here X/Y is simply a quotient space.

- Applying the long exact sequence to Lie groups, we have

$$\pi_1(G/H) = \pi_0(H) \text{ for } G \text{ simply connected}$$

and

$$\pi_2(G/H) = \pi_1(H) \text{ for } G \text{ compact, connected, simply connected.}$$

In the context of gauge theories, these conditions are automatically satisfied; G must always be compact, and may be taken to be connected and simply connected without loss of generality.

Example. Using higher homotopy groups, we can prove that $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^m$ if and only if $m = n$. To do this, note that $\mathbb{R}^n - \{p\}$ for any $p \in \mathbb{R}^n$ retracts onto S^{n-1}, and $\pi_{m-1}(S^{n-1})$ is trivial for $m < n$. Now if $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic with $m < n$, then so are $\mathbb{R}^n$ and $\mathbb{R}^m$ each with one point deleted, but only the latter has a nontrivial π_{m-1}, a contradiction.

Note. Homology vs. homotopy. While the n^{th} homology and homotopy groups roughly capture "n-dimensional holes", they provide rather different information.

- The fundamental group is larger than the first homology group, because a loop can get 'stuck' in ways that a 1-cycle can't. There are even spaces with trivial first homology group but nontrivial fundamental group!

- For the 2-torus, we have $H_1(\mathbb{T}^2) = \pi_1(\mathbb{T}^2) = \mathbb{Z}^2$, because the torus has "two one-dimensional holes". The second homology group is $\mathbb{Z}$, since the torus is 'hollow', but the second homotopy group is trivial, as a sphere can't wrap around a torus.

- The homology groups of the spheres are simple: $H^0(S^n) = H^n(S^n) = \mathbb{Z}$, and all others are trivial. But the homotopy groups of spheres $\pi_m(S^n)$ for $m > n$ are extremely complex, because higher dimensional spheres can wrap around lower-dimensional ones, one example being the Hopf fibration.

3.5 Topological Defects

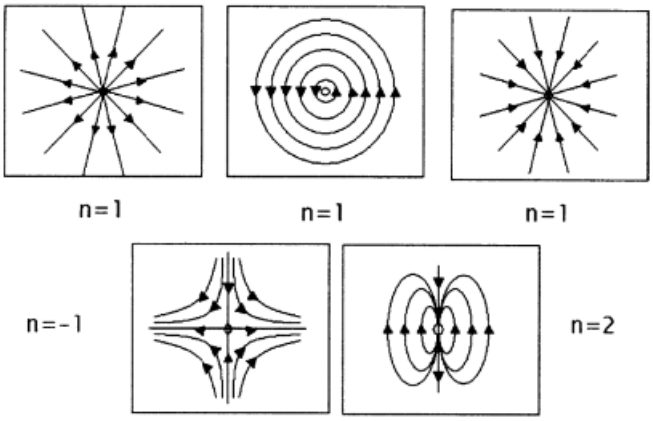
Our background above now allows us to classify topological defects in condensed matter systems. For further physical context, see the notes on Condensed Matter.

Example. A planar magnet in the XY model. In the continuum limit, the configuration of the magnet is described by a spin vector $\mathbf{S}(\mathbf{x})$ in the plane with unit magnitude. The range of the spin field, which is called the order parameter space, is S^1. (Note that since the punctured plane is homotopic to S^1, our homotopy results would also hold if we only demanded the spin be nonvanishing.)

Consider a closed loop in the plane. Then mapping the distance along the loop to the spin vector at that point gives a map $f\colon S^1 \to S^1$. Since the fundamental group of S^1 is $\mathbb{Z}$, we may assign this map an integer, called the winding number; physically, this is the number of times the spin vector rotates around the path.

A nonzero winding number implies a singularity in the spin field. To see this, suppose the field were continuous everywhere. Then continuously deforming the loop to a point yields a continuous deformation of f to the identity, a contradiction. We conclude the field has a 'point defect' inside the loop. Physically, this defect is a localized excitation in the magnet.

The classification of point defects is somewhat subtle. It is tempting to conclude that positive and negative defects have circulation in opposite directions, but both clockwise and counterclockwise circulation are homotopic (the fields are related by a 180° rotation). In fact, sources, sinks, and vortices all have the same winding number (which we call +1), and winding number −1 looks qualitatively different, as shown below.



It is also tempting to conclude that a source and a sink annihilate, because as we move them on top of each other, the singularities they produce should smoothly 'cancel out', leaving a field configuration which must be topologically trivial. However, the dipole field has winding number +2, not 0. The reason is that there's a discontinuous switch in the direction of the field inside the dipole, so the limit is singular.

Note. The existence of an order parameter space at all is linked to symmetry breaking. The symmetry breaking process yields a set of field values M with the same (free) energy, i.e. soft modes. Then we can have low-energy nontrivial field configurations if their values are in M.

In three dimensional space, the fundamental group detects line defects going through the loop; the second homotopy group π_2 detects point defects.

Example. Superfluid He-4 in 3D. The superfluid is described by a complex-valued field $\psi(\mathbf{x}) = Ae^{i\varphi}$, corresponding to the classical expectation value of the superfluid quantum field. In general, A is nonzero throughout the superfluid, so the order parameter space is homotopic to S^1. However, the fundamental group now detects line defects instead of point defects. We can show

$$\mathbf{v} = \frac{\hbar}{m}\nabla\varphi$$

which implies that the fluid circulates around these lines; we thus call them vortices.

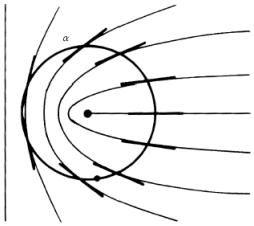
Example. The Heisenberg model. These are the 3D version of the XY model, i.e. magnets where the magnetization is a unit vector in $\mathbb{R}^3$, so the order parameter space is S^2. There are no line defects, since $\pi_1(S^2)$ is trivial, but there are point defects, as $\pi_2(S^2) = \mathbb{Z}$. However, unlike the case of the XY model, sources and sinks are now topologically distinct, with opposite charge. These are also called "hedgehog" defects.

Example. A cubic crystal lattice. Deformations of the lattice can be parametrized by the vector $\mathbf{u}$ from an atom's position to its corresponding unperturbed position. However, $\mathbf{u}$ is equivalent up to the addition of a lattice vector, so $M = T^3$. We have $\pi_1(M) = \mathbb{Z}^3$ and $\pi_2(M)$ is trivial. The homotopy classes $\pi_1(M)$ correspond to the number of "missing" lattice planes in each direction.

There is an important subtlety we have suppressed. We're considering maps without base points, so we really are indexing *free* homotopy classes. For line defects, these correspond to the conjugacy classes of the fundamental group. This makes no difference for an abelian fundamental group, but more generally, this set is not even a group at all, as the product of two conjugacy classes is not well-defined. Physically, the homotopy class of two line defects together cannot be determined from their charges alone, but instead depends on the global structure of the field. It also turns out that for higher homotopy groups, the free homotopy classes are given by $\pi_n(M)/\pi_1(M)$, by an action of $\pi_1(M)$ on $\pi_n(M)$ that we do not specify here. Note that none of these subtleties have applied to any of the examples we've considered so far.

Example. Nematic liquid crystals. These crystals contain long molecules which behave like rigid rods and try to align with their neighbors. We may specify the orientation of a molecule by a vector $\mathbf{v}$, but since the molecules are symmetric, $\mathbf{v} \sim -\mathbf{v}$. Therefore the order parameter space is $\mathbb{R}P^2$, the set of directors.

Nematic liquids support line defects, but unlike in the XY model, these line defect annihilate each other because $\pi_1(\mathbb{R}P^2) = \mathbb{Z}_2$. To see this, consider the below line defect, shown in the plane.

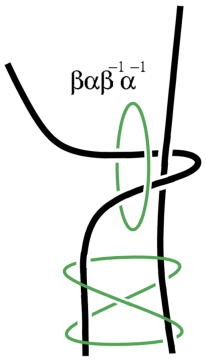


Placing two of these defects side-by-side gives a dipole field, as shown in the planar magnet example. However, this field is *not* singular, because the discontinuous direction switch of the vectors in the middle of the dipole corresponds to no change at all for the directors. Then merging the two defects produces a field with no singularities, which must be topologically trivial.

Nematic liquid crystals also have point defects, since $\pi_2(\mathbb{R}P^2) = \mathbb{Z}$. However, the sources and sinks of the Heisenberg model are now identical. In fact, the action of $\pi_1(\mathbb{R}P^2)$ on $\pi_2(\mathbb{R}P^2)$ is to flip

the charge, so the point defects are classified by *nonnegative* integers, and there is no natural group multiplication law. For example, the combination of two defects of charge 1 could have charge 0 or 2, depending on the global structure of the field. For further discussion, see *Disclination Loops, Hedgehogs, and All That*. More exotic examples are given in the notes on Quantum Field Theory.

Note. Can two line defects be pulled past each other? Ignoring the subtlety mentioned above, suppose two line defects have homotopy classes α and β. If this is allowed topologically, then the upper green loop shown must be topologically trivial.



This can be deformed into the lower green loop, which takes the form of the commutator $\beta\alpha\beta^{-1}\alpha^{-1}$. So if α and β do not commute, there must be an energetic obstruction to passing the line defects through each other. (This is not a particularly powerful result, because few systems have noncommutative fundamental groups, and those that do are subject to the subtlety mentioned above, making this analysis oversimplified. Also, note that the converse is not true; line defects that do commute often do repel each other, for nontopological reasons. Much cannot be inferred from topology alone.)

We can also identify defects using π_0. The "zeroth homotopy group" $\pi_0(M)$ does not have a group structure; it is simply a set. It consists of homotopy classes of maps $S^0 \to M$, where $S^0 = \{-1, 1\}$. Taking -1 to be the base point, such a map is characterized by a single point, and homotopy allows us to move that point continuously. Thus $|\pi_0(M)|$ is the number of connected components of M, and physically π_0 detects domain walls.

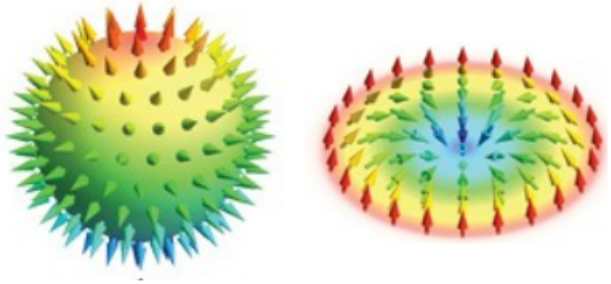
Example. The Ising model. In the Ising model of magnetism, the spin of a particle can only be $+1$ or -1. Then the order parameter space is S^0, and the only topological defects are domain walls. These are the boundaries between regions with upward spin and downward spin.

In 3D, the third homotopy group $\pi^3(M)$ classifies textures, which are topological properties of entire field configurations. (In particle physics, they are called skyrmions, as they appeared in an old model for the nucleon by Skyrme.) Specifically, suppose we are studying a singularity-free field configuration which approaches a common limiting value at infinity, e.g. a locally perturbed magnet in a strong external field. Then we may compactify $\mathbb{R}^3$ to S^3 by adding a point at infinity. The field configuration as a whole is then a map $f\colon S^3 \to M$ which may be classified with $\pi^3(M)$.

Example. Consider a linear magnet with order parameter space S^1. Since $\pi^1(S^1) = \mathbb{Z}$, textures exist. One example is given by $\theta(x) = \pi \tanh(x)$. It is clear that this texture cannot be unwound, even though it contains no singularity.

Example. Textures in a planar magnet. We return to the plane, where textures are detected with π^2, but allow the magnetization to be a unit vector in three dimensions. The order parameter space

is S^2, and $\pi^2(S^2) = \mathbb{Z}$, so textures exist. The resulting nontrivial field configurations are shown below, both on $\mathbb{R}^2$ and its stereographic projection onto S^2.



Example. Superfluid He-3 has order parameter space $\mathbb{R}P^3$, and

$$\pi_3(\mathbb{R}P^3) = \pi_3(S^3) = \mathbb{Z}.$$

The resulting textures are called Shankar's monopoles.

Example. In high energy physics, it is also useful to classify the entire *spacetime* profile of a field. This is important because each topologically distinct sector will contain at least one local minimum of the action, so this procedure classifies instantons. Heuristically, instantons are classified by $\pi^4(M)$, but this is a bit of a simplification because the relevant field is a gauge field; as we'll see below and in the notes on Quantum Field Theory, they're instead classified by the topologically distinct G-bundles over S^4, which are in turn classified by $\pi^3(G)$.

Note. There are some physical caveats when considering topological defects. The first is that the smooth fields in our examples above do not exist in reality; they are extrapolated from a discrete lattice. On scales on the order of the lattice spacing (or more precisely, the coherence length), topological defects can simply fall apart.

The second subtlety is that topological stability does not guarantee energetic stability. For example, Derrick's theorem forbids the existence of stable textures in dimension $n > 1$. In the case of the previous example, scaling the texture down by a factor of λ increases the energy density by λ^2, but decreases the volume by λ^3. Thus the texture shrinks further and further down until it hits the coherence length and vanishes. More subtle models are required to circumvent Derrick's theorem.

Yet another subtlety comes in calculating the energy of a topological defect. When writing down free energies, we usually throw away total derivative terms, but this is not valid if topological defects are present. In high energy physics, this is the precise reason why the QCD θ-term can have a physical effect.

4 Manifolds

4.1 Smooth Manifolds

We define manifolds and related quantities informally.

- A topological manifold is a second countable Hausdorff topological space that is locally homeomorphic to $\mathbb{R}^n$. The number n is the dimension of the manifold. This is sufficient to talk about the continuity of functions on the manifold, but not their derivatives, as we don't have coordinates.
- A differentiable manifold M is a topological manifold with coordinate systems. More specifically, a chart on M is a pair (U, ϕ) where $U \subset M$ is open and
$$\phi \colon U \to V \subset \mathbb{R}^n$$
is a homeomorphism. Given a point $p \in U$, this chart gives it coordinates $\phi(p) = (x^1, \dots, x^n)$.
- Now, given a function $f \colon M \to \mathbb{R}$, we can define its smoothness by that of $f \circ \phi^{-1} \colon V \to \mathbb{R}$ and use this coordinate representation to differentiate f. From this point on we will assume all such maps are smooth (i.e. C^∞) for convenience. (In particular, the definition we are giving for a differentiable manifold is technically the definition of a smooth manifold.)
- In general, M cannot be covered by a single chart, so we need a set of charts $\{(U_i, \phi_i)\}$ called an atlas. We require that the U_i cover M, and that the maps ϕ_i are compatible, meaning that the transition functions
$$\psi_{ij} = \phi_j \circ \phi_i^{-1}$$
are smooth everywhere they are defined. Note that ψ_{ij} has a smooth inverse, namely ψ_{ji}.
- Technically, different sets of atlases can yield different manifolds. Define a differentiable structure on M to be an equivalence class of atlases that agree on which functions are smooth. Then differentiable manifolds are actually in correspondence with differentiable structures.
- In physical applications, there will be an obvious correct differentiable structure. Other structures, such as exotic $\mathbb{R}^4$ or exotic spheres, play no role in physics. In fact, in the great majority of physical calculations, one doesn't even need to work with more than one chart/coordinate system. Multiple charts play a greater role when talking about topological effects.

Example. The sphere S^2. When working informally, we often use coordinates that are singular, such as spherical coordinates (θ, ϕ). In this case, ϕ changes discontinuously from 2π to 0 at the meridian, and is not defined at the poles. This does not indicate a singularity in the manifold itself; it simply means that we need more charts.

We may cover S^2 with two charts using stereographic projection. The standard stereographic projection will map everything except a neighborhood of the North pole onto a finite subset of $\mathbb{R}^2$. Projection the opposite way gives everything but a neighborhood of the South pole.

Example. A manifold with boundary is the same as a differentiable manifold, except that open sets are mapped homeomorphically to $\mathbb{R}^n_{\geq 0}$. The points mapped to the boundary of that space are called the boundary of the manifold.

The presence of a boundary makes results slightly more complicated, as there are now 'special' points. We have to be careful with defining smoothness, since the boundary points in our coordinate space don't have open neighborhoods.

From this point on, 'manifold' implicitly means 'smooth manifold without boundary' and all maps are assumed to be smooth.

Example. A surface. Consider k smooth functions $f_1, \ldots, f_k \colon \mathbb{R}^n \to \mathbb{R}$ and let

$$M = \{\mathbf{x} \in \mathbb{R}^n | f_1(\mathbf{x}) = \ldots = f_k = 0\}.$$

One can show that if the rank of the $k \times N$ matrix $\partial f_i / \partial x^\mu$ is maximal (i.e. equal to k at all points), then M is a smooth manifold of dimension $N - k$. For example, the sphere is a manifold with $f(\mathbf{r}) = 1 - r$. It is called a hypersurface since it has codimension one. More generally, Whitney's embedding theorem states that any n-dimensional manifold may be realized as a surface in $\mathbb{R}^m$ with $m \leq 2n$.

Example. Real projective space $M = \mathbb{RP}^n$ is defined as

$$\mathbb{RP}^n = (\mathbb{R}^{n+1}/\{0\})/\sim, \quad \mathbf{x} \sim a\mathbf{x} \text{ for } a \in \mathbb{R}^*$$

where $\mathbb{R}^* = \mathbb{R}/\{0\}$. This is the space of all lines in $\mathbb{R}^{n+1}$. We will explicitly show that M is a manifold. Consider the $n+1$ open sets

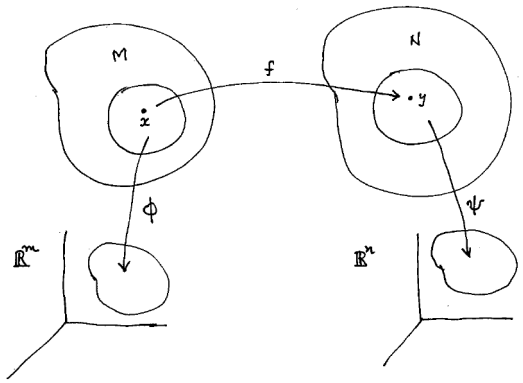
$$U_\alpha = \{\mathbf{x} \in \mathbb{R}^{n+1} | X^\alpha \neq 0\}$$

and define the charts

$$\phi_\alpha \colon U_\alpha \to V_\alpha \in \mathbb{R}^n, \quad \phi_\alpha(\mathbf{x}) = (x^1/x^\alpha, \ldots, x^{\alpha-1}/x^\alpha, x^{\alpha+1}/x^\alpha, \ldots, x^{n+1}/x^\alpha).$$

Each chart is smooth, so it suffices to show that the transition functions are smooth. As an example, $\mathbb{RP}^3 = SO(3)$. To see this, note that $\mathbb{RP}^3$ is equal to S^3 with opposite points identified, $SO(3)$ is equal to $SU(2)/\{I, -I\}$, and $SU(2)$ is equal to S^3.

4.2 The Tangent Space



Now consider maps $f \colon M \to N$ where M and N are manifolds with dimension m and n.

- We say f is smooth at a point if the corresponding coordinate representation $\phi \circ f \circ \psi^{-1}$ is smooth at the corresponding point, where the charts ϕ and ψ are shown above.
- We may also compute the derivative of f, which is now an $m \times n$ matrix

$$D^i{}_j = \frac{\partial y^i}{\partial x^j}.$$

- As special cases, we might have $N = \mathbb{R}$, in which case $f\colon M \to \mathbb{R}$ is a scalar field. We let $\mathcal{F}(M)$ stand for the set of scalar fields on M. Note that it is tempting to say the coordinate functions $x^i\colon U \to \mathbb{R}$ are scalar fields, but they are manifestly not coordinate independent.
- If a map $f\colon M \to N$ is bijective and has a smooth inverse, then f is a diffeomorphism. Diffeomorphisms are isomorphisms for manifolds. In particular, if M and N are isomorphic, then we must have $\dim M = \dim N$, though this is somewhat hard to show rigorously.
- We may also define maps $c\colon I \to M$, in which case c is a parametrized curve. These will be useful for defining the tangent space.

Now, we turn to defining vectors on manifolds.

- Intuitively, a vector in $\mathbb{R}^n$ is a displacement from one point to another, and vectors may be added and multiplied by scalars. On a manifold, one can talk about displacements from one point to another, but it is unclear how to add them or multiply them by scalars.
- However, if the displacements are "small", we can approximate them as taking place in a plane, the 'tangent plane' to M at the base point, and perform vector operations inside it.
- This intuitive idea is nice, but unsatisfactory: there is no real way to define 'smallness', and we want an intrinsic definition of 'tangent plane', i.e. one that only involves M itself and not any embedding space.
- Geometrically, we are trying to take smaller and smaller pieces of a curve going through a point p. But there *is* a quantity associated with this motion that is defined solely at the point p, namely the velocity at p! Heuristically, we will define the tangent space at p as the set of possible velocities of curves through p.

Given the above motivation, we can lay down formal definitions.

