Notes for Math Methods of Physics, CMI, Spring 2024
Govind S. Krishnaswami, April 30, 2024

Please let me know at govind@cmi.ac.in of any comments or corrections

Course website http://www.cmi.ac.in/~govind/teaching/math-meth-e24

Contents

1 The concept of a manifold

By a manifold, we have in mind a space like a circle (denoted S^1), the plane, the surface of a sphere (S^2) or the 3d Euclidean space $\mathbb{R}^3$ in which a particle can move. A manifold is a space where every point[1] has an open neighborhood[2] that looks like[3] Euclidean space $\mathbb{R}^n$ for some fixed positive integer n, which is called the dimension of the manifold. By considering sufficiently many such overlapping open neighborhoods

[1]In §8 we will extend the concept of a manifold to one with a boundary. The points on the boundary will not have open neighborhoods and need to be treated differently.

[2]The open neighborhoods we have in mind are simple ones: they must come in one piece and be contractible to a point. In 1d they are open intervals on the real line (a, b) or bent (continuously deformed) versions thereof. In 2d they are open disks $x^2 + y^2 < 1$ or stretched/bent (continuously deformed) versions of disks. In 3d they are continuous deformations of open balls $x^2 + y^2 + z^2 < 1$. The open interval, disk and ball is each continuously deformable into $\mathbb{R}, \mathbb{R}^2$ and $\mathbb{R}^3$. Similarly, we have open balls in higher dimensions. They are our neighborhoods. By contrast, an annulus $1 < x^2 + y^2 < 2$ is not contractible to a point, it cannot be continuously shrunk to a point.

[3]By 'looks like', we mean *continuously deformable into*. A rubber balloon undergoes continuous deformation as it is inflated. More precisely, by 'looks like', we mean *homeomorphic to*. A homeomorphism is a continuous map with a continuous inverse. An untied balloon is homeomorphic to a disc-shaped rubber sheet since the latter can be stretched into a balloon without tearing the rubber sheet.

of points, we obtain an open covering of the space. Thus, roughly, a manifold is a space that can be covered by charts or 'coordinate' patches, as in Fig. 1a and Fig. 2a. The charts together are said to furnish an atlas for the manifold. The terminology is borrowed from cartography, where an atlas consists of several overlapping charts which can, for instance, together describe a continent.

The idea is to use existing notions on smooth functions, vector fields, differentials and Cartesian tensors on $\mathbb{R}^n$ to develop corresponding notions for the manifold, via a combination of patchwork and consistency conditions between overlapping charts. It took a long time for a satisfactory definition of a manifold to be arrived at (in the work of Hermann Weyl (1912) and Hassler Whitney (1930s)), with examples playing a key role. In this appendix, we will introduce a number of concepts and technical terms from the theory of manifolds. The reader who is meeting these for the first time should not despair, as they are invariably accompanied by illustrative examples.

Analogy with cell phone networks and cartography. The idea of covering a space with overlapping patches of a simple sort is practically realized in cell phone networks, which we caricature now. For instance, a city is covered by cells (say disks), each serviced by a cell phone tower. Each point in the city lies in at least one such cell and communication to/from the cell phone is transmitted via some protocols associated to the corresponding tower (manner of storage, encryption etc.). If a phone lies in the intersection of two cells, then two towers can simultaneously communicate with it. The data received by the two towers can be related to each other via a suitable transformation between the protocols followed by each tower. This is crucial when a person is traveling in the back seat of a car and speaking on a cell phone. When moving from one cell to another, the two towers must agree on what the person is saying when the phone is in the intersection, before the 'future' tower takes over from the 'past' tower. Evidently, the city is our manifold and the cells are our coordinate patches. The transformations between data received by two towers from the overlap of two cells play the role of transition functions that we will soon encounter. A similar analogy, which explains much of the terminology, may be made with cartography, where the charts or maps prepared by two explorers have to be related (e.g., the scales of magnification may be different) in regions they both explore. □

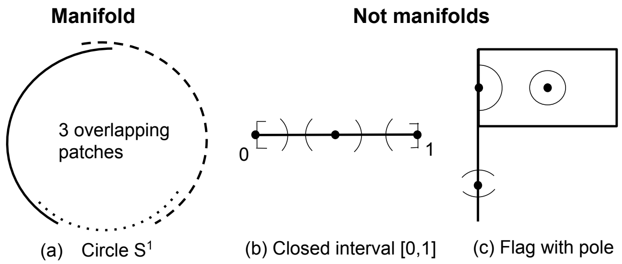


Figure 1: (a) The circle manifold S^1 covered here by three overlapping open neighborhoods. (b) The closed interval $[0, 1]$ is not a manifold as 0 and 1 do not have open neighborhoods: it is a manifold with boundary. (c) A flag with pole is not a manifold: open neighborhoods do not all have the same dimension and the neighborhoods of all points do not look alike.

Coordinate charts and transition functions. Returning to our definition of a manifold, why do we insist on open neighborhoods of the same dimension to cover a manifold? Examples of spaces we do not want to regard as manifolds will reveal why. For instance, the closed interval $C = [0, 1] \subset \mathbb{R}^1$ is *not* a manifold[4]. The points 0 and 1 do not have any open neighborhoods[5] lying in C while all other points $0 < x < 1$ have open neighborhoods (see Fig. 1) of the form $(x - \epsilon, x + \epsilon) \subset C$ for some sufficiently small ϵ [we could take ϵ to be the smaller of $x/2$ and $(1 - x)/2$]. Intuitively, C looks different in the vicinity of 0 and 1 from how it looks elsewhere. We do not want to allow such 'inhomogeneities' in a manifold. This is why we insist on open neighborhoods. Similarly, the space that is shaped like the multiplication sign $\times$ is not a manifold: it looks different at the center and the extremes compared to how it looks elsewhere. A cloth flag attached to a pole is also not a manifold: points on the lower part of the pole look different from points on the cloth: the former typically have 1d open neighborhoods while the latter typically have 2d open neighborhoods (see Fig. 1). This is why we insist that all neighborhoods have the same dimension.

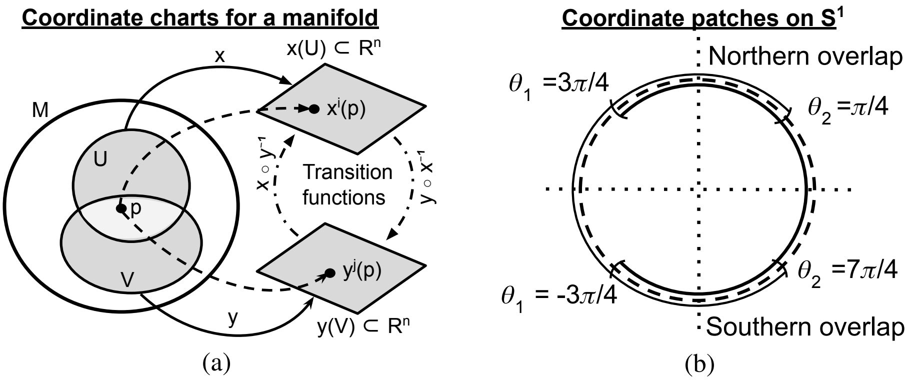


Figure 2: (a) Coordinate charts and transition functions for a manifold M (which can be thought of as the surface of a sphere S^2). (b) Overlapping coordinate patches for the circle S^1 (dashed curve) as a manifold. A minimum of two (open) coordinate patches is needed to cover the circle: here they are the Eastern and Western patches indicated by thick and thin solid curves. The angle θ is measured counterclockwise from the the horizontal axis.

On the other hand, the unit circle S^1 defined as the set of points (x, y) on the plane with $x^2 + y^2 = 1$ is a one-dimensional manifold. As shown in Fig. 2b, the circle can be covered by two patches: the Eastern and Western neighborhoods $-3\pi/4 < \theta_1 < 3\pi/4$ and $\pi/4 < \theta_2 < 7\pi/4$ defined in terms of a polar angle measured counterclockwise with respect to the positive x-axis. θ_1 and θ_2 are called local[6] coordinates in their respective patches. Thus, the circle is one-dimensional. Each of these angular patches is continuously deformable (by stretching) into the real line $\mathbb{R}$ since every open interval can be continuously mapped to the whole real line (e.g., $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$).

[4] However, $[0, 1]$ may be viewed as the manifold $(0, 1)$ with boundary included, see §8.

[5] Neighborhoods of 0 such as $[0, 1/2)$ are 'closed-open'.

[6] *Local* means coordinates are defined on a patch rather than *globally* on the whole manifold.

These patches intersect in a pair of 'upper' and 'lower' intervals: running from North-East to North-West and from South-West to South-East. When a point lies in such an intersection, either of the coordinates can be expressed in terms of the other via a 'transition function' or coordinate transformation. For the circle, we have in the upper intersection $\theta_1 = \theta_2$ and in the lower intersection $\theta_1 = \theta_2 - 2\pi$. The manifold is differentiable if these transition functions between coordinate systems in each such intersection is a differentiable map from $\mathbb{R}^n \to \mathbb{R}^n$ (or between open subsets of $\mathbb{R}^n$ which are homeomorphic to $\mathbb{R}^n$). In the circle example, the transition functions are linear maps of one real variable, so the circle is a differentiable manifold of dimension one. If the transition functions are infinitely differentiable (as is the case here), we say the manifold is smooth[7] or C^∞. Note that the circle cannot be covered by a single open chart: the largest ones such as $-\pi < \theta_1 < \pi$ unfortunately exclude one point while $-\pi < \theta_1 \leq \pi$ or $0 \leq \theta_1 \leq 2\pi$ fail to be open subsets of $\mathbb{R}^1$.

Sometimes we are lucky, and a single coordinate patch is sufficient to cover the whole manifold or the portion we are interested in. This is the case with the plane or a disk ($x^2 + y^2 < 1$) or 3d Euclidean space $\mathbb{R}^3$, which can be covered by a single patch with, say, Cartesian coordinates. In particular, $\mathbb{R}^n$ for each[8] $n = 1, 2, 3, \ldots$ is automatically a smooth manifold, as are all the open subsets of $\mathbb{R}^n$ that may be continuously shrunk to a single point. Unfortunately, the circle, which is a 1d manifold, is not an open subset of $\mathbb{R}^1$ and S^2 is not an open subset of $\mathbb{R}^2$, so we cannot cover them with a single chart[9] and need to work harder to find an atlas for these manifolds. The circle or the 2-sphere S^2 require a minimum of two coordinate patches. For S^2, the patches (each continuously deformable into a disk) consisting of all latitudes strictly above the Tropic of Capricorn and all latitudes strictly below the Tropic of Cancer furnish one possible atlas. These patches intersect over the tropics.

Given a manifold M, suppose a point $p \in M$ lies in the intersection of two coordinate patches so that p may be assigned the coordinates $x = (x^1, \cdots, x^n)$ or $y = (y^1, \cdots y^n)$. Then the 'transition function' from x to y is given by the equations for the coordinate transformation $y^i = y^i(x)$ and conversely $x^j = x^j(y)$ for $1 \leq i, j \leq n$. For the manifold to be smooth, both the transformation $x \mapsto y$ and its inverse $y \mapsto x$ must be smooth maps between open subsets of $\mathbb{R}^n$ (see Fig. 2a).

Refining an atlas. Given a smooth manifold (which must necessarily come with an atlas of smoothly interrelated coordinate charts), we are free to add a chart to the atlas, provided we are consistent. For instance, if the new chart with coordinate y overlaps

[7] If the transition functions are continuous we call it a C^0 manifold or a topological manifold. If they are once differentiable with a continuous derivative, we call it a C^1 manifold. More generally we have the notion of a C^k manifold if the transition functions are continuously differentiable k times for some $k = 0, 1, 2, \ldots$.

[8] For some purposes, it is convenient to regard $\mathbb{R}^0$ as a zero-dimensional manifold with only one point. A zero-dimensional manifold is either a point or a discrete set of points. For example, the zero-dimensional sphere S^0 is the pair of points $\{-1, 1\}$ satisfying $x^2 = 1$ in $\mathbb{R}^1$.

[9] If we could cover S^1 with a single chart, the chart (and hence S^1) would be an open subset of $\mathbb{R}^1$. Note, however, that merely being an open subset of $\mathbb{R}^n$ does not mean we can cover a manifold with a single chart, since our charts are assumed to be homeomorphic to open balls. For instance, the annulus $1 < x^2 + y^2 < 2$ is an open subset of $\mathbb{R}^2$, but we need a minimum of two charts to cover it.

an existing chart with coordinate x, the transformation $y = y(x)$ and its inverse must be smooth. For the Euclidean plane, the Cartesian coordinates $(x^1 \in \mathbb{R}, x^2 \in \mathbb{R})$ furnish a one-chart atlas. Suppose we wish to add a chart consisting of plane polar coordinates $(y^1 = r, y^2 = \phi)$. We can do this provided we choose the polar coordinate chart to be an open set on which the transformation to/from Cartesian coordinates is smooth. This is the case, for instance, if we choose the polar coordinate chart to be defined on $\mathbb{R}^2$ with the origin and negative horizontal axis excluded so that $y^1 = r \in (0, \infty)$ and $y^2 = \phi \in (-\pi, \pi)$. The Cartesian product $(0, \infty) \times (-\pi, \pi)$ is clearly an open subset of $\mathbb{R}^2$ continuously deformable into $\mathbb{R}^2$. Note that if we retained the negative real axis, the patch consisting of the punctured plane would not be continuously deformable into $\mathbb{R}^2$. In this region of overlap we have the familiar coordinate transformation $y^1 = \sqrt{(x^1)^2 + (x^2)^2}$ and $y^2 = \arctan(x^2/x^1)$ and the inverse transformation $x^1 = y^1 \cos y^2$ and $x^2 = y^1 \sin y^2$ which are both seen to be smooth. The smoothness fails at the origin.

Maps between manifolds. Having defined differentiable manifolds, we can now consider maps between manifolds. We will use such maps to say when two manifolds are to be considered the same. Two manifolds are topologically equivalent (or homeomorphic) if they are related by an invertible continuous map. The surface of a cube can be continuously deformed into that of a sphere, so they are homeomorphic. If two differentiable or smooth manifolds can be related via an invertible differentiable or smooth map, then they are called diffeomorphic.

To make precise the notion of a continuous, differentiable or smooth map between manifolds, we make use of the corresponding concept for maps between Euclidean spaces or open subsets thereof. So to begin with, a map $f : \mathbb{R}^p \to \mathbb{R}^q$ given by $y^i = f^i(x)$ is differentiable if all the first partials $\frac{\partial y^i}{\partial x^j}$ exist (it is continuously differentiable if these partial derivatives exist and are continuous). It is smooth if partial derivatives of all orders exist. It is continuous if y^i are continuous functions of x^j. Next, a map between manifolds $\phi : M^p \to N^q$ is continuous/differentiable/smooth if in each coordinate patch, the corresponding maps between $\mathbb{R}^p$ and $\mathbb{R}^q$ are continuous/differentiable/smooth. For consistency, if coordinate patches overlap, then the individual maps should agree on the overlap. Thus, we have the notion of smooth maps between manifolds. Two manifolds M, N are said to be diffeomorphic if there is a smooth bijective (1-1 and onto) map $f : M \to N$ with smooth inverse. They are homeomorphic if smoothness is replaced with continuity. Diffeomorphic or homeomorphic manifolds must have the same dimension and cannot be distinguished in so far as their smooth/topological structure is concerned. The circles $x^2 + y^2 = 1, x^2 + y^2 = 2$ and the ellipse $x^2 + 2y^2 = 1$ are all diffeomorphic (see Prob. **??**) as are the sphere $x^2 + y^2 + z^2 = 1$ and the ellipsoid $x^2 + y^2 + 2z^2 = 1$ or the open interval $(0, 1)$ and the real line $\mathbb{R}$ (see Prob. **??**). On the other hand, the surface of a cube is not smooth (due to the sharp edges and corners), so it is not diffeomorphic to the sphere. Similarly, the surface of a sphere and that of a torus (inflated tyre tube or vadai) are not homeomorphic: one can show that there is no continuous bijection between them since the latter has a 'handle' which the former lacks.

The concept of a manifold that we have defined does not possess any notion of distances between points or lengths of tangent vectors or angles between tangent vectors. To define these 'geometric' concepts we need additional structure on the manifold, such as a metric (see §6). At present, our manifolds are either topological manifolds (if the transition functions are continuous) or differentiable/smooth manifolds (if the transition functions are differentiable/smooth). Thus, our manifolds currently lack any intrinsic geometric rigidity of shape or size. In particular, the surface of a triaxial ellipsoid ($x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ with no two among a^2, b^2, c^2 equal) and that of a round sphere ($x^2 + y^2 + z^2 = 1$) are identical as topological or smooth manifolds since they can be continuously or smoothly deformed into each other.

2 Submanifolds, connected and simply connected manifolds

It is tempting to think of a submanifold as a subset of a manifold M that acquires a manifold structure when charts of M are suitably restricted. However, this is a little too restrictive for some purposes. While the unit circle $x^2 + y^2 = 1$ is a submanifold of the Euclidean x-y plane $\mathbb{R}^2$ in this sense, we would like to admit the examples of the cubic curve and 3-petaled rose shown in Fig. 3 (c) and (d) as suitable submanifolds of $\mathbb{R}^2$. On the face of it, these curves are not manifolds due to the self-intersections. However, there is a simple way to view them as images of bona fide manifolds sitting inside ('included in') the plane. Without attempting to be very precise, we outline a framework for the idea of a submanifold. Given an n-dimensional manifold M, an s-dimensional manifold S ($s \leq n$) and an 'inclusion' map $i : S \hookrightarrow M$ we can specify what we mean by immersed and embedded submanifolds.

For example, S could be $\mathbb{R}$, thought of as an infinitely long (or open stretch of) rope and M could be the plane $\mathbb{R}^2$. The inclusion map is some way of laying the rope on the plane. The question is one of whether S sits inside M in a sufficiently nice way. For example, we readily admit the x-axis contained in $\mathbb{R}^2$ and the interval $(0, 1) \subset \mathbb{R}$ as submanifolds. Questions arise, for instance, when the image of S in M involves sharp corners (as in the curve that looks like the character V or the cardioid and cycloid of Fig. 3 (a) and (b)) or self-intersections (as in the curve that looks like α or the 3-petaled rose of Fig. 3(d)). Very roughly, if the tangent to S behaves nicely (sharp corners are absent), we will say that S is an immersed submanifold while we will call it an embedding if it has neither sharp corners nor self-intersections. The manner in which S sits inside M can be encoded in properties of the inclusion map $i : S \hookrightarrow M$ which takes any point $x \in S$ to the corresponding point $i(x) \in M$. If the derivative of the inclusion map, which is the $n \times s$ Jacobian matrix of first partials, has the maximum possible rank[10] s everywhere, then S is said to be an immersed submanifold (this eliminates sharp corners but allows for self-intersections, as in the symbol α or the planar cubic curve $y^2 = x^3 + x^2$ of Fig. 3(c)). Thus, in an immersion, the inclusion map need not be 1-1, though its derivative must be 1-1. An immersion where the inclusion map is also 1-1 (this eliminates self-intersections) is called an embedding.

[10]The row (column) rank of a matrix is the number of linearly independent rows (columns). The rank of a matrix is the larger of its row and column ranks.

The n-sphere $S^n = \{\boldsymbol{x} \in \mathbb{R}^{n+1} \text{ such that } (x^1)^2 + (x^2)^2 + \cdots + (x^{n+1})^2 = 1\}$ is an embedded submanifold of $\mathbb{R}^{n+1}$ for $n = 0, 1, 2, \ldots$. An important theorem of Whitney states that essentially any smooth n-dimensional manifold M (defined as above using an atlas) can be realized as a smoothly embedded submanifold[11] of $\mathbb{R}^{2n}$.

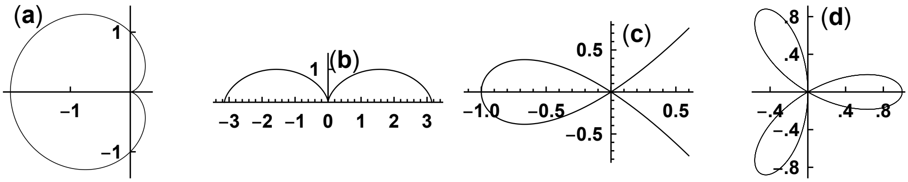


Figure 3: Plane curves (a) cardioid $x = \cos t(1 - \cos t), y = \sin t(1 - \cos t)$ (b) cycloid $x = (t - \sin t)/2, y = (1 - \cos t)/2$ (c) cubic $y^2 = x^3 + x^2$ and (d) 3-petaled rose $r = \cos 3\theta$ given in parametric, implicit and polar forms. The cardioid and cycloid fail to be immersed submanifolds of the plane since the rank of their Jacobians drop from 1 to 0 at the sharp corners. E.g., for the cycloid, the transpose of the Jacobian is $J^t = (\dot{x}, \dot{y}) = ((1 - \cos t)/2, (\sin t)/2) = (0, 0)$ at $t = 2n\pi$ where n is an integer. The cubic curve and the rose are immersions but not embeddings: they have no sharp corners but display self-intersections. If we think of the image curve as the path traced by an ant walking on the plane, the Jacobian is the velocity vector. If the ant does not momentarily come to rest, its path may be modeled as an immersion (self-intersections occur when an ant returns to an earlier location while at a sharp corner, the ant must momentarily come to rest and abruptly change direction). If the curve is the world line of a massive particle in space-time parametrized by proper time, then it must be an embedding since the 4-velocity (**??**) cannot vanish and the world line cannot have self-intersections.

