Chapter 11

Differential Calculus on Manifolds

In this section we will apply what we have learned about vectors and tensors in linear algebra to vector and tensor *fields* in a general curvilinear co-ordinate system. Our aim is to introduce the reader to the modern language of advanced calculus, and in particular to the calculus of differential forms on surfaces and manifolds.

11.1 Vector and covector fields

Vector fields — electric, magnetic, velocity fields, and so on — appear everywhere in physics. After perhaps struggling with it in introductory courses, we rather take the field concept for granted. There remain subtleties, however. Consider an electric field. It makes sense to add two field vectors at a single point, but there is no physical meaning to the sum of field vectors $\mathbf{E}(x_1)$ and $\mathbf{E}(x_2)$ at two distinct points. We should therefore regard all possible electric fields at a single point as living in a vector space, but each different point in space comes with its own field-vector space.

This view seems even more reasonable when we consider velocity vectors describing motion on a curved surface. A velocity vector lives in the *tangent space* to the surface at each point, and each of these spaces is a differently oriented subspace of the higher-dimensional ambient space.

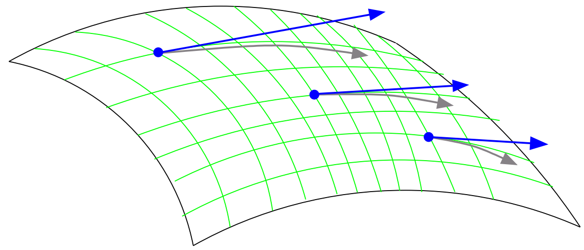

Figure 11.1: *Each point on a surface has its own vector space of tangents.*

Mathematicians call such a collection of vector spaces — one for each of the points in a surface — a *vector bundle* over the surface. Thus, the *tangent bundle* over a surface is the totality of all vector spaces tangent to the surface. Why a bundle? This word is used because the individual tangent spaces are not completely independent, but are tied together in a rather non-obvious way. Try to construct a smooth field of unit vectors tangent to the surface of a sphere. However hard you work you will end up in trouble somewhere. You cannot comb a hairy ball. On the surface of torus you will have no such problems. You can comb a hairy doughnut. The tangent spaces collectively know something about the surface they are tangent to.

Although we spoke in the previous paragraph of vectors tangent to a curved surface, it is useful to generalize this idea to vectors lying in the tangent space of an n-dimensional *manifold*, or n-manifold. A n-manifold M is essentially a space that locally looks like a part of $\mathbb{R}^n$. This means that some open neighbourhood of each point $x \in M$ can be parametrized by an n-dimensional co-ordinate system. A co-ordinate parametrization is called a *chart*. Unless M is $\mathbb{R}^n$ itself (or part of it), a chart will cover only part of M, and more than one will be required for complete coverage. Where a pair of charts overlap, we demand that the transformation formula giving one set of co-ordinates as a function of the other be a smooth (C^∞) function, and to possess a smooth inverse.[1] A collection of such smoothly related co-ordinate charts covering all of M is called an *atlas*. The advantage of thinking in terms of manifolds is that we do not have to understand their properties as arising from some embedding in a higher dimensional space. Whatever structure they have, they possess in, and of, themselves

[1]A formal definition of a manifold contains some further technical restrictions (that the space be *Hausdorff* and *paracompact*) that are designed to eliminate pathologies. We are more interested in doing calculus than in proving theorems, and so we will ignore these niceties.

Classical mechanics provides a familiar illustration of these ideas. Except in pathological cases, the configuration space M of a mechanical system is a manifold. When the system has n degrees of freedom we use generalized co-ordinates q^i, $i = 1, \ldots, n$ to parametrize M. The tangent bundle of M then provides the setting for Lagrangian mechanics. This bundle, denoted by TM, is the $2n$-dimensional space each of whose whose points consists of a point $q = (q^1, \ldots, q^n)$ in M paired with a tangent vector lying in the tangent space TM_q at that point. If we think of the tangent vector as a velocity, the natural co-ordinates on TM become $(q^1, q^2, \ldots, q^n\,; \dot{q}^1, \dot{q}^2, \ldots, \dot{q}^n)$, and these are the variables that appear in the Lagrangian of the system.