- Given a parametrized curve $c\colon [a, b] \to M$ which passes through p at $t = 0$, we may define the differential operator
$$Xf = \lim_{\Delta t \to 0} \frac{f(c(\Delta t)) - f(p)}{\Delta t} = \left.\frac{d(f \circ c)}{dt}\right|_{t=0}.$$
This is a linear operator $X\colon \mathcal{F}(M) \to \mathbb{R}$.
- The set of curves c is enormously redundant, since the operator X only depends on behavior at $t = 0$, so each X is associated with an equivalence class of curves. We call X a tangent vector, and the set of all X based at p the tangent space T_pM.
- Thinking of vectors more abstractly as linear operators $X\colon \mathcal{F}(M) \to \mathbb{R}$ shows the tangent space is a vector space. Formally, T_p is the set of first order differential operators at p, i.e. its elements satisfy the Leibniz rule $X(fg) = X(f)g + fX(g)$ where everything is evaluated at p.
- Given coordinates $\{x^i\}$, the chain rule yields
$$\frac{df}{dt} = \sum_i \frac{\partial f}{\partial x^i}\frac{dx^i}{dt}$$

where df/dt on the left stands for $d(f \circ c)/dt$. Now, we can break a tangent vector X into components by its action on the coordinate functions. Define the component X^i by

$$X^i = Xx^i = \left.\frac{dx^i}{dt}\right|_{t=0}.$$

Then we may write the chain rule as

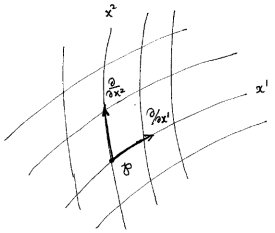
$$Xf = \sum_i X^i \frac{\partial f}{\partial x^i}.$$

Since f is arbitrary, we have

$$X = X^i \frac{\partial}{\partial x^i}$$

where we are using the summation convention, and the derivatives are evaluated at p.

- The basis vectors $\partial/\partial x^i$ are associated with curves along the coordinate axes, as shown below.



The existence of this basis shows that the dimension of the tangent space T_pM is the same as the dimension of M.

- Finally, we can see how components of vectors transform:

$$X'^i = Xx'^i = \frac{dx'^i}{dt} = \frac{\partial x'^i}{\partial x^j}\frac{dx^j}{dt} = \frac{\partial x'^i}{\partial x^j}X^j.$$

This is the standard transformation law for contravariant vector components.

Note. We check that the definition of T_p as derivations at p only depends on local information: let $f = g$ in a neighborhood of p. Then

$$0 = X(\phi(f-g)) = (X\phi)(f(p)-g(p)) + \phi(p)X(f-g) = X(f) - X(g)$$

where ϕ is a bump function which is one inside the region where $f = g$ and zero elsewhere. This step uses the existence of bump functions (constructed using e^{-1/x^2}, etc.), which don't exist for complex manifolds by analyticity; thus this is the beginning of the divergence between real manifold and complex manifold theory.

Another way to enforce locality is to define T_p to act on germs of smooth functions at p. A third way is to define T_p^* as the set of germs of smooth functions at p mod constant functions and T_p to be its dual. We could also go back to our earlier picture and define T_p in terms of equivalence classes of germs of curves through p. The point is that there are many equivalent ways to define T_p, and we just favor the one that's easier to calculate in.

4.3 The Cotangent Space

Every real vector space V is associated with a dual space V^* consisting of real-valued linear maps on V. The dual space of the tangent space T_pM is the cotangent space T_p^*M, and elements of the cotangent space are called covectors, cotangent vectors, or one-forms.

- Recall that to get a rate of change, we need to combine a curve $c\colon \mathbb{R} \to M$ with a scalar function $f\colon M \to \mathbb{R}$. We then defined tangent vectors as equivalence classes of the curves. Analogously, cotangent vectors are equivalence classes of functions.

- More specifically, the covector associated with a function f at p is

$$df|_p\colon T_pM \to \mathbb{R}\colon X \to Xf = X^i \left.\frac{\partial f}{\partial x_i}\right|_p.$$

  We call df the differential of f. It represents the 'slicing of space' associated with the level sets of f. Note that conceptually, there is nothing small or 'infinitesimal' about df.

- We may break a covector α into components by its action on the basis vectors,

$$\alpha_i = \alpha(\partial/\partial x^i).$$

  This gives the general action of covectors on vectors,

$$\alpha(X) = X^i \alpha(\partial/\partial x^i) = X^i \alpha_i.$$

- The basis vectors corresponding to these components are the differentials of the coordinate functions, because

$$(dx^i)(X) = Xx^i = X^i, \quad \alpha(X) = \alpha_i (dx^i)(X)$$

  Then we may write $\alpha = \alpha_i dx^i$. This basis is the dual basis of the tangent vector basis, as

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

- Finally, we note that the components of df are given by the standard chain rule,

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

  and that covector components transform covariantly,

$$\alpha'_i = \frac{\partial x^j}{\partial x'^i} \alpha_j.$$

  The d above can also be thought of as the exterior derivative.

Example. Examples of one-forms include row vectors and bras. The infinite-dimensional case is more subtle, as V^* can be larger than V. For example, the Dirac delta can be regarded as one-form on a space of functions by $\langle \delta(x), f(x) \rangle = f(0)$. However, there is no such thing as a Dirac delta 'function'. As another example, if V is the vector space of sequences with finitely many nonzero elements, V^* can contain covectors which are infinite linear combinations of basis covectors.

We now define vector and covector fields.

- Intuitively, a (co)vector field is simply a (co)vector at every point in the manifold. A vector field X will be written with the same letter as a single tangent vector, so the meaning must be inferred from context.
- The components $X^i(x)$ become functions which depend on the coordinate chart used, and we say X is smooth if the component functions $X^i(x)$ are. (This isn't a coordinate-dependent statement, as transition functions are smooth.) From this point forward we assume all vector fields are smooth.
- More formally, define the tangent bundle TM as the union of all tangent spaces

  $$TM = \bigcup_{p \in M} T_p M$$

  and define the cotangent bundle T^*M similarly. The projection map $\pi \colon TM \to M$ takes a tangent vector and returns its base point. The tangent bundle is locally trivial; its restriction to a chart decomposes as a product. Note that the topology on TM is inherited from the manifold, i.e. it can be read off from the charts.
- Now, a vector field is a map $X \colon M \to TM$ satisfying $\pi(X(p)) = p$. We denote the set of all vector fields by $\mathfrak{X}(M)$ and the set of covector fields by $\mathfrak{X}^*(M)$.
- Similarly, we may define (r, s) tensor fields, which are functions from M to $(T^*M)^r(TM)^s$. We define the tensor components in the usual way,

  $$T^{i_1 \dots i_r}{}_{j_1 \dots j_s} = T(dx^{i_1}, \dots, dx^{i_r}, \partial/\partial x^{j_1}, \dots, \partial/\partial x^{j_s}).$$

  The components transform in the natural generalization of the previous transformations.
- An equivalent definition is that a vector field is a linear operator on $\mathcal{F}(M)$ satisfying the Leibniz rule

  $$X(fg) = X(f)g + fX(g)$$

  which specifies that it's a first derivative.
- Similarly, a covector field is a linear operator $\mathfrak{X}(M) \to \mathcal{F}(M)$. The apparent asymmetry here is just a consequence of the order we defined things; we could just as easily have defined covectors first, and then vectors as their dual vectors. A deeper link is that vectors are associated with differentiation (i.e. the velocity of a curve) while covectors (and more generally differential forms) are associated with integration.

Note. The mathematical definition of the tensor product.

- Tensor products are appropriate for studying bilinear maps $U \times V \to W$, where $W = \mathbb{R}$ in physical applications. These are distinct from linear maps $U \times V \to W$. For example, for a bilinear map we must map $(\alpha u, v)$ to $\alpha(u, v)$.
- The tensor product $U \otimes V$ is uniquely characterized by the existence of a bilinear map

  $$\pi \colon U \times V \to U \otimes V$$

  so that for any bilinear map $\alpha \colon U \times V \to W$, there is a unique linear map $\hat{\alpha} \colon U \otimes V \to W$ with $\hat{\alpha} \circ \pi = \alpha$.

- To explicitly construct $U \otimes V$, take $F(U \times V)$ to be the free vector space on elements of $U \times V$, then quotient out by the subspace generated by elements of the form

  $$(u_1+u_2, v)-(u_1, v)-(u_2, v), \quad (u, v_1+v_2)-(u, v_1)-(u, v_2), \quad (\alpha u, v)-\alpha(u, v), \quad (u, \alpha v)-\alpha(u, v)$$

  to manually impose bilinearity.

- Another concrete definition is to let

  $$(U \otimes V)^* = \text{Bilinear}(U \times V, \mathbb{R})$$

  where we think of W as $\mathbb{R}$.

- A final concrete definition is to define $U \otimes V$ to have basis vectors $u_i \otimes v_j$ (where $\otimes$ is just an abstract symbol) for bases $\{u_i\}$ and $\{v_j\}$. This is the usual definition in physics texts; explicitly $\pi(u, v) = u \otimes v$.

- Given the abstract definition we can show that the tensor product is commutative and associative, and $\text{Hom}(U, V) \cong U^* \otimes V$.

4.4 Pushforward and Pullback

Suppose we have a smooth map $f\colon M \to N$ with $f(p) = q$. The map is not necessarily injective or surjective, and M and N need not have the same dimension.

- Given a function $\phi\colon N \to \mathbb{R}$, we can 'pull it back' to a function $M \to \mathbb{R}$ by

  $$f^*\phi = \phi \circ f.$$

  We call f^* the pullback map.

- We say that f is smooth if ϕ being smooth implies that $f^*\phi$ is smooth. This is our general definition for a smooth map between manifolds, and subsumes earlier definitions of smoothness as special cases. From here on, we always assume all maps are smooth.

- We can take a vector $X \in T_pM$ and associate it with a vector $Y \in T_pN$. Heuristically, viewing tangent spaces as small pieces of a manifold, we just apply f to the tangent space T_pM. More rigorously, we can map curves with f, defining

  $$f_*\colon T_pM \to T_qN, \quad [c] \mapsto [f \circ c].$$

  This is called the tangent map, the differential of f, or the pushforward map.

- To write f_* in components, note that $\partial/\partial x_j$ maps to $(\partial y_i/\partial x_j)(\partial/\partial y_i)$, so vector components transform as

  $$Y^i = \frac{\partial y^i}{\partial x^j} X^j.$$

  Thinking in terms of parametrized curves, this is really just the chain rule: we just multiply by the Jacobian matrix of f. We can also think of this as a generalization of change of coordinates, which is the special case where $M = N$ and f is a diffeomorphism.

- Another equivalent definition of f_* is

$$(f_*X)(\phi) = X(f^*\phi).$$

  This makes it clear that vectors are pushed forward because they act on functions, and functions are pulled back.

- Similar logic applies to covectors. Suppose we want to associate $\alpha \in T_q^*N$ with $\beta \in T_p^*M$. Then we define

$$f^*\colon T_q^*N \to T_p^*M, \quad (f^*\alpha)(X) = \alpha(f_*X).$$

  The covector $\beta = f^*\alpha$ is called the pullback of α, and f_* and f^* are each others' dual maps.

- Plugging in components, we have

$$\beta_i = \frac{\partial y^j}{\partial x^i}\alpha_j.$$

  Like the previous formula, this result is the only way to 'line up the indices'.

- Note that we may also pullback covector fields. To pullback the value of a covector field at a point, we simply pushforward its vector argument,

$$(f^*\alpha)_p(X) = \alpha_q(f_*X).$$

  The definition clearly generalizes to the pullback of $(0,r)$ tensors, and $(0,r)$ tensor fields.

- Using the definition of the pullback of a covector, we can similarly define the pushforward of an $(s,0)$ tensor. However, we cannot pushforward $(s,0)$ tensor fields (e.g. vector fields) because f may not be bijective. This is a key asymmetry between pushforward and pullback. If f^{-1} exists, we may define the pushforward and pullback of arbitrary tensor fields.

- In math, we say that things which can be pushed forward are covariant and things which can be pulled back are contravariant (where 'co' and 'contra' are with respect to the function f). This is completely unrelated to the physical definition, which says that vectors are invariant, basis vectors are covariant, and vector components are contravariant.

- Formally, the tangent map is a functor from pointed manifolds to real vector spaces; given a map $f\colon (M,p) \to (N,q)$, the functor gives the map $f_*\colon T_pM \to T_qN$.

Note. More terminology for maps.

- Given a map $f\colon M \to N$, compose with charts to get the map

$$\phi \circ f \circ \psi^{-1}\colon \mathbb{R}^m \to \mathbb{R}^n, \quad \psi\colon M \to \mathbb{R}^m, \quad \phi\colon N \to \mathbb{R}^n.$$

  We can define a Jacobian matrix on M by

$$Df|_p = J(\phi \circ f \circ \psi^{-1})|_{\psi(p)}$$

  where Df doesn't depend on the charts, essentially by the chain rule.

- We say f is an immersion at p if Df is injective; intuitively, f looks like inclusion of M in N.

- We say f is an embedding if M is diffeomorphic to its image. This is stronger than immersion. For example, a curve embedded in $\mathbb{R}^2$ cannot cross itself, but a curve immersed in $\mathbb{R}^2$ can.
- We say f is a submersion at p if Df is surjective. Locally, this looks like projection of M down to N.
- We say $q \in N$ is a regular value of f if $Df|_p$ is surjective for all points $p \subset f^{-1}(q)$. The preimage theorem states if q is a regular value, then $f^{-1}(q)$ is either empty or an $m-n$ dimensional submanifold of M. This shows surfaces defined as constrained subsets of $\mathbb{R}^n$ and matrix Lie groups defined as subsets of $\mathrm{Mat}_n(\mathbb{R})$ are indeed manifolds; all we have to check is regularity.
- Sard's theorem states that if f is smooth, the nonregular values of f have measure zero. That is, regular values are 'generic', so the preimage theorem only fails at a few points.

4.5 Vector Fields and Flows

We now turn to the geometric picture of a vector field.

- Since a vector is the velocity of a parametrized curve, a vector field X assigns a velocity to every point on the manifold M. A parametrized curve γ whose tangent vector at time t is $X(\gamma(t))$ is called an integral curve of X.
- In terms of the components $x^i(t)$ of γ, we have

$$\frac{dx^i}{dt} = X^i(x).$$

- The flow equation above is the generalization of an n-dimensional system of first-order ODEs from $\mathbb{R}^n$ to a manifold. The solutions obey similar existence-uniqueness theorems, which hold here due to our implicit assumption that X is smooth.
- One technical point is that existence of a solution is only guaranteed in a neighborhood about our initial point. For example, the ODE

$$\dot{x} = x^2, \quad x_0 = 1$$

  on the real line blows up in finite time. For simplicity, we will assume this does not happen; one can show it never happens on a compact manifold, a complex manifold, or a Lie group.
- The integral curves define a map from M to itself, by following the curves for a fixed time t. More specifically, we have $\Phi\colon \mathbb{R} \times M \to M$ satisfying the properties

$$\Phi^i(0, x_0) = x_0^i, \quad \frac{\partial \Phi^i}{\partial t}(t, x_0) = X^i(\Phi(t, x_0)).$$

- Now define the advance map $\Phi_t = \Phi(t, \cdot)$. This clearly satisfies the composition property

$$\Phi_s \Phi_t = \Phi_{s+t}$$

  which may be formally proven using the EUT. Therefore, we have a group structure on the Φ_t maps, where the identity element is Φ_0 and the inverse of Φ_t is Φ_{-t}. Since inverses exist, Φ_t is a diffeomorphism, and the set of Φ_t constitute a one-parameter group of diffeomorphisms. This set is called the flow generated by X. It is an example of an action of $\mathbb{R}$ on M.

- On an analytic manifold, where Taylor series converge, we have

$$\Phi_t = e^{tX} \equiv 1 + tX + \frac{t^2}{2!}X^2 + \dots.$$

The terms on the right make sense: X maps $\mathcal{F}(M) \to \mathcal{F}(M)$, X^n does the same, and hence so does e^{tX}. Acting on a function f,

$$e^{tX} f = f + tX^i \frac{\partial}{\partial x^i} f + \frac{t^2}{2} X^i \frac{\partial}{\partial x^i} X^j \frac{\partial}{\partial x^j} f + \dots = f + t\frac{df}{dt} + \frac{t^2}{2}\frac{d^2 f}{dt^2} + \dots$$

where the derivatives of f are stand for the derivative of $f \circ c$ where c is an integral curve passing through x at $t = 0$. Since the manifold is analytic,

$$(e^{tX} f)(x_0) = f(\Phi_t(x_0)) = (\Phi_t^* f)(x_0).$$

Since x_0 and f are arbitrary, we thus have

$$\Phi_t^* = e^{tX}.$$

That is, the pullback map on functions corresponding to Φ_t is equal to e^{tX}. If we do not include the pullback, the equation is not strictly true. (This distinction isn't necessary if we always work in coordinates, as in that case we never mention Φ_t at all. We are always really talking about Φ_t^*, which acts on the coordinate functions.)

- The exponential notation also has other advantages. It makes the group structure of Φ_t apparent, and it also behaves correctly under differentiation, giving

$$\frac{d}{dt}\Phi_t^* = Xe^{tX}.$$

Combining this with our previous result, we have

$$\frac{d}{dt}\Phi_t^* = X\Phi_t^*.$$

This is a coordinate-free form of the flow equations above.

Note. A set of coordinates defines a family of vector fields, $\{\partial/\partial x^i\}$, and a family of covector fields, $\{dx^i\}$. However, the converse is not necessarily true.

- By equality of mixed partials, the commutator of coordinate vector fields $[\partial/\partial x^i, \partial/\partial x^j]$ is always zero. This isn't true for generic sets of vector fields.
- If the commutators are zero, then we can construct a coordinate system from the integral curves of the vector fields, as long as the vector fields are nonzero and everywhere independent, and the manifold has trivial topology.
- The exterior derivative of all coordinate-based covector fields dx^i is zero, as $d^2 = 0$. However, this isn't true for a generic covector field α.
- If we have $d\alpha = 0$, then we can find a coordinate function f so that $df = \alpha$ as long as the manifold has trivial topology. The first cohomology group tells us when this doesn't hold.

Note. More on the commutator. The commutator of two vector fields is

$$[V, W] = (V^i \partial_i W_j - W^i \partial_i V_j)\partial_j.$$

Note that the result is also a vector field; the second-derivative term drops out by equality of mixed partials. To interpret the commutator, note that

$$[e^{\epsilon V}, e^{\epsilon W}] = \epsilon^2 [V, W] + O(\epsilon^3).$$

Then the commutator tells us the difference between flowing along V and then W, or vice versa.

Note. Constructing coordinates from commuting vector fields. In 2D, consider the vector fields V and W, and use initial coordinates x^i. We define a candidate new coordinate system (α, β) by

$$x^i(\alpha, \beta) = e^{\beta W} e^{\alpha V} x^i|_P$$

where P is some arbitrary base point. That is, we define coordinates by simply flowing along the vector fields for durations α and β. Intuitively, the result is well-defined as long as flows commute, which is equivalent to having zero vector field commutator.

Formally, we would like to show that $V = \partial/\partial\alpha$ and $W = \partial/\partial\beta$. The first is always true; to prove the second, note that

$$\frac{\partial}{\partial\beta} x^i = e^{\beta W} W e^{\alpha V} x^i|_P = e^{\beta W} e^{\alpha V} (W x^i)|_P = (W x^i)|_{(\alpha,\beta)}$$

where in the second equality we used the commutation relations, and in the final equality we used the fact the $e^{\beta W} e^{\alpha V}$ translates any analytic function by (α, β). We must also prove that the (α, β) are actually a coordinate system; note that the map from the x^i to the new coordinates has Jacobian

$$J = \begin{pmatrix} \partial x^1/\partial\alpha & \partial x^2/\partial\alpha \\ \partial x^1/\partial\beta & \partial x^2/\partial\beta \end{pmatrix} = \begin{pmatrix} V x^1 & V x^2 \\ W x^1 & W x^2 \end{pmatrix}.$$

Then an inverse function exists if $\det J$ is nonzero by the inverse function theorem. But $\det J \neq 0$ is just the condition that V and W be nonzero and independent, as stated earlier.

5 Lie Theory

5.1 The Lie Derivative

In tensor analysis in $\mathbb{R}^n$, the convective derivative measures the rate of change of a tensor being transported in a velocity field. The Lie derivative does the same for manifolds.

- Given a vector field X, we define the Lie derivative of a scalar field as
$$(\mathcal{L}_X f)(x_0) = \lim_{t\to 0} \frac{f(x_1) - f(x_0)}{t}, \quad x_1 = \Phi_t x_0.$$
- It is clear that $\mathcal{L}_X f$ is just the rate of change of f along integral curves, so
$$\mathcal{L}_X f = Xf = X^i \frac{\partial f}{\partial x^i}$$
which is analogous to the convective derivative term $\vec{v}\cdot\nabla f$ in fluid mechanics.
- More formally, we can directly use the limit definition
$$(\mathcal{L}_X f)(x_0) = \lim_{t\to 0} \frac{1}{t}\left((\Phi_t^* f)(x_0) - f(x_0)\right) = \left(\left(\frac{d\Phi_t^*}{dt}\bigg|_{t=0}\right) f\right)(x_0) = (Xf)(x_0)$$
as desired, where we used $\Phi_t^* = e^{tX}$.
- Next, we can define the Lie derivative of a vector field. Generally, there's no way to compare vectors at different points on a manifold, but given a vector field, we can transport vectors using the flow; intuitively, a transported vector behaves like a stick moving in a stream.
- To formalize this, we use the pullback map, as we did for the scalar field. This is valid because Φ_t is invertible, so we can get a pullback through the inverse of the pushforward map,
$$(\mathcal{L}_X Y)(x_0) = \lim_{t\to 0} \frac{(\Phi_{t*}^{-1} Y)(x_0) - Y(x_0)}{t}.$$
- To simplify this expression, we expand all terms in the brackets to first order in t. Using the standard pullback formulas, the pulled back components just pick up a Jacobian factor,
$$(\Phi_{t*}^{-1} Y)(x_0) = \frac{\partial x_0^i}{\partial x_1^j} Y^j(x_1)$$
Expanding the flow equation $dx^i/dt = X^i(x)$ to first order,
$$x_0^i = x_1^i - tX^i(x_0), \quad \frac{\partial x_0^i}{\partial x_1^j} = \delta_j^i - t\frac{\partial X^i(x_0)}{\partial x_1^j} \approx \delta_j^i - t\frac{\partial X^i(x_0)}{\partial x_0^j}$$
There is another first-order component from the fact that Y^j is evaluated at x_1.
- Collecting all first-order terms and suppressing position arguments,
$$(\Phi_{t*}^{-1} Y)^i = (\delta_j^i - t\partial_j X^i)(Y^j + tX^k \partial_k Y^j) + O(t^2).$$
We thus conclude
$$(\mathcal{L}_X Y)^i = X^j \partial_j Y^i - Y^j \partial_j X^i.$$
The first term is what we would naively expect if we just followed the flow. The second term is more subtle and arises from how the flow affects the vector: a stick in a circulating current rotates, so the Lie derivative of a vector field with constant components can be nonzero.