Connected and simply connected manifolds. A manifold is connected if it comes in one piece. For example, the disjoint union of two open real intervals $(0, 1) \cup (2, 3)$ is not connected. To define the concept of connectedness, we imagine a point-like ant walking on the manifold M. If it can reach any point from any other point via a continuous path ($\gamma(t)$ parametrized by time) that lies in M, then M is connected. More precisely, M is path connected if any two points $p, q \in M$ can be joined by a continuous path $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\gamma(1) = q$. For example, $\mathbb{R}^n$ and S^n for $n = 1, 2, 3, \ldots$ are connected. If a manifold is not path connected, then it is called disconnected. The real line punctured[12] at the origin is a disconnected manifold. For a disconnected manifold, the connected component of $p \in M$ is the submanifold consisting of points $q \in M$ that can be reached from p via continuous paths lying in M. The line punctured at the origin has two connected components $(-\infty, 0)$ and $(0, \infty)$. The connected component of the point 2 is the right half-line. Similarly, $S^0 = \{-1, 1\}$ is disconnected, it has two connected components. A connected manifold M is called simply connected if any nontrivial closed curve in M can be continuously deformed (shrunk) to a point while remaining in M. The two-sphere S^2 is simply connected, a rubber band on a globe can always be shrunk to a point while remaining on the globe.

[11] In favorable cases, one may be able to embed the n-dimensional manifold M in a Euclidean space of dimension less than $2n$, as is the case with $S^n \hookrightarrow \mathbb{R}^{n+1}$.

[12] To puncture the line $\mathbb{R}$ is to remove (or excise) one point from it.

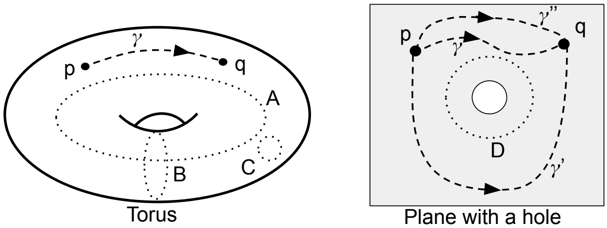


Figure 4: (a) Any two points p, q on a torus can be joined by a continuous path γ: it is path connected. It is not simply connected: the closed curves A, B winding around the torus cannot be shrunk to a point, though C can. C is a *contractible* closed curve or one that is *homotopic* to a point. (b) The plane with a hole (disk excised) is homeomorphic to the once punctured plane. It too is connected but not simply connected: the closed curve D cannot be continuously shrunk to a point. It is *multiply connected*: there are many paths from p to q that are not continuously deformable into each other: direct (γ), around the hole (γ'), winding twice around the hole, etc. Though γ is *not* homotopic to γ' (they cannot be continuously deformed into each other), γ and γ'' are homotopic to each other (a rubber band stretched from p to q along γ can be deformed to γ''). A homotopy between $\gamma, \gamma'' : [a, b] \to M$ is a continuous map $\Gamma : [a, b] \times [0, 1] \to M$ with $\Gamma(t; 0) = \gamma(t)$ and $\Gamma(t; 1) = \gamma''(t)$.

On the other hand, S^1, the torus, the surface of an infinite circular cylinder and the punctured plane are all connected but not simply connected (see Fig 4).

3 Smooth functions or scalar fields

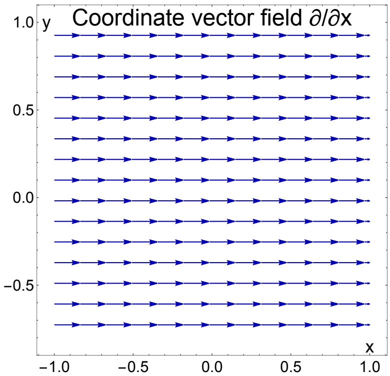
In essence, a smooth real function f on a smooth manifold M is a way of assigning a smoothly varying real number to each point on the manifold. Smooth functions $f : M \to \mathbb{R}$ are also called scalar fields. They are smooth real-valued functions of the coordinates in any given patch with the consistency condition that the value of the function at a point $p \in M$ must be the same irrespective of which coordinate system is used to describe p, in the event that p lies in the intersection of coordinate patches. In other words, if the function is described by the formulae $f(x)$ and $g(y)$ in two coordinate patches, then we must have $g(y(x)) = f(x)$ at each point p of the overlap. Sometimes, we turn this around and say that given a scalar field $f(x)$ in one coordinate system, under a change from $x \mapsto y$ given by the transformation $y = y(x)$ (and its inverse $x = x(y)$), the formula for the function becomes $F(y) = f(x(y))$. Henceforth, we will mostly take this second viewpoint and speak in terms of how objects transform under a change of coordinates in a region of overlap between two coordinate patches. The space of smooth functions on M is denoted $C^\infty(M)$ or $\mathcal{F}(M)$. If M is the phase space of a classical system, then the set of observables is given by $\mathcal{F}(M)$. It is a commutative algebra: it is closed under real linear combinations $af + bg \in \mathcal{F}$ and pointwise products $(fg)(x) = f(x)g(x) = g(x)f(x) = (gf)(x)$ for all $f, g \in \mathcal{F}(M)$, with products distributing over sums. The property $fg = gf$ encodes the fact that classical observables commute under multiplication, a feature that is not always true in the quantum theory, where observables are hermitian operators.

4 Vector fields

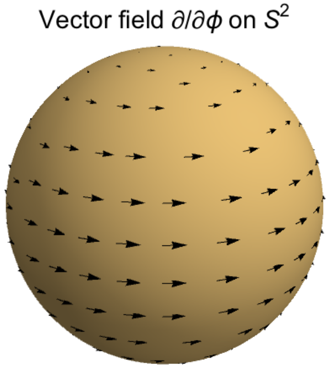
Given local coordinates x^i in a chart on an n-dimensional manifold M, we have the notion of coordinate vector fields. These are defined as the first order partial differential operators $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \cdots, \frac{\partial}{\partial x^n}$, which are often abbreviated ∂_{x^i} or ∂_i for $i = 1, \cdots, n$. Geometrically, we may think of the coordinate vector fields at a point p as tangent vectors to M at p. For instance, ∂_1 is the tangent vector to the coordinate curve parametrized by x^1 passing through p holding $x^2, \cdots, x^n$ fixed. As a consequence, ∂_x is a tangent vector field on $\mathbb{R}^2$ that points rightward at every point, as shown in Fig. 5a. These coordinate vector fields furnish a basis for more general vector fields on M. A general vector field is given by a linear combination

$$v = \sum_{i=1}^{n} v^i(x) \frac{\partial}{\partial x^i} \equiv v^i(x) \partial_i. \tag{1}$$

A vector field restricted to a point $p \in M$ is called a tangent vector at p. The set of tangent vectors at p is the tangent space $T_p(M)$, a real vector space of dimension n. The coordinate tangent vectors $\partial_1, \cdots, \partial_n$ at p furnish a basis for $T_p(M)$. E.g., the tangent space to the 2-sphere at a point on the equator may be visualized as a vertical tangent plane spanned by the coordinate tangent vectors $\frac{\partial}{\partial \phi}$ and $\frac{\partial}{\partial \theta}$ (see Fig. 5b).



(a)



(b)

Figure 5: (a) Coordinate vector field ∂_x on the x-y plane. (b) Azimuthal coordinate vector field $\partial_\phi = -y\partial_x + x\partial_y$ on the unit sphere with the z-axis pointing vertically upwards. At the North and South poles $x = y = 0$ and $z = \pm 1$. At the poles, ∂_ϕ vanishes, they are zeros of ∂_ϕ. In fact, there is no nonvanishing smooth vector field on a sphere: loosely speaking, it is not possible to comb hair on the sphere. Here, a smooth distribution of hair combed tangent to a sphere may be regarded as a vector field on the sphere. The vector field has a zero at a bald spot where there is no hair.

The set of n functions $v^i(x)$ are called the components of v in the coordinate basis. Though each is a function within a coordinate patch, they do not transform as scalar functions under a change of coordinates. The components of a vector field have a special transformation law that follows from the chain rule in multivariable calculus.

Suppose the same vector field is expressed in another coordinate system y^i:

$$v = \tilde{v}^j(y)\frac{\partial}{\partial y^j}. \tag{2}$$

Since $y = y(x)$, we may relate the two sets of coordinate vector fields via a Jacobian:

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j} = J^j_i \frac{\partial}{\partial y^j} \quad \text{so that} \quad v = v^i \frac{\partial}{\partial x^i} = v^i \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j}. \tag{3}$$

Comparing with (2) we find how the components of a vector field transform:

$$\tilde{v}^j(y) = v^i(x(y))\frac{\partial y^j}{\partial x^i} \quad \text{or} \quad \tilde{v}^j = J^j_i v^i. \tag{4}$$

Thus, the components of a vector field transform via the Jacobian matrix[13]: the new $j^{\rm th}$ component is a linear combination of all the old components (quite unlike how n scalar fields would transform). Such a transformation is called *contravariant*. The prefix *contra*[14] arises from the manner in which the coordinate vector fields transform, i.e., via the inverse of the Jacobian matrix[15]:

$$\frac{\partial}{\partial y^j} = \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial x^i} = (J^{-1})^i_j \frac{\partial}{\partial x^i}. \tag{5}$$

Thus, tangent vector fields are also called contravariant vector fields. A vector field on a smooth manifold is called smooth if the components $v^i(x)$ are smooth functions of the coordinates in each patch. The matrix elements J^j_i entering the transformation formula for v^i between overlapping coordinate patches are automatically smooth since the manifold is smooth.

A vector field can act on a differentiable function on M and give its derivative along the vector field:

$$v(f) = v^i \frac{\partial f}{\partial x^i}. \tag{6}$$

$v(f)$ is another function on M and generalizes the concept of the directional derivative[16] $\boldsymbol{v} \cdot \boldsymbol{\nabla} f$ from vector calculus in $\mathbb{R}^3$. Evidently, a vector field acts linearly on the space of functions: $v(af + bg) = av(f) + bv(g)$ for any pair of scalar fields f, g and real numbers a, b. Since a vector field is a first order differential operator, v acts as a derivation on the space of functions: verify that the Leibniz rule

[13] We may view this as the product of a matrix with a column vector by regarding J^j_i as the entry in the $j^{\rm th}$ row and $i^{\rm th}$ column of a square matrix J and v^i as the element in the $i^{\rm th}$ row of a column vector.

[14] In §5 we will meet covector fields. Coordinate covector fields (9) transform via J rather than J^{-1}, so covector fields are called *covariant*.

[15] The Jacobians for $x \mapsto y$ and $y \mapsto x$ are inverse matrices. This is seen by using the chain rule to differentiate $x^i(y(x)) = x^i$ with respect to x^k to get $\frac{\partial x^i}{\partial y^j}\frac{\partial y^j}{\partial x^k} = \delta^i_k$.

[16] $v(f)$ is also called the Lie derivative of f along v and denoted $\mathcal{L}_v f$. To find $\mathcal{L}_v f$ at a point $x_0 \in M$, we consider the integral curve $x(t)$ (7) of v through x_0 with $x(0) = x_0$. Then $\mathcal{L}_v f = \lim_{s\to 0}[\{f(x(s)) - f(x(0))\}/s]$.

$v(fg) = fv(g) + v(f)g$ is satisfied. Thus, we may view a vector field simply as a linear map from $\mathcal{F}(M) \to \mathcal{F}(M)$ that satisfies the Leibniz rule. The set of vector fields on M is denoted Vect(M), it is an infinite-dimensional real vector space since the coefficients v^i can be arbitrary smooth functions in any given patch. For instance, a large class of vector fields on the real line can be written as $\sum_{l=0}^{\infty} c_l x^l \partial_x$ where $c_0, c_1, \ldots$ are suitable real coefficients. Vect(M) may also be viewed as a module[17] over the ring of smooth functions on M: $fv + gw \in$ Vect(M) if $v, w \in$ Vect(M) and $f, g \in \mathcal{F}(M)$. In other words, we may multiply a vector field by a scalar field to get another vector field.

Integral curves of a vector field. Given a vector field v on a manifold, it defines a flow on the manifold. By this, we mean that there is a family of curves on M that are everywhere tangent to v. Precisely, the integral curve through the point $x_0 \in M$ is the solution $x^i(t)$ to the system of first order ODEs

$$\frac{dx^i}{dt} = v^i(x) \quad \text{with} \quad x^i(0) = x_0. \tag{7}$$

For examples, figures and much more on integral curves of vector fields, see Chapt. **??**.

Commutator of vector fields. Given a pair of differentiable vector fields on a manifold M, we may define their commutator, which is another vector field. In local coordinates, suppose $u = u^i \partial_i$ and $v = v^i \partial_i$. Then their commutator $[u, v]$ is

$$[u, v] = (u^j \partial_j v^i - v^j \partial_j u^i)\partial_i. \tag{8}$$

By making a change of coordinates, one may check that this first order differential operator transforms as a contravariant vector field. Given a function $f : M \to \mathbb{R}$, both $u(v(f))$ and $v(u(f))$ are functions on M. The commutator $[u, v]f$ measures the extent to which the two differ (see Prob. **??**). Notably, $u(v(f))$ and $v(u(f))$ involve both first and second order derivatives, so the composition of vector fields is not a vector field. Pleasantly, we verify that these second order derivatives cancel out in the commutator. The latter is also called the Lie bracket of vector fields since it is linear[18] $[au + bv, w] = a[u, w] + b[v, w]$, antisymmetric $[u, v] + [v, u] = 0$ and satisfies the Jacobi identity $[u, [v, w]] + [w, [u, v]] + [v, [w, u]] = 0$ (Prob. **??**) for any three vector fields u, v and w and real numbers a, b. Consequently, the linear space of vector fields

[17] A module over a ring is a generalization of the concept of a vector space over a field like the real or complex numbers. Here, a field is an algebraic concept distinct from the differential geometric concepts of scalar, vector and tensor fields. Multiplication, addition and division by nonzero elements are defined in a field. Elements of the ring, like those of a field can be added and multiplied in a manner satisfying the distributive law. A key difference is that not all nonzero elements of the ring may have reciprocals (multiplicative inverses) and division is therefore not defined. The integers $\mathbb{Z}$ form a ring but not a field. For the ring of functions, only those functions that are nowhere zero have reciprocals. As with vectors in a vector space, elements of a module can be multiplied (say from the left) by scalars that come from the ring, with multiplication distributing over addition. Our main example is the module of vector fields over the commutative ring of functions on a manifold. We cannot always divide a vector field by a function since the latter may have zeros. Sometimes, we may circumvent this by adding the 'point at infinity', define $1/0 = \infty$ and turn the ring of rational functions into a field.

[18] Note that $[fu, v] \neq f[u, v]$ in general for a nonconstant smooth function f, see Prob. **??**.

$\mathrm{Vect}(M)$ equipped with the commutator Lie bracket is called a real Lie algebra (see §**??** for a definition and more examples of Lie algebras). The commutator $[u, v]$ is also called the Lie derivative of v along u and is written $\mathcal{L}_u v = [u, v]$. From (8), we see that the Lie derivative[19] of v along u includes two contributions: the first is the 'obvious' change in v in the direction of u while the second accounts for the fact that the components of u themselves change with location. Finally, we note that given a smooth function f, the Lie derivative of a vector field satisfies the Leibniz rule: $\mathcal{L}_u(fv) = (\mathcal{L}_u f)v + f\mathcal{L}_u v$, as we verify in Prob. **??**.

5 Covector fields or 1-forms

On the Euclidean plane, the differentials dx and dy are examples of covector fields or 1-forms. They are to be thought of as dual to the coordinate vector fields $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ via the 'pairing' $dx(\partial_x) = 1, dx(\partial_y) = 0, dy(\partial_x) = 0$ and $dy(\partial_y) = 1$ which is defined to be linear: for instance, $dx(f\partial_x + g\partial_y) = f dx(\partial_x) + g dx(\partial_y) = f(x, y)$ for any two smooth functions f and g. A general covector field is a linear combination $\phi = a(x, y)dx + b(x, y)dy$ where a and b are smooth functions. A 1-form is also called a Pfaffian differential expression after the German mathematician J F Pfaff who studied equations[20] of the form $a(x, y, z)dx + b(x, y, z)dy + c(x, y, z)dz = 0$. Physically, for a particle moving on a plane, while the velocity $\dot{q}(t) = \dot{q}^1(t)\partial_x + \dot{q}^2(t)\partial_y$ is a tangent vector at each point $(x(t), y(t))$ on a trajectory, the momentum $p(t) = p_1 dx + p_2 dy$ is a covector at each such point (see §**??**) on the configuration plane. Another example of a 1-form is the Liouville 1-form $pdq = p_i dq^i$. This is to be regarded as a 1-form on the phase space of a mechanical system that is furnished with canonical (Darboux) coordinates (q^i, p_j). Covector fields are also encountered on the thermodynamic state space. According to the 1$^{\text{st}}$ law of thermodynamics, the infinitesimal heat added to a gas is given by the action of the heat 1-form $\phi = dU + pdV$ (U, p and V are the internal energy, pressure and volume of the gas) on the tangent vector representing the infinitesimal process[21]. If the process is reversible, the second law postulates that $\phi = TdS$, where T, S are the absolute temperature and entropy.

More generally, a covector field or covariant vector field or 1-form is simply a (smoothly varying) assignment of a covector at each point of a manifold. In more detail, given local coordinates x^i in a patch, we have the coordinate basis 1-forms

[19] It is tempting to mimic the geometric approach to the Lie derivative of a function given in Footnote 16 to define $\mathcal{L}_u v$ as the $s \to 0$ limit of a difference quotient $\{v(x(s)) - v(x(0))\}/s$, where $x(t)$ is the integral curve (7) of u through x_0 with $x(0) = x_0$. However, there is a difficulty since $v(x(s))$ and $v(x(0))$ live in different tangent spaces and cannot be subtracted. One needs a way to 'push' one of the vectors to the tangent space where the other lives before subtracting. This can be done using the push forward defined in §7.

[20] The Pfaffian differential equation $\phi = a\,dx + b\,dy + c\,dz = 0$ is said to be integrable if it admits an integrating denominator $T(x, y, z)$ (or integrating factor $1/T$) such that $\phi/T = dS$ is an exact differential. Then $dS = 0$ and the solutions of the Pfaffian differential equation are given by $S(x, y, z) = \sigma$ for any constant σ.

[21] The thermodynamic state space is a 3d manifold with coordinates U, V, S. An infinitesimal process is represented by a tangent vector $v = a\partial_U + b\partial_V + c\partial_S$. The work done in this infinitesimal process is $pdV(v) = pb$ and the heat added is $a + pb$. Equilibrium states form a 2d hypersurface determined by an equation of state (EOS). Tangent vectors to this EOS surface represent infinitesimal reversible processes.

given by the differentials of the coordinates $dx^1, \cdots, dx^n$. At a point $p \in M$, the basis 1-forms $dx^i(p)$ are said to span the cotangent space to M at p. The cotangent space is denoted $T_p^*(M)$ and is the vector space dual to the tangent space $T_p(M)$. Indeed, $\{dx^i\}$ is the dual basis to $\{\partial_i\}$ defined via the pairing $dx^i(\partial_j) = \delta^i_j$. In general, a covector field on M is a linear combination of the basis 1-forms $\phi = \phi_i(x) dx^i$. The n real-valued quantities $\phi_i(x)$ in a coordinate patch are called the components of the covector field. As with the components of a vector field, they are not scalar functions on M but satisfy a special transformation law. Indeed, suppose y^j is another local coordinate system defined on a chart that has an overlap with that of the x^i. On the overlap, the coordinate 1-forms are related by the chain rule

$$dy^j = \frac{\partial y^j}{\partial x^i} dx^i = J^j_i \, dx^i. \tag{9}$$

We see that coordinate 1-forms transform via the Jacobian matrix (as opposed to its inverse, as was the case for coordinate vector fields in (5)). For this reason, covector fields are called covariant vector fields. Now suppose the same covector field ϕ is expressed in the y basis: $\phi = \tilde{\phi}_j(y) dy^j = \tilde{\phi}_j J^j_i \, dx^i$. Comparing, we see that the components of a covector field transform via the inverse of the Jacobian[22]:

$$\phi_i = \tilde{\phi}_j \frac{\partial y^j}{\partial x^i} \quad \text{or} \quad \tilde{\phi}_j = (J^{-1})^i_j \phi_i. \tag{10}$$

Compare this with the corresponding formula (4) for components of a vector field. If the components of ϕ in all charts are smooth functions of the local coordinates on a smooth manifold, then ϕ is called a smooth covector field. Since covectors are dual to vectors at each point of M, covector fields are linear functions on the space of vector fields. The value of $\phi = \phi_i dx^i$ on the vector field $v = v^j \partial_j$ is the smooth function or scalar field

$$\phi(v) = \phi_i dx^i(v^j \partial_j) = \phi_i v^j dx^i(\partial_j) = \phi_i v^j \delta^i_j = \phi_i(x) v^i(x). \tag{11}$$

We used linearity of the action of a covector on a vector to pull the components $v^j(x)$ out. $\phi(v)$ is called the contraction of ϕ with v. More generally, for vector fields v, w,

$$\phi(fv + gw) = f\,\phi(v) + g\,\phi(w) \quad \text{for any} \quad f, g \in \mathcal{F}(M). \tag{12}$$

The space of covector fields on M is denoted $\Omega^1(M)$ and is dual to Vect(M) over $\mathcal{F}(M)$. In particular, if ϕ and ψ are 1-forms and f, g scalar fields, then $f\phi + g\psi$ is also a 1-form. Note that $f\phi = \phi f$, the order does not matter.

On the other hand, we can also evaluate a vector field v on a 1-form ϕ to get a scalar field. In fact, since they are dual bases, we also have $\partial_i(dx^j) = \delta^j_i$ so that $v(\phi) = v^j \partial_j(\phi_i dx^i) = v^i \phi_i = \phi(v)$. This allows us to reinterpret vector fields

[22] If we view $(J^{-1})^i_j$ as the entry in the i^{th} row and j^{th} column of a matrix and ϕ_i as the entry in the i^{th} column of a row vector, then this is the product (from the left) of a row vector with the square matrix J^{-1} producing the row vector with j^{th} column entry $\tilde{\phi}_j$.

as linear functions on the space of 1-forms. This viewpoint will soon be useful in generalizing vector fields to contravariant tensor fields.