If we consider a vector tangent to some curved surface, it will stick out of it. If we have a vector tangent to a manifold, it is a straight arrow lying atop bent co-ordinates. Should we restrict the length of the vector so that it does not stick out too far? Are we restricted to only infinitesimal vectors? It is best to avoid all this by adopting a clever notion of what a vector in a tangent space is. The idea is to focus on a well-defined object such as a derivative. Suppose that our space has co-ordinates x^μ. (These are *not* the contravariant components of some vector) A *directional derivative* is an object such as $X^\mu \partial_\mu$, where ∂_μ is shorthand for $\partial/\partial x^\mu$. When the components X^μ are functions of the co-ordinates x^σ, this object is called a tangent-vector field, and we write[2]

$$X = X^\mu \partial_\mu. \tag{11.1}$$

We regard the ∂_μ at a point x as a basis for TM_x, the tangent-vector space at x, and the $X^\mu(x)$ as the (contravariant) components of the vector X at that point. Although they are not little arrows, what the ∂_μ are is mathematically clear, and so we know perfectly well how to deal with them.

When we change co-ordinate system from x^μ to z^ν by regarding the x^μ's as invertable functions of the z^ν's, *i.e.*

$$\begin{aligned} x^1 &= x^1(z^1, z^2, \ldots, z^n), \\ x^2 &= x^2(z^1, z^2, \ldots, z^n), \\ &\vdots \\ x^n &= x^n(z^1, z^2, \ldots, z^n), \end{aligned} \tag{11.2}$$

[2] We are going to stop using bold symbols to distinguish between intrinsic objects and their components, because from now on almost everything will be something other than a number, and too much black ink would just be confusing.

then the chain rule for partial differentiation gives

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \frac{\partial z^\nu}{\partial x^\mu}\frac{\partial}{\partial z^\nu} = \left(\frac{\partial z^\nu}{\partial x^\mu}\right)\partial'_\nu, \tag{11.3}$$

where ∂'_ν is shorthand for $\partial/\partial z^\nu$. By demanding that

$$X = X^\mu \partial_\mu = X'^\nu \partial'_\nu \tag{11.4}$$

we find the components in the z^ν co-ordinate frame to be

$$X'^\nu = \left(\frac{\partial z^\nu}{\partial x^\mu}\right) X^\mu. \tag{11.5}$$

Conversely, using

$$\frac{\partial x^\sigma}{\partial z^\nu}\frac{\partial z^\nu}{\partial x^\mu} = \frac{\partial x^\sigma}{\partial x^\mu} = \delta^\sigma_\mu, \tag{11.6}$$

we have

$$X^\nu = \left(\frac{\partial x^\nu}{\partial z^\mu}\right) X'^\mu. \tag{11.7}$$

This, then, is the transformation law for a contravariant vector.

It is worth pointing out that the basis vectors ∂_μ are *not* unit vectors. As we have no metric, and therefore no notion of length anyway, we cannot try to normalize them. If you insist on drawing (small?) arrows, think of ∂_1 as starting at a point $(x^1, x^2, \ldots, x^n)$ and with its head at $(x^1+1, x^2, \ldots, x^n)$. Of course this is only a good picture if the co-ordinates are not too "curvy."

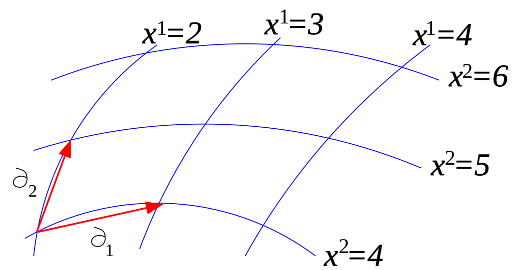

Figure 11.2: *Approximate picture of the vectors ∂_1 and ∂_2 at the point* $(x^1, x^2) = (2, 4)$.

Example: The surface of the unit sphere is a manifold. It is usually denoted by S^2. We may label its points with spherical polar co-ordinates, θ measuring the co-latitude and ϕ measuring the longitude. These will be useful everywhere except at the north and south poles, where they become singular because at $\theta = 0$ or π all values of of the longitude ϕ correspond to the same point. In this co-ordinate basis, the tangent vector representing the velocity field due to a rigid rotation of one radian per second about the z axis is

$$V_z = \partial_\phi. \tag{11.8}$$

Similarly

$$\begin{aligned} V_x &= -\sin\phi\,\partial_\theta - \cot\theta\cos\phi\,\partial_\phi, \\ V_y &= \cos\phi\,\partial_\theta - \cot\theta\sin\phi\partial_\phi, \end{aligned} \tag{11.9}$$

respectively represent rigid rotations about the x and y axes.