- Similarly, we can find the Lie derivative of a covector field, using the usual pullback, giving

$$(\mathcal{L}_X \alpha)_i = X^j \partial_j \alpha_i + \alpha_j \partial_i X^j.$$

- Now we define the Lie derivative for arbitrary tensor fields. One way is to define the derivative as above, using the pullback map induced by Φ^* and Φ_*^{-1}. However, it is equivalent to use the Leibniz rule for tensor products, e.g.

$$\mathcal{L}_X(\alpha \otimes Y) = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left[(\Phi_\epsilon^* \alpha) \otimes (\Phi_{\epsilon *}^{-1} Y)\right] = (\mathcal{L}_X \alpha) \otimes Y + \alpha \otimes (\mathcal{L}_X Y).$$

  This is sufficient to define the Lie derivative for all tensors, because scalar multiplication is a special case of the tensor product,

$$f \otimes T = fT, \quad \mathcal{L}_X(fT) = (\mathcal{L}_X f)T + f(\mathcal{L}_X T)$$

  and every tensor field may be written as a linear combination of scalars times tensor products of vector and covector fields.

The Lie derivative of vector fields has some special properties.

- Comparing to our previous work, the Lie derivative is just the commutator,

$$\mathcal{L}_X Y = [X, Y].$$

  We call this operation the Lie bracket of two vector fields.

- Using the fact that the bracket is the commutator, we can easily show the Lie bracket is bilinear, antisymmetric, and satisfies the Jacobi identity

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0 \leftrightarrow \mathcal{L}_{[X,Y]} = [\mathcal{L}_X, \mathcal{L}_Y].$$

  Therefore, the set of vector fields $\mathfrak{X}(M)$ is a Lie algebra; the Lie group is $\mathrm{Diff}(M)$.

- If $f \colon M \to N$ is a diffeomorphism, then the bracket commutes with pushforward,

$$f_*[X, Y] = [f_* X, f_* Y].$$

  This follows because advance maps commute with diffeomorphisms, as diffeomorphisms are just isomorphisms of differentiable structure.

Note. Intuition for the Lie derivative. As we saw earlier, the commutator $[X, Y]$ measures the difference between traversing integral curves of X or Y first. This translates into intuition for Lie dragging the *vectors* X and Y, because we can think of vectors as tiny pieces of integral curves.

Specifically, let points a and b be linked by flowing along the integral curves of X for an infinitesimal time. Then we can think of the vector $X(a)$ as pointing from a to b. Now let

$$X \colon a \to b, \quad Y \colon a \to c, \quad X \colon c \to d, \quad Y \colon b \to d'.$$

Then $c \to d$ represents $X(c)$, but $c \to d'$ represents $X(a)$ Lie dragged along Y. The difference between d and d' measures the Lie derivative $\mathcal{L}_Y X$, and it is also the commutator because XY takes $a \to d$ and YX takes $a \to d'$.

5.2 Frobenius' Theorem

Next, we apply the Lie bracket to submanifolds.

- An m-dimensional (embedded) submanifold S of an n-dimensional manifold M is a set of points in M so that, in a neighborhood of any point of S, there exists a coordinate system where the points of S are described by $x^1 = \ldots = x^{n-m} = 0$.
- The above definition allows S to inherit all smoothness properties of M. It is natural for applications, because the solutions of differential equations are usually relations between the x^i, so they are already in the desired form.
- Note that any open subset of M is trivially a submanifold of M. Moreover, submanifolds are not allowed to intersect themselves.
- Using the inclusion map from S to M, we can restrict forms from M to S by pullback, and move vectors on S to M. Both of these facts make sense geometrically, thinking of forms as contour surfaces and vectors as small arrows.
- A set of vector fields $X^{(i)}$ is involutive if it is closed under the Lie bracket, i.e. if the commutators $[X^{(i)}, X^{(j)}]$ may be written in terms of linear combinations of the $X^{(i)}$, where the coefficients may be functions.
- Using the identity
$$\mathcal{L}_X(fY) = (\mathcal{L}_X f)Y + f\mathcal{L}_X Y$$
we can show that this implies that the commutators $[f_i X^{(i)}, g_j X^{(j)}]$ are also linear combinations of the $X^{(i)}$, where the coefficients f_i and g_j are functions.
- Defining coordinates y^a on S, we have m vector fields $\partial/\partial y^a$ which are clearly involutive. Since this closure property is also true for linear combinations of the vector fields, the set of all vector fields on S is closed under the Lie bracket. Intuitively, no combination of tangent vectors could yield anything besides another tangent vector; flows on S can't take us off S.
- Frobenius' theorem states that the converse is true: if the vector fields $V^{(i)}$ are involutive, then the integral curves of the vector fields mesh together to form a family of submanifolds that foliate M. That is, every point of M lies on one such submanifold (except for a small number of degenerate points), and at each point the vector fields $V^{(i)}$ span the tangent space of that submanifold.
- As a simple example, for one vector field, the submanifolds are just the integral curves.
- Proof sketch: by taking appropriate linear combinations, we can set the commutators to zero, so we have a set of commuting flows. As we've seen earlier, such a set defines a coordinate system, and this is the desired coordinate system for the submanifold.

Example. Consider $X = \partial_x + y\partial_z$ and $Y = \partial_y$. If Frobenius' theorem applied, we would expect the submanifold going through the origin to be tangent to the xy plane, since X and Y span it there. However, $[X, Y] = -\partial_z$, so it is possible to move along the z-axis at the origin, so the integral curves do not combine into a family of two-dimensional submanifolds. Instead, we can get to *any* point by flowing along X and Y.

Example. Define the vector fields ℓ_i as

$$\ell_z = -y\partial_x + x\partial_y$$

and ℓ_x and ℓ_x similarly. These generate rotations about the x, y, and z axes. The commutation relations are $[\ell_x, \ell_y] = -\ell_z$ along with cyclic permutations, so Frobenius' theorem applies. The resulting submanifolds are spheres centered about the origin. To show this formally, define $r = \sqrt{x^2 + y^2 + z^2}$. Then one can show that

$$dr(\ell_i) = 0.$$

Then the submanifolds lie within surfaces of constant r. Since the ℓ_i span a two-dimensional tangent space at every point, the submanifolds must in fact be these surfaces.

Note. There is a dual formulation to Frobenius' theorem. Consider a set of p linearly independent one-forms $\omega^{(i)}$. At each point, the annihilator is the subset of the tangent space annihilated by all of these forms; then we might ask if these tangent spaces mesh together to form submanifolds of codimension p. This is clearly true if the $\omega^{(i)}$ are exact, $\omega^{(i)} = df^i$, because then the submanifold is simply $f^i = \text{const}$. By our definition of a submanifold, the converse is true: the one-forms are 'surface-forming' if they can be written as linear combinations of a set of p exact forms.

Frobenius' theorem for forms states that the forms are surface-forming if and only if they are closed, which means that for V and W in the annihilator, $d\omega^{(i)}(V, W) = 0$. The name of the condition is because this is a generalization of ordinary closure, $d\omega = 0$.

One important application of the Lie derivative is to express the symmetries of a physical problem.

- A tensor field T is said to be invariant under V if

  $$\mathcal{L}_V T = 0.$$

  As a simple example, functions $f(r)$ are invariant under the ℓ_i defined above.

- The set of vector fields V under which T is invariant forms a Lie algebra, where the operation is the usual Lie bracket. To prove this, we need to check closure under linear combinations with constant coefficients, which is straightforward, and closure under the Lie bracket, which follows from the Jacobi identity $[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X,Y]}$.

- The dimension of a Lie algebra is equal to its dimension as a vector space. Note that we only allow scalar multiplication by constants, not functions, in a Lie algebra. Thus the ℓ_i form a three-dimensional Lie algebra, even though they span only a two-dimensional space at every point. Moreover, the set of all vector fields on a manifold is an infinite-dimensional Lie algebra, even though the dimension of the manifold's tangent spaces are finite.

- Geometrically, it is natural to allow linear combinations with functions as coefficients for Frobenius' theorem, because multiplying a vector field by a function leaves its integral curves invariant, simply changing the speed at which they are traversed. But it is unnatural to do the same for symmetries: invariance under translation is very different from invariance under any deformation of space.

Killing vectors are vector fields under which the metric is invariant, and are useful in relativity.

Example. Killing vectors in $\mathbb{R}^3$ with the Euclidean metric. Note that in general,

$$(\mathcal{L}_{\partial_i} T)_{i\ldots j}{}^{k\ldots\ell} = \partial_i T_{i\ldots j}{}^{k\ldots\ell}.$$

Therefore, since the metric components are independent of x, y, and z, we have three Killing vectors ∂_x, ∂_y, and ∂_z. The rotation operator $\ell_z = \partial_\phi$ is also a Killing vector, as can be seen by switching to spherical coordinates where the metric components are independent of ϕ. Similarly, ℓ_x and ℓ_y are also Killing vectors. We'll prove later that these are all the independent Killing vectors.

Note. Killing vectors appear in the background even in classical mechanics. For example, angular momentum is conserved in a spherically symmetric potential. However, no analogous conserved quantity exists for a potential which is constant on ellipsoids, even though such a potential does have a symmetry. To see why, note that the equation of motion is

$$m\dot{v}^i = -g^{ij}\partial_j \Phi.$$

Since the metric is involved, symmetries must be derived from both Φ and g, and g is spherically symmetric but not 'ellipsoidally symmetric'.

Example. Axial symmetry. Suppose we wish to solve the equation $L\psi = 0$, where L is a linear differential operator and ψ is a function, e.g. a wavefunction or a field. If there is an axial symmetry, then L is independent of the angular coordinate ϕ, though it may contain derivatives with respect to ϕ. Thus, as maps on smooth functions, L commutes with $\mathcal{L}_{\partial/\partial\phi}$ so we can simultaneously diagonalize them. (Note that L need not be a first order differential operator, so it may not be interpreted as a vector field.)

Therefore we can consider solutions with axial eigenvalue m,

$$\mathcal{L}_{\partial/\partial\phi}\psi = im\psi.$$

Such a solution can be written as $\psi = \psi_m e^{im\phi}$, and the equation $L(\psi) = 0$ can be simplified to $L_m(\psi_m) = 0$, where L_m contains no derivatives with respect to ϕ. Thus the axial symmetry allows separation of variables. This procedure is familiar from quantum mechanics, but here we see it applies in general.

We call the functions $e^{im\phi}$ the scalar axial harmonics; any solution to $L\psi = 0$ may be written as a sum of solutions ψ_m, each of whose angular dependence is a scalar axial harmonic. If ψ is instead a vector field V, then the analogous procedure uses vector axial harmonics, which satisfy

$$\mathcal{L}_{\partial/\partial\phi} V = imV.$$

If the space is n-dimensional, there are n independent vector axial harmonics for each value of m. For example, in $\mathbb{R}^3$ with axial symmetry about the z axis, one basis for the $m = 0$ vector axial harmonics is $\hat{\mathbf{z}}$, $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, as these basis vectors are Lie dragged by $\mathcal{L}_{\partial/\partial\phi}$.

5.3 Lie Groups and Lie Algebras

Axial symmetries, and more generally rotations, form a Lie group, a manifold with a smooth group operation. The vector fields generating the symmetry form the Lie algebra. In this section we formally define these objects, beginning by reviewing group actions.

- Given a space X, let $\mathrm{Bij}(X)$ be the group of all bijections of X. If $X = M$ is a differentiable manifold, the analogous group is $\mathrm{Diff}(M)$, the group of diffeomorphisms of M onto itself.

- Given a group G, a homomorphism $G \to \mathrm{Bij}(X)$ is an action of G on X. We write the bijection corresponding to g as Φ_g.
- If X is a vector space and the Φ_g are linear operators, we call the group action a representation of G. Sometimes, we also call X itself the representation.

  A slightly different convention is to call group actions 'realizations' of the group, and any realization that is not a representation a 'nonlinear realization'.
- The orbit of $x \in X$ is the set $\{\Phi_g x \,|\, g \in G\}$. It is straightforward to show that the orbits partition X into subsets; we write the orbit of x as $[x]$. The action is transitive if $[x] = X$, i.e. it takes each element to all others.
- Define the stabilizer of x as $I_x = \{g \in G \,|\, \Phi_g x = x\}$. For any $y \in [x]$, the set of group elements that map x to y is a coset of I_x. This gives the orbit-stabilizer theorem

  $$|G/I_x| = |[x]|$$

  which is useful for some combinatorics problems. Since I_x need not be normal, G/I_x need not be a group, so we should only this of this relation as set equality.
- If all transformations except for $\Phi_e = \mathrm{id}_X$ move all points of X, the action is free. Then all stabilizers are trivial, so $G \cong [x]$. Then X consists of copies of G in orbits.
- If all transformations except for $\Phi_e = \mathrm{id}_X$ move some point of X, then the action is effective. Then the kernel of the action $G \to \mathrm{Bij}(X)$ is trivial, and G is isomorphic to the set of $\{\Phi_g\}$. If the action is a representation, it is called faithful.

Example. The group $SO(3)$ acts on $\mathbb{R}^3$ by rotations. The orbits are spheres, unless $\mathbf{x} = 0$, in which case $[\mathbf{x}]$ is a point. The stabilizer $I_{\mathbf{x}}$ of any $\mathbf{x} \neq 0$ is the set of rotations about the axis $\hat{\mathbf{x}}$, an $SO(2)$ subgroup. The orbit-stabilizer theorem says $SO(3)/SO(2) \cong S^2$.

Example. An arbitrary group G acts on itself by left or right translation,

$$L_a \colon G \to G, \quad L_a(g) = ag, \quad R_a \colon G \to G, \quad R_a(g) = ga.$$

Then the mapping $a \to L_a$ is an action. The mapping $a \to R_a$ is not since it doesn't obey the homomorphism condition $R_a R_b = R_{ab}$, but $a \to R_{a^{-1}}$ is. Note that left and right translations always commute, e.g. $L_a R_b = R_b L_a$. However, left/right translations don't commute among themselves if the group is not Abelian.

Example. A group G also acts on itself by conjugation,

$$I_a \colon G \to G, \quad g \mapsto aga^{-1}.$$

That is, $I_a = L_a R_{a^{-1}}$. This action is automatically trivial if the group is Abelian.

We now turn to defining the Lie algebra of a Lie group.

- The actions of left and right translation are diffeomorphisms of the group manifold. We define a left-invariant vector field (LIVF) to satisfy

  $$L_{a*} X = X$$

  for all $a \in G$. These vector fields are 'constant' along the group manifold.

- The set of all LIVFs is a Lie algebra under the commutator, because

$$L_{a*}[X,Y] = [L_{a*}X, L_{a*}Y] = [X,Y]$$

  where we used that fact that pushforward commutes with the commutator. We call this set the Lie algebra $\mathfrak{g}$ of G.

- Note that every LIVF X is determined by its value at the identity element X_V. Conversely, every vector at the T_eG can be extended to an LIVF. Therefore we can identity $\mathfrak{g}$ with the tangent space T_eG, where the bracket operation is inherited from the LIVFs,

$$[X_V, X_W] = X_{[V,W]}.$$

  Since we've already proven that the vector field commutator satisfies the Jacobi identity, the bracket operation here does as well.

- If $\mathfrak{h} \subset \mathfrak{g}$ is a Lie subalgebra, then $\mathfrak{h}$ corresponds to a Lie subgroup $H \subset G$. This follows by Frobenius' theorem, since $\mathfrak{h}$ defines a family of involutive vector fields.

Note. As we've seen above, the group structure on the manifold is quite powerful. It automatically gives a way to identity distinct tangent spaces, as well as a privileged point, the identity element.

As an application, suppose we pick a basis of T_eG. Using the diffeomorphisms L_a, we can extend this smoothly to a set of vector fields on G that are linearly independent at every point, called a field of frames. Using these to define coordinates for the tangent space at each point shows that the tangent bundle of a Lie group must be trivial!

Conversely, it is impossible to find even a single independent (i.e. nonvanishing) vector field on an even-dimensional sphere, which is an indication that their tangent bundles are nontrivial. Thus even-dimensional spheres cannot be given a Lie group structure.

Note that a field of frames generally does not define a coordinate basis. We have shown earlier that in $\mathbb{R}^n$ they define local coordinates if and only if their bracket vanishes; this construction can still fail globally due to topology.

We can relate $\mathfrak{g}$ back to G by the exponential map.

- Given $V \in \mathfrak{g}$, define $\Phi_{V,t}$ to be the advance map of the vector field X_V. Let $\sigma(t)$ be the integral curve going through e, so $\sigma(t) = \Phi_t e$.

- Because advance maps and pushforward commute,

$$\Phi_{V,t} g = \Phi_{V,t} L_g e = L_g \Phi_{V,t} e = L_g \sigma(t) = g\sigma(t).$$

  Therefore, the advance map is simply a right translation. This is another way the fact that all elements of a group "look the same" constrains geometry on the group manifold.

- Using this fact, we can prove that $\sigma\colon \mathbb{R} \to G$ is a homomorphism,

$$\sigma(s+t) = \Phi_{s+t} e = \Phi_s \Phi_t e = \Phi_s \sigma(t) = \sigma(t)\sigma(s).$$

  Such a homomorphism is called a one-parameter subgroup, and this construction shows that elements of $\mathfrak{g}$ are in correspondence with them. By contrast with the general case, our vector field flows are always defined for all $t \in \mathbb{R}$ thanks to the group structure.

- Define the exponential map $\exp\colon \mathfrak{g} \to G$ by simply following the integral curves for a unit time, i.e. mapping V to $\sigma(1)$ in the notation above. Note that the differential of exp is just the identity on $\mathfrak{g}$.
- It can be shown that the exponential map is surjective for connected, compact Lie groups.
- The exponential map is bijective in a neighborhood of the identity, so we can define a coordinate system there by taking a basis $\{e_\mu\}$ of $\mathfrak{g}$ and assigning the point $\exp(V^\mu e_\mu)$ the coordinates V^μ. For example, for $SO(3)$, one set of exponential coordinates are Cartesian coordinates, when we embed D^3 in $\mathbb{R}^3$. Riemann normal coordinates in relativity are similar, though they use geodesics of connections, not integral curves of vector fields.
- It can be shown that every Lie algebra $\mathfrak{g}$ is the Lie algebra of exactly one simply-connected Lie group G. Moreover, if G' also has Lie algebra $\mathfrak{g}$ and is connected, then G is its universal cover. One example of this is $\mathfrak{g} = \mathfrak{su}(2)$, $G = SU(2)$, and $G' = SO(3)$.

Example. If all the brackets in a Lie algebra are zero, it is abelian. Now, we know that $\mathbb{R}^n$ has an abelian Lie algebra, and since $\mathbb{R}^n$ is simply connected, it covers any other group with the same Lie algebra. Since $\mathbb{R}^n$ is abelian, all groups with an abelian Lie algebra are abelian groups!

Example. One-parameter subgroups in $SO(3)$ are rotations with a fixed angular velocity. To visualize them, recall that $SO(3)$ is D^3 with antipodal points identified. Then these subgroups are lines through the origin, which wrap back around when they hit the edge.

Note. Often, a Lie group describes the symmetries of a manifold. We should not confuse the action of the Lie group on that manifold with the actions of the Lie group on itself. While the elements of $SO(3)$ represent rotations, left-translation in $SO(3)$ looks nothing like a rotation of $D^3/\{\pm 1\}$.

We now cover the Maurer-Cartan structure equations.

- Let V_μ be a basis for $\mathfrak{g}$ with corresponding LIVFs X_μ. We define the structure constants

$$[V_\mu, V_\nu] = c_{\mu\nu}{}^\sigma V_\sigma.$$

  The structure constants are the components of a $(1,2)$ tensor on $\mathfrak{g}$. We must be careful not to confuse the Greek indices above with components.

- By left-translating the above equation, we find

$$[X_\mu, X_\nu] = c_{\mu\nu}{}^\sigma X_\sigma$$

  where the structure constants $c_{\mu\nu}{}^\sigma$ do not depend on position (hence the name).

- We can also shift attention from vectors to one-forms. We let $\mathfrak{g}^*$ be the dual space of $\mathfrak{g}$, with basis β^μ dual to V_μ, and define the left-invariant one-forms

$$\theta^\mu|_a = L^*_{a^{-1}}\beta^\mu$$

  in analogy with X_μ. We pullback using a^{-1} because pullback runs opposite to pushforward.

- The left-invariant vector fields and one-forms are dual everywhere,

$$\theta^\mu(X_\nu)|_a = (L^*_{a^{-1}}\beta^\mu)(L_{a*}V_\nu) = \beta^\mu(L_{a^{-1}*}L_{a*}V_\nu) = \beta^\mu V_\nu = \delta^\mu_\nu.$$

- By expanding in coordinates, we can show that for any one-form field α,

$$d\alpha(X,Y) = X\alpha(Y) - Y\alpha(X) - \alpha([X,Y]).$$

- Applying this identity, we have

$$(d\theta^\mu)(X_\nu, X_\sigma) = X_\nu \delta^\mu_\sigma - X_\sigma \delta^\mu_\nu - \theta^\mu([X_\nu, X_\sigma]).$$

  The first two terms are zero as they are the derivatives of a constant, giving

$$(d\theta^\mu)(X_\nu, X_\sigma) = -c_{\nu\sigma}{}^\mu.$$

  It's important to avoid being distracted by the Greek indices here. We conclude that

$$d\theta^\mu = -\frac{1}{2} c_{\nu\sigma}{}^\mu \theta^\nu \wedge \theta^\sigma.$$

  These are the Maurer-Cartan structure equations.