An important class of 1-forms are **differentials of functions** on M: $\phi = df = \frac{\partial f}{\partial x^i} dx^i$. So the partial derivatives of a function should be thought of as components of a covector rather than a vector. Sometimes, it is convenient to regard functions as covector fields of degree zero and call $\mathcal{F}(M)$ $\Omega^0(M)$. Thus, the differential d is a linear map (over $\mathbb{R}$) from $\Omega^0(M)$ to $\Omega^1(M)$ satisfying the Leibniz rule.

The tangent space $T_p(M)$ and the cotangent space $T_p^*(M)$ are both n-dimensional real vector spaces and are therefore isomorphic. However, there is no preferred or canonical isomorphism between them. If a basis, such as a coordinate basis is chosen for vector fields then one gets an isomorphism that maps ∂_i to dx^i and vice versa. However, this depends on the choice of coordinates[23]. Thus, given a smooth manifold, there is no distinguished or natural way to relate vectors to covectors, there are many ways to do this, but none of them is special. The situation changes if the manifold is equipped with a metric tensor. In this case, there is a standard way (called *lowering an index*) of mapping vectors to covectors, which does not depend on the coordinates chosen, as we will see in §6.

6 Tensors of rank two and 2-forms

Vector fields $v = v^i \partial_i$ are called contravariant tensor fields of rank one (or of type (1,0) as their components (v^i) have one upper index), while 1-forms are called covariant tensor fields of rank one (or of type (0,1)). More generally, one may define tensors of higher rank.

At a point $p \in M$ lying in a patch with local coordinates x^i, we may consider the tensor product of the tangent space with itself $T_p(M) \otimes T_p(M)$. This is the space of dimension n^2 with basis consisting of $\partial_i \otimes \partial_j$ for $1 \leq i, j \leq n$. A type (2,0) tensor field or second rank contravariant tensor field is then a linear combination

$$t = t^{ij}(x) \partial_i \otimes \partial_j. \tag{13}$$

Without further ado, we note that upon changing coordinates $x \mapsto y$, the components t^{ij} transform via the Jacobian matrix, just as for contravariant vector fields, except that there are now two Jacobian factors

$$t = \tilde{t}^{kl} \frac{\partial}{\partial y^k} \otimes \frac{\partial}{\partial y^l} \quad \text{where} \quad \tilde{t}^{kl} = J_i^k J_j^l t^{ij} \quad \text{and} \quad J_i^k = \frac{\partial y^k}{\partial x^i}. \tag{14}$$

Just as vector fields act linearly on 1-forms to produce functions, second rank con-

[23]For example, on $\mathbb{R}$ with coordinate x, we have an isomorphism mapping $dx \leftrightarrow \partial_x$. If we change to a new coordinate $y = 2x$, then $dy = 2dx$ and $\partial_y = \frac{1}{2}\partial_x$. The new isomorphism between cotangent and tangent spaces $dy \leftrightarrow \partial_y$ is *different* since it takes dx to $\frac{1}{4}\partial_x$. Thus, there are many isomorphisms between the spaces of covectors and vectors, none of which can be considered coordinate-independent or standard.

travariant tensors act bilinearly[24] on a pair of 1-forms to produce functions:

$$t(\phi,\psi) = t^{ij}\partial_i \otimes \partial_j(\phi_k dx^k, \psi_l dx^l) = t^{ij}\phi_k\psi_l\partial_i(dx^k)\partial_j(dx^l) = t^{ij}\phi_i\psi_j. \quad (15)$$

Poisson tensor. A physically important example of a (2,0) tensor is the Poisson tensor on the phase space of a mechanical system: $r = r^{ij}\partial_i \otimes \partial_j$, which has the further property of antisymmetry: $r^{ij} = -r^{ji}$. The Poisson bracket of a pair of smooth functions (observables) is the function $\{f, g\} = r(df, dg) = r^{ij}\partial_i f \partial_j g$. For a particle moving on a line, $M = \mathbb{R}^2$ with canonical coordinates $\xi = (q, p)$ and $r^{ij} = (0, 1| - 1, 0)$. Given a Hamiltonian function H on phase space, the Poisson tensor allows us to define the Hamiltonian vector field V_H. It is the vector field which acts on any 1-form ϕ via $V_H(\phi) = r(\phi, dH)$. The Hamiltonian vector field defines time evolution of any observable through $\dot{f} = V_H(df)$. Trajectories on phase space are the integral curves of V_H. They are governed by the ODEs $\dot{\xi}^i = V_H^i = r^{ij}\partial_j H$. For the canonical Poisson tensor on $\mathbb{R}^2$, they reduce to Hamilton's canonical equations $\dot{\xi}^1 = \dot{q} = r^{12}\partial_2 H = \frac{\partial H}{\partial p}$ and $\dot{\xi}^2 = \dot{p} = r^{21}\partial_1 H = -\frac{\partial H}{\partial q}$.

Similarly, we have covariant tensor fields of rank two or tensors of type (0,2):

$$t = t_{ij} dx^i \otimes dx^j, \quad (16)$$

which are linear combinations of the tensor products of the coordinate basis covector fields. Such a tensor transforms via two factors of the inverse Jacobian:

$$\tilde{t}_{kl} = (J^{-1})^i_k (J^{-1})^j_l \, t_{ij}. \quad (17)$$

In summary, each upper index on a tensor transforms via J and each lower one via J^{-1}. Covariant tensors of rank two can act on a pair or vector fields and produce a scalar function, they are bilinear maps from Vect$(M)\times$ Vect(M) to $\mathcal{F}(M)$.

Metric tensor. An important example of a 2$^{\text{nd}}$ rank covariant tensor field is the metric tensor $g = g_{ij} dx^i \otimes dx^j$, which has the further property of being symmetric $g_{ij} = g_{ji}$ and nondegenerate [g_{ij} an invertible matrix]. A metric allows us to generalize the concept of the dot product of vectors in Euclidean space to tangent vectors at a point p on a manifold M. We encounter it as the mass metric in a Lagrangian quadratic in velocities (**??**). A metric is called Riemannian if g_{ij} is a positive-definite matrix at every point on the manifold. If the positive-definiteness condition is dropped, it is called pseudo-Riemannian (this case includes the Lorentzian metric tensor of space-time; e.g., Minkowski space in Cartesian coordinates $x^\mu = (ct, x, y, z)$ has the metric given by the constant diagonal matrix $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ where $\mu, \nu = 0, 1, 2, 3$). A virtue of an invertible metric tensor is that it defines an isomorphism from vectors to covectors: $v \mapsto v'$ where $v'_i = g_{ij}v^j$. The inverse metric with components g^{ij} maps covectors to vectors $g^{ij}v'_j = v^i$. Thus, on a Riemannian manifold, the

[24] $t(\phi, \psi)$ is bilinear if it is linear in *both* the entries. For instance, $t(f\phi_1 + g\phi_2, \psi) = ft(\phi_1, \psi) + gt(\phi_2, \psi)$ for any functions f, g and 1-forms ϕ_1, ϕ_2, ψ. Bilinearity is a consequence of the definition of a dual space: vector fields are dual to 1-forms and act linearly on 1-forms (as discussed in §5). So pairs of vectors fields (written as a tensor product $\partial_i \otimes \partial_j$) act linearly on pairs of 1-forms (ϕ, ψ).

tangent and cotangent spaces are canonically isomorphic. We say that the metric and its inverse can be used to lower and raise indices. In particular, we may use the inverse metric g^{ij} to define the *gradient of a function* (see §4) by raising the index of the components of the 1-form df: $(\text{grad}\, f)^i = (\boldsymbol{\nabla} f)^i = g^{ij}\partial_j f$.

A metric tensor gives a manifold a rigid geometric shape. The square of the length of the vector $v = v^i\partial_i \in T_pM$ is defined as

$$g(v,v) = g_{ij}dx^i \otimes dx^j(v^k\partial_k, v^l\partial_l) = g_{ij}v^iv^j. \tag{18}$$

Given a pair of tangent vectors $u, v \in T_pM$, their inner product is defined as $g(u,v) = g_{ij}u^iv^j$. The cosine of the angle between them is $g(u,v)/\sqrt{g(u,u)g(v,v)}$.

Two-forms. An antisymmetric second rank covariant tensor is called a 2-form. To make the antisymmetry manifest, one defines the wedge product[25]

$$dx^i \wedge dx^j = dx^i \otimes dx^j - dx^j \otimes dx^i \tag{19}$$

and writes a 2-form with antisymmetric components ω_{ij} as (show the 2^{nd} equality!)

$$\omega = \omega_{ij}dx^i \otimes dx^j = \frac{1}{2}\omega_{ij}dx^i \wedge dx^j. \tag{20}$$

Note that $dx^1 \wedge dx^1 = 0$, etc. Geometrically, two-forms are related to area elements in a manifold. The familiar area 'element' $dx\,dy$ on a plane is more precisely the 2-form $dx \wedge dy$. The antisymmetry of the wedge product allows us to encode the orientation of the area element, which in vector calculus is conveyed by the inward/outward normal in an 'infinitesimal area vector' $dx\,dy\,\hat{\boldsymbol{n}}$ on a surface parametrized by x and y. Moreover, in Euclidean space $\mathbb{R}^3$ with Cartesian coordinates, the components of the wedge product of 1-forms df and dg are related to those of the cross product $\boldsymbol{\nabla} f \times \boldsymbol{\nabla} g$ whose magnitude measures the area of a parallelogram.

The space of 2-forms is denoted $\Omega^2(M)$. Recall that functions can also be regarded as 0-forms and that we could go from functions to 1-forms by taking the differential: $df = (\partial_i f)dx^i$. The differential of a function is also called its exterior derivative. Interestingly, there is a similar way of going from 1-forms to 2-forms by (exterior) differentiation. Given a 1-form $\phi = \phi_j dx^j$, we *define* its exterior derivative

$$\omega = d\phi = d\phi_j \wedge dx^j, \tag{21}$$

which is a 2-form. To find its components we write

$$d\phi = \frac{\partial \phi_j}{\partial x^i}dx^i \wedge dx^j = \frac{1}{2}(\partial_i\phi_j - \partial_j\phi_i)dx^i \wedge dx^j \quad \text{whence} \quad \omega_{ij} = \partial_i\phi_j - \partial_j\phi_i. \tag{22}$$

[25] The wedge product can be written as a sum over permutations of two objects: $dx^1 \wedge dx^2 = \sum_{\sigma \in S_2} \text{sgn}(\sigma) dx^{\sigma(1)} \otimes dx^{\sigma(2)}$. Here S_2 is the permutation or symmetric group consisting of two elements, the identity $(\sigma(1) = 1, \sigma(2) = 2)$ and the exchange transposition $(\sigma(1) = 2, \sigma(2) = 1)$. $\text{sgn}(\sigma)$ is the sign of the permutation: -1 to the power of the number of pairwise transpositions needed to write σ as a product of exchanges. The identity has sign $+1$ and the exchange has sign -1.

We used the antisymmetry of the wedge product in the second step, relabelled indices and used the definition (20) to identify the antisymmetric tensor ω_{ij}.

However, unlike ordinary differentiation that can be done repeatedly to produce higher order derivatives of a function, the square of the exterior derivative vanishes[26]. Indeed, using the definition in (21), the exterior derivative of the 1-form df is

$$d(df) = d(\partial_j f)dx^j = (\partial_i \partial_j f)dx^i \wedge dx^j = 0. \tag{23}$$

Here $\partial_i \partial_j f$ is symmetric under $i \leftrightarrow j$ exchange due to the equality of mixed partials, while the wedge product $dx^i \wedge dx^j$ is antisymmetric, so the sum vanishes. Thus, $d^2 f = 0$. This identity is a generalization of the vector identity $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = 0$ valid for real-valued functions on $\mathbb{R}^3$.

Just as a 1-form acts linearly on vector fields to produce functions $\phi(v) = \phi_i v^i$, a 2-form acts as a skew-symmetric bilinear map from pairs of vector fields to $\mathcal{F}(M)$:

$$\omega(u,v) = \omega_{ij} dx^i \otimes dx^j (u^k \partial_k, v^l \partial_l) = \omega_{ij} dx^i (u^k \partial_k) dx^j (v^l \partial_l) = \omega_{ij} u^i v^j. \tag{24}$$

Here, the 1st (2nd) factor in a tensor product acts on the 1st (2nd) entry of the ordered pair (u, v). We used linearity of the action of forms on vector fields (11) and the pairing $dx^i(\partial_k) = \delta^i_k$.

A 2-form can be used to define an area for infinitesimal parallelograms in each tangent space to a manifold. For example, if ∂_i, ∂_j are two coordinate tangent vectors at x, then the area of the parallelogram they span is defined as $\omega(\partial_i, \partial_j) = \omega_{ij}(x)$.

Physical examples of 2-forms. (i) An interesting example of a 2-form is the electromagnetic field strength tensor F, which is a 2-form on the 4-dimensional Minkowski space-time. It is the exterior derivative of the 1-form ‘gauge potential’:

$$A = A_\mu dx^\mu \quad \text{and} \quad F = dA = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \quad \text{where} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{25}$$

as in (22). Here $\mu, \nu = 0,1,2,3$ and $A_\mu = (\phi, -\boldsymbol{A})$ is a combination of the scalar and vector potentials of electrodynamics. The electric and magnetic fields appear as the components of F. (ii) An example of a 1-form in mechanics is the so-called canonical or Liouville 1-form on the $2n$-dimensional phase space $M = \mathbb{R}^{2n}$ of a system with n-degrees of freedom:

$$\alpha = p_i dq^i = p_1 dq^1 + p_2 dq^2 + \cdots + p_n dq^n. \tag{26}$$

$\xi = (q^1, \cdots, q^n, p_1, \cdots p_n)$ together furnish coordinates on M. Notice that α has no components along the dp_i. Its exterior derivative is a 2-form

$$\begin{aligned} \omega &= d\alpha = dp_i \wedge dq^i = -dq^i \wedge dp_i = -(dq^1 \wedge dp_1 + \cdots + dq^n \wedge dp_n) \\ &= \frac{1}{2}(-dq^1 \wedge dp_1 - \cdots - dq^n \wedge dp_n + dp_1 \wedge dq^1 + \cdots + dp_n \wedge dq^n). \end{aligned} \tag{27}$$

[26] We say that the exterior derivative is nilpotent of degree two: $d^2 = 0$

From this we may read off the (antisymmetric) components of $\omega = \frac{1}{2}\sum_{a,b=1}^{2n} \omega_{ab} d\xi^a \wedge d\xi^b$. The only nonzero ones are:

$$\omega_{i,n+i} = -1 \quad \text{and} \quad \omega_{n+i,i} = 1 \quad \text{for} \quad i = 1, 2, \ldots, n. \tag{28}$$

ω is called the canonical symplectic 2-form (the inverse of the canonical Poisson tensor). E.g., for one degree of freedom, $\alpha = pdq$ and

$$\omega = \frac{1}{2}(-dq\wedge dp + dp\wedge dq) = \frac{1}{2}(\omega_{11}dq\wedge dq + \omega_{12}dq\wedge dp + \omega_{21}dp\wedge dq + \omega_{22}dp\wedge dp) \tag{29}$$

so that $\omega_{12} = -\omega_{21} = -1$ and $\omega_{11} = \omega_{22} = 0$ and $\omega = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Referring back to our discussion of the Poisson tensor earlier in this section, we observe that given a Hamiltonian function H on phase space, the Hamiltonian vector field is defined via $\omega(\cdot, V_H) = dH(\cdot)$. In components, $\omega_{ab}V_H^b = \partial_a H$ or inverting, $V_H^c = r^{ca}\partial_a H$.

Mixed second rank tensors. Aside from contravariant and covariant tensors, we also have mixed second rank tensors[27] of type (1,1): $t = t^i_j \partial_i \otimes dx^j$. They transform via one Jacobian and one inverse Jacobian factor: $\tilde{t}^k_l = J^k_i (J^{-1})^j_l t^i_j$. A (1,1) tensor restricted to a point $p \in M$ can be viewed as a linear transformation on the tangent space $T_p(M)$. Indeed, contracting it with a tangent vector gives another tangent vector:

$$t(\cdot, v) = t^i_j \partial_i(\cdot) dx^j (v^k \partial_k) = t^i_j v^j \partial_i(\cdot) \quad \text{or} \quad v^i \mapsto v'^i = t^i_j v^j. \tag{30}$$

The $\cdot$ is a placeholder for an unspecified 1-form that t could act on via the first slot. Similarly, $t^i_j v^j \partial_i(\cdot)$ is the action of a vector field on an unspecified 1-form. The components t^i_j of a (1,1) tensor define a matrix in the coordinate basis, and the above transformation rule written in matrix notation, $\tilde{t} = JtJ^{-1}$ is just a similarity transformation!

7 Higher rank tensor fields & forms

More generally, we have tensor fields of type (p, q) for $p, q \geq 0$ which, in local coordinates, are given by the linear combinations

$$t = t^{i_1 \cdots i_p}_{j_1 \cdots j_q} \partial_{i_1} \otimes \cdots \otimes \partial_{i_p} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_q}. \tag{31}$$

Their components transform via p factors of J for the upper indices and q factors of J^{-1} for the lower indices. Such a tensor field can act linearly on p one-forms and q vector fields to produce a function: $t(\phi, \psi, \cdots, u, v, \cdots) = t^{i_1 \cdots i_p}_{j_1 \cdots j_q} \phi_{i_1} \psi_{i_2} \cdots u^{j_1} v^{j_2} \cdots$. Thus, algebraically, (p, q) tensor fields are simply multilinear maps from p copies of $\Omega^1(M)$ and q copies of Vect(M) to the space of smooth functions on M.

[27] We have arbitrarily chosen to place the ∂_i ahead of the dx^j in the tensor product. The opposite order can also be followed throughout.

Of particular importance are the p-forms, which are covariant antisymmetric tensor fields[28] of rank p (or of type $(0,p)$),

$$\omega = \omega_{i_1 \cdots i_p} dx^{i_1} \otimes \cdots \otimes dx^{i_p}. \tag{32}$$

By antisymmetric, we mean that the components are antisymmetric under interchange of *any* pair of indices. As a consequence, a p-form on an n dimensional manifold must be identically zero if $p > n$ (at least one basis 1-form must appear twice in the tensor product, which when contracted with an antisymmetric coefficient, must vanish). They can be written as linear combinations of p-fold wedge products of coordinate 1-forms, which are obtained by antisymmetrizing the p-fold tensor product:

$$\omega = \frac{1}{p!} \omega_{i_1 \cdots i_p} \, dx^{i_1} \wedge \cdots \wedge dx^{i_p}. \tag{33}$$

For instance, a three-fold wedge product is a sum over all permutations of three objects (which comprise the symmetric group[29] S_3) weighted by the signs of the permutations (see Footnote 25):

$$\begin{aligned} dx^1 \wedge dx^2 \wedge dx^3 &= \sum_{\sigma \in S_3} \operatorname{sgn}(\sigma) dx^{\sigma(1)} \otimes dx^{\sigma(2)} \otimes dx^{\sigma(3)} \\ &= dx^1 \otimes dx^2 \otimes dx^3 - dx^2 \otimes dx^1 \otimes dx^3 - dx^1 \otimes dx^3 \otimes dx^2 \\ &\quad - dx^3 \otimes dx^2 \otimes dx^1 + dx^2 \otimes dx^3 \otimes dx^1 + dx^3 \otimes dx^1 \otimes dx^2. \end{aligned} \tag{34}$$

An example of a 3-form on $\mathbb{R}^3$ is the Euclidean volume form whose components in Cartesian coordinates are given in terms of the Levi-Civita symbol:

$$\Omega = \frac{1}{3!} \epsilon_{ijk} dx^i \wedge dx^j \wedge dx^k. \tag{35}$$

Combining the six nonzero terms using the antisymmetry of the wedge product, we verify that Ω is simply the familiar volume element $\Omega = dx^1 \wedge dx^2 \wedge dx^3$. The Levi-Civita symbol generalizes to $\mathbb{R}^n$: $\epsilon_{i_1 \cdots i_n}$ is antisymmetric under every exchange of indices and satisfies $\epsilon_{12\cdots n} = 1$. For example, on $\mathbb{R}^2$ with Cartesian coordinates $x^i = (x, y)$ we have the volume form $\Omega = \frac{1}{2}\Omega_{ij} dx^i \wedge dx^j = dx \wedge dy$ where $\Omega_{ij} = \epsilon_{ij}$. Thus $\Omega = dx \wedge dy$.

Now suppose we change to polar coordinates $\tilde{x}^i = (r, \theta)$ via $x = r\cos\theta, y = r\sin\theta$. Then the inverse Jacobian is $(J^{-1})_i^k = \frac{\partial x^k}{\partial \tilde{x}^i} = (x/r, -y|y/r, x)$. The

[28] We may take linear combinations of p-forms ω, ψ: $f\omega + g\psi$ for any smooth functions f, g to produce other p-forms. The space of p-forms is denoted $\Omega^p(M)$.

[29] The symmetric group on 3 letters (§**??**) has $3! = 6$ elements. The identity $\sigma(i) = i$ denoted $(1)(2)(3)$ has sign 1. There are three pairwise transpositions $(12)(3), (1)(23)$ and $(2)(31)$ which have sign -1. For example (23) means 2 and 3 are mapped to each other. Thus, $(1)(23)$ means $\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 2$. There are also two cyclic permutations $(123) = (12)(23)$ and $(132) = (12)(13)$ which have been written as products of pairwise exchanges composed from right to left. Here (132) means $\sigma(1) = 3, \sigma(3) = 2$ and $\sigma(2) = 1$. In the composition $(12)(13)$, 3 is mapped to 1 which is then mapped to 2, so that 3 is on the whole mapped to 2. On the other hand, 2 is directly mapped to 1. The cyclic permutations have sign $+1$ as they are products of an even number of transpositions. See also Footnote 25.

components of Ω in the new coordinates are $\tilde{\Omega}_{ij} = (J^{-1})_i^k (J^{-1})_j^l \Omega_{kl}$. One finds $\tilde{\Omega}_{ij} = (0, r| - r, 0)$ so that the volume form in polar coordinates is $\Omega = \frac{1}{2}(rdr \wedge d\theta - rd\theta \wedge dr) = rdr \wedge d\theta$. We notice that the prefactor r is $\det J^{-1}$. This is generally true: volume elements transform via a Jacobian determinant.