We now know how to think about vectors. What about their dual-space partners, the covectors? These live in the *cotangent bundle* T^*M, and for them a cute notational game, due to Élie Cartan, is played. We write the basis vectors dual to the ∂_μ as $dx^\mu(\ )$. Thus

$$dx^\mu(\partial_\nu) = \delta^\mu_\nu. \tag{11.10}$$

When evaluated on a vector field $X = X^\mu\partial_\mu$, the basis covectors dx^μ return its components:

$$dx^\mu(X) = dx^\mu(X^\nu\partial_\nu) = X^\nu dx^\mu(\partial_\nu) = X^\nu\delta^\mu_\nu = X^\mu. \tag{11.11}$$

Now, any smooth function $f \in C^\infty(M)$ will give rise to a field of covectors in T^*M. This is because a vector field X acts on the scalar function f as

$$Xf = X^\mu\partial_\mu f \tag{11.12}$$

and Xf is another scalar function. This new function gives a number — and thus an element of the field $\mathbb{R}$ — at each point $x \in M$. But this is exactly what a covector does: it takes in a vector at a point and returns a number. We will call this covector field "df." It is essentially the gradient of f. Thus

$$df(X) \stackrel{\text{def}}{=} Xf = X^\mu\frac{\partial f}{\partial x^\mu}. \tag{11.13}$$

If we take f to be the co-ordinate x^ν, we have

$$dx^\nu(X) = X^\mu\frac{\partial x^\nu}{\partial x^\mu} = X^\mu\delta^\nu_\mu = X^\nu, \tag{11.14}$$

so this viewpoint is consistent with our previous definition of dx^ν. Thus

$$df(X) = \frac{\partial f}{\partial x^\mu} X^\mu = \frac{\partial f}{\partial x^\mu} dx^\mu(X) \tag{11.15}$$

for any vector field X. In other words, we can expand df as

$$df = \frac{\partial f}{\partial x^\mu} dx^\mu. \tag{11.16}$$

This is *not* some approximation to a change in f, but is an exact expansion of the covector field df in terms of the basis covectors dx^μ.

We may retain something of the notion that dx^μ represents the (contravariant) components of a small displacement in x provided that we think of dx^μ as a machine into which we insert the small displacement (a vector) and have it spit out the numerical components δx^μ. This is the same distinction that we make between sin( ) as a function into which one can plug x, and $\sin x$, the number that results from inserting in this particular value of x. Although seemingly innocent, we know that it is a distinction of great power.

The change of co-ordinates transformation law for a covector field f_μ is found from

$$f_\mu \, dx^\mu = f'_\nu \, dz^\nu, \tag{11.17}$$

by using

$$dx^\mu = \left(\frac{\partial x^\mu}{\partial z^\nu}\right) dz^\nu. \tag{11.18}$$

We find

$$f'_\nu = \left(\frac{\partial x^\mu}{\partial z^\nu}\right) f_\mu. \tag{11.19}$$

A general tensor such as $Q^{\lambda\mu}{}_{\rho\sigma\tau}$ transforms as

$$Q'^{\lambda\mu}{}_{\rho\sigma\tau}(z) = \frac{\partial z^\lambda}{\partial x^\alpha}\frac{\partial z^\mu}{\partial x^\beta}\frac{\partial x^\gamma}{\partial z^\rho}\frac{\partial x^\delta}{\partial z^\sigma}\frac{\partial x^\epsilon}{\partial z^\tau} Q^{\alpha\beta}{}_{\gamma\delta\epsilon}(x). \tag{11.20}$$

Observe how the indices are wired up: Those for the new tensor coefficients in the new co-ordinates, z, are attached to the new z's, and those for the old coefficients are attached to the old x's. Upstairs indices go in the numerator of each partial derivative, and downstairs ones are in the denominator.