- To remove the coordinates completely, define the Lie-algebra valued one-form $\theta = V_\mu \otimes \theta^\mu$, called the Maurer-Cartan form. Unlike a regular one-form, which maps vectors to real numbers, θ maps vectors to $\mathfrak{g}$. It is similar to the gauge potential A in Yang–Mills.

- The form $\theta|_a$ maps T_aG to $\mathfrak{g} = T_eG$. Geometrically, it simply moves a vector based at a over to the identity by left translation. It encodes the structure of the Lie group in the same way the structure constants do.

- We define the operations

$$d\theta = V_\mu \otimes d\theta^\mu, \quad [\theta, \theta] = [V_\mu, V_\nu] \otimes \theta^\mu \wedge \theta^\nu.$$

  Then the Maurer-Cartan structure equation reduces to

$$d\theta + \frac{1}{2}[\theta, \theta] = 0.$$

Note. Another proof that fields of frames locally correspond to coordinates if and only if the vector field brackets vanish. For any field of frames, the Maurer-Cartan structure equations hold, except that the coefficients $c_{\nu\sigma}{}^\mu$ are no longer constant. If the brackets vanish, the structure constants vanish, so $d\theta^\mu = 0$. Then the θ^μ are locally exact by the Poincare lemma, giving the desired coordinates.

Next, we consider the adjoint representation.

- Every group acts on itself by conjugation I_g, and conjugation always fixes the identity element. Since every path through the identity remains a path through the identity, conjugation maps a Lie algebra to itself.

- More formally, the map is

$$\mathrm{Ad}_g = I_{g*}|_e \colon \mathfrak{g} \to \mathfrak{g}, \quad I_g = R_{g^{-1}} L_g.$$

  Using the composition rule for pushforward, $(fg)_* = f_* g_*$, we have $\mathrm{Ad}_g = R_{g^{-1}*}|_g L_{g*}|_e$.

- Using the fact that left-translations and right-translations commute, we can show that $\mathrm{Ad}_g \mathrm{Ad}_h = \mathrm{Ad}_{gh}$, so the adjoint is a group action. It is a representation of G on the vector space $\mathfrak{g}$.

Next, we consider how Lie groups act on manifolds, our original motivation for studying them.

- Let G be $SO(3)$ and act on $M = \mathbb{R}^3$ by spatial rotations. Then V is an angular velocity, and we can associated it with an induced vector field V_M on M equal to $\boldsymbol{\omega} \times \mathbf{r}$. More generally, for any G acting on any M, the induced vector field is the infinitesimal symmetry corresponding to a Lie algebra element.
- Formally, the infinitesimal symmetry V generates the diffeomorphisms $\Phi_{\exp(tV)}$. Then it is natural to define the induced vector field as

$$V_M = \frac{d}{dt} \Phi^*_{\exp(tV)} \bigg|_{t=0}$$

  where both sides should be regarded as maps $\mathcal{F}(M) \to \mathcal{F}(M)$.

5.4 Matrix Groups

The group $GL(n, \mathbb{R})$ of invertible $n \times n$ real matrices is a Lie group with dimension n^2. Many important groups in physics, such as $SO(3)$, $SO(3,1)$, etc. are subgroups of this group. Matrix groups come with a natural embedding in $\mathbb{R}^{n^2}$ which makes some concrete computations easier.

- To start, we need to show that $GL(n, \mathbb{R})$ is a submanifold of $\mathrm{Mat}(n, \mathbb{R}) \cong \mathbb{R}^{n^2}$. Note that it is the inverse image of $\mathbb{R} \setminus \{0\}$ under the determinant, and this is open in $\mathbb{R}$. Then by continuity $GL(n, \mathbb{R})$ is open in $\mathbb{R}^{n^2}$ and hence locally looks like $\mathbb{R}^{n^2}$. Since the group operation is clearly smooth, we conclude $GL(n, \mathbb{R})$ is a Lie group.
- It can be shown that every closed subgroup H of a Lie group G is also a Lie group.

  As an application, note that $SL(n, \mathbb{R})$ it is the kernel of the homomorphism $f(A) = \det A$. Since f is continuous and $\{1\}$ is closed in $\mathbb{R}$, the kernel is closed, so $SL(n, \mathbb{R})$ is a Lie group.
- The operation in a matrix Lie group is simply matrix multiplication, and the one-parameter subgroups are matrix exponentials $\sigma(t) = e^{At}$. Conversely, given a one-parameter subgroup, we can compute a Lie algebra element by evaluating $\sigma'(0)$. Equivalently, the Lie algebra elements are the matrices X so that $I + \epsilon X$ is in the group.
- The Lie bracket/vector field commutator turns out to just be the matrix commutator. To see this, note that the vector field commutator encodes the noncommutativity of flows, and

$$e^{\epsilon A} e^{\epsilon B} - e^{\epsilon B} e^{\epsilon A} = \epsilon^2 [A, B] + O(\epsilon^3).$$

  This is how the Lie bracket encodes 'second-order' information about the group.
- The adjoint representation is just matrix conjugation.
- Ado's theorem states that all finite-dimensional Lie algebras can be viewed as matrix Lie algebras, so we actually lose little generality with this approach.

We can quickly calculate the Lie algebras of matrix Lie groups.

- For $GL(n,\mathbb{R})$, note that the determinant is continuous. Then every matrix near the identity has determinant near one, so $\mathfrak{gl}(n,\mathbb{R})$ is the set of all real $n \times n$ matrices.
- For $SL(n,\mathbb{R})$, we use the identity
$$\det e^X = \exp \operatorname{tr} X$$
which shows that $\mathfrak{sl}(n,\mathbb{R})$ is the set of matrices with zero trace.
- For $SO(n)$, X is in the Lie algebra if $I + \epsilon X$ is a rotation matrix up to $O(\epsilon^2)$, which means
$$(I + \epsilon X)^T (I + \epsilon X) = I + O(\epsilon^2).$$
Expanding this shows $X = -X^T$, so $\mathfrak{so}(n)$ is the set of skew-symmetric matrices. Note that $\mathfrak{so}(n)$ is the same as $\mathfrak{o}(n)$.
- Similarly, $\mathfrak{u}(n)$ contains skew-Hermitian matrices. In quantum mechanics, we throw in an extra factor of i to consider Hermitian observables instead.
- The Lie algebra $\mathfrak{su}(n)$ is the subset of $\mathfrak{u}(n)$ with zero trace. This is different from the relationship between $\mathfrak{so}(n)$ and $\mathfrak{o}(n)$ because restricting to unit determinant when the determinant can be complex sets the phase to zero, taking a dimension away from the group manifold.

Example. The group $SO(3,1)$ acts on Minkowski space by Lorentz transformations. The group $SL(2,\mathbb{C})$ does as well. Points in Minkowski space correspond with Hermitian matrices by
$$X = x^\mu \sigma_\mu, \quad x^\mu = \frac{1}{2} \operatorname{tr} \sigma_\mu X.$$
In particular, note that
$$\det X = x^\mu x_\mu.$$
We define the action of $SL(2,\mathbb{C})$ on Minkowski space by
$$x \to AX(x)A^\dagger, \quad A \in SL(2,\mathbb{C}).$$
These are Lorentz transformations, because they preserve $x^\mu x_\mu$. We only get Lorentz transformations connected to the identity (the proper orthochronous ones), which we call $SO(3,1)_+$, because $SL(2,\mathbb{C})$ is connected. We get all of $SO(3,1)_+$ because
$$A = \exp\left(-i\frac{\theta}{2}(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma})\right)$$
represents a rotation by θ about the $\hat{\mathbf{n}}$ axis, while
$$A = \exp\left(\frac{\alpha}{2}(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma})\right)$$
represents a boost of rapidity α along the $\hat{\mathbf{n}}$ axis, and boosts and rotations generate everything. Using this setup, we can define a map of $SL(2,\mathbb{C})$ onto $SO(3,1)_+$, but it is a double cover since A and $-A$ map to the same Lorentz transformation. The situation here is closely analogous to the relationship between $SU(2)$ and $SO(3)$.

Example. Often, we consider the action of a matrix Lie group on $\mathbb{R}^n$ by matrix multiplication. In this case, induced vector fields are easy to compute. The flow generated by $V \in \mathfrak{g}$ is $\exp(tV)$, so

$$V_M f|_x = \frac{d}{dt} f(e^{tV} x)\bigg|_{t=0} = (Vx)_i \frac{\partial f}{\partial x_i}$$

which implies that $V_M = V_{ij} x_j \partial_i$.

We give some examples of Lie group actions.

- Let a Lie group G act on a manifold M. The little group of a point $p \in M$ is the stabilizer $H(p)$ of p. We claim that the little group is a Lie subgroup.

  To see this, consider the map $g \mapsto gp$. The little group is the inverse image of p, and since a single point is closed, the little group is closed as well, giving the result.

- If G acts transitively on a space M, the orbit-stabilizer theorem states

  $$G/H(p) \cong M$$

  which holds given some compactness conditions. Moreover, the coset space $G/H(p)$, called a homogeneous space, is a manifold.

- We've already seen that $SO(3)/SO(2) = S^2$. This generalizes to

  $$SO(n+1)/SO(n) = O(n+1)/O(n) = S^n, \quad U(n+1)/U(n) = SU(n+1)/SU(n) = S^{2n+1}$$

  where we interpret unitary matrices as rotations in complex space.

- As another example, $O(n+1)$ acts on $\mathbb{R}^{n+1}$ in the usual manner, but it preserves lines through the origin. Thus $O(n+1)$ acts on $\mathbb{RP}^n$, and it is clear this action is transitive. The stabilizer of a line is $O(n) \times O(1)$, where the elements of $O(1)$ flip the line, so

  $$O(n+1)/(O(1) \times O(n)) = \mathbb{RP}^n.$$

  This is consistent with our earlier work, as we know that $\mathbb{RP}^n = S^n/\mathbb{Z}_2$.

- More generally, let the Grassmannian $G_k(\mathbb{R}^n)$ be the set of k-dimensional subspaces of $\mathbb{R}^n$, so that $G_1(\mathbb{R}^n) = \mathbb{RP}^{n-1}$. Then

  $$G_k(\mathbb{R}^n) = O(n)/(O(k) \times O(n-k))$$

  where the denominator comes from rotations in the subspace and its orthogonal complement.

6 Differential Forms

6.1 Geometric Intuition

Definition. A differential form of rank r is a completely antisymmetric $(0, r)$ tensor.

Example. A 0-form is a scalar function, a 1-form is a covector, and a 2-form is an antisymmetric $(0, 2)$ tensor, i.e. $\omega(X, Y) = -\omega(Y, X)$.

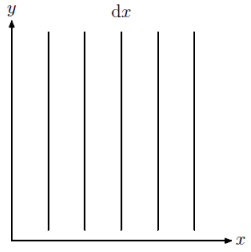
Example. In n dimensions, an r-form has $\binom{n}{r}$ independent components. We may also consider r-form fields. The components of these fields are functions, and an r-form field has $\binom{n}{r}$ independent component functions. Denote the set of smooth r-form fields on M as $\Lambda^r(M)$. By convention, we let all r-forms with $r > n$ be equal to zero.

Example. An n-form has a single independent component, and hence may be written as

$$\phi_{i_1 \dots i_n}(x) = \sigma(x)\epsilon_{i_1 \dots i_n}$$

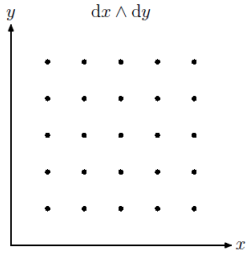
where $\sigma(x)$ is called the scalar density, and $\epsilon_{i_1 \dots i_n}$ is the Levi–Civita symbol.

We use differential forms to define integration on manifolds. More specifically, an r-form is the integrand in an integral over an r-dimensional submanifold. This operation has a natural geometric picture; for simplicity, work in the plane $\mathbb{R}^2$. Then a 1-form is a 'slicing of space', i.e. a set of 1-dimensional surfaces in the plane. For example, dx is shown below.



An integral along a curve may be evaluated with respect to a 1-form by counting the number of these slices the 1-form passes through. More generally, an r-form on an n-dimensional manifold is a set of $(n - r)$-dimensional surfaces (also called $(n - r)$-leaves).

Next, the picture corresponding to the exterior product of two forms $\alpha \wedge \beta$ is the set of intersections of the surfaces of α and β. For example, the exterior product of two 1-forms is a 2-form, which in two dimensions is a set of points. Below we show the 2-form $dx \wedge dy$.



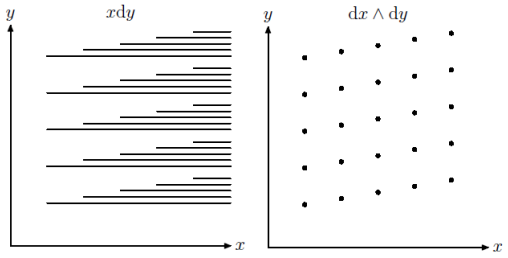
A surface integral over this 2-form is the number of points inside the surface. Generally, an n-form on an n-dimensional manifold is represented by a density of points.

One subtlety that our pictures cannot capture is that the surfaces associated with a differential form are signed, so that an integral over forms can be negative. In particular, for a non-orientable manifold, there is no nonvanishing n-form, though this is impossible to see in terms of a density of

points. For $(n-1)$-forms, we may indicate the sign by giving directions to the lines, in which case the integral is a signed flux.

Note. Now we consider the exterior derivative d. This operation takes an r-form ω and yields an $(r+1)$-form $d\omega$. Geometrically, the surfaces in $d\omega$ are the boundaries of those in ω. As we know from the homology chapter, the boundary of a boundary is zero, so $d^2 = 0$.

As a specific example, given a function f, df is the set of its contour lines, which are closed curves. Then $d(df) = 0$ automatically. A less trivial case is $d(xdy) = dx \wedge dy$, shown below.



Now, Stokes' theorem tells us that integration on forms satisfies the identity

$$\int_C d\omega = \int_{\partial C} \omega$$

where ∂ is the boundary operator introduced earlier. This has a simple interpretation in terms of the example above: the number of 0-leaves of $dx \wedge dy$ contained in C is equal to number of 1-leaves of xdy piercing the surface ∂C.

There are several limitations of our visualization. The associated pictures are hard to see in higher than three dimensions. Worse, they don't behave nicely under addition; the form $dx + dy$ in the plane has diagonal lines, while dx and dy individually contain horizontal/vertical lines. One hack is to represent $dx+dy$ as the union of the horizontal and vertical lines, which makes integration work, but then it's hard to tell this is the same form as $dx + dy$ drawn with diagonal lines.

Note. We can also visualize the Hodge dual $*$. This is a bijective map between r-forms and $(n-r)$-forms, and geometrically, it is performed by replacing r-leaves with their orthogonal complements. For example, in two dimensions, $*dx = dy$, and in three dimensions, $*dx = dy \wedge dz$. For $r = n/2$, the density of the r-leaves of the dual is equal to the density of the r-leaves of the original form. Note that we've dropped some signs above, since our diagrams can't show sign.

Example. Maxwell's equations in 2D. The equations are

$$dF = 0, \quad d * F = 0$$

where F is a one-form. We find a rotationally symmetric solution. Since $dF = 0$, we need the curves of F to form closed circles or head out to infinity; we take the former. Then $*F$ contains lines that head out to infinity. The density of lines in $*F$ falls as $1/r$, so the same must hold for the lines of F. Thus $F \propto dr/r$, and electromagnetic fields in 2D fall off as $1/r$.

Note. Finally, let's return to the issue of integration. We have seen that r-forms can be used to integrate over r-dimensional regions. To connect this with our definition of an r-form as a $(0, r)$ antisymmetric tensor, we need to relate the region itself to the argument of the r-form.

An r-form maps $(r,0)$ tensors to numbers, but since the contraction of a symmetric and antisymmetric object yields zero, only the fully antisymmetric part of the $(r,0)$ tensor contributes. Thus r-forms act on antisymmetric $(r,0)$ tensors, which we call multivectors or r-blades.

Multivectors are constructed from the exterior products (i.e. antisymmetrized tensor products) of vectors, and we claim that they represent signed volumes. For example, $\mathbf{a} \wedge \mathbf{b}$ corresponds to the signed area of the parallelogram formed by $\mathbf{a}$ and $\mathbf{b}$, and in n dimensions, the exterior product of n vectors $\mathbf{a}_i$ is equal to the signed volume spanned by those vectors. This volume is equal to the determinant of the matrix with columns $\mathbf{a}_i$. On a more abstract level, multivectors correspond to signed volumes because the exterior product is linear and antisymmetric in its arguments.

This gives our last piece of intuition for integration of forms: an r-form acting on an r-blade is the number of r-leaves passing through the corresponding r-dimensional region.

Example. Physically, differential forms are 'things that go under integral signs', and they represent densities. For example, the magnetic flux density is a two-form, and in fluid dynamics, we use both the volume form and the mass density form, whose integral gives the mass inside a region.

6.2 Operations on Forms

We begin with basic operations to construct differential forms.

- Given any $(0,r)$ tensor, we can antisymmetrize it to get an r-form, with components

$$A_{[i_1 \ldots i_r]} = \frac{1}{r!} \sum_{j=\sigma(i)} |\sigma| A_{j_1 \ldots j_r}.$$

  Here, we are summing over permutations σ and $|\sigma|$ is the sign of the permutation. The normalization constant ensures that this operation leaves differential forms invariant.

- Given two one-forms, we define the wedge product or exterior product as

$$p \wedge q = p \otimes q - q \otimes p.$$

  The wedge product of two one-forms is a two-form.

- Similarly, we can generalize this definition to many one-forms by completely antisymmetrizing,

$$p_1 \wedge \ldots \wedge p_r = \sum_{i'=\sigma(i)} |\sigma| \, p_{1'} \otimes \cdots \otimes p_{r'}.$$

  Note that there is no $1/r!$ factor for normalization, by convention.

- Since every differential form can be written as a linear combination of exterior products of one-forms, this definition extends by linearity to arbitrary forms. The exterior product makes the set of all differential forms

$$\Lambda^*(M) = \Lambda^0(M) \oplus \cdots \oplus \Lambda^n(M)$$

  into a Grassmann algebra, called the exterior algebra, with dimension 2^n. Another definition of $\Lambda^*(M)$ is to start with the full tensor algebra and quotient out symmetric tensors; the tensor product then becomes the wedge product.

- The exterior product is associative, and for an r-form α and an s-form β,

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

  To see this, we can decompose α and β into linear combinations of exterior products of one-forms. Each term then requires rs anticommutations to swap the positions of α and β.

- Conventionally, we write an r-form ω as

$$\omega = \frac{1}{r!} \omega_{i\ldots j}(x)\, dx^i \wedge \ldots \wedge dx^j.$$

  The factor of $1/r!$ cancels the 'missing' $1/r!$ in the definition of the exterior product.

We can also contract differential forms with vectors.

- For any vector X, we define

$$\alpha(X) = \alpha(X, \cdot, \ldots, \cdot), \quad [\alpha(X)]_{j\ldots k} = \alpha_{i\ldots k} \xi^i.$$

  We arbitrarily choose to contract X with the first slot; choosing other slots would just change the definition by a sign. We also call this operation the 'interior product' and write it as i_X. By antisymmetry, $i_X^2 = 0$.

- In particular, if we expand α in the wedge product convention introduced above, we have

$$\alpha(X) = \frac{1}{(r-1)!} X^i \alpha_{ij\ldots k}\, dx^j \wedge \ldots \wedge dx^k.$$

  The $1/r!$ is canceled down to $1/(r-1)!$ because there are r separate terms in the contraction, from the vector being contracted with each one-form.

- The interior product is a 'graded derivation'. If β is a p-form, then

$$(\beta \wedge \alpha)(X) = \beta(X) \wedge \alpha + (-1)^p \beta \wedge \alpha(X)$$

  as can be shown by expanding in components.

All of our results generalize directly to differential form fields.

- As seen earlier, $(0, r)$ tensor fields can always be pulled back, so differential forms can be pulled back. Since pullback distributes over tensor products, we also have

$$f^*(\alpha \wedge \beta) = f^*\alpha \wedge f^*\beta.$$

- A differential form at a point with tangent space V can be restricted to a subspace W simply by restricting its arguments to lie in W. Geometrically, we replace the curves representing the differential form with their intersection with W. If W has lower dimension than the rank of the form, the result is automatically zero; we can also get a zero result if W is 'parallel' to the differential form.

- The same reasoning applies for differential form fields, which can be restricted to submanifolds S by performing this operation at every tangent space. More formally, this is pullback by the inclusion map. This also provides some intuition for why vectors can't be pulled back: they can 'stick out' of the submanifold.

6.3 Volume Forms

First, we define orientability.

- Consider a nonzero n-form ω defined at a point. Then ω acts on a basis $\{e_i\}$ and gives a nonzero number. We say the basis is right-handed if the number is positive. We define an oriented atlas to be one containing only coordinate systems with the same handedness; then the Jacobian determinant for switching between coordinate patches is always positive.
- This definition clearly depends on the form used, but the *classes* are invariant: two bases either have the same handedness or don't.
- Given a smooth, nowhere-vanishing n-form ω, we can also globally define the orientation across the entire manifold. Manifolds that admit such an n-form are called orientable.
- Alternatively, we can flip these definitions and think of an orientation as being specified by an oriented atlas, which then defines a volume form.