Pullback and pushforward. Given a pair of smooth manifolds X and Y and a smooth map $\phi : X \to Y$, we may (in favorable cases) use ϕ to move tensor fields between the manifolds (Nb. the manifolds need not have the same dimensions, we take $\dim X = n$ and $\dim Y = n'$). However, this only works in certain directions. Forms and more generally covariant tensor fields on Y may be 'pulled back' to X, the pullback being denoted ϕ^*. On the other hand, vector fields (and more generally contravariant tensor fields) on X may (in some cases) be 'pushed forward' to Y via ϕ_*. Combining these, if ϕ is a diffeomorphism (*invertible* smooth map with smooth inverse) then the pullback and pushforward via ϕ and ϕ^{-1} may be used to move arbitrary tensor fields in either direction.

The simplest tensor field is a scalar function. Given a smooth function $f : Y \to \mathbb{R}$, its pullback is the function $\phi^* f : X \to \mathbb{R}$ defined as $(\phi^* f)(x) = f(\phi(x))$ for any $x \in X$. In other words, we simply compose f with ϕ to go from X to $\mathbb{R}$ in two steps, $\phi^* f : X \xrightarrow{\phi} Y \xrightarrow{f} \mathbb{R}$. For example, suppose $\phi : S^2 \to \mathbb{R}^3$ is the map $\phi(\theta, \varphi) = (x = \sin\theta\cos\varphi, y = \sin\theta\sin\varphi, z = \cos\theta)$ and let $f(x, y, z) = z$ be the height function. Then the pullback is $(\phi^* f)(\theta, \varphi) = \cos\theta$ is the function that assigns the cosine of the polar angle to any point on the sphere. On the other hand, it is generally not possible to define the pushforward of a function. For this reason, we will view scalar functions as covariant (rather than contravariant) tensors of rank zero.

More generally, the gadget that helps us do this pushing and pulling is the linearization or differential $d\phi$ of the map ϕ. Suppose x^i and y^j are local coordinates on X and Y respectively and $y = \phi(x)$ or $y^j = \phi^j(x)$. Then the linearization at the point x is represented by the $n' \times n$ Jacobian matrix with entries $\frac{\partial \phi^j}{\partial x^i}$. Next, given a 1-form $\omega_j dy^j$ on Y we define its pullback at a point $x \in X$, denoted $(\phi^* \omega)(x)$ via

$$(\phi^* \omega)_i(x) = \frac{\partial \phi^j}{\partial x^i} \omega_j(\phi(x)). \tag{36}$$

Notice that no assumption on the invertibility of ϕ has been made. This makes it apparent why it is not possible, in general, to pushforward a differential form. If ϕ is invertible (say when $X = Y$ and ϕ is a diffeomorphism), we may multiply by the inverse Jacobian and formally recover the coordinate transformation formula of (10). However, there is a conceptual difference: while a map $\phi : X \to X$ actively moves points around, a coordinate transformation only relabels them. The generalization to the pullback of covariant rank-p tensor fields (including p-forms) is:

$$(\phi^* \omega)_{i_1 \cdots i_p}(x) = \frac{\partial \phi^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial \phi^{j_p}}{\partial x^{i_p}} \omega_{j_1 \cdots j_p}(\phi(x)). \tag{37}$$

Evidently, the pullback of a smooth function is the special case when $p = 0$.

Pushing forward vector fields (or contravariant tensors) is not so straightforward. To begin with, we note that the linearization of ϕ defines a linear transformation $d\phi$ between tangent spaces. If $y = \phi(x)$, then $d\phi(x) : T_x X \to T_y Y$. Once coordinates are chosen, this map is represented by the $n' \times n$ Jacobian matrix. Now if $v = v^i \partial_{x^i} \in T_x X$ then it is natural to define its pushforward to be the image of the vector v under the linear transformation $d\phi$. Thus, we are tempted to define the pushforward $\phi_* v$ as the vector field whose components at $y = \phi(x)$ are given by

$$(\phi_* v)^j(y) = \frac{\partial \phi^j}{\partial x^i} v^i(x) \quad \text{for} \quad j = 1, 2, \ldots, n'. \tag{38}$$

However, if ϕ is not surjective (say, if $n' > n$) then this does not define a vector field on all of Y but at best on the image of X. There is a further difficulty with (38): how do we express the x that appears on the RHS in terms of y? If ϕ is one-to-one, then there is a unique y in the image $\phi(X)$ corresponding to any $x \in X$, so that we may write $x = \phi^{-1}(y)$. Thus, if ϕ is injective, we may define the pushforward $\phi_* v$ as a vector field on $\phi(Y)$ with components given in (38). See Prob. **??** for a simple example. The definition has a straightforward generalization to rank p contravariant tensor fields for any $p = 1, 2, \ldots$:

$$(\phi_* t)^{j_1 \cdots j_p}(y) = \frac{\partial \phi^{j_1}}{\partial x^{i_1}} \frac{\partial \phi^{j_2}}{\partial x^{i_2}} \cdots \frac{\partial \phi^{j_p}}{\partial x^{i_p}} t^{i_1 \cdots i_p}(x). \tag{39}$$

As before, we notice the formal similarity with the coordinate transformation laws [e.g., (14)] for contravariant tensor fields when X and Y are the same manifold. What is more, if ϕ is a diffeomorphism then it is both injective and surjective so that (39) unambiguously defines a pushforward tensor field on all of Y.

Pullback of a metric: induced metric via an example. Consider the unit sphere S^2 ($x^2 + y^2 + z^2 = 1$) embedded as a submanifold of $\mathbb{R}^3$. If we use polar coordinates ($\xi^1 = \theta, \xi^2 = \varphi$) on S^2, the embedding is defined by a smooth map $\phi : S^2 \to \mathbb{R}^3$ given by $\phi(\theta, \varphi) = (x = \sin\theta\cos\varphi, y = \sin\theta\sin\varphi, z = \cos\theta)$. Now, $\mathbb{R}^3$ has the standard flat Euclidean metric whose components in Cartesian coordinates are $g_{ij} = \delta_{ij}$. We may pullback this rank-2 covariant symmetric tensor field to get an 'induced' metric h_{ab} on S^2 with components

$$h_{ab} = (\phi^* g)_{ab} = \frac{\partial \phi^i}{\partial \xi^a} \frac{\partial \phi^j}{\partial \xi^b} g_{ij}. \tag{40}$$

This formula defines the induced metric and is of course not special to the above example. In the case of the embedding $\phi : S^2 \hookrightarrow \mathbb{R}^3$, the induced metric $h_{ab}(\theta, \varphi)$ is the familiar 'round sphere' metric. Work out its components.

8 Exterior algebra, exterior derivative and Bianchi's identity

Exterior algebra. In §5 and §6 we introduced 1- and 2-forms. The former can be used to describe the momentum of a particle on the configuration space of a mechanical system, the Liouville form '$p\,dq$' on phase space, the infinitesimal heat added to a gas in a thermodynamic process or the electromagnetic 'scalar' and 'vector' potentials. Two-forms are used to model infinitesimal area elements, the electromagnetic field strength tensor $F_{\mu\nu}$ and the symplectic form ω_{ij} in mechanics. Furthermore, the wedge product of two 1-forms was seen to produce a 2-form. On the other hand, the differential or exterior derivative of a function df was shown to give a 1-form, while the exterior derivative of a 1-form led to a 2-form. In this section, we extend the wedge product and exterior derivative to forms of any rank (introduced in §7) and also discuss an analog of the Leibniz rule for the exterior derivative of a wedge product. The space of differential forms with these algebraic properties is called the exterior algebra. These developments are then applied to understand some properties of the symplectic form ω of Hamiltonian mechanics introduced in (**??**) and (27) and further discussed in §**??**.

Recall from §7, that a differential form of order $p = 0, 1, 2, \ldots$ or p-form ω in a patch with coordinates x^i is a linear combination of p-fold wedge products of coordinate 1-forms:

$$\omega = \frac{1}{p!}\omega_{i_1 \cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p}. \tag{41}$$

Given any smooth p-forms ω, ψ and smooth functions f, g, $f\omega + g\psi$ is also a smooth p-form. Thus, the space of p-forms denoted $\Omega^p(M)$ is said to be a module (see Footnote 17) over the ring of smooth real-valued functions on M ($\mathcal{F}(M)$ of §3).

Owing to the antisymmetry of $dx^{i_1} \wedge \cdots \wedge dx^{i_p}$ there are no nonzero p forms for $p > n$ on a manifold of dimension n. For instance on $\mathbb{R}$, there is only one coordinate 1-form dx and the only possible coordinate basis 2-form $dx \wedge dx$ vanishes by antisymmetry (there is no concept of area on a line). On $\mathbb{R}^2$, we have two coordinate basis 1-forms dx and dy, one independent basis 2 form $dx \wedge dy = -dy \wedge dx$ and no nonzero 3-forms as $dx \wedge dy \wedge dx$ etc., all vanish. In fact, the number of linearly independent p-forms at a point is the binomial coefficient $\binom{n}{p}$ since each choice of p distinct coordinate 1-forms $dx^{i_1}, \ldots, dx^{i_p}$ furnishes one coordinate basis p-form $dx^{i_1} \wedge \ldots \wedge dx^{i_p}$. In particular, there is only one ($= \binom{n}{0}$) independent 0-form and one ($= \binom{n}{n}$) independent n-form. What we mean is that any 0-form is some smooth function times the constant function 1 and any n-form is some smooth function times the volume form $dx^1 \wedge \cdots \wedge dx^n$.

We may take the direct sum of the spaces of p-forms to obtain the $\sum_{p=0}^{n} \binom{n}{p} = 2^n$ dimensional space of all differential forms on M:

$$\Omega(M) = \oplus_{p=0}^{n} \Omega^p(M). \tag{42}$$

In addition to taking linear combinations of forms, we may take their wedge product. For the coordinate basis forms

$$(dx^{i_1} \wedge \cdots \wedge dx^{i_p}) \wedge (dx^{i_{p+1}} \wedge \cdots \wedge dx^{i_{p+q}}) = dx^{i_1} \wedge \cdots \wedge dx^{i_{p+q}}. \tag{43}$$

For example, $(dx \wedge dy) \wedge (dz \wedge dw) = dx \wedge dy \wedge dz \wedge dw$. By repeated use of the antisymmetry property $dx^i \wedge dx^j = -dx^j \wedge dx^i$, we may show that the wedge product of a p-form ω and a q-form ψ is (anti)commutative:

$$\omega \wedge \psi = (-1)^{pq} \psi \wedge \omega. \tag{44}$$

Let us explain the origin of the sign. Suppose $\omega = dx$ and $\psi = dy \wedge dz$ so that $p = 1$ and $q = 2$. Then dx has to 'pass through' dy and dz producing two minus signs resulting in $dx \wedge (dy \wedge dz) = (-1)^{1 \cdot 2} (dy \wedge dz) \wedge dx$. Similarly, suppose we consider $(dx \wedge dy) \wedge (du \wedge dv \wedge dw)$. Here we move dy first through the 3-form picking up a $(-1)^3$ and then move dx and get another $(-1)^3$. Thus we see the emergence of p factors of $(-1)^q$ leading to the sign $(-1)^{pq}$.

Equipped with this wedge product, $\Omega(M)$ is called the exterior algebra. A special case is the wedge product of a p-form ω and a 0-form f: $\omega \wedge f = (-1)^0 f \wedge \omega = f\omega$.

Exterior derivative. The exterior derivative may be extended to a map from p-forms to $(p+1)$-forms: $d : \Omega^p(M) \to \Omega^{p+1}(M)$ for any $p = 0, 1, 2, \ldots$ satisfying the three axioms:

1. Linearity: $d(a\omega + b\psi) = a d\omega + b d\psi$ for any $a, b \in \mathbb{R}$ and p-forms ω, ψ.

2. Leibniz (antiderivation) rule: $d(\omega \wedge \phi) = d\omega \wedge \phi + (-1)^p \omega \wedge d\phi$ where $\omega \in \Omega^p(M)$ and ϕ is any form.

3. Nilpotent[30] of degree two: $d^2 \omega = 0$ for any p-form ω.

The need for the minus sign in this Leibniz rule is already evident if we consider the wedge product of a 1-form $\phi = \phi_j dx^j$ and a zero form f. Now $\phi \wedge f = f \wedge \phi = f\phi$. We will calculate $d(\phi \wedge f)$ from first principles and see the emergence of the minus sign. In fact, using $d\phi = \partial_i \phi_j dx^i \wedge dx^j$, we get

$$\begin{aligned} d(\phi \wedge f) &= d(f\phi) = \partial_i (f \phi_j) dx^i \wedge dx^j = \left((\partial_i f) \phi_j + f \partial_i \phi_j \right) dx^i \wedge dx^j \\ &= df \wedge \phi + f d\phi = -\phi \wedge df + d\phi \wedge f = d\phi \wedge f - \phi \wedge df. \end{aligned} \tag{45}$$

Closed and exact forms on a manifold M. A p-form ω such that $d\omega = 0$ is said to be closed. On the other hand, if $\omega = d\phi$ for some $(p-1)$-form ϕ, then ω is said to be exact (generalizing the idea of an exact differential). Since $d^2 = 0$, an exact form is automatically closed. The quotient linear space of closed p-forms modulo exact p-forms is called the p^{th} cohomology (group) of the manifold M. The homogeneous Maxwell equations $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$ and $\frac{1}{c} \frac{\partial \boldsymbol{B}}{\partial t} + \boldsymbol{\nabla} \times \boldsymbol{E} = 0$ are together the statement that the Faraday 2-form $F = (1/2) F_{\mu\nu} dx^\mu \wedge dx^\nu$ on Minkowski space-time $\mathbb{R}^4$ (with $x^\mu = (ct, x, y, z)$ for $\mu = 0, 1, 2, 3$) is a closed 2-form. Here, $F_{0i} = E_i$ and $F_{ij} = -\epsilon_{ijk} B_k$ for $1 \leq i, j, k \leq 3$. Since $\mathbb{R}^4$ has trivial cohomology groups, F must be exact and expressible as $F = dA$ for some 'gauge potential' 1-form $A = A_\mu dx^\mu$. This is why we may express $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$ and $\boldsymbol{E} = -\boldsymbol{\nabla}\phi - c^{-1} \frac{\partial \boldsymbol{A}}{\partial t}$ in terms of the scalar and vector potentials which are the components of $A_\mu = (\phi, -\boldsymbol{A})$ (see §**??** and Prob. **??**). □

[30] $d^2 = 0$ is a generalization of the vector calculus identities $\boldsymbol{\nabla} \times \boldsymbol{\nabla} f = 0$ and $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{v}) = 0$ for any smooth function f and vector field $\boldsymbol{v}$ in $\mathbb{R}^3$.

Symplectic form and Bianchi's identity. Suppose $\alpha = p_i dq^i$ is the canonical Liouville 1-form on the phase space $\mathbb{R}^{2n}$ of a mechanical system with n degrees of freedom. Then we have seen that the canonical symplectic form is given by $\omega = d\alpha = dp_i \wedge dq^i$. It follows that $d\omega = d^2\alpha = 0$. In other words, the canonical symplectic form is closed. More generally, the Jacobi identity implies that the inverse ω of any invertible (but not necessarily canonical) Poisson tensor r satisfies the Bianchi identity $\partial_i \omega_{jk} + \text{cyclic} = 0$. We now verify that the Bianchi condition is simply the statement that $d\omega = 0$. Indeed, using linearity and the Leibniz rule,

$$d\omega = \frac{1}{2} d(\omega_{jk} dx^j \wedge dx^k) = \frac{1}{2}(\partial_i \omega_{jk}) dx^i \wedge dx^j \wedge dx^k. \tag{46}$$

Since $dx^i \wedge dx^j \wedge dx^k$ is antisymmetric under exchange of any pair of indices, only the similarly antisymmetric part of $\partial_i \omega_{jk}$ can contribute. Antisymmetrizing as in (34), we write

$$\begin{aligned} d\omega &= \frac{1}{12}\left(\partial_i \omega_{jk} - \partial_j \omega_{ik} - \partial_k \omega_{ji} - \partial_i \omega_{kj} + \partial_k \omega_{ij} + \partial_j \omega_{ki}\right) dx^i \wedge dx^j \wedge dx^k \\ &= (1/3!) \left(\partial_i \omega_{jk} + \partial_k \omega_{ij} + \partial_j \omega_{ki}\right) dx^i \wedge dx^j \wedge dx^k, \end{aligned} \tag{47}$$

where we used the antisymmetry of ω in the last step. Thus $(d\omega)_{ijk} = \partial_i \omega_{jk} + \partial_k \omega_{ij} + \partial_j \omega_{ki}$, so that $d\omega = 0$ iff the Bianchi condition is satisfied. We begin to see the economy and clarity that the use of differential forms can bring to tensor calculus. What is more, given a smooth 'Hamiltonian' function H on M, we may use ω to define a vector field v_H called the Hamiltonian vector field via the formula

$$v_H(\cdot) = \omega^{-1}(\cdot, dH). \tag{48}$$

Here ω^{-1} is a contravariant (antisymmetric) second rank tensor that can act on a pair of 1-forms, one of which is chosen to be dH. The resulting object is a vector field as it can act linearly on an (unspecified) 1-form.

More generally, even if we do not have canonical (q-p-type) coordinates on phase space and do not have available the canonical Liouville 1-form $\alpha = p_i dq^i$, we may still wish to define a symplectic form using physical or geometric considerations. From the foregoing, the essential conditions it must satisfy are invertibility and the Bianchi identity. Thus, one defines a symplectic manifold as a sufficiently smooth manifold that is equipped with a closed nondegenerate (i.e., invertible) two-form ω called the symplectic form. Though it is required to be closed, ω need not be exact[31].

9 Integration on manifolds and Stokes' theorem

We now move from the exterior differential calculus to the integral calculus on a manifold. This will allow us to generalize the concepts of line, surface and volume integrals to manifolds. To do this, we first need the idea of an oriented manifold.

[31]The symplectic form (**??**) on the 2-sphere S^2, given by $\omega = -l \sin\theta \, d\theta \wedge d\phi$ (for $l \neq 0$) is closed (being a top degree form) but not exact. To show this, we note that $\int_{S^2} \omega = -4\pi l$ is proportional to the surface area of the unit sphere. If $\omega = d\alpha$, then by Stokes' theorem (53), this integral must vanish, since S^2 has no boundary: $\int_{S^2} d\alpha = \int_{\partial S^2} \alpha = 0$. Thus, ω cannot be exact. However, locally in a coordinate patch, it can be written as $\omega = d\alpha$ for $\alpha = l \cos\theta \, d\phi$ (local exactness is called the Poincaré lemma). The problem is that α cannot be smoothly extended to a 1-form on all of S^2.

Orientability of a manifold. We may orient a curve $\boldsymbol{\gamma}$ in 3d space by placing arrows on it so that the curve is traversed in only one direction. A parametrized curve $\boldsymbol{\gamma}(s) : [0,1] \to \mathbb{R}^3$ with $\dot{\boldsymbol{\gamma}} \neq 0$ everywhere has a natural orientation, namely the direction in which $\boldsymbol{\gamma}$ points (which is the same as that of increasing s). If $\dot{\boldsymbol{\gamma}}$ vanishes somewhere, we would not know which way the arrow points there. Worse still, if γ retraces the curve (goes back and forth), then there would be places where the direction of the arrow is ambiguous. This is why we assume $\dot{\boldsymbol{\gamma}} \neq 0$. Given a vector field $\boldsymbol{v}$ and such a curve $\boldsymbol{\gamma}$, we may define the line element $d\boldsymbol{\gamma} = \dot{\boldsymbol{\gamma}}(t)dt$ and the line integral of $\boldsymbol{v}$ along $\boldsymbol{\gamma}$: $\int_{\boldsymbol{\gamma}} \boldsymbol{v} \cdot d\boldsymbol{\gamma}$. Note that $\boldsymbol{\gamma}$ need not be an integral curve of $\boldsymbol{v}$. One verifies that this line integral is reparametrization invariant. This means the line integral can depend on the route the curve follows but not on how it is parametrized[32]. Indeed, suppose $s = s(t) : [0,1] \to [0,1]$ is a reparametrization (invertible map) and let $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(s(t))$. Then

$$\int \boldsymbol{v} \cdot d\tilde{\boldsymbol{\gamma}} = \int_0^1 v_j \frac{d\tilde{\gamma}^j}{dt} dt = \int_0^1 v_j \frac{d\gamma^j}{ds}\frac{ds}{dt} dt = \int_0^1 v_j \frac{d\gamma^j}{ds} ds = \int \boldsymbol{v} \cdot d\boldsymbol{\gamma}. \quad (49)$$

For a 2d surface Σ embedded in $\mathbb{R}^3$, we usually speak of an outward or inward pointing unit normal at each point of Σ. When Σ is defined by the condition $C(x,y,z) = 0$, the normal in the direction of increasing C is given by the unit vector along the gradient $\boldsymbol{\nabla} C$. To be well-defined (unambiguous), when the normal is followed around any closed loop on Σ, it must return to its original direction. When this happens, we say that the surface is oriented. In vector calculus, this normal to the surface is used to define a vectorial area element $(\hat{n} dS)$ that goes into the definition of surface integrals. These concepts can be generalized to manifolds of any dimension and are used to define integration on manifolds. An n-dimensional manifold is orientable if it admits a nonvanishing[33] form of top degree n (called a volume form). The choice of such a form is called an orientation. On $\mathbb{R}^2$ we usually choose the orientation as given by $dx \wedge dy$, the choice $dy \wedge dx$ is equally valid, but would correspond to reversing the orientation. On $\mathbb{R}^n$, the standard volume form is $dx^1 \wedge \cdots \wedge dx^n$. Admitting a volume form is equivalent to the Jacobian determinants of all the transition functions between coordinate charts being positive, so that all the coordinate charts have a common orientation[34] and the atlas may be called an oriented atlas. The two-sphere is orientable since the symplectic form (**??**) on S^2 is a nonvanishing 2-form

[32] When we use a line integral (**??**) to model the work $W = \int \boldsymbol{F} \cdot d\boldsymbol{\gamma}$ done by a force field as a particle moves along a path, we are asserting that the work done is independent of how fast the particle moves at various places along the path.

[33] A p-form ω is nonvanishing if at each point on the manifold, $\omega(u, v, \ldots) \neq 0$ for any p linearly independent tangent vector fields $u, v, \cdots$.

[34] It is possible to concoct a nonoriented atlas. Consider the Euclidean plane $\mathbb{R}^2$ and define a new manifold via two overlapping patches. The left patch $x < 1$ and the right patch $x > -1$. On the left patch we define local coordinates $\xi^1 = x, \xi^2 = y$ while on the right patch we define local coordinates $\eta^1 = y, \eta^2 = x$. They overlap along the strip $-1 < x < 1$ where the point (x,y) has two addresses or sets of coordinates: (ξ^1, ξ^2) and (η^1, η^2). The transition functions are $\eta^1 = \xi^2$ and $\eta^2 = \xi^1$ resulting in an off-diagonal Jacobian matrix $\frac{\partial \eta^i}{\partial \xi^j} = (0,1|1,0)$ with determinant -1. Through this atlas, we have defined a manifold that is not orientable, the two charts have opposite orientations.

(proportional to the standard area form). For a surface in $\mathbb{R}^3$, orientability allows us to unambiguously distinguish two sides of the surface. The Möbius strip is not orientable[35]: one can go from the 'upper' side of the surface to the 'lower' side at the same point by taking a walk on the strip; this is not possible on a cylindrical surface or on a sphere, which are orientable.

Riemannian volume form. On an oriented n-dimensional Riemannian or pseudo-Riemannian manifold M with nondegenerate metric g, one has a natural volume form ω_g. In local coordinates x^i, it is $\omega_g = \sqrt{|\det g|}\, dx^1 \wedge \cdots \wedge dx^n$. Since g_{ij} is invertible, $\det g \neq 0$ so that this is a nonvanishing form. Let us check that this formula holds in any coordinate system. As noted in §7, under a coordinate change $x^i \to y^i$ a volume form $f(x) dx^1 \wedge \cdots \wedge dx^n$ (where f is a scalar function) transforms to $\det J^{-1} f(x(y)) dy^1 \wedge \cdots \wedge dy^n$ where $(J^{-1})^i_j = \frac{\partial y^i}{\partial x^j}$ is the inverse Jacobian matrix. If the transformation is orientation-preserving, then $\det J^{-1} > 0$. However, $\sqrt{|\det g|}$ is not a scalar function since the metric components transform to $\tilde{g}_{ij} = g_{kl}(J^{-1})^k_i (J^{-1})^l_j$. Hence, $\det \tilde{g} = \det g \det((J^{-1})^t) \det J^{-1}$. Consequently,

$$\sqrt{|\det g|} dx^1 \wedge \cdots \wedge dx^n = \frac{\sqrt{|\det \tilde{g}|}}{\det J^{-1}} \det J^{-1} dy^1 \wedge \cdots \wedge dy^n = \sqrt{|\det \tilde{g}|} dy^1 \wedge \cdots \wedge dy^n. \tag{50}$$

We see that in any coordinate system, the Riemannian volume form has the same expression. What is more, if one takes any orthonormal basis $\phi^1, \phi^2, \cdots, \phi^n$ for 1-forms on M, then $\omega = \pm \phi^1 \wedge \phi^2 \wedge \cdots \wedge \phi^n$. For example, consider 3d Euclidean space $\mathbb{R}^3$. In Cartesian coordinates, the Euclidean metric has components $g_{ij} = \delta_{ij}$ with unit determinant and the Euclidean volume form is $\omega = dx^1 \wedge dx^2 \wedge dx^3$. In spherical polar coordinates, the nonzero metric components are $g_{rr} = 1, g_{\theta\theta} = r^2, g_{\phi\phi} = r^2 \sin^2\theta$ so that $\det g = r^4 \sin^2\theta$ and the volume form becomes $\omega = r^2 \sin\theta dr \wedge d\theta \wedge d\phi$. On the other hand, the inverse metric is $g^{ij} = \mathrm{diag}(1, 1/r^2, 1/(r^2 \sin^2\theta))$ so that the coordinate 1-forms have squared-lengths $g^{-1}(dr, dr) = 1, g^{-1}(d\theta, d\theta) = 1/r^2, g^{-1}(d\phi, d\phi) = 1/r^2 \sin^2\theta$. It follows that $\{dr, r d\theta, r \sin\theta d\phi\}$ is an orthonormal basis for 1-forms. We see that their wedge product is the volume form ω.

Integration of forms. In vector calculus, we define line integrals, surface integrals and volume integrals. These are examples of the integration of a 1-form along a curve, a 2-form over a surface and a 3-form over a 3d manifold. More generally, a p-form ψ may be integrated over an oriented p-dimensional manifold M to obtain a real number denoted $\int_M \psi$. To evaluate the integral, the manifold is covered by nonoverlapping cells and their boundaries (each lying within a coordinate chart) and the integral is a sum of contributions from each cell[36]. In each cell, the integral is evaluated as in multivariable calculus. In more detail, the p-form ψ can be written as $\psi = f\omega$ where f is a

[35] A cylinder is constructed by taking a rectangular page from a tall book and pasting the two short edges together: the left of the bottom edge to the left of the top edge. The Möbius strip is obtained by twisting the bottom edge once before pasting it onto the top edge, so that the left of the bottom edge is joined to the right of the top edge.

[36] If ψ has bounded components, the boundaries between cells do not contribute to $\int_M \psi$ so it does not matter if these boundaries are omitted or counted a finite number of times.

scalar and ω the volume form. Moreover, within a patch with coordinates $x^1, \cdots, x^p$, the volume form may be written as $\omega = \mu(x) dx^1 \wedge \cdots \wedge dx^p$ for some nonvanishing function μ. Then the contribution of the cell C is $\int_C \psi = \int_{x(C)} f(x)\mu(x) dx^1 \cdots dx^p$ where $x(C)$ is the image of the cell in $\mathbb{R}^p$. Examples: (i) We may integrate the 1-form $\psi = f(x)dx$ over the submanifold $I = (1,2) \cup (3,6)$ of $\mathbb{R}$:

$$\int_I \psi \equiv \int_1^2 f(x)dx + \int_3^4 f(x)dx + \int_4^6 f(x)dx. \tag{51}$$

Here we have chosen the 'increasing' orientation $\omega = dx$ (as opposed to $-dx$) and broken I into three cells. (ii) The polar coordinate patch $x^i = (\theta, \phi)$ along with its boundary covers the unit sphere S^2. So the integral of the 2-form $\psi = f\omega$ on S^2 may be expressed as

$$\int_{S^2} \psi = \int_{x(S^2)} f(\theta, \phi) \sin\theta \, d\theta \wedge d\phi \equiv \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, f(\theta, \phi) \sin\theta. \tag{52}$$

The orientation has been chosen so that for $f = 1$, the integral of $\omega = \sin\theta d\theta \wedge d\phi$ over S^2 gives the area 4π of the unit sphere.

Manifold with boundary. To discuss Stokes' theorem, we need to generalize the notion of a manifold to include manifolds with boundary. By the definition of §1, the closed unit disk D ($x^2 + y^2 \le 1$) contained in the plane is not a manifold, since points of D on the rim (with $x^2 + y^2 = 1$) do not have open neighborhoods lying within D. For points on the rim, we will allow neighborhoods of a different sort: roughly those shaped like a half Moon that include nearby points on the rim. There is an obvious sense in which the unit circle S^1 is the boundary of D, which we indicate via $\partial D = S^1$. The set theoretic difference $D \setminus \partial D$ is the open unit disk, it is called the interior of D. More generally, a manifold with boundary is a (topological) space M with two types of points: (a) *interior points* which together comprise an n-dimensional manifold (i.e., which have open neighborhoods homeomorphic to $\mathbb{R}^n$ or the n-ball $B_n : x_1^2 + \cdots + x_n^2 < 1$) and (b) *boundary points* which together comprise an $n-1$ dimensional manifold called the boundary (∂M) consisting of points of M which have a neighborhood homeomorphic to a half space ($\boldsymbol{x} \in \mathbb{R}^n$ with $x_1 \ge 0$) or half ball ($\boldsymbol{x} \in B_n$ with $x_1 \ge 0$) with the homeomorphism taking the boundary points to points with $x_1 = 0$.

Stokes' theorem. Suppose $\omega = d\phi$ is an exact p-form on a p-dimensional manifold M with boundary denoted ∂M. Then Stokes' theorem

$$\int_M d\phi = \int_{\partial M} \phi \tag{53}$$

expresses the integral of ω over M as that of the $(p-1)$-form ϕ over the $(p-1)$-dimensional boundary ∂M. In particular, the integral of an exact form over a manifold without boundary vanishes. This is a generalization of Gauss' divergence theorem and

Stokes' theorem from vector calculus (see Prob. **??**):

$$\int_\Omega \boldsymbol{\nabla} \cdot \boldsymbol{v}\, d^3r = \int_{\partial\Omega} \boldsymbol{v} \cdot d\boldsymbol{S} \quad \text{and} \quad \int_S (\boldsymbol{\nabla} \times \boldsymbol{v}) \cdot d\boldsymbol{S} = \oint_{\partial S} \boldsymbol{v} \cdot d\boldsymbol{l}. \tag{54}$$

Here, Ω is a 3d region in $\mathbb{R}^3$ while S is a surface in $\mathbb{R}^3$. In fact, (53) is also a generalization of the fundamental theorem of calculus for the integral of a 1-form over an interval $M = [a,b] \subset \mathbb{R}$: $\int_M f'(x)dx = \int_{\partial M} f = f(b) - f(a)$. Here f is a zero form and the boundary ∂M is the 0-dimensional disconnected manifold consisting of two points a, b. The orientation of the interval gives ∂M an orientation ($+1$ at b and -1 at a) leading to the relative sign on the RHS.

10 Geodesic equation

Geodesic equation from extremizing length of curve. Geodesics generalize the concept of straight lines on the Euclidean plane and are governed by equations that generalize those of a straight line ($\ddot{x}^i = 0$). More precisely, a geodesic is a curve of extremal length joining two points on a Riemannian manifold M. Suppose x^i are coordinates on a patch where the metric tensor (positive-definite symmetric matrix depending on location) has components $m_{ij}(x)$. Let γ be a sufficiently smooth curve lying in this patch, given by $x^i(\tau)$ for some parameter $0 \le \tau \le 1$. Its length is defined as[37]

$$\ell(\gamma) = \int_0^1 \sqrt{m_{ij}\dot{x}^i\dot{x}^j}\, d\tau. \tag{55}$$

The arc length parameter $s(\tau)$ is defined as the length of the curve up to parameter value τ:

$$s(\tau) = \int_0^\tau \sqrt{m_{ij}\dot{x}^i\dot{x}^j}\, d\tau'. \tag{56}$$

Evidently, $\ell(\gamma) = s(1)$. Since the parametrization that appears in the simplest form of the geodesic equation is the arc length[38], we will change parametrizations from τ to s. This is facilitated by introducing a symbol for the speed

$$c(x(\tau), \dot{x}(\tau)) = \sqrt{m_{ij}\dot{x}^i\dot{x}^j}, \tag{57}$$

which loosely plays the role of a Lagrangian in the length functional (55). By the fundamental theorem of calculus,

$$\frac{ds}{d\tau} = c \quad \text{or} \quad \frac{d}{d\tau} = c\frac{d}{ds}. \tag{58}$$

[37] The length of the curve is independent of the choice of parametrization τ. An invertible reparametrization $\tau \to \sigma(\tau)$ (with $\sigma'(\tau) > 0$) does not alter the form of the integrand: $\sqrt{m_{ij}\dot{x}^i(\tau)\dot{x}^j(\tau)}\, d\tau = \sqrt{m_{ij}\dot{x}^i(\sigma)\dot{x}^j(\sigma)}\, d\sigma$ since the Jacobian factor $d\sigma/d\tau$ cancels out.

[38] An affine transformation is a linear transformation plus a shift. Parameters s' that are related to arc length via $s' = as + b$ are said to be affine parameters. The geodesic equation takes the same form for any affine parameter. The geodesic equation would look more complicated if we used parametrizations that are not affinely related to the arc length.

Now, the Euler-Lagrange condition for $\ell(\gamma)$ to be stationary to first order in variations $x \to x + \delta x$ holding the endpoints at $\tau = 0, 1$ fixed is

$$\frac{d}{d\tau}\frac{\partial c}{\partial \dot{x}^k} = \frac{\partial c}{\partial x^k}. \tag{59}$$

The RHS is

$$\frac{\partial c}{\partial x^k} = \frac{1}{2c} m_{ij,k}\dot{x}^i\dot{x}^j = \frac{c}{2} m_{ij,k}\frac{dx^i}{ds}\frac{dx^j}{ds}. \tag{60}$$

Similarly,

$$\frac{\partial c}{\partial \dot{x}^k} = \frac{1}{c} m_{kj}\dot{x}^j = m_{kj}\frac{dx^j}{ds}. \tag{61}$$

So the LHS of (59) is[39]

$$\frac{d}{d\tau}\frac{\partial c}{\partial \dot{x}^k} = c\frac{d}{ds}\left(m_{kj}\frac{dx^j}{ds}\right) = c\left[m_{kj,i}\frac{dx^i}{ds}\frac{dx^j}{ds} + m_{kj}\frac{d^2x^j}{ds^2}\right]. \tag{62}$$

Equating the two, we get the condition for $\ell(\gamma)$ to be extremal:

$$m_{kj}\frac{d^2x^j}{ds^2} = \frac{1}{2}\left(m_{ij,k} - 2m_{kj,i}\right)\frac{dx^i}{ds}\frac{dx^j}{ds}. \tag{63}$$

Symmetrizing the second term on the right under $i \leftrightarrow j$ and contracting with the inverse metric, we get the geodesic equation in terms of the Christoffel symbols and arc length parametrization:

$$\ddot{x}^l + \Gamma^l_{ij}\dot{x}^i\dot{x}^j = 0 \quad \text{where} \quad \Gamma^l_{ij} = \frac{1}{2}m^{lk}\left[m_{ki,j} + m_{kj,i} - m_{ij,k}\right]. \tag{64}$$

We notice that the Christoffel symbols vanish if the metric components are constant. In particular, in Cartesian coordinates on Euclidean space (where $m_{ij} = \delta_{ij}$), the geodesic equation reduces to that for a straight line: $\ddot{x}^l = 0$.

Geodesic equation in terms of covariant derivative. It is possible to write the geodesic equation in a nice way if we introduce the so-called covariant derivative. The covariant derivative $\boldsymbol{\nabla}_u v$ of a vector v along the vector u is a generalization of the directional derivative of vector calculus. In addition to the partial derivative, it includes a 'correction' term involving the Christoffel symbols (64):

$$(\boldsymbol{\nabla}_u v)^i = u^j \boldsymbol{\nabla}_j v^i = u^j\partial_j v^i + u^j\Gamma^i_{jk}v^k. \tag{65}$$

We will soon show (see §11) that this correction term ensures that $\boldsymbol{\nabla}_u v$ transforms as a contravariant vector field just as v itself does. By contrast, $u^j\partial_j v^i$ does not (in

[39]Notice that if we did not use arc length parametrization (or one that is affinely related to it), the resulting formulae would be more complicated. In fact, if we try to work with τ instead of s, to calculate the τ derivative of the 'momentum' conjugate to $\dot{x}^k$ in (61), we would need to calculate $\frac{d(1/c)}{d\tau}$. This would produce additional terms in the condition for ℓ to be extremal. Thus, the geodesic equation takes the simple form (64) only for affine parameters.

general) transform as a vector field. Now taking $u = v = \dot{x}$ and using the chain rule $\frac{d}{ds} = \frac{dx^j}{ds}\frac{\partial}{\partial x^j} = \dot{x}^j \frac{\partial}{\partial x^j}$ we get

$$(\nabla_{\dot{x}} \dot{x})^i = \dot{x}^j \partial_j \dot{x}^i + \dot{x}^j \Gamma^i_{jk} \dot{x}^k. \tag{66}$$

Thus, the geodesic equation is simply the statement that $\nabla_{\dot{x}} \dot{x} = 0$. This means the covariant derivative of the tangent vector along the tangent to a geodesic is zero. We say that the tangent vector to a geodesic is covariantly constant along the geodesic. Thus, a geodesic may also be viewed as a curve parametrized by arc length that parallel transports its own tangent vector.

11 Covariant derivative

The laws governing the dynamics of physical systems (especially continuum mechanical systems with infinitely many degrees of freedom) are often formulated in terms of (partial) differential equations involving scalar, vector and tensor fields. Given a p-form, we have learned (in §8) how to take its exterior derivative using partial derivatives and the Leibniz rule to arrive at a p+1-form. What about the derivative of a vector field v? If one is given another vector field u, then as in §4, one may differentiate v along the integral curves of u to arrive at the Lie derivative $\mathcal{L}_u v = [u, v]$ (an example of this is the Lie derivative of the vorticity along the velocity field of a fluid). In the absence of such a u, one could consider the partial derivatives $\partial_i v^j$. Unfortunately, the partial derivatives of a vector field do not form the components of a tensor. As we will soon show, although $\partial_i v^j$ has one upper and one lower index, it does not transform as a (1,1) tensor field. In the absence of additional structure (like another vector field), it is not possible to differentiate a vector field (or other contravariant tensors or covariant tensors that are not antisymmetric) in a manner that produces a tensor. An advantage of working with tensors is that they have a coordinate-independent meaning (as multilinear maps) although one may choose to write their components in a particular coordinate system with simple transformation properties under a change of coordinates. However, if the manifold is equipped with a metric tensor g_{ij} (as may arise from a kinetic energy quadratic in velocities or from the geometry of space or space-time), then it is possible to 'covariantly' differentiate a vector field to arrive at a (1,1) tensor field $\nabla_i v^j \, dx^i \otimes \partial_j$.

This notion of covariant differentiation (also called a metric connection) can then be extended to other tensor fields in such a way that it satisfies the Leibniz rule and reduces to the partial derivative for functions. It turns out that in Euclidean space ($\mathbb{R}^n$ with the metric δ_{ij} in Cartesian coordinates), the covariant derivative reduces to the partial derivative. However, the covariant derivative is essential if we wish to use curvilinear coordinates in Euclidean space or work on a curved manifold such as a sphere or hyperboloid. In fact, we have already met the covariant derivative in our discussion of the geodesic equation in (65) of §10, where we noted that the tangent vector to a geodesic is covariantly constant along the geodesic.

There are several ways of introducing the covariant derivative such as by (a) postulating axioms (such as linearity and the Leibniz rule) it must satisfy, (b) using the geometric notion of parallel transport and (c) combining the partial derivative with the Christoffel symbols from the geodesic equation and requiring tensorial behavior under coordinate transformations. Since we already have a provisional formula (65) for the covariant derivative[40]:

$$\nabla_j v^i = \partial_j v^i + \Gamma^i_{jk} v^k, \tag{67}$$

we will follow approach (c). In other words, we will examine how $\partial_j v^i$ transforms under a coordinate change and observe that it fails to behave like a (1,1) tensor. Then, we will use our definition (64) of the Christoffel symbols to similarly observe that they do not transform as a tensor. The offending terms in the transformation laws for the partial derivative and Christoffel symbols will be seen to cancel, allowing $\nabla_j v^i$ to transform as the components of a (1,1) tensor.

Thus, suppose we make a change of coordinates $x^i \mapsto \tilde{x}^i(x)$ under which the components of the vector field v^i become $\tilde{v}^i = \frac{\partial \tilde{x}^i}{\partial x^j} v^j$. Then the components of the partial derivatives of $\tilde{v}^i$ are

$$\frac{\partial \tilde{v}^i}{\partial \tilde{x}^j} = \frac{\partial}{\partial \tilde{x}^j}\left(\frac{\partial \tilde{x}^i}{\partial x^k} v^k\right) = \frac{\partial \tilde{x}^i}{\partial x^k}\frac{\partial x^l}{\partial \tilde{x}^j}\frac{\partial v^k}{\partial x^l} + \frac{\partial^2 \tilde{x}^i}{\partial x^l \partial x^k}\frac{\partial x^l}{\partial \tilde{x}^j} v^k. \tag{68}$$

The first term on the RHS is all there would have been if $\frac{\partial v^i}{\partial x^j}$ transformed as a (1,1) tensor. The second term with the second derivative is an inhomogeneous nontensorial term. In particular, unlike a tensor, which vanishes in all coordinate systems, if it vanishes in one, $\partial_j v^i$ fails to have this property.

Next, we use the transformation law for the metric to find how the term involving Christoffel symbols

$$\Gamma^i_{jk} = \frac{1}{2} g^{il}\left(\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}\right) \tag{69}$$

transforms. To make a long story short (see Prob. **??**) one finds that

$$\tilde{\Gamma}^i_{jk}\tilde{v}^k = \frac{\partial \tilde{x}^i}{\partial x^k}\frac{\partial x^l}{\partial \tilde{x}^j}\Gamma^k_{lm} v^m - \frac{\partial^2 \tilde{x}^i}{\partial x^l \partial x^k}\frac{\partial x^l}{\partial \tilde{x}^j} v^k. \tag{70}$$

Again, the first term is all there would have been if $\Gamma^i_{jk} v^k$ transformed as a (1,1) tensor. Interestingly, the second (inhomogeneous) term cancels the corresponding one in (68) so that the covariant derivative transforms as a (1,1) tensor.