Next, we define the integral of an n-form over a region.

- Consider an n-form ω in a region with coordinates x^i, so that

  $$\omega = f(x)\, dx^1 \wedge \ldots \wedge dx^n.$$

  Note that our earlier convention would have had $\omega_{1\ldots n}/n!$ in place of $f(x)$. Our new convention is nice because there's only one independent component of ω anyway.
- Now consider a small cell spanned by the vectors $\Delta x^i\, \partial/\partial x^i$. Intuitively, the volume of the cell according to the form is given by the form acting on these vectors,

  $$\Delta x^1 \ldots \Delta x^n\, \omega(\partial/\partial x^1, \ldots, \partial/\partial x^n) = f(x) \Delta x^1 \ldots \Delta x^n.$$

  Therefore, over a region R of the manifold, we are motivated to define

  $$\int_R \omega = \int_R f(x^1, \ldots, x^n)\, dx^1 \ldots dx^n$$

  where the right-hand side is an ordinary integral from calculus performed in $\mathbb{R}^n$.
- Making all the maps explicit, our definition really says

  $$\int_R \omega = \int_{\phi(R)} (\phi^*)^{-1} \omega$$

  where ϕ is a coordinate chart mapping R into $\mathbb{R}^n$.
- As a check, we consider coordinate transformations. On the right-hand side, we know from calculus that dx^i factors pick up a Jacobian. Therefore the corresponding wedge product of differential forms should pick up the same Jacobian, and indeed

  $$dx^1 \wedge \ldots \wedge dx^n = \sum_\sigma dy^{\sigma(1)} \wedge \ldots \wedge dy^{\sigma(n)}\, J_{i,\sigma(i)} = dy^1 \wedge \ldots \wedge dy^n \sum_\sigma |\sigma| J_{i,\sigma(i)}$$

  where $J_{i,j} = \partial x^i / \partial y^j$ and the sum on the right-hand side gives J as desired.

- Despite this, there is still an arbitrary choice in the definition, from the identification

$$dx^1 \wedge \ldots \wedge dx^n \to dx^1 \ldots dx^n.$$

  The left-hand side is anticommutative, while the right-hand side is commutative. Then the definition can be changed up to a sign, which is equivalent to a choice of positive orientation (conventionally called 'right-handed') on the manifold.

- The above definition holds for a region contained in a single chart on the manifold. For larger regions, there's no problem as long as the manifold is orientable, as we can transfer the choice of orientation across the overlaps.

- The above procedure only defines the integral of top-dimensional forms. It also thereby defines the integral of a scalar function f, as the integral of the top-dimensional form $f\omega$. We can also define the integral of an r-form over an r-dimensional submanifold by pulling the form back to it. However, we generally can't define the integral of an r-form over an s-dimensional submanifold for $r \neq s$ because we would have to pick arbitrary tangent vectors to plug in.

Note. It's also possible to define integration on a nonorientable manifold. The key reason that differential forms require an orientation is that they give signed quantities, and the sign must be defined. For example, even a simple one-dimensional area integral $\int f\,dx$ gives a signed area and requires an orientation of the real line. A related object called a density can be used to find unsigned quantities, such as volume and arc length, and doesn't require orientation.

Note. Orientation is present in ordinary calculus, though hidden. For example, consider

$$I = \int_0^1 dx_1 \int_0^1 dx_2.$$

Under the substitution $(y_1, y_2) = (x_2, x_1)$, we pick up a factor of $J = -1$, flipping the sign of the result. The reason is that the region we're integrating over is now negatively oriented, so to get back to a positive orientation we need another sign flip.

Thus it was always necessary to define an orientation to evaluate I. To avoid explicitly mentioning it, we always integrate over positively oriented regions, tacitly flipping negatively orientated regions as needed, and then multiply by $|J|$ instead of J. The exception is the case of one-dimensional integrals, where the orientation is obvious: $[a, b]$ is positively oriented if $a < b$. We thus abandon the convention and allow the Jacobian factor du/dx to be negative. These contradictory ad hoc rules in ordinary calculus are unified in the more geometric formulation we have here.

Note. Orientation for submanifolds. Given an n-form on $M \supset S$ that is nonvanishing on S, we can define a volume form for S by choosing normal vectors $n_1, \ldots, n_{n-p}$ continuously and plugging those into the form. (The result clearly depends on the choice of the n_i.) If the n_i can be chosen nonvanishing, then S has a nonvanishing volume form, and we say we have given it an 'external orientation'. 'Internal orientability' of S is just the usual notion of orientability, where we forget about M.

- If M is orientable, one can show internal and external orientability of S are equivalent.

- More generally, external orientability implies internal orientability, but not vice versa. For example, a closed curve is always internally orientable, but a circle drawn around a Mobius strip is not externally orientable. However, a small circle on a Mobius strip is externally orientable because it doesn't 'feel' the twist.

- If M is orientable and has a boundary ∂M, then ∂M is always canonically internally/externally orientable; we simply plug the outward normal into the volume form.
- Note that we can't simply pull the volume form back to S, because we need a p-form, not an n-form. Since $n > p$, the pullback of the volume form is identically zero.

Note. Generally, a manifold cannot be covered by a single chart, so we need a little more machinery to define integration. The idea is to split the manifold using a "partition of unity", turning an integral over the manifold into a sum of integrals, each of which takes place on a single chart.

6.4 Duals of Forms

Next, we define duals of differential forms. A p-vector is a totally antisymmetric $(p, 0)$ tensor; they may be constructed by the exterior product just like p-forms and form an algebra as well. To distinguish vectors and forms, we write vector names in bold.

- Given a volume form ω, we can associate a p-vector with an n-p-form by

$$A = *\mathbf{T}, \quad A_{j,\dots l} = \frac{1}{p!}\omega_{i\dots kj\dots l}T^{i\dots k}.$$

- To go backwards, define the N-vector $\boldsymbol{\omega}$ by

$$\omega^{i\dots k}\omega_{i\dots k} = n!, \quad \omega^{1\dots n}\omega_{1\dots n} = 1.$$

  Then we can analogously define the dual of a p-form by

$$\mathbf{S} = *B, \quad S^{i\dots k} = \frac{1}{p!}\omega^{l\dots mi\dots k}B_{l\dots m}.$$

- The normalizing factors are chosen so that when the vectors and forms are written in terms of wedge products (and the equivalent for vectors), there are no extra numerical factors. For example, in 4D and coordinates where $\omega^{i\dots j} = \epsilon^{i\dots j}$, $*(dx^1 \wedge dx^2) = \partial_3 \wedge \partial_4$.
- It is convenient to define the Levi–Civita symbol

$$\epsilon_{i\dots k} = \epsilon^{i\dots k} = \operatorname{sign}(i \dots k).$$

  Note that the Levi–Civita symbol is not a tensor, as we have defined it to have the same components in all coordinates. It's simpler to write the components of volume forms with this symbol, as it 'factors out' the antisymmetry,

$$\omega_{i\dots k} = f\epsilon_{i\dots k}, \quad \omega^{i\dots k} = \frac{1}{f}\epsilon^{i\dots k}.$$

- Define the p-delta symbol as

$$\delta^{i\dots j}_{k\dots \ell} = p!\,\delta^i_{[k}\dots\delta^j_{\ell]}.$$

  The interpretation of the symbol is as follows: the $p!$ is present to cancel the $1/p!$ in the antisymmetrization. Then we have all possible delta functions linking $i \dots j$ with a permutation of $k \dots \ell$, with appropriate signs, so

$$\delta^{i\dots j}_{k\dots \ell} = |\sigma| \text{ if } (i, \dots j) = \sigma(k, \dots, \ell).$$

  In particular, the antisymmetrization ensures that all of the indices $(k, \dots, \ell)$ must be distinct.

- As a result, the product of Levi–Civita symbols is an n-delta,

$$\epsilon_{i\dots k}\epsilon^{\ell\dots r} = \delta^{\ell\dots r}_{i\dots k}.$$

  This is because the product on the left-hand side is only nonzero if the $(i,\dots,k)$ are distinct, and the $(\ell,\dots,r)$ are distinct as well. But then $(\ell,\dots,r)$ must be a permutation of $(i,\dots k)$, because the Levi–Civita symbol has n indices. The sign on the left-hand side is positive if $(i,\dots k)$ and $(\ell,\dots,r)$ have the same sign as permutations of $(1,\dots,n)$, which means that the permutation linking them is even.

- Similarly, the general formula for the contraction of Levi–Civita symbols is

$$\epsilon_{i\dots jk\dots\ell}\epsilon^{i\dots jm\dots r} = (n-p)!\,\delta^{m\dots r}_{k\dots l}$$

  where $n-p$ is the number of contracted indices.

Example. We can simply read off identities for Levi–Civita contraction in three dimensions.

$$\epsilon_{ijk}\epsilon^{imn} = \delta^m_j\delta^n_k - \delta^m_k\delta^n_j, \quad \epsilon_{ijk}\epsilon^{ijn} = 2\delta^n_k, \quad \epsilon_{ijk}\epsilon^{ijk} = 6.$$

It is similarly easy to get the coefficients in four dimensions.

Example. Taking the dual twice gives back the original form/vector, up to a sign. For a p-form B, we have

$$(**B)_{j\dots\ell} = \frac{1}{(n-p)!\,p!}\,\omega_{i\dots kj\dots\ell}\omega^{r\dots s\,i\dots k}B_{r\dots s} = \frac{(-1)^{p(n-p)}}{(n-p)!\,p!}\,\epsilon_{i\dots kj\dots\ell}\epsilon^{i\dots kr\dots s}B_{r\dots s}$$

where we used antisymmetry and canceled factors of f. Applying our earlier identities, we get

$$(**B)_{j\dots\ell} = \frac{(-1)^{p(n-p)}}{p!}\delta^{r\dots s}_{j\dots\ell}B_{r\dots s} = (-1)^{p(n-p)}B_{j\dots\ell}$$

where in the second step we used the fact that the contraction generates $p!$ identical terms.

Example. The determinant can be simply written as

$$|A| = \epsilon_{i\dots k}A^{1i}\dots A^{nk} = \frac{1}{n!}\epsilon_{a\dots c}\epsilon_{i\dots k}A^{ai}\dots A^{ck}.$$

More formally, any linear operator $T\colon V\to V$ defines an induced map on the space of top-dimensional multivectors; it turns out to be multiplication by $\det T$.

Example. In cases where there are multiple n-forms, it can be useful to relate all forms to the coordinate-dependent n-form $dx^1\wedge\ldots\wedge dx^n$, which by construction has components $\epsilon_{i\dots k}$ in all coordinate systems. Then we can write a general n-form as

$$\omega = \mathfrak{w}dx^1\wedge\ldots\wedge dx^n$$

in all coordinate systems. Under a coordinate transformation, we know that $dx^1\wedge\ldots\wedge dx^n$ should pick up a Jacobian factor J for ω to stay invariant, so we must have

$$\mathfrak{w}' = J\mathfrak{w}.$$

We say that $\mathfrak{w}$ is a scalar density of weight 1. In general, a tensor density of weight k is simply a quantity that transforms like a tensor, with an extra factor of J^k. For example, the coordinate-dependent form $dx^1\wedge\ldots\wedge dx^n$ could be regarded as a rank n tensor of weight -1. Ordinary tensors are densities with weight zero.

Example. Metric volume forms. When a metric g is provided, there is a preferred volume form

$$\omega = \omega^1 \wedge \ldots \wedge \omega^n$$

where the ω^i are an orthonormal basis. (Note that this definition is ambiguous up to orientation, as usual.) Now, in matrix form, the metric transforms as

$$g \to J^T g J$$

where J is the Jacobian matrix. Taking determinants, the metric determinant g is a scalar density of weight 2, and thus the metric volume form is, explicitly,

$$\omega = |g|^{1/2}\, dx^1 \wedge \ldots \wedge dx^n$$

where we have taken the absolute value because g can be negative for Lorentzian signature. This volume form is also sometimes called the Levi–Civita tensor.

Example. The cross product in three dimensions is

$$\mathbf{U} \times \mathbf{V} = *(U \wedge V)$$

where $\mathbf{U}$ and $\mathbf{V}$ are regular 1-vectors and U and V are the corresponding 1-forms.

This explains the strange transformation behavior of cross products. In the 'passive' view, consider flipping the sign of a basis vector. The vectors $\mathbf{U}$ and $\mathbf{V}$ are invariant, but their cross product picks up a sign because the volume form flips sign, as the orientation of the basis has flipped. In the 'active' view, consider negating both $\mathbf{U}$ and $\mathbf{V}$. Then the cross product receives two sign flips, so it stays the same.

Using a metric volume form, we can define a duality between p-forms and n-p-forms.

- Define the Hodge dual/star $\star\alpha$ of a p-form α to satisfy

  $$\beta \wedge \star\alpha = \langle \beta, \alpha \rangle \omega$$

  where ω is the metric volume form, and the inner product is taken with the metric. It is equivalent to taking the dual $*$ defined earlier, then lowering all the indices with the metric.

- To compute the Hodge star more easily, let $e^1, \ldots, e^n$ be a basis of one-forms satisfying

  $$\langle e^\mu, e^\nu \rangle = \delta^{\mu\nu} \epsilon(\mu), \quad \epsilon(\mu) = \pm 1.$$

  Then given a permutation $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$,

  $$\star(e^{i_1} \wedge \ldots \wedge e^{i_p}) = \text{sign}(i_1, \ldots, i_n) \epsilon(i_1) \cdots \epsilon(i_p).$$

  As a simple example, in $\mathbb{R}^3$ this implies $\star dx = dy \wedge dz$, and $\star(dy \wedge dz) = dx$.

- More generally, $\star^2 = \pm 1$. We pick up a minus sign for each of the s minus signs in the metric. Moreover, the second time we take the dual, the permutation used is related to the original one by $p(n-p)$ transpositions. Therefore

  $$\star^2 = (-1)^{p(n-p)+s}.$$

We will generally stick to using the dual, but the Hodge star will also be useful.

6.5 The Exterior Derivative

Since we have defined integration for forms, we now turn to defining differentiation. We already know how to turn a 0-form into a 1-form, namely by the gradient $f \to df$. The exterior derivative generalizes this to general forms.

- We define the exterior derivative to take p-forms to $(p+1)$-forms, satisfying the properties

$$d(\beta+\gamma) = d\beta + d\gamma, \quad d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta, \quad d^2 = 0$$

  where α is a p-form. The first two properties are just linearity and a modified Leibniz rule that accounts for anticommutation.

- These properties, in addition to the existing definition of df, uniquely specify the exterior derivative. In coordinates, the result is

$$\alpha = \frac{1}{p!}\alpha_{j\ldots k}\, dx^j \wedge \ldots \wedge dx^k, \quad d\alpha = \frac{1}{p!}\partial_i \alpha_{j\ldots k}\, dx^i \wedge dx^j \wedge \ldots \wedge dx^k$$

  or alternatively

$$(d\alpha)_{ij\ldots k} = (p+1)\partial_{[i}\alpha_{j\ldots k]}.$$

- Heuristically, the exterior derivative is "$d = \partial\wedge$". The result $d^2 = 0$ just follows by antisymmetry: partial derivatives commute, but wedge products anticommute.

- The exterior derivative, like the Lie derivative and covariant derivative, generalizes the partial derivative to a map from tensors to tensors. In the 1-form case, it works because the 'error' term from the partial derivative is symmetric, so the antisymmetrization removes it. While d requires no additional structure to define, the downside is it only works on differential forms.

- The exterior derivative d commutes with pullbacks,

$$f^*(d\alpha) = d(f^*\alpha).$$

  The proof is by induction; the base case (α is a scalar) is just the definition of the pullback map. Intuitively this statement is clear given our geometric intuition for d.

- One coordinate-free identity the exterior derivative satisfies is

$$d\omega(X,Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]).$$

  This can also serve as a starting point for defining d without ever invoking coordinates.

Example. The exterior derivative generalizes familiar operations from calculus. In three dimensional Euclidean space, let $\mathbf{a}$ be a vector field and let a be the corresponding one-form. Then

$$*da = *(\partial_i a_j) dx^i \wedge dx^j = (\partial_i a_j)\epsilon^{ijk}\partial_k = \nabla \times \mathbf{a}.$$

Similarly, for the divergence we have

$$d * \mathbf{a} = d(a^i \epsilon_{ijk} dx^j \wedge dx^k) = (\partial^\ell a^i)\epsilon_{ijk} dx^\ell \wedge dx^j \wedge dx^k = (\partial_\ell a^i)\epsilon_{ijk}\epsilon^{\ell jk}\omega = (\partial_i a^i)\omega$$

where ω is the metric volume element. Taking the dual of both sides gives

$$*d * \mathbf{a} = \nabla \cdot \mathbf{a}.$$

The results $\nabla \cdot \nabla \times \mathbf{a} = 0$ and $\nabla \times \nabla f = 0$ are immediate.

Example. Consider the partial differential equation

$$\frac{\partial f}{\partial x} = g(x,y), \quad \frac{\partial f}{\partial y} = h(x,y).$$

In differential form notation, this states that $df = a$ for some one-form a. Then if a solution for f exists, then we must have $da = 0$, i.e. zero curl, as we know from multivariable calculus. Working in the manifold $\mathbb{R}^n$, the converse is also true: if $da = 0$, then a solution for f exists.

We now discuss closed and exact forms.

- The property $d^2 = 0$ is analogous to the property $\partial^2 = 0$ we saw in homology. Thus we define a form α to be closed if $d\alpha = 0$ (in analogy with cycles) and a form α to be exact if $\alpha = d\beta$ (in analogy with boundaries). All exact forms are closed.
- To understand this geometrically, note that in our visualization scheme, the exterior derivative takes the surfaces of α to their boundaries. Then a differential form if closed if the surfaces forming its visual representation are closed (e.g. closed contour lines for one-forms).
- Not all closed forms are exact. For example, consider $\mathbb{R}^2 \setminus \{0\}$ and

  $$\alpha = \frac{x\,dy - y\,dx}{x^2 + y^2} = d\theta.$$

  It is defined everywhere and obeys $d\alpha = 0$, but it is not exact. (In particular, θ doesn't count because it's multivalued.)
- The failure of closed forms to be exact is due to topological obstructions. Thus we expect that in a small enough neighborhood, closed forms are always locally exact.

Lemma (Poincare)**.** Closed forms are locally exact. Specifically, if $d\alpha = 0$ for a p-form α in a region U of M diffeomorphic to the unit open ball, then we can write $d\alpha = \beta$ in this region.

Proof. Since the exterior derivative and pullback commute, we map to the unit open ball and construct β there. Let

$$\alpha = \alpha_{i\ldots k}(\mathbf{x})\, dx^i \wedge \ldots \wedge dx^k.$$

The idea is that there should be nothing 'in the way' of constructing β. In multivariable calculus, we can construct β by just integrating α over any desired path, e.g. along straight lines parallel to the axes. In this case, it's most convenient to integrate α from the origin, i.e.

$$\beta_{j\ldots k}(\mathbf{x}) = \int_0^1 t^{p-1} \alpha_{ij\ldots k}(t\mathbf{x})\, x^i\, dt.$$

This is the Volterra formula. We now show that $\alpha = d\beta$. We have

$$d\beta_{i\ldots k} = p\partial_{[i}\beta_{j\ldots k]}, \quad \partial_i \beta_{j\ldots k} = \int_0^1 t^{p-1} \alpha_{ij\ldots k}(t\mathbf{x})\, dt + \int_0^1 t^p x^\ell \partial_i \alpha_{\ell j\ldots k}(t\mathbf{x})\, dt.$$

We still must antisymmetrize the $i\ldots k$ indices. We use the fact that α is closed. In components, this means $\partial_{[k}\alpha_{i,\ldots j]} = 0$. In the case where α is a 2-form, we can expand this out for

$$\partial_{[k}\alpha_{ij]} = \partial_k \alpha_{[ij]} + \partial_{[i}\alpha_{j]k} + \partial_{[j}\alpha_{|k|i]} = \partial_k \alpha_{[ij]} + 2\partial_{[i}\alpha_{j]k}$$

where the bar denotes exclusion from the antisymmetrization. More generally, we have

$$\partial_\ell \alpha_{[i\dots k]} = p\partial_{[i}\alpha_{|\ell|j\dots k]}.$$

This allows us to 'swap the indices' on the second term above, so that both terms involve $\alpha_{[i\dots k]}$,

$$d\beta_{i\dots k} = \int_0^1 (pt^{p-1} + t^p x^\ell \partial_\ell)\alpha_{ij\dots k}(t\mathbf{x})\,dt$$

where we have dropped the $i\dots k$ antisymmetrization as it now does nothing. Now, this quantity is just the total time derivative of $t^p\alpha_{i\dots k}(tx^1,\dots,tx^n)$. Integrating gives $\alpha_{i,\dots k}(\mathbf{x})$, as desired.

Next, we consider how the Lie derivative and exterior derivative interact.

- Cartan's magic formula states that

  $$\mathcal{L}_\mathbf{V}\omega = d(\omega(\mathbf{V})) + (d\omega)(\mathbf{V}) = (i_X \circ d + d \circ i_X)\,\omega$$

  for any p-form ω and vector field $\mathbf{V}$. This is natural, in the sense that the right-hand side contains the only p-forms that can be constructed using d, ω, and one power of $\mathbf{V}$.

- For a zero-form f, the function reads

  $$\mathcal{L}_\mathbf{V} f = (df)(\mathbf{V})$$

  which is true, as both sides are $V^i\partial_i f$.

- For a one-form, the result can be obtained by direct expansion. We have

  $$d(\omega(\mathbf{V})) = d(\omega_i V^i) = \partial_j(\omega_i V^i)dx^j, \quad d\omega = \partial_j\omega_i dx^j \wedge dx^i = \partial_j\omega_i(dx^j \otimes dx^i - dx^i \otimes dx^j).$$

  Contracting the latter with $\mathbf{V}$ and simplifying gives the result.