Remarks. (a) As mentioned, we define the covariant derivative of a scalar function as its partial derivative: $\nabla_i f = \partial_i f$. Thus $\nabla_i f$ are the components of a 1-form. (b) Equipped with formula (67), it is now easy to check that the covariant derivative of a vector field fv satisfies the Leibniz rule (see Prob. **??**)

$$\nabla_i (f v^j) = (\partial_i f) v^j + f \nabla_i v^j \quad \text{for any scalar function} \quad f. \tag{71}$$

[40]It is noteworthy that the proposed formula (67) for the covariant derivative is linear in the components of v, this will allow it to satisfy appropriate Leibniz-type rules of differentiation.

(c) In Riemannian geometry, the analog of the divergence of a vector field is the covariant divergence: $\nabla \cdot v = \partial_i v^i + \Gamma^i_{ij} v^j$. (d) The covariant derivative of v along a vector field u is defined as having the components $(\nabla_u v)^i = (u \cdot \nabla v)^i = u^i \nabla_i v^j$. We check that it is linear in u in the sense that $\nabla_{fu+hw} v = f \nabla_u v + h \nabla_w v$ for any scalar functions f, h and vector fields u, v, w. (e) A vector field v whose covariant derivative along itself vanishes ($v^i \nabla_i v^j = 0$) is called a *geodesic vector field*. The integral curves of a geodesic vector field are geodesics. This is not a surprise. We have already verified in (66) that the tangent to a geodesic is covariantly constant along the geodesic: $\dot{x}^i \nabla_i \dot{x}^j = 0$. □

Covariant derivative of other tensors. The covariant derivative can be generalized to other tensor fields. To begin with, one postulates that it reduces to the partial derivative on smooth functions $\nabla_i f = \partial_i f$. Next, proceeding as for vector fields, we propose a formula for the covariant derivative of a 1-form $\phi_i dx^i$:

$$\nabla_i \phi_j = \partial_i \phi_j - \Gamma^k_{ij} \phi_k, \tag{72}$$

and check that the RHS transforms as a (0,2) tensor field.

Since a 1-form and a vector field can be contracted to obtain a function (on which the covariant derivative reduces to the partial derivative), we should expect a Leibniz-like identity among covariant derivatives of 1-forms, vector fields and their contraction. In fact, we verify in Prob. **??** that

$$\nabla_i (\phi_j v^j) = \partial_i (\phi_j v^j) = (\nabla_i \phi_j) v^j + \phi_j \nabla_i v^j. \tag{73}$$

The extension to higher rank tensor fields is obtained by requiring ∇ to be a linear map (over $\mathbb{R}$) from the space of (p, q) tensors to $(p, q+1)$ tensors that satisfies the Leibniz rule: $\nabla(s \otimes t) = \nabla s \otimes t + s \otimes \nabla t$ where s, t are a pair of tensor fields. This leads, for instance, to a formula for the covariant derivative of a (0,2) tensor (which can be regarded as a linear combination of tensor products of pairs of (0,1) tensors)[41]:

$$\nabla_i t_{jk} = \partial_i t_{jk} - \Gamma^l_{ij} t_{lk} - \Gamma^l_{ik} t_{jl}. \tag{74}$$

We show in Prob. **??** that the metric tensor g_{jk} is special. It is covariantly constant: $\nabla_i g_{jk} = 0$. This has a geometric meaning, it ensures that the operation of parallel transport that one can define using the Christoffel connection preserves the inner products of vectors.

12 Curvature on a Riemannian manifold

Recall (from Sect. 10) that geodesics are curves that extremize the distance between a pair of points on a Riemannian manifold (manifold M with a metric tensor g). They play the same role that straight lines play in Euclidean space. The separation between straight lines on the Euclidean plane typically grows linearly with time.

[41] In general, aside from the partial derivative, the covariant derivative of a tensor involves one term with Christoffel symbols for each index of the tensor (with a negative sign for lower/covariant indices).

Based on this, we say that the Euclidean plane is flat. On the other hand, the separation between geodesics on the round sphere (which are great circles like longitudes on the Earth) oscillates in time. We say that the round sphere is positively curved. Interestingly, on a hyperboloid or saddle, geodesics typically separate exponentially fast: these surfaces are said to be negatively curved. Somewhat more precisely, curvature measures the behavior of the separation between geodesics that begin nearby. Gauss gave a way of quantifying the curvature of surfaces embedded in $\mathbb{R}^3$ without any reference to geodesics, that nevertheless captures this behavior of geodesics. This notion of curvature is called Gaussian curvature[42].

It would be nice to have a notion of curvature for general Riemannian manifolds. Remarkably, Riemann developed such a generalization of Gaussian curvature to manifolds of any dimension. As with Gaussian curvature, Riemannian curvature is defined without reference to geodesics. Following Christoffel, in Sect. 12.1, we will formulate it in terms of the covariant derivative of Sect. 11. Then in Sect. 12.2 we will relate it to the separation between geodesics by deriving the geodesic deviation equation and finding Riemann's curvature hidden inside it.

12.1 Riemann-Christoffel curvature tensor

Here we define the Riemann-Christoffel notion of curvature on an n-dimensional manifold M with metric tensor g. Suppose u and v are a pair of vector fields: we may think of these as the velocity vector of a geodesic and the separating vector to a nearby geodesic. The Riemannian curvature $R(u, v)$ takes vector fields to vector fields via

$$R(u, v)w = [\boldsymbol{\nabla}_u, \boldsymbol{\nabla}_v]w - \boldsymbol{\nabla}_{[u,v]}w. \tag{75}$$

Here, $\boldsymbol{\nabla}_u$ is the covariant derivative along the vector field u introduced in Sect. 11, $[\boldsymbol{\nabla}_u, \boldsymbol{\nabla}_v]w = \boldsymbol{\nabla}_u(\boldsymbol{\nabla}_v w) - \boldsymbol{\nabla}_v(\boldsymbol{\nabla}_u w)$ and $[u, v]$ is the commutator of vector fields. It is clear that $R(u, v) = -R(v, u)$ is antisymmetric. Remarkably, despite the appearance of all these derivatives, $R(u, v)w$ does not depend on the derivatives of u, v or w. In fact, R behaves linearly (over the ring of functions on M) in all three arguments[43]. This establishes that $R(u, v)$ is a $(1, 1)$ tensor and that R is a tensor of type $(1, 3)$: it acts multilinearly on triples of vector fields to produce another vector field (see Sect. 7). This means that when one changes coordinates, the components of R transform via some Jacobian factors implying that if R vanishes in one coordinate system, it also vanishes in any other coordinate system. In particular, the concept of a

[42]Let Σ be a surface embedded in $\mathbb{R}^3$. At each point $p \in \Sigma$, we have the tangent plane $T_p\Sigma$. A normal plane N_p through p is one that is orthogonal to $T_p\Sigma$. The osculating circle associated to a normal plane N_p is the best quadratic approximation (near p) to the curve of intersection between N_p and Σ. The maximum and minimum radii $R_{1,2}$ of osculating circles through p are called the principle radii of curvature at p. The Gaussian curvature $K(p)$ is defined to have a magnitude equal to $1/R_1 R_2$. Its sign is positive/negative according as the centers of the corresponding osculating circles lie on the same/opposite sides of Σ. Gauss showed that his curvature is an intrinsic property of Σ, it does not depend on the embedding in $\mathbb{R}^3$. It is invariant under isometries [deformations of the surface that preserve lengths and angles].

[43]In particular, using $\boldsymbol{\nabla}_{fu}v = f\boldsymbol{\nabla}_u v$ and $\boldsymbol{\nabla}_u(fv) = u(f)v + f\boldsymbol{\nabla}_u v$, one finds that, $R(fu, v)w = fR(u, v)w$ and $R(u, v)(fw) = fR(u, v)w$ for any scalar function f.

flat manifold (one with zero curvature such as the Euclidean spaces $\mathbb{R}^n$ with $g_{ij} = \delta_{ij}$) becomes independent of the choice of coordinates.

The components of the Riemann-Christoffel tensor $R^l{}_{kij}$ in a basis for vector fields $\{e_1, \cdots, e_n\}$ are defined via

$$R(e_i, e_j)e_k = \left([\boldsymbol{\nabla}_{e_i}, \boldsymbol{\nabla}_{e_j}] - \boldsymbol{\nabla}_{[e_i,e_j]}\right) e_k = R^l{}_{kij} e_l. \tag{76}$$

Thus, for any three vector fields u, v, w,

$$R(u, v)\, w = w^k\, u^i\, v^j R^l{}_{kij}\, e_l. \tag{77}$$

Let us specialize to a coordinate basis $e_i = \frac{\partial}{\partial x^i}$. Since coordinate vector fields commute ($[\partial_i, \partial_j] = 0$), the Riemann tensor simplifies to $R(\partial_i, \partial_j) = [\boldsymbol{\nabla}_i, \boldsymbol{\nabla}_j]$. We now wish to express its components in terms of the Christoffel symbols introduced in Sect. 11. From (67),

$$\begin{aligned}
(R(\partial_i, \partial_j)w)^l &= (\boldsymbol{\nabla}_i(\boldsymbol{\nabla}_j w))^l - (\boldsymbol{\nabla}_j(\boldsymbol{\nabla}_i w))^l \\
&= \partial_i(\boldsymbol{\nabla}_j w)^l + \Gamma^l_{ik}(\boldsymbol{\nabla}_j w)^k - (i \leftrightarrow j) \\
&= \partial_i(\partial_j w^l + \Gamma^l_{jk} w^k) + \Gamma^l_{ik}(\partial_j w^k + \Gamma^k_{jm} w^m) - (i \leftrightarrow j) \\
\Rightarrow \quad w^k R^l_{kij} &= \partial_i \partial_j w^l + \Gamma^l_{jk,i} w^k + \Gamma^l_{jk} \partial_i w^k + \Gamma^l_{ik} \partial_j w^k + \Gamma^l_{ik} \Gamma^k_{jm} w^m \\
&- \quad \partial_j \partial_i w^l - \Gamma^l_{ik,j} w^k - \Gamma^l_{ik} \partial_j w^k - \Gamma^l_{jk} \partial_i w^k - \Gamma^l_{jk} \Gamma^k_{im} w^m.
\end{aligned} \tag{78}$$

As expected from the tensor character of R, all terms involving derivatives of w cancel out. Relabeling $m \leftrightarrow k$ in terms quadratic in Γ, we get

$$w^k R^l_{kij} = \left[\Gamma^l_{jk,i} + \Gamma^l_{im} \Gamma^m_{jk} - (i \leftrightarrow j)\right] w^k \tag{79}$$

with '$, i$' denoting ∂_i. Thus, the components of the $(1, 3)$ Riemann tensor are

$$R^l{}_{kij} = \Gamma^l{}_{jk,i} - \Gamma^l{}_{ik,j} + \Gamma^m_{jk} \Gamma^l{}_{im} - \Gamma^m_{ik} \Gamma^l{}_{jm}. \tag{80}$$

By virtue of (69), they depend on the first two derivatives of the metric g_{ij} and its inverse g^{ij}.

Symmetries. $R^l{}_{kij}$ is clearly antisymmetric in the last two indices $R^l{}_{kij} = -R^l{}_{kji}$. Lowering l, we define a $(0, 4)$ tensor $R_{mkij} = g_{ml} R^l{}_{kij} = g(R(e_i, e_j)e_k, e_m)$ and verify using (80) and (67) that R_{mkij} is antisymmetric in the first two indices and symmetric under exchange of the first pair of indices with the last pair (see Prob. **??**):

$$R_{mkij} = -R_{kmij} \quad \text{and} \quad R_{mkij} = R_{ijmk}. \tag{81}$$

The antisymmetry in the first pair is related to the behavior of a vector when parallel transported around a small parallelogram spanned by u, v. Since parallel transport preserves lengths and angles, such a vector can only undergo a rotation. The antisymmetry in question arises because infinitesimal rotations differ from the identity by an antisymmetric matrix, as shown in Eq. (107). The significance of the pair exchange symmetry will be clarified in the context of the geodesic deviation equation (92).

Ricci tensor and scalar. For some purposes, the $(1, 3)$ Riemann tensor has more information/components than we need, and it is useful to define its second rank and scalar contractions. The Ricci curvature is the $(0, 2)$ tensor defined as $\mathrm{Ric}_{kl} = R^i{}_{kil}$ while the scalar curvature or Ricci scalar is $R = g^{kl}\mathrm{Ric}_{kl}$.

Sectional or Gaussian curvature. The information in the (1,3) Riemann tensor may be packaged in sectional curvatures. These are the curvatures associated to each tangent plane through any point of M. Given a pair of tangent vectors u, v, the sectional curvature in the plane they span is defined as

$$K(u, v) = \frac{\mathscr{k}(u, v)}{\mathrm{Ar}(u, v)^2} = \frac{g(R(u, v)v, u)}{g(u, u)g(v, v) - g(u, v)g(v, u)}. \tag{82}$$

The numerator $\mathscr{k}(u, v)$ is called the curvature biquadratic form and the denominator is the square of the area $(|\boldsymbol{u} \times \boldsymbol{v}|^2)$ of the parallelogram spanned by u and v. It can be shown that K does not depend on the choice of vectors that span a given plane (for instance, rescaling either u or v leaves K unchanged). What is more, the sectional curvature of the tangent plane at any point of a surface embedded in $\mathbb{R}^3$ reduces to its Gaussian curvature.

12.2 Geodesic deviation and Riemannian curvature

Suppose $x(t)$ is a geodesic with affine parameter t, so that $\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k = 0$. Let us look for a nearby geodesic $z^i = x^i + y^i$ where y^i are small. We impose the condition that z^i also solves the geodesic equation up to terms quadratic in y^i to get

$$\ddot{x}^i + \ddot{y}^i + \Gamma^i_{jk}(x + y)\,(\dot{x}^j + \dot{y}^j)(\dot{x}^k + \dot{y}^k) = 0. \tag{83}$$

Using $\Gamma^i_{jk}(x + y) = \Gamma^i_{jk}(x) + \Gamma^i_{jk,l}y^l + \mathcal{O}(y^2)$ and the symmetry of Γ, this becomes

$$\ddot{y}^i + 2\Gamma^i_{jk}\dot{x}^j\dot{y}^k + \Gamma^i_{jk,l}\dot{x}^j\dot{x}^k y^l = 0. \tag{84}$$

This is the geodesic deviation equation (GDE) for the time evolution of the separating vector y. It is also called the Jacobi equation and its solutions y are called Jacobi vector fields (defined along the geodesic $x(t)$). To find the Riemann tensor hidden inside (84), we wish to rewrite it in terms of covariant derivatives.

To this end, we calculate the covariant derivative of the separating vector along the unperturbed geodesic:

$$(\boldsymbol{\nabla}_{\dot{x}}y)^i = \dot{x}^k\partial_k y^i + \Gamma^i{}_{jk}\dot{x}^j y^k = \dot{y}^i + \Gamma^i{}_{jk}\dot{x}^j y^k. \tag{85}$$

Our search for $\ddot{y}$ leads us to evaluate the second covariant derivative:

$$\begin{aligned}
\left(\boldsymbol{\nabla}^2_{\dot{x}}y\right)^i &= \frac{d}{dt}\left(\boldsymbol{\nabla}_{\dot{x}}y\right)^i + \Gamma^i{}_{mn}\dot{x}^m(\boldsymbol{\nabla}_{\dot{x}}y)^n \\
&= \ddot{y}^i + \frac{d}{dt}\left(\Gamma^i{}_{jk}\dot{x}^j y^k\right) + \Gamma^i_{mn}\dot{x}^m\left(\dot{y}^n + \Gamma^n_{jk}\dot{x}^j y^k\right) \\
&= \ddot{y}^i + 2\Gamma^i{}_{jk}\dot{x}^j\dot{y}^k + \Gamma^i_{jk,l}\dot{x}^l\dot{x}^j y^k
\end{aligned}$$

$$-\Gamma^i_{jk} y^k \Gamma^j_{mn} \dot{x}^m \dot{x}^n + \Gamma^i_{mn} \Gamma^n_{jk} \dot{x}^j y^k \dot{x}^m. \tag{86}$$

We used the geodesic equation in the last equality. Thus, the Jacobi equation becomes

$$(\boldsymbol{\nabla}^2_{\dot{x}} y)^i + \Gamma^i_{jk,l} \dot{x}^j \dot{x}^k y^l - \Gamma^i_{jk,l} \dot{x}^l \dot{x}^j y^k + \Gamma^i_{jk} \Gamma^j_{mn} \dot{x}^m \dot{x}^n y^k - \Gamma^i_{mn} \Gamma^n_{jk} \dot{x}^j y^k \dot{x}^m = 0. \tag{87}$$

After relabeling indices, we find that it may be written in terms of the Riemann tensor:

$$(\boldsymbol{\nabla}^2_{\dot{x}} y)^i = R^i{}_{jkl} \dot{x}^j \dot{x}^k y^l. \tag{88}$$

Using the definition $(R(u,v)w)^i = u^k v^l w^j R^i{}_{jkl}$, we get the Jacobi equation

$$\boldsymbol{\nabla}^2_{\dot{x}} y = R(\dot{x}, y)\dot{x} = -R(y, \dot{x})\dot{x}. \tag{89}$$

On a flat manifold, $R^i{}_{jlk} = 0$ and the GDE becomes $\dot{x}^j \dot{x}^k \partial_j \partial_k y^i = 0$ or $\ddot{y}^i = 0$ with solution $y^i(t) = y^i(0) + \dot{y}^i(0)t$. So on flat space, components of a Jacobi vector field grow linearly with time (if $\dot{y}^i(0) \neq 0$) or remain constant (if $\dot{y}^i(0) = 0$). These correspond to two straight lines starting nearby but in different directions or with parallel initial velocities.

Relating geodesic separation to sign of sectional curvature. To understand and interpret solutions of (89), it is convenient to expand the Jacobi vector field $y(t) = \sum_k c_k(t) e_k(t)$ in an orthonormal basis $\{e_1(t), e_2(t), \ldots, e_n(t)\}$ for $T_{x(t)}M$ that is parallel transported along the geodesic, i.e., $\boldsymbol{\nabla}_{\dot{x}} e_k = 0$. Then the Jacobi equation becomes

$$\ddot{c}_k e_k = -c_k R(e_k, \dot{x})\dot{x} \tag{90}$$

with a sum on k implied. Taking an inner product with e_j and using orthonormality $(e_j, e_k) = \delta_{jk}$, the GDE becomes $\ddot{c}_j(t) = -g(R(e_k, \dot{x})\dot{x}, e_j) c_k(t)$ or

$$\ddot{c}_j = -\sum\nolimits_k \mathcal{R}_{jk}(t) c_k \quad \text{where} \quad \mathcal{R}_{jk}(t) = g(e_j(t), R(e_k(t), \dot{x}(t))\dot{x}(t)). \tag{91}$$

So the components of the separating vector satisfy a coupled system of linear ODEs with time-dependent coefficients. The geodesic deviation matrix $\mathcal{R}_{jk}(t)$ is a real symmetric matrix[44], so it can be diagonalized by a time-dependent orthogonal transformation to an orthonormal eigenbasis $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$ where $\mathcal{R}_{jk}$ is diagonal with eigenvalues κ_j along the diagonal:

$$\tilde{\mathcal{R}}_{jk} = g(\tilde{e}_j, R(\tilde{e}_k, \dot{x})\dot{x}) = \kappa_j \delta_{jk}. \tag{93}$$

[44]The matrix $\mathcal{R}_{jk}(t)$ is obviously real. Symmetry follows from definition (77) and the index pair exchange symmetry of the (0,4) Riemann tensor (81):

$$\begin{aligned} \mathcal{R}_{ij} &= g(R(e_j, \dot{x})\dot{x}, e_i) = g_{kl} (R(e_j, \dot{x})\dot{x})^k (e_i)^l = g_{kl} R^k{}_{pmn} (e_j)^m \dot{x}^n \dot{x}^p (e_i)^l \\ &= R_{lpmn} (e_i)^l \dot{x}^p (e_j)^m \dot{x}^n = R_{mnlp} (e_i)^l \dot{x}^p (e_j)^m \dot{x}^n \\ &= R_{lpmn} (e_i)^m \dot{x}^n (e_j)^l \dot{x}^p = \mathcal{R}_{ji}. \end{aligned} \tag{92}$$

What is more, the eigenvalue κ_j is simply the curvature biquadratic form (82) in the plane spanned by the tangent $\dot{x}$ to the unperturbed geodesic and the eigenvector $\tilde{e}_j$:

$$\kappa_j = g(R(\tilde{e}_j, \dot{x})\dot{x}, \tilde{e}_j) = \mathscr{r}(\tilde{e}_j, \dot{x}) = (\text{Area}\langle \tilde{e}_j, \dot{x} \rangle)^2 \, K(\tilde{e}_j, \dot{x}). \tag{94}$$

The separating vector may also be expanded in the eigenbasis, $y = \sum_j \tilde{c}_j \tilde{e}_j$. The Jacobi equation for the components $\tilde{c}_j$ is decoupled:

$$\ddot{\tilde{c}}_j = -\kappa_j(t)\tilde{c}_j \quad \text{for} \quad j = 1, 2, \ldots, n. \tag{95}$$

Each is the equation for a linear oscillator (**??**) with a time-dependent squared frequency. However, both signs of κ_j are possible. Thus, we should expect the coefficients $\tilde{c}_j$ to oscillate when κ_j are positive and to grow exponentially for negative κ_j. Thus, at least for short times (till the linearized approximation breaks down), the behavior of neighboring geodesics is determined by the sign of the curvature biquadratic (or equivalently, the sign of the sectional curvature) in the plane spanned by the tangent to the original geodesic and the initial separating vector. Positive sectional curvatures lead to oscillatory behavior of nearby geodesics while negative sectional curvatures lead to exponential growth in separation.