- For general forms, the result can be proven by induction. It suffices to prove the result for a form of the form $\omega = fa \wedge b$ by linearity, where the result holds for a and b by the inductive hypothesis. The proof can be completed by using simple properties of d and $\mathcal{L}_\mathbf{V}$.

- Using Cartan's formula twice gives the important result

  $$\mathcal{L}_\mathbf{V}(d\omega) = d(\mathcal{L}_\mathbf{V}\omega).$$

  That is, the exterior derivative and Lie derivative commute. This is natural because the exterior derivative essentially commutes with any smooth map. Using our geometric intuition, it's computed by taking the boundaries of surfaces, which can be done before or after a map.

- Alternatively, we can run the steps backwards, showing that $\mathcal{L}_\mathbf{V}$ and $i_X \circ d + d \circ i_X$ are both linear derivations that commute with d and agree on functions. To show that $\mathcal{L}_\mathbf{V}$ commutes with d, it suffices to show pullback commutes with d, which can be done in components.

Example. Moser's theorem. For a closed manifold M and two volume forms ω_0 and ω_1 with the same total volume, there exists a diffeomorphism $\psi\colon M \to M$ with $\psi^*\omega_1 = \omega_0$ where ψ is isotopic to the identity, where isotropy is the equivalent of homotopy in differential geometry.

To see this, let ω_t linearly interpolate between ω_0 and ω_1. Then ω_t is always a valid volume form because ω_0 and ω_1 always have the same sign at every point. We would like to define a family of diffeomorphisms ψ_t so that $\psi_t^*\omega_t = \omega_0$. There is an associated time-dependent vector field X_t describing the 'velocity' of every point, with

$$\frac{d}{dt}(\psi_t^*\omega_t) = \psi_t^*(\mathcal{L}_{X_t}\omega_t + \frac{d\omega_t}{dt}) = \psi_t^*(d(i_{X_t}\omega_t) + (\omega_0 - \omega_1))$$

where we used Cartan's formula. Now note that $\omega_0 - \omega_1 = d\alpha$ so

$$\frac{d}{dt}(\psi_t^*\omega_t) = \psi_t^*(d(i_{X_t}\omega_t + \alpha)).$$

Then $\psi_t^*\omega_t = \omega_0$ for all t if $i_{X_t}\omega_t + \alpha = 0$, where α is known. Now, the map $X \mapsto i_X\omega$ is an isomorphism if ω_t is a volume form. Then there is always a solution for X_t for each t, and we can define the diffeomorphisms by flow along these vector fields. This is valid on a closed manifold, where flows exist for arbitrary times, and the idea of the proof is called Moser's method.

6.6 Stokes' Theorem

In this section, we prove Stokes' theorem and examine its consequences.

Theorem (Stokes)**.** Let α be an $(n-1)$-form on an n-dimensional manifold M. Let U be a region of M with a smooth orientable boundary ∂U. Then

$$\int_U d\alpha = \int_{\partial U} \alpha.$$

Note that we should properly include the restriction of α to ∂U (by pullback under inclusion), but we suppress them for notational simplicity. Technically, we also require α to be smooth and compactly supported.

Proof. Let ω be an arbitrary n-form and let $\mathbf{V}$ be an arbitrary vector field. Let $U(\epsilon)$ be the region generated by flowing U along $\mathbf{V}$ for a parameter ϵ. We will compute

$$\frac{d}{d\epsilon}\int_{U(\epsilon)} \omega$$

in two different ways. First, we consider the motion of the boundary $\partial U(\epsilon)$. Assuming that $\mathbf{V}$ is not tangent to $\partial U(0)$, we can construct local coordinates where $\mathbf{V} = d/dx^1$ and $\partial U(0)$ is the surface $x^1 = 0$. Let $V \subset \mathbb{R}^{n-1}$ be the coordinates of $\partial U(0)$ and let $\omega = f dx^1 \wedge \ldots \wedge dx^n$. Then

$$\int_{\partial U(\epsilon) - \partial U(0)} \omega = \int_V \left[\int_0^\epsilon f\, dx^1\right] dx^2 \ldots dx^n \approx \epsilon \int_V f(0, x^2, \ldots, x^n)\, dx^2 \ldots dx^n = \epsilon \int_V \omega(\mathbf{V})$$

where we kept the restriction of $\omega(\mathbf{V})$ to ∂U implicit. Therefore

$$\frac{d}{d\epsilon}\int_{U(\epsilon)} \omega = \int_{\partial U} \omega(\mathbf{V}).$$

We can also compute the integral by thinking about how the integrand changes. To first order, the change in the integrand is $\epsilon\mathcal{L}_{\mathbf{V}}\omega$ at every point. Then

$$\frac{d}{d\epsilon}\int_{U(\epsilon)} \omega = \int_U \mathcal{L}_{\mathbf{V}}\omega.$$

Applying Cartan's formula gives the result

$$\int_U d(\omega(\mathbf{V})) = \int_{\partial U} \omega(\mathbf{V}).$$

Finally, since ω and $\mathbf{V}$ are arbitrary, we can set $\alpha = \omega(\mathbf{V})$, concluding the proof.

Note. Using Stokes' theorem twice tells us that $\partial^2 = 0$ if and only if $d^2 = 0$. This link will be made formal when we introduce singular homology, where ∂ becomes the boundary operator.

Example. In two dimensions, given a one-form α, Stokes' theorem becomes

$$\int (\partial_y \alpha_x - \partial_x \alpha_y)\, dxdy = \int_{\partial U} \alpha_i dx^i$$

which is the usual Stokes' theorem. In three dimensions, a similar calculation gives the three-dimensional Stokes' theorem.

Example. The divergence theorem. Given a volume form ω, we define the divergence of $\mathbf{V}$ to be

$$\nabla \cdot_\omega \mathbf{V} = * d * \mathbf{V}$$

where the dual is with respect to ω. We can also write this in terms of ω explicitly,

$$\omega \nabla \cdot_\omega \mathbf{V} = d(\omega(\mathbf{V})).$$

Using $\omega(\mathbf{V})$ as the differential form in Stokes' theorem gives

$$\int_U (\nabla \cdot_\omega \mathbf{V})\, \omega = \int_{\partial U} \omega(\mathbf{V})$$

which is the divergence theorem. To make this more familiar, let $\omega = n \wedge \alpha$ where n is a one-form normal to ∂U, i.e. $n(\boldsymbol{\eta}) = 0$ for any $\boldsymbol{\eta}$ on ∂U. Then the surface integrand is $n(\mathbf{V})\, \alpha$, which reduces to the familiar $(\hat{n} \cdot \mathbf{V})\, dS$ in coordinates.

Note. The decomposition $\omega = n \wedge \alpha$ says that ω measures volumes by letting α measure surface area and n measure distance along the normal. The forms α and n are not unique, as they can be scaled by reciprocal factors. But if we're using a metric volume form, we can canonically scale n so that $n(\mathbf{n}) = 1$, as is done in multivariable calculus.

Example. Divergence in curvilinear coordinates. If $\omega = f dx^1 \wedge \ldots \wedge dx^n$, then we can show

$$\nabla \cdot_\omega \mathbf{V} = \frac{1}{f} \partial_i (f V^i).$$

For example, in spherical coordinates, the Euclidean volume form becomes

$$\omega = r^2 \sin\theta\, dr \wedge d\theta \wedge d\phi$$

from which we can easily read off the divergence.

Note. It is essential that the differential forms be compactly supported. As a simple example, consider integrating the form $d(rd\theta) = dr \wedge d\theta$ on the annulus $1 \leq r \leq 2$. Stokes' theorem will give two boundary terms. However, if we change the annulus to $1 < r \leq 2$ or $1 \leq r < 2$, the manifold will lose one of its boundaries.

While it's obvious here that a boundary is 'missing', our proposed shapes are homeomorphic to a punctured circle, and an annulus that stretches to infinity. In this case, there still are 'missing' terms, though they're harder to see, so Stokes' theorem still fails.

6.7 de Rham Cohomology

To understand de Rham cohomology, we first need to introduce singular homology.

- So far, we have shown how to integrate an n-form over an n-dimensional region of an n-dimensional manifold M. We extended this definition to integrating r-forms on M over r-dimensional submanifolds by pulling the form back to the submanifold by inclusion.
- However, for physical purposes we need a more general definition. For instance, if we wanted to compute the work done on a particle, we would have to integrate over its path, but this path may cross itself, retrace or repeat itself, or temporarily stop.
- More concretely, we wish to define an integral over the map $f\colon I \to M$ where $I = [0,1]$ is an interval and f represents the path of the particle. The obvious answer is
$$\int_f \omega = \int_0^1 dt\, \omega_\mu(x(t)) \frac{dx^\mu}{dt}$$
which motivates the definition
$$\int_f \omega = \int_I f^*\omega.$$
- More generally, we call a smooth map $\sigma\colon I^r \to M$ a singular r-cube. Here I^r is a cube in $\mathbb{R}^r$, and σ need not be injective, and σ_* need not have maximal rank. In general any standard region in $\mathbb{R}^r$ works; for instance one may also use a "singular simplex". We then define
$$\int_\sigma \omega = \int_{I^r} \sigma^*\omega$$
for an r-form ω.
- We may take real linear combinations of the singular r-cubes to yield r-chains,
$$c^r = \sum_i a_i \sigma_i^r.$$
Integration is defined over r-chains by linearity,
$$\int_{c^r} \omega = \sum_i a_i \int_{\sigma_i^r} \omega.$$
The set of r-chains is called the r^{th} chain group $C_r(M)$. It is an abelian group under addition.
- The singular chain groups defined here are the analogues of the chain groups in simplicial homology, and we will use them to define singular homology. We started with simplicial homology because it could be straightforwardly computed. However, singular homology is nicer on general manifolds because it doesn't require constructing a triangularization. On the other hand, the chain group is now extremely large.
- To define the boundary operator ∂ for r-chains, we focus on r-cubes and extend by linearity. The boundary of an r-cube is the sum of the $2r$ $r-1$-cubes defined by each face, where the orientation is defined using the outward normal vector. It is straightforward to show that $\partial^2 = 0$, and to establish Stokes' theorem
$$\int_{\partial c} \omega = \int_c d\omega.$$

- We thus define the chain complex

$$\ldots \xrightarrow{\partial_{r+2}} C_{r+1}(M) \xrightarrow{\partial_{r+1}} C_r(M) \xrightarrow{\partial_r} C_{r-1}(M) \xrightarrow{\partial_{r-1}} \ldots$$

  and the cycle, boundary, and singular homology groups

$$Z_r(M) = \ker \partial_r, \quad B_r(M) = \operatorname{im} \partial_{r+1}, \quad H_r(M) = \frac{Z_r(M)}{B_r(M)}.$$

  One can show that the singular and simplicial homology groups are equivalent, but this is beyond the tools we have.

We now turn to de Rham cohomology.

- The exterior derivative yields a cochain complex

$$0 \to \Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \Omega_2(M) \xrightarrow{d_2} \ldots$$

  where $d^2 = 0$. The closed and exact forms yield the cocycle and coboundary groups

$$Z^r(M) = \ker d_r, \quad B^r(M) = \operatorname{im} d_{r-1}.$$

  We define the r^{th} de Rham cohomology group

$$H^r(M) = Z^r(M)/B^r(M)$$

  which measures the failure of the cochain complex to be exact at $\Omega^r(M)$.

- By the Poincare lemma, the open ball has trivial cohomology (except for the zeroth cohomology group, as we'll see below), as does any space homeomorphic to it.

- An r-chain acts on an r-form, or vice versa, by

$$(\omega, c) \equiv \int_c \omega.$$

  This suggests that $\Omega^r(M)$ and $C_r(M)$ are dual, though this is a bit difficult to make precise. Moreover, Stokes' theorem states $(d\omega, c) = (\omega, \partial c)$ which implies d is the pullback of ∂.

- By Stokes' theorem, the cycles $Z_r(M)$ annihilate the coboundaries $B^r(M)$, while the boundaries $B_r(M)$ annihilate the cocycles $Z^r(M)$. Now we consider the space dual to $H_r(M)$. An element $[c]$ of $H_r(M)$ is an r-cycle defined up to the addition of an r-boundary. This latter implies that, for the integral to be well-defined, it can only act on a cocycle. But nothing changes if we add a coboundary to this cocycle. Hence the dual space to $H_r(M)$ should be cocycles defined up to the addition of coboundaries, i.e. $H^r(M)$.

- The above is not a proof, since we haven't shown that independent elements of $H^r(M)$ yield independent linear maps on $H_r(M)$. However, it serves as motivation for de Rham's theorem,

$$H_r(M)^* = H^r(M).$$

  Proving it rigorously is well beyond our scope.

- Comparing homology and cohomology, we note that forms map out of M, while chains map into M. The boundary operators d and ∂ also go in opposite directions. A deeper difference is that, as we'll see later, the cohomology groups may be given a ring structure.

For concreteness, we now consider some cohomology computations.

Example. The zeroth cohomology group. We define $\Omega^{-1}(M)$ to be empty, so $B^0(M)$ is trivial. Then $Z^0(M) = H^0(M)$, so the group consists of functions f such that $df = 0$. Then f is constant on every connected component of M, so $H^0(M) = \mathbb{R}^n$ where n is the number of connected components of M. This exhibits de Rham's theorem for $r = 0$.

Example. The first cohomology group. Consider a closed one-form α integrated over an arbitrary closed chain C. Furthermore, suppose that for all such closed chains are boundaries, $C = \partial S$. Then

$$\int_C \alpha = \int_S d\alpha = 0$$

which implies that the integral of α between two points is path independent. This implies α is exact, where we define f with $df = \alpha$ by integration. This is consistent with de Rham's theorem for $r = 1$.

Another simple example is $M = S^1$, where $H^1(S^1) = \mathbb{R}$. The representative one-form is "$d\theta$", which is not exact since θ is not a function.

Example. The second cohomology group of $M = S^2$. For any exact two-form $\alpha = d\beta$, we have

$$\int_M \alpha = \int_{\partial M} \beta = 0$$

because $\partial M = 0$. However, the closed two-form $\omega = \sin\theta d\theta \wedge d\phi$ has a nonzero integral, so it cannot be exact, and the second cohomology group is nontrivial.

Conversely, we claim that any two-form α with zero integral over S^2 is exact. To see this, note that we can define β_+ and β_- so that $\alpha = d\beta_\pm$ on the northern and southern hemispheres by the Poincare lemma. We would like to stitch them together; note that along the equator C,

$$\int_C \beta_+ - \beta_- = 0$$

by Stokes' theorem. Therefore $\beta_+ - \beta_- = df$ on the equator by the previous example. Extending f to the northern hemisphere smoothly, we can define $\beta'_+ = \beta_+ - df$. Then combining β'_+ and β_- gives the desired form, so $H^2(S^2) = \mathbb{R}$. Alternatively, in terms of singular homology, S^2 itself is the 2-cycle that is not a boundary, so $H_2(S^2) = \mathbb{R}$.

Note. We've already seen very similar results in multivariable calculus. We know that for topologically trivial spaces, every curl-free vector function is a gradient, and every divergence-free vector function is a curl. These correspond to the triviality of the first and second cohomology groups.

A physical example is the two-form magnetic field of a monopole on $\mathbb{R}^3 - \{0\}$, which cannot be written as $B = dA$. However, it can be written in this form for $\mathbb{R}^3$ minus a Dirac string, a space with trivial second cohomology group.

Next we consider the structure of cohomology groups.

- Choose a basis $[e_i]$ of $H_r(M)$, where the e_i are r-cycles up to the addition of r-boundaries. Then any $z \in Z_r(M)$ can be written as

$$z = \sum_i c_i e_i + \partial c.$$

Similarly, if we choose a basis $[\theta_i]$ of $H^r(M)$, then any $\omega \in Z^r(M)$ can be written as

$$\omega = \sum_i a_i \theta_i + d\psi.$$

- Now, invoking de Rham's theorem, we may choose $[\theta_i]$ to be the dual basis of $[e_i]$,

$$([\theta_i], [e_j]) = (\theta_i, e_j) = \delta_{ij}.$$

The coefficients a_i can hence be found by integration,

$$a_i = (\omega, e_i).$$

Hence the closed form ω is exact if and only if $(\omega, e_i) = 0$ for all e_i, a convenient criterion.

- Let $\Omega(M)$ be the ring of differential forms on M, where the product is the wedge product. We may similarly define a cohomology ring

$$H^*(M) = H^0(M) \oplus \ldots \oplus H^m(M)$$

where the product, called the cup product, is induced by the wedge product,

$$[\omega] \wedge [\phi] = [\omega \wedge \phi].$$

To check this definition is consistent, first note that $\omega \wedge \phi$ is closed,

$$d(\omega \wedge \phi) = d\omega \wedge \phi + (-1)^r \omega \wedge d\phi = 0.$$

Next, note that under the addition of an exact form to ω,

$$(\omega + d\psi) \wedge \phi = \omega \wedge \phi + (d\psi) \wedge \phi = \omega \wedge \phi + d(\psi \wedge \phi)$$

so we add an exact form to $\omega \wedge \phi$. Similar reasoning holds for the addition of an exact form to ϕ, giving the result.

- Now consider a map $f \colon M \to N$ between manifolds. We may pullback forms and pushforward chains, and these turn out to be equivalent in the sense that

$$\int_{f_* c} \omega = \int_c f^* \omega.$$

Moreover, the pullback map induces a map on the cohomology rings, defined by

$$f^*[\omega] = [f^*\omega].$$

It is well defined because d and f^* commute, and it is a ring homomorphism because $\wedge$ and f^* commute.

- One can show that if $f, g\colon M \to N$ are homotopic, then they induce the same maps f^*, g^* on the cohomology rings. Then simple connectedness implies a trivial first cohomology group. To see this, take a closed one-form ω and any loop $c\colon S^1 \to M$. Then

$$\int_c \omega = \int_{S^1} c^*\omega.$$

  By the proposition, the integral on the right-hand side is homotopy invariant, and since M is simply connected it must be zero. But then $\int_c \omega = 0$ for all loops c, so ω is exact.

- Let M be an m-dimensional orientable closed manifold and let $\omega \in H^r(M)$ and $\eta \in H^{m-r}(M)$. Then since $\omega \wedge \eta$ is a top-dimensional form, we may define an inner product

$$\langle \omega, \eta \rangle = \int_M \omega \wedge \eta.$$

  This establishes the Poincare duality

$$H^r(M) \cong H^{m-r}(M).$$

  One immediate consequence is that the Euler characteristic of an odd-dimensional space is zero.

- For example, the n^{th} cohomology group of an n-dimensional connected orientable manifold is $\mathbb{R}$, indexed by the integral of the volume form. If the manifold is nonorientable, then the n^{th} cohomology group is trivial.

- In general, cohomology is more powerful than homology because of the additional ring structure. For example, one may distinguish spaces with the same cohomology groups if the cohomology ring differs. There are many important generalized cohomology theories, such as sheaf cohomology and K-theory.

- Note that taking the dual converts the cup product

$$H^*(M) \times H^*(M) \to H^*(M)$$

  to the map

$$H_*(M) \times H_*(M) \to H_*(M)$$

  by reversing the arrows. Hence we don't expect to have a ring structure for homology.

6.8 Physical Applications

Hamiltonian mechanics is covered in terms of differential forms in the notes on Undergraduate Physics. Here we turn to electromagnetism, an application which additionally requires a metric.

- The field strength F is a two-form, and two of Maxwell's equations are $dF = 0$. In terms of components, this tells us that

$$\partial_\sigma F_{\mu\nu} + \partial_\nu F_{\sigma\mu} + \partial_\mu F_{\nu\sigma} = 0$$

  which is called the Bianchi identity.

- The remaining two equations are $\partial_\nu F^{\mu\nu} = J^\mu$ in inertial Cartesian coordinates; the equations are not true in general due to the partial derivative. Recognizing a divergence, we have

$$\nabla \cdot_\omega \mathbf{F} = \mathbf{J}, \quad \star d \star F = J$$

  where J is the current one-form. These equations hold in all frames.

- The previous equation implies that charge is conserved, as

$$\nabla \cdot_\omega J = \star d \star J = 0.$$

  Since $\star J = d \star F$, it also gives the result

$$\int_D \star J = \int_{\partial D} \star F$$

  for any three-dimensional region D. In the case where D is purely spatial, this is Gauss's law, equating electric charge to electric flux.

- In terms of electric and magnetic fields, the Hodge star gives

$$\mathbf{B} \to \mathbf{E}, \quad \mathbf{E} \to -\mathbf{B}$$

  in vacuum. If we include sources, it must swap charge and magnetic charge.

- Alternatively, if we phrase $\mathbf{E}$ and $\mathbf{B}$ in terms of forms, then $\mathbf{E}$ is naturally a one-form and $\mathbf{B}$ is naturally a two-form, with

$$F = B + E \wedge dt.$$

  In this case, the operation of the Hodge star is

$$B \to \star_S E, \quad E \to - \star_S B$$

  where $\star_S$ is the Hodge star on space only.

Example. Self-duality. In Minkowski space, $\star^2 = -1$, so it has eigenvalues $\pm i$. Then every field strength can then be written in the form $F = F_+ + F_-$ where F_+ and F_- are (anti)self-dual, i.e.

$$\star F_\pm = \pm i F_\pm.$$

If F is (anti)self-dual and $dF = 0$, then it automatically satisfies $d \star F = 0$. This gives a shortcut for finding solutions to the vacuum Maxwell equations, which can be generalized to Yang–Mills theory to find instanton solutions.

In particular, the (anti)self-duality condition is $\mathbf{B} = \pm i \mathbf{E}$. Solving Maxwell's equations for a plane wave $\mathbf{E}$ shows that the solutions are circularly polarized plane waves.