13 Groups

Definition. A group is a mathematical construct that, among other things, helps us express and work with symmetries. Groups occur in various parts of physics such as crystallography, atomic physics, relativity and particle physics [**?**, **?**]. They help to recognize and organize patterns, but can also enter dynamical principles that constrain or determine the nature of forces. For instance, the angular distribution of possible locations of an electron in a hydrogen atom can be understood using the spherical symmetry of the electric potential felt by the electron. On the other hand, the strong nuclear force among quarks and gluons is determined by a gauge symmetry principle based on a so-called color symmetry. In mechanics, groups typically arise as families of symmetry transformations among states or configurations or solutions of the equations of a system. For example, rotations act on the possible locations of a planet in the Kepler problem while the $x \to -x$ reflection acts as a symmetry of an even harmonic oscillator potential $V(x) = \frac{1}{2}kx^2$ felt by a particle attached to a spring. However, it is advantageous to separate the algebraic concept of a group from its action on a space. Thus, we will begin by defining an 'abstract' group and later discuss how it may be realized via an action on an auxiliary space like the state space of a mechanical system. Precisely, a group G is a set of elements $g, h, k, \ldots$ among which a law of composition $G \times G \to G$ is defined: if $g, h \in G$, then their product or composition $gh \in G$. The product must satisfy the following properties. (i) It must be associative, i.e., $g(hk) = (gh)k$ for any $g, h, k \in G$. (ii) G must include an identity element e (sometimes denoted 1 or I) with $ge = eg = g$ for any $g \in G$. (iii) Every element g must have a two-sided inverse g^{-1}, i.e., $gg^{-1} = g^{-1}g = e$. Useful consequences are $(gh)^{-1} = h^{-1}g^{-1}$ and the cancellation law: if $gh = gk$ then $h = k$.

Cardinality, discrete and continuous groups. The number of elements $|G|$ in a group G is called its order or cardinality. A group of finite order is called a finite group. The 'trivial' group has just one element, the identity: $G = \{1\}$ with $1 \cdot 1 = 1$. The set $C_2 = \{1, -1\}$ under multiplication is a group of order two. While 1 is the identity, $(-1)(-1) = (-1)^2 = 1$. We say that -1 generates C_2 since -1 and $(-1)^2$ account for all the distinct elements. C_2 is called the cyclic group of order two. Notice that ± 1 are the two square-roots of unity. More generally, for $n = 1, 2, 3, \ldots$, we have the (multiplicative) cyclic group of order n consisting of the n^{th} roots of unity $\{1, e^{2\pi i/n}, e^{4\pi i/n}, \ldots, e^{2(n-1)\pi i/n}\}$. It is generated by $e^{2\pi i/n}$ and we write $C_n = \langle e^{2\pi i/n} \rangle$. For $n = 1, 2, 3, \ldots$, the set $\mathbb{Z}_n = \{0, 1, \cdots, n-1\}$ with composition given by addition modulo n (e.g., $2 + 3 \equiv 1 \pmod 4$) is also a cyclic group of order n. The identity element is 0 and 1 is its generator. Note that $1 + 1 + \cdots + 1$ (n summands) $= n1 \equiv 0 \pmod n$. We will soon see that $\mathbb{Z}_n$ and C_n are different presentations of the same group: up to 'isomorphism' there is just one cyclic group of a given order. Infinite groups could be discrete (like the additive group of integers $\mathbb{Z}$ or the infinite cyclic group) or continuous (like the multiplicative group of complex numbers of unit magnitude). The latter group is denoted $U(1)$ and its elements may be represented as $z(\theta) = e^{i\theta}$ for a real angle θ which is defined modulo 2π (see Fig. 6). Composition is given by $z(\theta_1)z(\theta_2) = e^{i(\theta_1+\theta_2)} = z(\theta_1 + \theta_2)$.

Subgroup. A subset H of a group G is called a *subgroup* if it satisfies the group axioms with respect to the operations inherited from G [see Prob. **??**]. The identity subgroup $H = \{e\}$ and $H = G$ are subgroups of any group G. Examples: (i) $C_2 = \{\pm 1\}$ and more generally C_n are subgroups of $U(1)$. (ii) Given any element g of a group G, it generates a (cyclic) subgroup, namely the set of its powers $\langle g \rangle = \{g^0 = e, g, g^{-1}, g^2, g^{-2}, \cdots\}$. If there is a smallest positive integer n such that $g^n = e$, then $\langle g \rangle$ is essentially the same as (i.e., isomorphic to) a cyclic group of order n and otherwise it is an infinite cyclic group. (iii) Every finite group may be realized as a subgroup of a group of permutations (see Prob. **??**).

Group homomorphisms. Given a pair of groups, a map $\phi : G \to G'$ is called a *homomorphism* if it preserves products: $\phi(g_1 g_2) = \phi(g_1)\phi(g_2)\ \forall\ g_1, g_2 \in G$. If ϕ preserves products then (see Prob. **??**) ϕ maps the identity in G to that in G' and maps inverses to inverses: $\phi(g^{-1}) = \phi(g)^{-1}$. The homomorphic image $\phi(G) \subseteq G'$ is a subgroup of G'. A homomorphism $\phi : G \to G'$ is an isomorphism if it is bijective (1-1 and onto, and hence invertible). Two groups G and G' are *isomorphic* (denoted $G \cong G'$) if there is an isomorphism between them. Isomorphic groups are algebraically identical but could arise or be presented differently. E.g., C_n and $\mathbb{Z}_n$ are isomorphic, with the isomorphism mapping the generators to each other: $\phi(e^{2\pi i/n}) = 1$, so that $\phi(e^{2\pi ij/n}) = j$ for $j = 0, 1, \cdots, n-1$. The group of unimodular complex numbers $U(1)$ is isomorphic to that of 2×2 orthogonal matrices (real A with $A^t A = I$) with unit determinant ($SO(2)$). Composition is given by matrix multiplication. Its elements are $A(\theta) = (\cos\theta, \sin\theta| - \sin\theta, \cos\theta)$ for a real angle θ defined modulo 2π. The isomorphism maps $z = e^{i\theta}$ to $A(\theta)$. Verify that under matrix multiplication, $A(\theta_1)A(\theta_2) = A(\theta_1 + \theta_2)$.

Isomorphisms and Automorphisms. An isomorphism ϕ from a group G to itself is called an *automorphism*. Every group has the identity or trivial automorphism defined by $\phi(g) = g$ for all $g \in G$. $C_2 = \{1, -1\}$ has no nontrivial automorphism since we cannot define $\phi(1) = -1$. Verify that C_3 has just one nontrivial automorphism given by $\phi(1) = 1, \phi(\omega) = \omega^2$ and $\phi(\omega^2) = \omega$ where $\omega = e^{2\pi i/3}$. Since an automorphism must preserve the algebraic structure, it must take a generator to another generator: we check that both ω and ω^2 are generators of C_3. $C_4 = \{1, i, -1, -i\}$ also has one nontrivial automorphism: it exchanges the two generators: $\phi(1) = 1, \phi(i) = -i, \phi(-1) = -1, \phi(-i) = i$.

Conjugation. Given a group G, we say that $k \in G$ is conjugate to $h \in G$ if $k = ghg^{-1}$ for some[45] $g \in G$. Conjugation $\phi_g(h) = ghg^{-1}$ by a fixed element g defines an automorphism of G. It is called an *inner automorphism*. It is a homomorphism since $\phi_g(h_1 h_2) = g h_1 h_2 g^{-1} = g h_1 g^{-1} g h_2 g^{-1} = \phi_g(h_1)\phi_g(h_2)$. It is $1-1$ since $\phi_g(h_1) = \phi_g(h_2)$ implies $g h_1 g^{-1} = g h_2 g^{-1}$ whence $h_1 = h_2$. It is surjective since given any $k \in G$ we can always find an $h \in G$ such that $\phi_g(h) = k$, in fact $h = g^{-1} k g$. What is more, the inverse of ϕ_g is just $\phi_{g^{-1}}$.

The *conjugacy class* of h is the set $C_h = \{ghg^{-1} | g \in G\}$. The identity element is always in a conjugacy class by itself $C_e = \{e\}$. Conjugacy is an equivalence relation[46]. This implies G is a disjoint union of conjugacy classes.

Abelian and nonabelian groups. The nature of conjugation and conjugacy classes are related to the notion of a commutative group. We begin by defining the *group commutator* of a pair of elements as $[g, h] = ghg^{-1}h^{-1}$. The commutator measures the extent to which gh and hg differ. If gh and hg are the same, then $[g, h] = e$ and they are said to commute. A group is called *abelian* or *commutative* if $[g, h] = e$ or $gh = hg$ or $ghg^{-1} = h$ for all $g, h \in G$. Otherwise, it is nonabelian. Evidently, a group is abelian iff all conjugacy classes are singleton sets or equivalently, if every inner automorphism is the identity. Roughly, conjugacy classes get longer the more nonabelian a group is. Only the first 4 groups below are abelian.

Examples. There are many elementary examples of groups that arise in interesting ways, some of which we have met: (i) the multiplicative group $C_2 = \{1, -1\}$ consisting of the identity and reflection symmetry $x \to -x$ of an even potential $V(x)$ in one dimension, (ii) the cyclic group C_5 of order 5, of rotational symmetries[47] of a regular pentagon (if one includes reflection symmetries, one obtains the dihedral group of order 10), (iii) the groups $\mathbb{R}^3$ and $\mathbb{R}$ of translations of 3d Euclidean space and time, (iv) the group $SO(2)$ of rotational symmetries of a circle or an axisymmetric (cylindri-

[45] If g works, so do gh^n for $n \in \mathbb{Z}$ and more generally gg' for any g' that commutes with h.

[46] Conjugacy is reflexive: $h = ehe^{-1}$ (h is conjugate to h), symmetric: $h = g'kg'^{-1}$ where $g' = g^{-1}$ (h is conjugate to k if k is conjugate to h) and transitive: k conjugate to h and h conjugate to l implies k conjugate to l. A binary relation with these properties is called an equivalence relation. It ensures that conjugacy classes either coincide or do not overlap. For instance, transitivity implies that C_{h_1} and C_{h_2} cannot have a 'partial' overlap.

[47] Cyclic and dihedral groups are *point groups* in 2d. They are symmetries of regular polygons and molecules with a fixed point and are discrete subgroups of the orthogonal group. *Space groups* are symmetries of an infinite crystal and include discrete translations.

cally symmetric) potential, (v) the group $SO(3)$ of proper rotations of 3d Euclidean space, (vi) the group $O(3)$ of rotations and reflections of $\mathbb{R}^3$, (vii) the Galilei group and (viii) the groups S_2 and S_3 of permutations of two and three objects encountered in Footnote 25 of §6 and Footnote 29 of §7.

Lie groups. While examples (i), (ii) and (viii) are discrete groups (in fact with finitely many elements), the rest are examples of continuous groups, where the group elements form continuous families and can be used to model continuous symmetries. Historically, discrete groups arose, in part, in modeling discrete symmetries of algebraic equations, while continuous groups arose via continuous symmetries of differential equations. Prominent among continuous groups are Lie groups, named after the Norwegian mathematician Sophus Lie. A Lie group is a group which is also a differentiable manifold, with the group operations of composition $(g, h) \mapsto gh$ and inversion $g \mapsto g^{-1}$ being smooth maps from $G \times G \to G$ and $G \to G$. The dimension $\dim G$ of a Lie group G is the dimension of the corresponding group manifold. Note that the Cartesian product $G \times G$ inherits a $2(\dim G)$-dimensional manifold structure from G upon using ordered pairs of charts and transition functions. The concept of smooth maps is as introduced in §1.

Matrix Lie groups. Natural examples of Lie groups are the matrix groups $GL_n(\mathbb{R})$ and $GL_n(\mathbb{C})$ of invertible $n \times n$ real and complex matrices with composition and inversion given by matrix multiplication and inversion (Nb. GL stands for general linear). Since matrix multiplication is generally noncommutative, these groups for $n > 1$ are nonabelian. Other examples of 'classical' Lie groups[48] such as the special linear, orthogonal, symplectic and unitary groups arise as closed subgroups of $GL_n(\mathbb{R})$ and $GL_n(\mathbb{C})$. The special linear groups $SL_n(\mathbb{R})$ and $SL_n(\mathbb{C})$ consist of invertible matrices with unit determinant. The orthogonal and special orthogonal groups $O(n)$ and $SO(n)$ consist of orthogonal matrices ($A^t A = I$) in $GL_n(\mathbb{R})$ and $SL_n(\mathbb{R})$ respectively. Similarly, the unitary and special unitary groups $U(n)$ and $SU(n)$ consist of unitary matrices ($U^\dagger U = I$) in $GL_n(\mathbb{C})$ and $SL_n(\mathbb{C})$. The symplectic group $Sp(2n, \mathbb{R})$ consists of $2n \times 2n$ matrices M that preserve the canonical symplectic structure: $M^t \omega M = \omega$ where $\omega = (0, -I|I, 0)$ and I is the $n \times n$ identity matrix. They are the linear canonical transformations of the phase space $\mathbb{R}^{2n}$. Later in this section, we will discuss some basic properties of Lie groups in the context of the orthogonal group.

Transformation group acting on a set. In physics, groups often arise as families of (often symmetry) transformations of a space M such as a configuration or state space or the space of solutions of equations of motion. For instance, the rotation group $SO(3)$ acts on the configuration space $\mathbb{R}^3$ of a particle in a spherically symmetric potential by rotating the radius vector about the force center: $\boldsymbol{r} \mapsto R\boldsymbol{r}$ for $R \in SO(3)$. Evidently, such a 'transformation group' is to be regarded as an action of an abstract group G on a set M. Precisely, an action of G on M is a map from $G \times M \to M$ taking $m \in M$ to $g \cdot m \in M$ such that $e \cdot m = m$ and $g \cdot (h \cdot m) = (gh) \cdot m$ for all $g, h \in G$ and $m \in M$. The set of points that a given point $m \in M$ can be mapped

[48] 'Classical' here is used to mean that these were among the first Lie groups to be studied.

to, $\mathcal{O}_m = \{g \cdot m | g \in G\}$ is called the orbit of m under the action of G. The action is said to be *transitive* if every point of M can be mapped to every other point of M by the action of some group element. In other words, the action is transitive if M is the orbit of any of its points. The action of rotations on $\mathbb{R}^3$ is not transitive: for instance, the origin cannot be mapped to any other point by a rotation. On the other hand, translations act transitively on $\mathbb{R}^3$: any point can be translated to any other point.

Lie group as a homogeneous manifold. A manifold M (or even just a topological space or set) is homogeneous for a group G if it carries a transitive action of G. To be meaningful, one needs to specify the nature of the manifold M (topological, smooth or geometrically rigid like a Riemannian manifold) and the action of the group must respect that structure. Roughly, all points of a homogeneous manifold look locally the same. For example, the unit circle $x^2 + y^2 = 1$ on the plane is homogeneous under the action of the group $SO(2)$ of rotations about the z axis. The time axis $\mathbb{R}$ is homogeneous under the action of the group of time-translations $t \mapsto t + s$. Euclidean space $\mathbb{R}^3$ is homogeneous under the action of the group of space-translations $\boldsymbol{r} \mapsto \boldsymbol{r} + \boldsymbol{s}$. The term *homogeneous* should ring a bell: recall the homogeneity of time and space. By contrast, the rigid toroidal surface of an inflated tube of a car tyre is *not* homogeneous for the action of the group of rotations about the axle. This action is not transitive since it cannot change the distance of a point on the tyre from the axle. In fact, near a point on the inner rim, the tubular surface looks like a saddle or mountain pass while near a point on the outer rim, it looks like a hill, so neighborhoods of points do not all look the same. Though not a group[49], the round unit sphere S^2 is homogeneous for the rotation group in 3d as it carries a transitive action of the latter: any point can be rotated to any other point on S^2. An ellipsoid of revolution $E = \{x^2 + y^2 + 2z^2 = 1\}$ regarded as a rigid surface in $\mathbb{R}^3$ is not homogeneous under 3d rotations since they do not preserve E. Rotations about the z-axis act on the ellipsoid, though not transitively.

Any group is homogeneous under its own action: G acts on itself transitively via both left and right multiplication. We define the left action of G on itself via $L_g h = gh$ for any $g, h \in G$. The action is transitive since, given any $h, k \in G$, we have $L_{kh^{-1}} h = k$. The right action $R_g h = hg$ is similarly transitive. The right and left actions coincide if G is abelian. For a Lie group, L_g and R_g are diffeomorphisms of G. Thus, a Lie group G is a homogeneous manifold under the action of G.

Lie algebra of a Lie group. Among the points of G, the identity is distinguished by its simplicity. It makes sense to begin a detailed study of G by focusing on the linear neighborhood of the identity[50]. This leads to the idea of the Lie algebra $\underline{G}$, which, as a vector space, is the tangent space at the identity $T_e G$. Each tangent vector at the identity is a Lie algebra element. We may use left translations L_g (by all elements of

[49] Any Lie group has at least one nonvanishing vector field: the left-invariant vector field obtained by pushing forward a nonzero tangent vector at the identity. However, as noted in Fig. 5b, S^2 does not admit a nonvanishing vector field.

[50] By homogeneity, the linear neighborhood $T_g G$ of any other point $g \in G$ may be studied by left- or right-translating the tangent space at the identity via L_g or R_g. This idea is also used in studying the rotational dynamics of a rigid body.

G) to pushforward (38) any fixed tangent vector $u \in T_e G$ to obtain a 'left-invariant' vector field $L_{g*}u$ on G. Thus, for each $u \in \underline{G}$ we get an associated left-invariant vector field on G. Consequently, the Lie algebra may also be regarded as the space of left-invariant (or right-invariant) vector fields on G. The algebraic structure of the group endows $T_e G$ (or the space of left-invariant vector fields) with the additional structure of a linear Poisson algebra (i.e., with a bilinear antisymmetric product or 'Lie bracket' satisfying the Jacobi identity). In fact, the Lie bracket is simply the commutator of left-invariant vector fields (which is known to be antisymmetric and to satisfy the Jacobi identity, see Prob. **??**). For the Lie algebra of a matrix Lie group, the Lie bracket may be realized concretely in terms of the commutator of matrices, as we shall soon see in the context of the rotation group (**??**). In fact, we may write a group element in the infinitesimal neighborhood of the identity as $g = e^{su} \approx I + su + s^2u^2/2$ for a real s with $|s| \ll 1$. The matrix u is then an element of the Lie algebra. The group commutator $[g, h]$ of two such elements $g = e^{su}$ and $h = e^{tv}$ may be shown to be $[g, h] \approx I + st[u, v]$ (see Prob. **??**). Thus, the matrix commutator of Lie algebra elements is the first nontrivial approximation to the group commutator.

Coset spaces. The idea of a group acting on itself is extremely useful and can be used to 'subdivide' a group. Given a subgroup H of G, we may consider all its left translates, i.e., the subsets $gH = \{gh | h \in H\}$ where g ranges over elements of G. The subsets gH are called left cosets of G by H. Note that distinct elements of G may produce the same coset. For instance, all elements $h_1, h_2, \ldots$ of H give rise to the same coset[51] $h_1 H = h_2 H = eH = H$. Moreover, all cosets have the same cardinality as H. In fact, the elements of the list gH are all distinct.

Additionally, two cosets are either the same or disjoint: $g_1 H = g_2 H$ or $g_1 H \cap g_2 H = \{\}$. The former happens if $g_1 = g_2 h$ for some $h \in H$ and the latter happens if there is no such $h \in H$. Thus, a group is a disjoint union of (left) cosets[52]. The set of left cosets forms the left coset space denoted G/H and pronounced 'G mod H'. Similarly, the right translates Hg of the subgroup by elements of G leads to the right coset space, denoted $H\backslash G$. It is often convenient to pick an element from each coset and use it as a representative for the coset. For example, the even integers $2\mathbb{Z}$ form a subgroup of the additive group of integers $\mathbb{Z}$. There are only two cosets: the sets of even and odd integers: $2\mathbb{Z}$ and $2\mathbb{Z} + 1$ (left and right cosets coincide since addition is commutative). In this case, we could pick 0 and 1 as the two coset representatives.

Normal subgroup and quotient or factor group. In general, neither the space of left nor right cosets is a group. However, they acquire the structure of a group if H is a so-called normal or invariant subgroup of G. Precisely, N is a normal subgroup (denoted $N \lhd G$) if each left coset gN is also a right coset Ng for the same $g \in G$.

[51] In formulae such as $h_1 H = h_2 H$ we mean that the two sets are the same, although the order of elements in the two lists may differ.

[52] (Left) cosets may be interpreted as equivalence classes. For any two elements of G, define the relation $g \sim g'$ if there is an $h \in H$ such that $gh = g'$. This relation is reflexive ($g \sim g$ since $ge = g$), symmetric ($g \sim g' \Rightarrow g' \sim g$ since $gh = g'$ implies $g'h^{-1} = g$) and transitive ($g \sim g'$ and $g' \sim g''$ implies $g \sim g''$ since $gh = g'$ and $g'h' = g''$ implies $ghh' = g''$) and therefore an equivalence relation. Evidently, the equivalence class of g is the left coset gH.

The name invariant subgroup comes from a reinterpretation of this condition. Indeed, consider the action of G on itself by conjugation: $h \mapsto A_g(h) = ghg^{-1}$. Evidently, conjugation is the composition of left and right actions: $A_g = L_g R_{g^{-1}} = R_{g^{-1}} L_g$. Now, a subgroup N is normal if $gN = Ng$ or $gNg^{-1} = N$, i.e., if it is invariant under conjugation by any element of G. It is then easy to see that the set of left (or right) cosets of G by N is a group with identity given by the coset $eN = N$. Indeed, the group multiplication and inversion (for left cosets) are given by

$$(gN)\,(g'N) = (gg')N \quad \text{and} \quad (gN)^{-1} = g^{-1}N \tag{96}$$

Here, we used the formulae: $gNg'N = gg'NN = gg'N$ and $(gN)^{-1} = N^{-1}g^{-1} = Ng^{-1} = g^{-1}N$ since $N = N^{-1}$ on account of it being closed under inverses. G/N is called the quotient group or factor group. Some elementary properties are worth noting. (i) Every subgroup of an abelian group is a normal subgroup. (ii) If G is finite, then the cardinality of the coset space G/H is $|G|/|H|$ (Lagrange's theorem). (iii) If N is an invariant Lie subgroup of the Lie group G, then the dimension of the coset space G/N is the difference between the dimensions of G and N. (iv) The kernel K (inverse image $\phi^{-1}(e')$ of the identity $e' \in G'$) of a group homomorphism $\phi : G \to G'$ is always a normal subgroup of G (see Prob. **??**) and the image $\phi(G)$ is isomorphic to G/K. (v) The center $Z(G)$ of a group G, consisting of elements that commute with all other elements, is an abelian normal subgroup. It is normal since every group element commutes with elements in the center: $Zg = gZ$.

Commutator subgroup. The commutator subgroup $[G, G]$ consists of products of any number of group commutators $[h_1, h_2][h_3, h_4] \cdots$ for $h_i \in G$. Notably, the inverse of a commutator is a commutator: $[h, k]^{-1} = [k, h]$, so this is a subgroup. It is normal since the conjugate of a commutator is the commutator of conjugates. The quotient $G/[G, G]$ is an abelian group since we have factored out all commutators. More formally, $gh = [g, h]hg$ and $gh = hg[g^{-1}, h^{-1}]$. Thus, gh and hg lie in the same right (and left) coset so that the composition of cosets is commutative:

$$(g[G, G])(h[G, G]) = gh[G, G] = hg[G, G] = (h[G, G])(g[G, G]). \tag{97}$$

In fact, $[G, G]$ is the smallest invariant subgroup such that the factor group is abelian.

Simple and semisimple groups. A simple group G is one that does not have any normal subgroups other than G and $\{e\}$. Simple groups are like prime numbers, they do not admit any nontrivial factor groups and can serve as building blocks for other groups. The cyclic group C_p for prime p is simple as every nontrivial element is a generator. By contrast, $C_4 = \{\pm 1, \pm i\}$ is not simple: $C_2 = \{\pm 1\}$ is a normal subgroup. More generally, G is semisimple if G has no nontrivial abelian invariant subgroups. If G is simple, then it is automatically semisimple. A connected nonabelian Lie group is called simple if it does not have any proper connected normal Lie subgroups (it can have discrete normal subgroups). $SO(3)$, $SU(2)$ and $SL_2(\mathbb{R})$ are simple Lie groups while $SO(4)$ is semisimple but not simple. The unitary groups $U(n)$ and general linear groups $GL_n(\mathbb{R})$ for $n \geq 2$ are neither simple nor semisimple since multiples of the identity ($e^{i\theta}I$ for real θ and λI for nonzero real λ) form nontrivial abelian connected invariant Lie subgroups. In fact, $U(2)$ also admits $SU(2)$ as a normal subgroup.

We now introduce two ways in which we may combine a pair of groups to synthesize a larger one: the direct and semidirect products.

Direct product. Suppose H and N are a pair of groups with identity elements e_H and e_N. Then the Cartesian product $H \times N$ consisting of all ordered pairs (h,n) with $h \in H$ and $n \in N$ can be given the structure of a group called the direct product of H and N. The composition law is defined as $(h,n)\cdot(h',n') = (hh', nn')$ and $(h,n)^{-1} = (h^{-1}, n^{-1})$. The subgroups consisting of elements of the form (e_H, n) and (h, e_N) are isomorphic to N and H respectively. $H \times N$ and $N \times H$ are isomorphic groups. When the groups are abelian, one tends to use additive rather than multiplicative notation. For example, the group $\mathbb{R}^2$ of translations of the Euclidean plane is the direct product (or sum) of two copies of the group $\mathbb{R}$ of translations of the real line: $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

$*$ **Semidirect product.** The semidirect product is a generalization of the direct product. Here, we suppose that we are given an action of H on N. More precisely, for each $h \in H$, we have an automorphism $\varphi_h : N \to N$ such that $\varphi_{h'}\varphi_h = \varphi_{h'h}$ and $\varphi_h^{-1} = \varphi_{h^{-1}}$. We may use this to define the composition law $(h,n)\cdot(h',n') = (hh', n\varphi_h(n'))$. We verify in Prob. **??** that $H \times N$ with this composition law is a group. It is called the semidirect product of H acting on N via φ and is denoted $H \rtimes N$. Evidently, the semidirect product reduces to the direct product if H acts trivially on N, i.e., $\varphi_h(n') = n'$ for all $h \in H$ and $n' \in N$. What is more, the set of elements (e_H, n) forms a normal subgroup of $H \rtimes N$ isomorphic to N. The Euclidean group is a semidirect product of 3d rotations acting on space translations. The Galilei and Poincaré groups are semidirect products of the group of rotations and boosts acting on space-time translations.

We now illustrate some of the concepts introduced so far in the context of the discrete permutation group and the continuous circle and orthogonal groups.

Permutation group. The permutation group S_n or symmetric group on n letters is the set of all permutations of n distinct objects, usually denoted $1, 2, \cdots, n$, with group multiplication given by composition of permutations. A permutation σ may be written in two-row notation as $\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 & \cdots \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots \end{smallmatrix}\right)$. The group has order $n!$ since $\sigma(1)$ can be chosen in n ways followed by $\sigma(2)$ in $n-1$ ways and so on. A permutation may also be written as a product of disjoint cycles: its cycle decomposition. For $k = 0, 1, \ldots$, a $(k+1)$-cycle is of the form $(i\ \sigma(i)\ \sigma^2(i) \cdots \sigma^k(i))$ with $\sigma^{k+1}(i) = i$. For example, S_2 consists of 2 elements: the identity $\sigma = e$ [with $e(1) = 1, e(2) = 2$] and exchange transposition $\sigma = \tau$ [$\tau(1) = 2, \tau(2) = 1$] with $\tau^2 = e$. Thus,

$$e = \left(\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix}\right) = (1)(2) \quad \text{and} \quad \tau = \left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right) = (12). \tag{98}$$

The group S_3 has 6 elements. The identity $\sigma(i) = i$ is denoted $(1)(2)(3)$. There are three pairwise transpositions[53] $(12)(3), (1)(23)$ and $(2)(31)$. Here, $(1)(23)$ means $\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 2$. There are also two cyclic permutations $(123) =$

[53]When clear from context, we suppress 1-cycles. So in S_3, (23) is short for $(1)(23)$.

$g \downarrow,\ \ h \rightarrow$	e	(12)	(23)	(31)	(123)	(132)
e	e	(12)	(23)	(31)	(123)	(132)
(12)	(12)	e	(123)	(132)	(23)	(31)
(23)	(23)	(132)	e	(123)	(31)	(12)
(31)	(31)	(123)	(132)	e	(12)	(23)
(123)	(123)	(31)	(12)	(23)	(132)	e
(132)	(132)	(23)	(31)	(12)	e	(123)

Table 1: Multiplication table of gh for $g, h \in S_3$

$(12)(23) = (13)(12)$ and $(132) = (12)(13)$ which have been written as products of pairwise exchanges composed from right to left. Here (132) means $\sigma(1) = 3, \sigma(3) = 2$ and $\sigma(2) = 1$. In the composition $\sigma = (12)(13)$, 3 is mapped to 1 which is then mapped to 2, so that $\sigma(3) = 2$. On the other hand, $\sigma(2) = 1$ and $\sigma(1) = 3$.

S_3 can be realized as the group of rigid motion *symmetries of an equilateral triangle* Δ with vertices labelled v_1, v_2, v_3, say counterclockwise, with horizontal base $v_1 v_2$ and apex v_3. The symmetries of Δ are counterclockwise rotations R_θ about the center by angles $\theta = 0, 2\pi/3, 4\pi/3$ and reflections about the perpendiculars through the vertices v_1, v_2 and v_3. The transformation group consisting of these 6 symmetries is called the *dihedral group* of order 6 and is isomorphic to S_3 via the following map. To R_0 we associate the identity element e. $R_{2\pi/3}$ is mapped to (123) since it takes $v_1 \to v_2$, $v_2 \to v_3$, $v_3 \to v_1$. Similarly, $R_{4\pi/3}$ corresponds to (132) as it takes $v_1 \to v_3$, $v_3 \to v_2$, $v_2 \to v_1$. In the same spirit, reflection through the perpendicular through v_1 is mapped to (23) and so on. Notice that the square of any reflection is the identity and that $R_{2\pi/3}^2 = R_{4\pi/3}$ while $R_{4\pi/3}^2 = R_{8\pi/3} = R_{2\pi/3}$. Correspondingly, the square of any transposition is the identity while $(123)^2 = (132)$ and $(132)^2 = (123)$. The 'multiplication table' of S_3 is displayed in Table. 1. Evidently, it is a nonabelian group. In general, reflections do not commute $[(12)(23) = (123)$ while $(23)(12) = (132)]$ nor do rotations commute with reflections: $(123)(12) = (31)$ while $(12)(123) = (23)$.

By Lagrange's theorem, since the order of a subgroup must divide that of the group, S_3 can only have *subgroups* of order $1, 2, 3$ and 6. There are 4 nontrivial subgroups, each is cyclic and is generated by a transposition or cyclic permutation:

$$\{e, (12)\}, \quad \{e, (23)\}, \quad \{e, (31)\} \quad \text{and} \quad \{e, (123), (132)\}. \tag{99}$$

The first 3 are reflection symmetries while the fourth consists of rotations of Δ. Pairwise transpositions are the building blocks: any permutation can be expressed as a product of transpositions, although the expression is not unique. However, a permutation σ requires an even or odd number of transpositions to be expressed this way. Thus, we define the *sign (or signature or parity) of a permutation* $\operatorname{sgn}(\sigma)$ as ± 1 in the even and odd cases. The identity has sign $+1$ and any exchange has sign -1. For S_3, cyclic permutations have sign $+1$ as $(123) = (31)(12)$ and $(132) = (12)(31)$.

The sign of a permutation gives a homomorphism: $S_n \to C_2$. The kernel is the *alternating group* A_n of even permutations, a normal subgroup[54] of S_n. For $n = 3$, $A_3 = \{e, (123), (132)\}$ consists of rotational symmetries of Δ. It has 2 left cosets

$$\begin{aligned} & (12)A_3 = (23)A_3 = (31)A_3 = \{(12), (23), (31)\} \\ \text{and} \quad & eA_3 = (123)A_3 = (132)A_3 = \{e, (123), (132)\} = A_3. \end{aligned} \tag{100}$$

As expected, the left cosets are also right cosets, i.e., $(12)A_3 = A_3(12)$ etc.

All members of a *conjugacy class* have cycle decompositions of the same structure. Cycle structure refers to the number of 1-cycles, 2-cycles etc. Hence, we should expect S_3 to have three conjugacy classes: the identity, the transpositions and the cyclic permutations: $\{e\}$, $\{(12), (23), (31)\}$ and $\{(123), (132)\}$. The members of a conjugacy class must have the same parity. For instance, the conjugates of (12) are

$$\begin{aligned} (23)(12)(23)^{-1} = (31), & \qquad (31)(12)(31)^{-1} = (23), \\ (123)(12)(123)^{-1} = (23) \quad \text{and} & \quad (132)(12)(132)^{-1} = (13). \end{aligned} \tag{101}$$

S_3 can be realized as a *semidirect product* $H \ltimes N$ of H acting on N, where H and N are cyclic groups of order 2 and 3. For instance, we take $H = \{e, (12)\}$ and $N = A_3$ regarded as subgroups of S_3 and consider the action of H on A_3 via conjugation: $\varphi_h(n') = hn'h^{-1}$. Thus, the semidirect product is

$$(h, n) \cdot (h', n') = (hh', nhn'h^{-1}). \tag{102}$$

The Cartesian product has six elements

$$(e, e),\ (e, (123)),\ (e, (132)),\ ((12), e),\ a = ((12), (123)) \quad \& \quad b = ((12), (132)). \tag{103}$$

The first 4 elements are identified with $e, (123), (132), (12) \in S_3$. If $a \leftrightarrow (31)$ and $b \leftrightarrow (23)$ then one finds that (102) agrees with the S_3 composition law. For instance,

$$((12), (123)) \cdot ((12), (123)) = ((12)^2, (123)(12)(123)(12)) = (e, (31)^2) = (e, e), \tag{104}$$

which agrees with $(31)^2 = e$ in S_3.

Circle group $U(1)$. Perhaps the easiest Lie group to understand is the group $U(1)$ of unimodular complex numbers $\{z \in \mathbb{C} | z^* z = 1\}$. As shown in Fig. 6, the group elements lie on the unit circle in the complex plane, so the group is also called the circle group S^1. This establishes that it is a differentiable manifold. It is called $U(1)$ since it is also the set of 1×1 unitary matrices[55]. Any unimodular z may be expressed as $z = e^{i\theta}$ where θ is defined modulo 2π. The identity element is $z = 1$, corresponding to $\theta \equiv 0$ modulo 2π. The multiplication law is abelian $e^{i\theta_1} e^{i\theta_2} = e^{i(\theta_1 + \theta_2)} = e^{i\theta_2} e^{i\theta_1}$. The inverse of $z = e^{i\theta}$ is the reciprocal $1/z = e^{-i\theta}$. Since $(\theta_1, \theta_2) \mapsto \theta_1 + \theta_2$ and

[54] Conjugation by any element $(g\sigma g^{-1})$ cannot change the parity of σ, so A_n invariant.

[55] The unitary group $U(n)$ consists of $n \times n$ complex matrices with $U^\dagger U = I$ where $U^\dagger = (U^t)^*$. Equivalently, it consists of linear maps on an n-dimensional complex vector space that preserve a Hermitian positive-definite inner product. It is a real Lie group of dimension n^2: transition functions are smooth real (not complex) functions.

$\theta \mapsto -\theta$ modulo 2π are smooth maps, $U(1)$ is a one-dimensional Lie group. It is compact (closed and bounded as a subset of the complex plane) and path connected though not simply connected. Its Lie algebra $\underline{U(1)}$ is the tangent space at $z = 1$, which is isomorphic to $\mathbb{R}$. $\underline{U(1)}$ can be taken to be the 1d vector space of imaginary numbers iy for $y \in \mathbb{R}$. A basis for the Lie algebra may be chosen as i. We notice that exponentiating a Lie algebra element such as πi gives us a group element $e^{\pi i} = \cos\pi + i\sin\pi = -1$. This map from Lie algebra to Lie group is called the exponential map. More generally, given a nonzero Lie algebra element (say i), exponentiating all its real multiples iy, we get a 1-parameter subgroup e^{iy}. In this case, the exponential map surjects onto the group but is many-to-one: $e^{iy} = e^{i(2n\pi+y)}$ for any $n \in \mathbb{Z}$. Since $U(1)$ is abelian, all its subgroups are also abelian, they are given by the cyclic groups C_n for $n = 1, 2, \cdots$. Here, $C_n = \{e^{2\pi ij/n} | j = 0, 1, 2, \cdots n-1\}$, so the elements of C_n lie at the vertices of a regular n-gon centered at the origin of the complex plane, with the identity as one of its vertices.

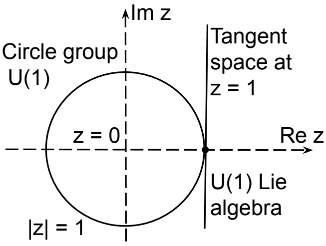


Figure 6: The group $U(1)$ of unimodular complex numbers and its Lie algebra $\underline{U(1)} \cong \mathbb{R}$.

The orthogonal group O(3). The orthogonal group $G = O(3)$ consists of the set of 3×3 real orthogonal matrices, i.e., matrices A that satisfy $A^tA = I$ (see Prob. **??** for examples). The generalization[56] to $n \times n$ orthogonal matrices for $n = 1, 2, 3, 4, \ldots$ is called $O(n)$. The group composition law is associative matrix multiplication: note that $(AB)^tAB = B^tA^tAB = I$ if A and B are orthogonal. The identity element is the identity matrix while the inverse of A is simply its transpose A^t. It is a nonabelian group since $AB \neq BA$ in general for a pair of orthogonal matrices (see Prob. **??**). The orthogonal group is a matrix group, it is a subgroup of the general linear group of all invertible real 3×3 matrices. The orthogonal group is important as it is the group of rotations and reflections of 3d Euclidean space. It frequently arises as a group of symmetries or as the configuration space of a mechanical system. We will soon view $O(3)$ as a manifold. First, what is its dimension? The condition $A^tA = I$ implies that a 3×3 orthogonal matrix is one whose columns furnish an orthonormal basis $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ for $\mathbb{R}^3$ (see below). The first basis vector $\boldsymbol{a}$ is any unit vector. The latter are parametrized by points on the unit sphere $S^2 \subset \mathbb{R}^3$. Thus, 2 real parameters are

[56] Alternatively, suppose V is an n dimensional real vector space with positive-definite inner product $\langle \cdot, \cdot \rangle$. Then $O(n)$ is the group of linear maps $A : V \to V$ that preserve the inner product $\langle Au, Av \rangle = \langle u, v \rangle$ for all $u, v \in V$ with product given by composition of maps. The definition is independent of the choice of V and inner product, it only depends on n.

needed to specify $\boldsymbol{a}$. Having picked $\boldsymbol{a}$, the second basis vector $\boldsymbol{b}$ can be any unit vector in the plane orthogonal to $\boldsymbol{a}$ and is specified by a point on the unit circle S^1 on this plane. Thus, one additional parameter is needed to specify $\boldsymbol{b}$. Having chosen $\boldsymbol{a}$ and $\boldsymbol{b}$, the third basis vector $\boldsymbol{c}$ must be perpendicular to both: $\boldsymbol{c} = \pm \boldsymbol{a} \times \boldsymbol{b}$. Thus, there is no additional continuous real parameter needed to specify $\boldsymbol{c}$. We conclude that $O(3)$ is a 3-parameter family of matrices. In fact, we may view it as a 3d submanifold of $\mathbb{R}^9$. Suppose we write A in terms of its columns,

$$A = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{pmatrix} \quad \text{so that} \quad A^t = \begin{pmatrix} \boldsymbol{a}^t \\ \boldsymbol{b}^t \\ \boldsymbol{c}^t \end{pmatrix}. \tag{105}$$

The constraint $A^t A = I$, becomes 6 conditions on the 9 matrix elements of A:

$$\boldsymbol{a}^t \boldsymbol{a} - 1 = \boldsymbol{b}^t \boldsymbol{b} - 1 = \boldsymbol{c}^t \boldsymbol{c} - 1 = \boldsymbol{a}^t \boldsymbol{b} = \boldsymbol{b}^t \boldsymbol{c} = \boldsymbol{c}^t \boldsymbol{a} = 0. \tag{106}$$

Thus, $O(3)$ is the common zero locus of these six independent quadratic functions of nine real variables. Hence, we may view $O(3)$ as a 3d algebraic submanifold of $\mathbb{R}^9$. It is bounded since $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{c}$ must each be a unit vector. It is closed[57] since it is the intersection of the inverse images of the closed one-element set $\{0\}$ under the continuous maps $\boldsymbol{a}^t \boldsymbol{a} - 1, \cdots, \boldsymbol{c}^t \boldsymbol{a}$ from $\mathbb{R}^9 \to \mathbb{R}$. Thus, $O(3)$ is a compact 3d manifold. However, it is not path connected. Taking the determinant of $A^t A = I$, we find $(\det A)^2 = 1$, so $\det A = \pm 1$. In Prob. **??** we show that there are orthogonal matrices with either sign of determinant. Since the determinant cannot jump discontinuously from 1 to -1 along a continuous path, we conclude that $O(3)$ is disconnected. It has two connected components. The identity I lies in the connected component where $\det A = 1$ and comprises proper rotations of $\mathbb{R}^3$. In fact, the connected component of the identity is a closed subgroup of $O(3)$ and is a Lie group in its own right, the special orthogonal group $SO(3)$ which is also the kernel of the determinant homomorphism from $O(3)$ to $\{\pm 1\}$. This subgroup of proper rotations and its Lie algebra play a key role in rigid body mechanics. The other component where $\det A = -1$ is not a subgroup as it is not closed under composition. It consists of so-called improper rotations and is a coset of $SO(3)$ by a reflection: product of a reflection and a proper rotation.

The orthogonal Lie algebra. Roughly, the Lie algebra (denoted $\underline{G}$ or $\mathfrak{g}$) is the linear approximation to the group in the neighborhood of the identity. More precisely, the Lie algebra as a vector space is defined as the tangent space to the group at the identity. To identify the orthogonal Lie algebra $\underline{O(3)}$, we suppose $A \approx I + u$ where u is treated to linear order. To this order, the orthogonality condition

$$(I + u)(I + u^t) \approx I \quad \text{becomes} \quad u + u^t = 0. \tag{107}$$

[57]For our purposes, a closed set C is one that contains all its limit points with respect to the Euclidean distance function. The inverse image $f^{-1}(C)$ of a closed set under a continuous map f is closed. The intersection of a finite number of closed sets is closed. A closed and bounded subset of Euclidean space is called compact.

Thus, the Lie algebra of the orthogonal group consists of 3×3 real antisymmetric matrices[58]. A real linear combination of antisymmetric matrices $\alpha u + \beta v$ remains antisymmetric, so this is indeed a vector space. The entries above the diagonal are the only linearly independent entries of an antisymmetric matrix, so $\underline{O(3)}$ is a 3-dimensional real vector space isomorphic to $\mathbb{R}^3$. We say that $\underline{O(3)}$ is a 3-dimensional Lie algebra. It is no surprise that G and $\underline{G}$ have the same dimension.

[58] Since $SO(3)$ is the identity component of $O(3)$, they have the same Lie algebra. We note in passing that the group $SU(2)$ of 2×2 unitary matrices with unit determinant has an isomorphic Lie algebra. Indeed, the Lie brackets among $(i/2)\times$ the Pauli matrices (**??**) [which furnish a basis for $\underline{SU(2)}$] are the same as those among e_1, e_2 and e_3 of (**??**).