Note. We can have a nonzero electric flux through a closed surface even without charge ($d \star F = 0$), if the space is topologically nontrivial. For example, consider a spacetime which contains a 'wormhole' at some fixed time t. Electric field lines can go in through one end and out the other, so that there can be nonzero electric flux through a sphere about one end. This doesn't contradict our result above, because the sphere is not a boundary. Topologically, these situations can arise if the second (co)homology group is nontrivial.

Electric flux can hence be topological or nontopological. Similarly, we can have nontopological magnetic flux by defining $\star dF = J_M$, or topological magnetic flux via 'wormholes' or more simply by removing the point of a monopole from spacetime. However, since electromagnetism is typically formulated in terms of a gauge potential $F = dA$, the nontopological option is ruled out. Demanding that $F = dA$ also forces A to be singular on a Dirac string. The proper way to avoid this singularity is to describe A by a more powerful object: a connection on a fiber bundle.

7 Fiber Bundles

7.1 Motivation

In this section, we motivate the correspondence between gauge fields and connections on a principal fiber bundle.

- In electrostatics, the electric field one-form obeys $dE = 0$. If the space is simply connected, i.e. has trivial first cohomology group H^1, we may write the field in terms of a scalar potential, $E = dV$, which implies that all closed loop integrals of E vanish by Stokes' theorem.
- Now consider a magnetostatic field. The magnetic field has zero divergence, so the magnetic field two-form obeys $dB = 0$. If the space has trivial second cohomology group H^2, we may write the field in terms of a vector potential, $B = dA$, which implies that all closed surface integrals of B vanish. These integrals represent magnetic fluxes.
- Now consider the field of a magnetic monopole,

  $$\mathbf{B} = \frac{g}{\rho^2}\hat{\boldsymbol{\rho}}.$$

  This field is defined on $\mathbb{R}^3 - \{0\}$, which has nontrivial H^2, reflected in the fact that the flux integral of $\mathbf{B}$ is nonzero. Hence $\mathbf{B}$ cannot be written as $\mathbf{B} = \nabla \times \mathbf{A}$.
- However, we can define $\mathbf{A}$ locally by further restricting the domain. Imagine removing a 'Dirac string' from $\mathbb{R}^3$, a line which begins at the origin and goes out to infinity. The resulting space has trivial H^2, so we may define a vector potential on it.
- Taking the Dirac string to point along $-\hat{\mathbf{z}}$ and $+\hat{\mathbf{z}}$, respectively, gives

  $$\mathbf{A}_+(\rho,\phi,\theta) = \frac{g}{\rho\sin\phi}(1-\cos\phi)\,\hat{\boldsymbol{\theta}}, \quad \mathbf{A}_-(\rho,\phi,\theta) = -\frac{g}{\rho\sin\phi}(1+\cos\phi)\,\hat{\boldsymbol{\theta}}.$$

  When both fields are defined, they differ by a gradient,

  $$\mathbf{A}_+ - \mathbf{A}_- = \nabla(2g\theta)$$

  which confirms they yield the same field. Alternatively, in terms of differential forms,

  $$\mathbf{A}_+ = g(1-\cos\phi)\,d\theta, \quad \mathbf{A}_- = -g(1+\cos\phi)\,d\theta.$$

Next, we connect the vector potential to dynamics.

- The Schrodinger equation with a vector potential has the gauge symmetry

  $$\mathbf{A} \to \mathbf{A} + \nabla\Omega, \quad \psi \to e^{iq\Omega}\psi.$$

  For example, such a gauge transformation can be used to transfer between $\mathbf{A}_+$ and $\mathbf{A}_-$.
- We also know that a localized particle moving through a field picks up a phase $\int \mathbf{A}\cdot d\mathbf{x}$. To compute this phase when $\mathbf{A}$ is not defined globally, we simply work in a patch, then perform a gauge transformation to switch over to the next patch. As such, the value of the phase of ψ is not physical, since it is gauge-dependent, but relative phases are, as seen in the Aharanov–Bohm effect.

- Now suppose a charge is transferred around the equator; then it picks up a phase relative to a charge that isn't moved. The difference of the phases calculated using $\mathbf{A}_+$ and $\mathbf{A}_-$ is $4\pi qg\theta$, so this must be a multiple of 2π. Thus we have the Dirac quantization condition

$$qg = \frac{n}{2}, \quad n \in \mathbb{Z}.$$

  This can be made more precise with the path integral.

- The original physical argument for the Dirac quantization imagines the Dirac string as a physical half-infinite solenoid; then the vector potential is well-defined everywhere and we can use ordinary electromagnetism. If a particle went around such a string, it would pick up a phase of $4\pi qg$, and such a phase should be easy to detect. Since we haven't observed this, we must have $4\pi qg = 2\pi n$, the same quantization condition.

- Intuitively, we imagine a copy of S^1, specifying the phase, sitting above every point of our domain. Then the job of the vector potential is to tell us how paths in space lift to paths in this bundle, so it is a connection.

- As an example, if the domain is all of $\mathbb{R}^3$, it is contractible and the bundle is automatically trivial. This doesn't mean that $\mathbf{A}$ has no effect; one can still have magnetic fields in $\mathbb{R}^3$. But magnetic monopoles can only exist if the bundle is nontrivial.

- In our example above, our domain is $\mathbb{R}^3 - \{0\}$, which retracts to S^2. We are thus motivated to study S^1 bundles over S^2.

We now take a mathematical detour to construct the Hopf bundle.

- We parametrize the sphere S^1 as $e^{i\xi}$. It is also the group $U(1)$.

- The sphere S^2 is homeomorphic to the extended complex plane $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$ by stereographic projection. Explicitly, if U_S is S^2 minus the North pole, we have

$$\varphi_S \colon U_S \to \mathbb{R}^2, \quad (p^1, p^2, p^3) \mapsto (p^1, p^2)/(1 - p^3)$$

  and the North pole itself maps to the point at infinity. In terms of complex notation the inverse map is

$$z \mapsto \frac{(z + \overline{z}, -i(z - \overline{z}), z\overline{z} - 1)}{z\overline{z} + 1}.$$

  Similarly, let U_N be S^2 minus the South pole. Then

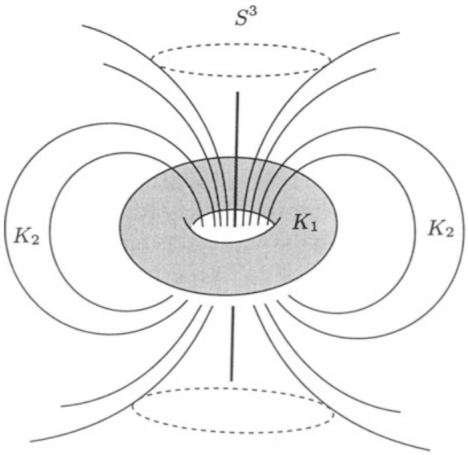
$$\varphi_N \colon U_N \to \mathbb{R}^2, \quad (p^1, p^2, p^3) \mapsto (p^1, p^2)/(1 + p^3) = \frac{1}{\varphi_S(p)^*}.$$

- Finally, S^3 is homeomorphic to $(\mathbb{R}^3)^*$ by similar reasoning. We can also identify it with a subset of $\mathbb{C}^2$ by

$$S^3 = \{(z^1, z^2) \,|\, |z^1|^2 + |z^2|^2 = 1\} = \left\{ \left(\cos\frac{\phi}{2} e^{i\xi_1}, \sin\frac{\phi}{2} e^{i\xi_2} \right) \,\middle|\, \phi \in [0, \pi] \right\}.$$

  Fixing any value of ϕ besides 0 and π yields a torus, since the ξ_i are invariant under a change by 2π, while $\phi = 0$ and $\phi = \pi$ yield circles.

- We can thus visualize S^3 as follows. We place the circle $\phi = 0$ in $(\mathbb{R}^3)^*$, so that the region $\phi \leq \pi/2$ forms a solid torus K_1.



  The region $\phi \geq \pi/2$ is another solid torus K_2 whose boundary is identified with that of K_1. It can be drawn as shown in the figure; the straight line at $\phi = \pi$ is indeed a circle in $(\mathbb{R}^3)^*$.

- Next, note that $U(1)$ acts on S^3 on the right by

$$p \cdot g = (z^1, z^2) \cdot g = (z^1 g, z^2 g), \quad g \in U(1).$$

  For any fixed $p \in S^3$, the orbit is a circle $U(1)$. The set of distinct orbits is the quotient space S^3/S^1. To understand this space, note that every orbit is identified by the ratio $z^1/z^2 \in \mathbb{C}^*$, so $S^2 \cong S^3/S^1$.

- We define the projection map

$$P\colon S^3 \to S^2, \quad (z^1, z^2) \mapsto (\varphi_S^*)^{-1}(z^1/z^2).$$

  This map is known as the Hopf fibration; it was originally constructed to show that $\pi_3(S^2)$ was nontrivial. For us, it gives S^3 the structure of a principle $U(1)$ bundle over S^2.

- Intuitively, if we think of S^3 as a normalized spinor state, the $U(1)$ is the phase ambiguity and the projection maps the state to the direction the spin 'points' in. The nontriviality of the bundle is reflected in the fact that it is impossible to define a continuous phase convention for the spinors; the usual conventions have singularities at the North and South poles.

- For completeness, we show local triviality. This means that we can cover S^2 with open sets V so that we have diffeomorphisms $\Psi\colon P^{-1}(V) \to V \times G$ of the form

$$\Psi(p) = (P(p), \psi(p)), \quad \psi(p \cdot g) = \psi(p) g.$$

  Explicitly, consider the subsets U_S and U_N, which satisfy

$$P^{-1}(U_S) = \{(z^1, z^2) \in S^3 \,|\, z^2 \neq 0\}, \quad P^{-1}(U_N) = \{(z^1, z^2) \in S^3 \,|\, z^1 \neq 0\}.$$

  Then we can use

$$\Psi_S^{-1}((z^1, z^2), g) = (z^1, z^2) \cdot \left(g \frac{|z^2|}{z^2}\right), \quad \Psi_N^{-1}((z^1, z^2), g) = (z^1, z^2) \cdot \left(g \frac{|z^1|}{z^1}\right)$$

  which satisfies all the requirements; we can verify smoothness by taking components.

- We may transfer between the local trivializations using transition functions,

$$\psi_{S,x} \circ \psi_{N,x}^{-1}(g) = g_{SN}(x)g, \quad \psi_{N,x} \circ \psi_{S,x}^{-1}(g) = g_{NS}(x)g$$

  where a computation shows that

$$g_{SN}(x) = \frac{z^2/|z^2|}{z^1/|z^1|} = e^{-i(\xi_1-\xi_2)} = e^{-i\theta}, \quad g_{NS}(x) = \frac{z^1/|z^1|}{z^2/|z^2|} = e^{i(\xi_1-\xi_2)} = e^{i\theta}$$

  where θ is a spherical coordinate on S^2.

Next, we link the Hopf bundle to the magnetic monopole.

- It can be shown that the $U(1)$ bundles over S^2 are classified by elements of $\pi_1(U(1))$. Thus, they are indexed by integers just like the charges of magnetic monopoles. In the case of the Hopf bundle, we have the homotopy class 1 because the transition function is $e^{i\theta}$.

- Taking the elementary charge to be $q = 1$, the weakest monopole has $g = 1/2$ and hence

$$\mathbf{A}_N = \frac{1}{2}(1 - \cos\phi)\, d\theta, \quad \mathbf{A}_S = \frac{1}{2}(1 + \cos\phi)\, d\theta, \quad A_N = A_S + d\theta.$$

  Now suppose we multiply both of these one-forms by $-i$. Then we have

$$\mathbf{A}_N = e^{i\theta}\mathbf{A}_S e^{-i\theta} + e^{i\theta}\, de^{-i\theta}.$$

- Now, on the mathematical end, a connection on a principal fiber bundle turns out to be a globally defined Lie algebra-valued one-form on the entire bundle space. It can be built out of locally defined Lie algebra-valued one-forms which are related by

$$\mathbf{A}_2 = g_{12}^{-1}\mathbf{A}_1 g_{12} + g_{12}^{-1}\, dg_{12}$$

  where g_{12} is a transition function; comparison with g_{NS} shows that $\mathbf{A}_N$ and $\mathbf{A}_S$ are related in just this way. Here, we are thinking of $\mathfrak{u}(1)$ as the set of pure imaginary numbers, so $\mathbf{A}_N$ and $\mathbf{A}_S$ are indeed $\mathfrak{u}(1)$–valued.

- Now we would like to use a connection to lift a path from S^2 into S^3. Consider the tangent space at some point in the bundle. It is sufficient to say which direction in the tangent space corresponds to the fiber (the 'vertical space'). Then path lifting is performed by moving in the bundle purely horizontally.

- A Lie algebra-valued one-form maps vectors in the tangent space to the Lie algebra $\mathfrak{u}(1) \cong \mathbb{R}$. Thus the kernel of the one-form is a two-dimensional subspace which identifies the 'horizontal subspace'.

- Finally, we identify the physical field F with the (covariant) exterior derivative of the connection. Then the field tells us about the holonomy associated with parallel transport in a loop, just as in physics, the Aharanov–Bohm phase picked up by a particle moving around a loop is the magnetic flux through the loop.

- All of these statements generalize to more complicated internal spaces, such as quark color. For example, the Hopf bundle can be generalized by replacing the complex numbers with quaternions; then the base space is S^4, the one-point compactification of $\mathbb{R}^4$, the fiber is $S^3 \cong SU(2)$, and the total space is $S^7 \subset \mathbb{R}^8$. This bundle is associated with the BPST instanton solutions to the Yang–Mills equations.

7.2 Definitions

We begin with the example of the Mobius strip.

- A fiber bundle is a manifold that looks locally like a product, but is not necessarily a product globally. For example, the cylinder is the product $S^1 \times L$ for a line segment L, and the Mobius strip M looks locally like $S^1 \times L$.
- The cylinder is a trivial bundle, so we can parametrize it with coordinates $(s, t) \in S^1 \times L$, while this is impossible for the Mobius strip.
- We would like to describe how the Mobius strip is twisted mathematically. For every open subset U of S^1, we can define a diffeomorphism

  $$\phi\colon U \times L \to \pi^{-1}(U)$$

  where $\pi\colon M \to S^1$ is the projection. This means that M is locally trivial on each U.
- Now cover the circle with two open sets U_1 and U_2 which overlap on the disjoint open intervals A and B. Then we may define

  $$\phi_1^{-1} \circ \phi_2\colon (A \cup B) \times L \to (A \cup B) \times L.$$

  Then at each point of S^1 in $A \cup B$, $\phi_1^{-1} \circ \phi_2$ defines a diffeomorphism from L to L. By scaling the coordinates, we can ensure that this diffeomorphism is either trivial or a sign flip.
- We can always choose the diffeomorphism to be trivial on A. The difference between the Mobius strip and the cylinder is that for the Mobius strip, we are forced to choose the diffeomorphism to be the sign flip on B. Thus the nontriviality of a fiber bundle is encoded in the nontriviality of its 'transition functions'.

We now proceed to the definition of a fiber bundle.

- A fiber bundle consists of a manifold E called the total space, a manifold M called the base space, and a manifold F called the fiber, equipped with a surjection $\pi\colon E \to M$ called the projection. For $x \in M$, the inverse image $\pi^{-1}(x) = F_x \cong F$ is called the fiber at x. To specify a bundle, we write $\pi\colon E \to M$.
- The fiber bundle is equipped with an open covering $\{U_i\}$ of M and a set of diffeomorphisms $\phi_i\colon U_i \times F \to \pi^{-1}(U_i)$ so that $\pi\phi_i(x, t) = x$, called local trivializations. The local trivializations relate each fiber to the standard fiber F, i.e. they provide local coordinates for the fibers.
- At each point $x \in M$, $\phi_{i,x}(t) \equiv \phi_i(x, t)$ is a diffeomorphism $\phi_{i,x}\colon F \to F_x$. On each point x in the overlap $U_i \cap U_j$, we require the transition function $t_{ij}(x) = \phi_{i,x}^{-1}\phi_{j,x}\colon F \to F$ to be an element of a Lie group G, called the structure group, which acts on the fiber on the left,

  $$\phi_j(x, t) = \phi_i(x, t_{ij}(x)t).$$

  Alternatively, for a fixed $u \in E$ with $\pi(u) = x$ we have

  $$\phi_i^{-1}(u) = (x, f_i), \quad \phi_j^{-1}(u) = (x, f_j), \quad f_i = t_{ij}(p)f_j.$$

  The final result is the transformation rule for sections.

- To be as general as possible, we can choose the structure group G to be $\mathrm{Diff}(F)$. However, we often instead find that G is a much smaller subset of $\mathrm{Diff}(F)$, or use the same G in various applications. For example, in particle physics G will usually be a gauge group.
- The bundle should not depend on the choice of open covering or local trivializations, so bundles are defined as equivalence classes of this data. Formally, a fiber bundle is an equivalence class of the 'coordinate bundles' defined above.
- By the definitions, the transition maps satisfy consistency conditions

$$t_{ij}t_{jk} = t_{ik} \text{ on } U_i \cap U_j \cap U_k, \quad t_{ij}^{-1} = t_{ji} \text{ on } U_i \cap U_j.$$

  Intuitively, the transition maps are simply the 'changes of coordinates' for the fibers required to pass from one patch to another.
- A fiber bundle is trivial if all transition functions can be chosen to be identity maps by adjusting the local trivializations. Specifically, suppose that the $\{U_i\}$ have two local trivializations $\{\phi_i\}$ and $\{\widetilde{\phi}_i\}$. Then if we define

$$g_i(x)\colon F \to F \text{ for } x \in U_i, \quad g_i(x) = \phi_{i,x}^{-1}\widetilde{\phi}_{i,x}$$

  where the $g_i(x)$ are in the structure group, then we have

$$\widetilde{t}_{ij}(x) = g_i(x)^{-1}t_{ij}(x)g_j(x)$$

  by the definitions. Then the transitions functions of a trivial bundle have the factorized form

$$t_{ij}(x) = g_i(x)^{-1}g_j(x).$$

  Conversely, if we can redefine the local trivializations so the transition functions do nothing, the bundle is trivial.
- In the case of gauge theories, the transition functions will be interpreted as gauge transformations. The t_{ij} are gauge transformations that link distinct patches, while the g_i are the more familiar gauge transformations within a single patch.

Next, we set up a bit more formalism.

- Given F, M with open cover $\{U_i\}$, and transition functions, we can always reconstruct a bundle $\pi\colon E \to M$. This is intuitive; formally we would take the union of the $U_i \times F$ and glue them together/define an equivalence relation using the transition functions.
- Consider two fiber bundles $\pi\colon E \to M$ and $\pi'\colon E' \to M'$. A smooth map $\overline{f}\colon E' \to E$ is a bundle map if it maps each fiber F'_p of E' onto a fiber F_q of E. (We also require some compatibility conditions for the transition functions.) Then $\overline{f}$ naturally induces a smooth map $f\colon M' \to M$ so that the diagram

$$\begin{array}{ccc} E' & \xrightarrow{\bar{f}} & E \\ {\scriptstyle\pi'}\downarrow & & \downarrow{\scriptstyle\pi} \\ M' & \xrightarrow{f} & M \end{array}$$

  commutes.

- Two bundles $\pi\colon E \to M$ and $\pi'\colon E' \to M$ are equivalent if there exists a bundle map $\overline{f}\colon E' \to E$ so that $\overline{f}$ is a diffeomorphism and f is the identity. In particular, bundles that differ only by a redefinition of the local trivializations, as considered above, are equivalent.
- Given a bundle $\pi\colon E \to M$ with fiber F and a map $f\colon N \to M$, we can define a pullback bundle f^*E over N with the same fiber F by

$$f^*E = \{(p, u) \in N \times E \mid f(p) = \pi(u)\}.$$

  Unpacking this, we define f^*E by pulling back the open cover and the transition functions; the rest of the bundle can be reconstructed using the reasoning above.
- It can be shown that if $f, g\colon N \to M$ are homotopic, then f^*E and g^*E are equivalent bundles over N. In particular, suppose M is contractible, so the identity map on M is homotopic to a constant map. Then E must be trivial.
- Given $\pi\colon E \to M$, a global section s is a smooth map $s\colon M \to E$ so that $\pi(s(x)) = x$ for all $x \in M$. The set of global sections is called $\Gamma(M, E)$, and depending on the bundle, there may not be any.
- A local section is a section only defined on an open set U of M. Local sections always exist by local triviality, and the set of local sections over U is called $\Gamma(U, E)$.

Next, we introduce vector bundles and give some examples.

- A vector bundle is a bundle E where the fiber is a vector space; tangent bundles are one example. If the fiber is $F = \mathbb{R}^n$, we say $\dim E = n$. The structure group is $GL(n, \mathbb{R})$. A line bundle is a one-dimensional vector bundle. Note that any vector bundle admits a global section called the null section, which is simply zero everywhere.
- The set of the tangent spaces in an n-dimensional manifold forms the tangent bundle TM. If the coordinates on a patch are x^i, then the local trivialization is to simply write a tangent vector in components in the basis $\partial/\partial x^i$. The transition function is the Jacobian.
  - $T\mathbb{R}^n$ is clearly trivial and equal to $\mathbb{R}^{2n}$, as we would expect since $\mathbb{R}^n$ is contractible.
  - TS^1 is also trivial. To see this, simply define the unit vector ∂_θ. Then we have the global trivialization $(\theta, t) \mapsto (\theta, t\partial_\theta)$.
  - TS^2 is not trivial. Note that a global trivialization implies the existence of a basis of nonvanishing global sections. However, the Poincare–Hopf theorem states that TM has a nonvanishing global section if and only if $\chi(M) = 0$, where χ is the Euler characteristic.

  Thinking exclusively in terms of tangent bundles can be a bit misleading, because we usually think of the fiber as sitting 'above' the base space, drawing the two perpendicular.
- There are also many examples of vector bundles that aren't tangent bundles. For example, let M be an m-dimensional manifold embedded in $\mathbb{R}^{m+k}$ and let N_pM be the vector space normal to T_pM in $\mathbb{R}^{m+k}$, under the Euclidean metric. Then the normal bundle

$$NM = \bigcup_{p \in M} N_pM$$

  is a vector bundle of dimension k.

  - For the sphere S^2 embedded in $\mathbb{R}^3$, NS^2 is a trivial line bundle, by spherical coordinates.
  - Consider a relativistic particle with spin. In its frame, its spin is a spacelike vector, orthogonal to its timelike path M. The normal bundle NM can be used to describe Thomas precession.

- Recall that an element of $\mathbb{C}P^n$ is a complex line in $\mathbb{C}^{n+1}$ through the origin. Then $\mathbb{C}P^n$ has a canonical line bundle where the fiber of a point is the corresponding line. To define this formally, let $I = \mathbb{C}P^n \times \mathbb{C}^{n+1}$ be a trivial bundle over $\mathbb{C}P^n$ with elements (p, v). Then the canonical line bundle L is the subset

$$L = \{(p, v) \in I | v = ap, a \in \mathbb{C}\}$$

  with projection $\pi(p, v) = p$.

- Given a vector bundle $\pi\colon E \to M$ with fiber F, we can define the dual bundle $\pi\colon E^* \to M$ whose fiber F^* is the set of linear maps from F to the field $\mathbb{R}$. Then the cotangent bundle is the dual bundle of the tangent bundle.

Now we give some more ways to combine bundles.

- Let $\pi\colon E \to M$ and $\pi'\colon E' \to M'$ be vector bundles with fibers F and F'. The product bundle

$$\pi \times \pi'\colon E \times E' \to M \times M'$$

  is a fiber bundle with fibers $F \oplus F'$. For example, if $M = M_1 \times M_2$ then $TM = TM_1 \times TM_2$.

- Now let $f\colon M \to M \times M$ be defined by $f(p) = (p, p)$. The Whitney sum bundle $E \oplus E'$ is the pullback bundle of $E \times E'$ by f. It is a bundle over M with fiber $F \oplus F'$. If the transition functions are matrices t_{ij}^E and $t_{ij}^{E'}$, then the transition function of $E \oplus E'$ is $\text{diag}(t_{ij}^E, t_{ij}^{E'})$.

- Let $\pi\colon E \to M$ and $\pi'\colon E' \to M$ be vector bundles with fibers F and F'. The tensor product bundle $E \otimes E'$ is obtained by taking the tensor product of fibers $F_p \otimes F'_p$ at every point $p \in M$. For example, bundles of differential forms are defined as antisymmetrized tensor products of the cotangent bundle.

Example. The Whitney sum bundle of two copies of the Mobius strip, with fiber $\mathbb{R}$. We cover the Mobius strip with two open sets; to have a trivial bundle, we must have $t_{12}(x) = g_1(x)^{-1}g_2(x)$. We choose $g_1(x)$ to be the identity without loss of generality, so we require $g_2(x_1) = I$ and $g_2(x_2) = -I$ where x_1 and x_2 are the two places the open sets overlap. This is impossible for the Mobius strip because $GL_1(\mathbb{R}) = \mathbb{R} - \{0\}$ is disconnected, but perfectly possible for the sum bundle.

Example. Consider the sphere S^2 embedded in $\mathbb{R}^3$. Then the Whitney sum bundle of TS^2 and NS^2 is simply a trivial bundle over S^2 with fiber $\mathbb{R}^3$.

7.3 Principal Bundles

Next, we turn to principal bundles.

- A principal bundle is a bundle whose fiber is equal to its structure group. They are written as $P(M, G)$ and called G-bundles over M.

- In general, we should think of M as a quotient of E, not a submanifold. This is clearest for principal bundles: identifying M as a submanifold would be equivalent to finding a global section, but this requires the bundle to be trivial.
- Unlike generic bundles, we have a natural action of G on P on the right. If we have $u = \phi_i(p, g_i)$ where $u \in P$, then we define
$$ua = \phi_i(p, g_i a)$$
for $a \in G$. It's straightforward to check this definition is independent of the local trivializations, because right actions commute with left actions.
- The same idea wouldn't work for, e.g. the tangent bundle because finding how an element of $GL_n(\mathbb{R})$ acts on a tangent space requires a basis choice. But for a frame, there is a natural action, since a frame *is* a basis.
- Conversely, given a group action of G on P, we can construct a principle bundle with $M = P/G$. This is how we constructed the Hopf bundle, through an action of $U(1)$ on S^3. Note that the fibers of the bundle will only be isomorphic to G if the group action is free.
- Note that the typical fiber F has a preferred element, the identity. We should not think of each individual fiber F_x as having a preferred element, since the mapping of elements of F_x to F depends on the local trivialization; instead the F_x are merely manifolds. However, given a section $s_i(p)$ over U_i, there is a preferred local trivialization ϕ_i where $s_i(p) = \phi_i(p, e)$.
- A principal bundle is trivial if and only if it admits a single global section. To show the backwards direction, let $s(p)$ be such a section. Then we have a map
$$\Phi\colon P \to M \times G, \quad \Phi\colon s(p)a \to (p, a)$$
which is a homeomorphism, giving the result.
- Every bundle has an associated principal bundle, by replacing the fiber with the structure group and keeping the same transition functions. Since the nontriviality of a bundle is encoded in the transition functions, the associated bundle is trivial if and only if the original bundle is.
- Conversely, consider a principle bundle $P(M, G)$ and an n-dimensional representation $\rho\colon G \to GL_n(\mathbb{R})$ which acts on $V = \mathbb{R}^n$ from the left. Then the vector bundle E_ρ associated to P is
$$E_\rho = P \times V/\sim, \quad (u, v) \sim (ug, \rho(g^{-1})v)$$
Essentially, we replace the fiber G with V and turn the transition function t_{ij} into $\rho(t_{ij})$. More generally, for any manifold F we can construct an associated bundle given any left-action of G on F.

Example. The Mobius strip revisited. If the fiber is $\mathbb{R}$, it is a vector bundle and there are no nonvanishing global sections by continuity. If the fiber is $\mathbb{Z}_2$ so that the bundle is a principal bundle, then no global sections exist because we pick up a sign flip if we try to go around.

Example. The frame bundle is the principal bundle associated with a vector bundle.

- A frame is an ordered set of basis vectors for F at a point. In the frame bundle, the fiber is the set of possible frames, and we keep the exact same transition functions, so the structure group remains $GL(n, \mathbb{R})$.

- Applying the orbit-stabilizer theorem shows the set of frames at a point is diffeomorphic to $GL(n, \mathbb{R})$, so the frame bundle is indeed a principle bundle, and it is not a vector bundle.
- As a corollary of our previous results, the tangent bundle is trivial if and only if there is a global frame. This is intuitively clear, since given a global frame we can assign vectors components at every point, giving a diffeomorphism between TM and $M \times \mathbb{R}^n$.
- The frame bundle FM of a manifold M is the frame bundle associated with the tangent bundle. In relativity, local sections of FM are called tetrads or vierbeins.
- The frame bundle FS^2 is nontrivial by the Poincare–Hopf theorem, since there are no nonvanishing global sections. In optics, FS^2 is used to describe the polarizations of spherical waves, but as we've just shown, a set of polarizations cannot be chosen continuously!
- If we restrict to orthonormal frames, the structure group and the fiber both changes to $O(n)$. This general procedure is called the 'reduction of the structure group'.
- To define a spin bundle, we start with a Lorentzian manifold, take the frame bundle, then reduce the structure group to the Lorentz group. We then take an associated vector bundle with fiber $\mathbb{C}^2$, lifting the structure group to $SL(2, \mathbb{C})$. Sections of such a bundle describe Weyl spinors. Topological obstructions in nontrivial spacetimes may prevent the lifting.

Example. Using our machinery, we can describe nonrelativistic quantum mechanics.

- We consider a quantum particle in $\mathbb{R}^3$ with a complex-valued wavefunction. We can measure the norm $|\psi(\mathbf{x})|^2$ but not the local phase, and we would like to express this geometrically.
- We define a Hermitian line bundle E, i.e. a complex line bundle with a metric, over $\mathbb{R}^3$. We restrict the structure group so it preserves this metric, so it is $U(1)$. Then the state of the particle is described by a global section $\psi(\mathbf{x})$ of E.
- In general, there is no natural isomorphism between a fiber F_x and $\mathbb{C}$, so $\psi(\mathbf{x})$ cannot be interpreted as a wavefunction. However, given a global orthonormal frame $e(\mathbf{x})$, we may define $\psi(\mathbf{x}) = e(\mathbf{x})\phi(\mathbf{x})$ and interpret $\phi(\mathbf{x})$ as a wavefunction. In ordinary language, we must pick a phase convention for the position basis to write down wavefunctions.
- A gauge transformation can be performed by changing $e(\mathbf{x})$. There will be a corresponding change in the gauge potential $A(\mathbf{x})$, which is a connection on a $U(1)$-bundle over $\mathbb{R}^3$, as we'll see in more detail below.
- In the case of $\mathbb{R}^3$ all bundles are automatically trivial, but in the case of defects such as magnetic monopoles, we work in subspaces of $\mathbb{R}^3$ which may be nontrivial.
- More generally, a matter field transforms in a representation R of a gauge group G. The gauge potential is a connection on a G-bundle over spacetime M, while the matter field is a section of an associated vector bundle with fiber R and gauge group G.

Example. Classifying $U(1)$ bundles over $\mathbb{R}^3 - \{0\}$. As we'll see, this is the topological setting of the magnetic monopole. We perform a deformation retraction of $\mathbb{R}^3 - \{0\}$ to S^2 for convenience, then cover it with a 'North' and 'South' chart. The charts overlap on a strip along the equator, which

is effectively S^1. Then the nontriviality of the bundle is entirely encoded in the single transition function,

$$t_{NS}(\theta) = e^{i\alpha} \in U(1).$$

By a change of the transition functions, which we interpret as a local gauge transformation,

$$\tilde{t}_{NS}(\theta) = g_N(\theta)^{-1} t_{NS}(\theta) g_S(\theta).$$

Now look at the North patch from above, as a disc. Then varying r in $g_N(\theta, r)$ provides a homotopy between $g_N(\theta)$ defined above and a constant map. Thus gauge transformations cannot change the homotopy class of $t_{NS}(\theta)$, so the $U(1)$ bundles over S^2 are classified by $\pi_1(S^1)$. For example, the transition function of the Hopf map corresponds to the homotopy class $n = 1$.

Example. Consider an $SU(2)$ bundle over $\mathbb{R}^4$. To find instanton solutions, we compactify $\mathbb{R}^4$ to S^4, which is non-contractible and hence admits nontrivial fiber bundles. By the same argument as above, the bundles are classified by $\pi_3(SU(2)) = \mathbb{Z}$. Parametrizing the overlap S^3 with unit vectors (x, y, z, t), the transition function

$$t_{NS}(p) = \left(t\mathbf{1} + i \sum_i x^i \sigma_i \right)^n$$

corresponds to the homotopy class n. Explicitly, we can generalize the Hopf map with quaternions to yield a map $\pi\colon S^7 \to S^4$, an S^3 bundle over S^4, whose transition function belongs to the homotopy class 1.

Example. In general, let H be a closed Lie subgroup of a Lie group G. Then H acts on the coset space $M = G/H$, so we have a principal H-bundle over M where the projection $\pi\colon G \to G/H$ just takes the coset. This is a general method of constructing principal bundles, and some useful examples include

$$O(n)/O(n-1) \cong SO(n)/SO(n-1) \cong S^{n-1}, \quad U(n)/U(n-1) \cong SU(n)/SU(n-1) \cong S^{2n-1}.$$

7.4 Connections on Fiber Bundles

Next, we add the additional structure of a connection.

- In general relativity, the connection allows us to parallel transport vectors along a path. In our fiber bundle language, we start with a curve γ in the base space M and want to construct a curve in the tangent bundle s^γ that projects down to γ.
- Given a principal bundle $P(M, G)$, we would like to lift a curve γ on M to a curve γ_P on P that projects down to γ. Equivalently, we want to lift vectors in $T_\gamma M$ to vectors in TP.
- For the Mobius strip, we can think of the circle as being 'horizontal' and the line segment as being 'vertical'. Then given a starting point for γ_P, we can simply move horizontally, i.e. make the tangent vector to γ_P always horizontal. Then a connection is just a choice of what 'horizontal' means. Note that we shouldn't think of 'vertical' and 'horizontal' as orthogonal directions, as there generally is no metric.

- Formally, we define a connection on P to be a smooth choice of horizontal subspaces $H_uP \subset T_uP$ so that
$$T_uP = V_uP \oplus H_uP$$
where $V_uP \cong \mathfrak{g}$ is the tangent space to the fiber, and
$$H_{ug}P = R_{g*}H_uP.$$
This compatibility condition requires every point in the group fiber to be equivalent.

- Note that a connection defines a distribution with dimension $\dim M$, in the sense of Frobenius' theorem. The distribution is integrable, i.e. it meshes together into surfaces, if and only if the curvature of the connection vanishes.

- Let $\gamma\colon [0,1] \to M$ be a curve in the base manifold. Then $\gamma_P\colon [0,1] \to P$ is the horizontal lift of γ if $\pi(\gamma_P) = \gamma$ and the tangent vector X_P to γ_P is always horizontal, $X_P \in H_{\gamma_P}P$, and one can show that γ_P is unique.

- Generally, a closed loop won't lift to a closed loop. Instead, we will have $\gamma_P(1) = \gamma_P(0)g$ for some group element g. The set of possible group elements attained by varying the loop and keeping the base point $p = \gamma_P(0)$ fixed is called the holonomy group $\mathrm{Hol}_p(P)$. The holonomy group depends on both the bundle and the connection. If M is connected, the holonomy group is the same at all points of M, so we call it $\mathrm{Hol}(P)$.

Next, we consider an alternative definition of a connection that is closer to a gauge potential.

- Consider a vector $Y \in T_uP$. We can decompose it into horizontal and vertical components using the projections
$$Y_v = \pi_v(Y), \quad Y_h = \pi_h(Y), \quad Y = Y_v + Y_h$$

- For convenience, we will regard a vector as an equivalence class of curves, so $X = [\sigma(t)]$ where $\sigma(0) = x$ and $X \in T_xM$.

- Consider a Lie algebra element $V \in \mathfrak{g}$, so
$$V = [\exp(tV)].$$
Taking the right action of G on F_x by right-multiplication, we may associate V with an induced vector field $V^\sharp$ on F_x, with
$$V^\sharp|_u = [R_{\exp(tV)}u] = [u\exp(tV)].$$
We thus have a map $\sharp\colon \mathfrak{g} \to V_uP$. Note that $\sharp$ does not depend on the choice of local trivialization for F_x, since left-multiplication and right-multiplication commute.

- Next, we define the Ehresmann form
$$\omega = \sharp^{-1} \circ \pi_v, \quad \omega|_u\colon T_uP \to \mathfrak{g}$$
which is a Lie algebra-valued one-form. Then by definition we have
$$\omega|_u(H_uP) = 0, \quad \omega|_u(V^\sharp|_u) = V.$$

- The Ehresmann form obeys another identity. We note that for $a \in G$ and $V \in \mathfrak{g}$,

$$R_{a*}(V^\sharp|_u) = R_{a*}[u\exp(tV)] = [R_a u\exp(tV)] = [u\exp(tV)a].$$

Using the identity

$$a^{-1}\exp(tV)a = \exp(t\,\mathrm{ad}_{a^{-1}}V)$$

which holds for matrix Lie groups by power series, we have

$$R_{a*}(V^\sharp|_u) = [ua\exp(t\,\mathrm{ad}_{a^{-1}}V)] = (\mathrm{ad}_{a^{-1}}V)^\sharp|_{ua}.$$

Therefore, we have

$$(R_a^*\omega)|_u(V^\sharp|_u) = \omega|_{ua}(R_{a*}(V^\sharp|_u)) = \omega|_{ua}((\mathrm{ad}_{a^{-1}}V)^\sharp|_{ua}) = \mathrm{ad}_{a^{-1}}V = \mathrm{ad}_{a^{-1}}\omega|_u(V^\sharp|_u)$$

which implies that $R_a^*\omega$ and $\mathrm{ad}_{a^{-1}}\omega$ agree on vertical vectors. They also both annihilate horizontal vectors, so they are equal.

- Conversely, we may define a connection by a $\mathfrak{g}$-valued one-form ω satisfying

$$\omega|_u(V^\sharp|_u) = V, \quad R_a^*\omega = \mathrm{ad}_{a^{-1}}\omega$$

which defines H_uP by $\omega|_u(H_uP) = 0$.

Example. A falling cat can turn over even though it has zero angular momentum at every moment, since its body is deformable. To describe this with fiber bundles, let $\mathcal{C}$ be the configuration space of a deformable body; we quotient out by center of mass positions since we won't care about them. Then the shape space is obtained by quotienting by rotations, $\overline{\mathcal{C}} = \mathcal{C}/SO(3)$, so $\mathcal{C}$ is a principal $SO(3)$ bundle over shape space. Then a local section of $\mathcal{C}$ can be used to define orientations, and it is geometrically obvious that there is no canonical choice of local section, and no way to compare the orientations of distinct shapes. The connection is defined by imposing conservation of angular momentum, and the analysis of the falling cat is a statement about the holonomy group of $\mathcal{C}$.

The same formalism can be applied to bacteria which swim in low Reynolds number by deforming their bodies. In this case, $\mathcal{C}$ is the configuration space including different center of mass conditions, and we define $\overline{\mathcal{C}} = \mathcal{C}/\mathbb{R}^3$ so $\mathcal{C}$ is a principal $\mathbb{R}^3$ bundle over shape space.

Finally, we link principal bundles to gauge theory.

- A gauge theory with gauge group G on a spacetime M is associated with a G-bundle over M, as well as a connection ω on it, called the gauge potential.

- On each patch, taking a local section σ_i gives a local description $A_i = \sigma_i^*\omega$ of ω, which is the coordinate expression of the gauge field familiar to physicists. When the bundle is nontrivial, one must work with multiple patches; if one tries to work naively in a global patch one will find singularities in the gauge field.

- If we change the local section in a patch $\sigma_i \to g_i\sigma_i$, where g_i is another local section, then we perform a gauge transformation

$$A_i \to g_i^{-1}A_ig_i + g_i^{-1}dg_i.$$

It does not change the abstract gauge field A, but merely its description.

- The local descriptions of the gauge field on different patches are related by
$$A_j = t_{ij}^{-1} A_i t_{ij} + t_{ij}^{-1} dt_{ij}$$
which has the same form as a gauge transformation, though it is conceptually distinct.
- If the bundle is trivial, we may define a global section g, which yields gauge transformations $\sigma_i \to g\sigma_i$. Now, since g is not defined on a topologically trivial base space, it may not be homotopic to the identity. Such a transformation is called a large gauge transformation. One can also find large gauge transformations if g is defined on a topologically nontrivial subset of M, such that the bundle is trivial when restricted to M.
- As we'll see below, the topology of the bundle can place constraints on the connections that can be put on it. If one isn't careful and simply computes naively in coordinates, it's possible to perform a 'false' gauge transformation that changes the bundle topology. This is not a gauge transformation in any sense, but is often mistaken for one.

Example. The Aharanov–Bohm effect, for a particle confined to a ring. This is described by a $U(1)$ bundle over S^1, but all such bundles are trivial. If we cover S^1 with two patches, then the transition functions live in $\pi_0(U(1))$, which is trivial; they can be taken to be trivial with an appropriate choice of σ_i. This fits with the fact that one can write a nonsingular global gauge potential, namely
$$\mathbf{A} = \frac{\Phi_0}{2\pi r}\hat{\boldsymbol{\theta}}.$$
There is an Aharanov–Bohm phase of Φ_0 associated with parallel transport of a particle around the loop; this is specifically the holonomy associated with the loop for an associated vector bundle over $\mathbb{C}$. This is clearest when the transition functions are trivial, but doesn't change under any gauge transformations, including large ones; when one totals up the integral $\int \mathbf{A} \cdot d\mathbf{s}$ along with the phases incurred via transition functions when switching between patches, the net phase is always Φ_0.

Example. The vacua of Yang–Mills theory. Compactifying space to S^3, the fiber bundles are described by G-bundles over S^3 and classified by $\pi_2(G)$. However, $\pi_2(G)$ is trivial for any Lie group, so all bundles here are trivial. Hence one may define a global gauge potential $A = g^{-1}dg$, where we have a map $g\colon S^3 \to G$. However, for simple Lie groups $\pi_3(G) = \mathbb{Z}$, so the vacua are indexed by integers. They are related by large gauge transformations, so they are completely equivalent classically (i.e. they correspond to the same gauge connection ω) but can be chosen to be distinct or equivalent as quantum states.

Example. Magnetic monopoles. As we've seen before, these are associated with nontrivial fiber bundles; we work on $\mathbb{R}^3 - \{0\}$ which retracts to S^2. Working with an abelian gauge group for simplicity, the bundle is classified by $\pi_1(U(1)) \in \mathbb{Z}$. The integer here is proportional to the first Chern class
$$\int d^2x\, F$$
which measures the magnetic charge of the monopole, yielding the quantization of magnetic charge.

Example. Instantons are associated with nontrivial fiber bundles on compactified Euclidean spacetime S^4, where bundles are classified by $\pi_3(S^3) \in \mathbb{Z}$ for all simple Lie groups. The integer here is proportional to the second Chern class
$$\int d^4x\ \mathrm{tr}(F \wedge F)$$
which measures the instanton number.