The language of bundles and sections

At the beginning of this section, we introduced the notion of a vector bundle. This is a particular example of the more general concept of a *fibre bundle*, where the vector space at each point in the manifold is replaced by a "fibre" *over* that point. The fibre can be any mathematical object, such as a set, tensor space, or another manifold. Mathematicians visualize the bundle as a collection of fibres growing out of the manifold, much as stalks of wheat grow out the soil. When one slices through a patch of wheat with a scythe, the blade exposes a cross-section of the stalks. By analogy, a choice of an element of the the fibre over each point in the manifold is called a *cross-section*, or, more commonly, a *section* of the bundle. In this language, a tangent-vector field becomes a section of the tangent bundle, and a field of covectors becomes a section of the cotangent bundle.

We provide a more detailed account of bundles in Chapter 16.

11.2 Differentiating tensors

If f is a function then $\partial_\mu f$ are components of the covariant vector df. Suppose that a^μ is a contravariant vector. Are $\partial_\nu a^\mu$ the components of a type $(1,1)$ tensor? The answer is *no*! In general, differentiating the components of a tensor does not give rise to another tensor. One can see why at two levels:

a) Consider the transformation laws. They contain expressions of the form $\partial x^\mu/\partial z^\nu$. If we differentiate both sides of the transformation law of a tensor, these factors are also differentiated, but tensor transformation laws never contain second derivatives, such as $\partial^2 x^\mu/\partial z^\nu \partial z^\sigma$.
b) Differentiation requires subtracting vectors or tensors at different points — but vectors at different points are in different vector spaces, so their difference is not defined.

These two reasons are really one and the same. We need to be cleverer to get new tensors by differentiating old ones.

11.2.1 Lie bracket

One way to proceed is to note that the vector field X is an *operator*. It makes sense, therefore, to try to compose two of them to make another. Look at

XY, for example:

$$XY = X^\mu \partial_\mu (Y^\nu \partial_\nu) = X^\mu Y^\nu \partial^2_{\mu\nu} + X^\mu \left(\frac{\partial Y^\nu}{\partial x^\mu}\right) \partial_\nu. \tag{11.21}$$

What are we to make of this? Not much! There is no particular interpretation for the second derivative, and as we saw above, it does not transform nicely. But suppose we take a *commutator*:

$$[X, Y] = XY - YX = (X^\mu(\partial_\mu Y^\nu) - Y^\mu(\partial_\mu X^\nu))\,\partial_\nu. \tag{11.22}$$

The second derivatives have cancelled, and what remains is a directional derivative and so a *bona-fide* vector field. The components

$$[X, Y]^\nu \equiv X^\mu(\partial_\mu Y^\nu) - Y^\mu(\partial_\mu X^\nu) \tag{11.23}$$

are the components of a new contravariant vector field made from the two old vector fields. This new vector field is called the *Lie bracket* of the two fields, and has a geometric interpretation.

To understand the geometry of the Lie bracket, we first define the *flow* associated with a tangent-vector field X. This is the map that takes a point x_0 and maps it to $x(t)$ by solving the family of equations

$$\frac{dx^\mu}{dt} = X^\mu(x^1, x^2, \ldots), \tag{11.24}$$

with initial condition $x^\mu(0) = x^\mu_0$. In words, we regard X as the velocity field of a flowing fluid, and let x ride along with the fluid.

Now envisage X and Y as two velocity fields. Suppose we flow along X for a brief time t, then along Y for another brief interval s. Next we switch back to X, but with a minus sign, for time t, and then to $-Y$ for a final interval of s. We have tried to retrace our path, but a short exercise with Taylor's theorem shows that we will fail to return to our exact starting point. We will miss by $\delta x^\mu = st[X, Y]^\mu$, plus corrections of cubic order in s and t. (See figure 11.3)

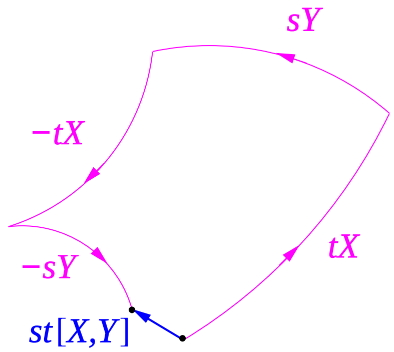

Figure 11.3: *We try to retrace our steps but fail to return by a distance proportional to the Lie bracket.*

Example: Let

$$\begin{aligned} V_x &= -\sin\phi\,\partial_\theta - \cot\theta\cos\phi\,\partial_\phi, \\ V_y &= \cos\phi\,\partial_\theta - \cot\theta\sin\phi\,\partial_\phi, \end{aligned}$$

be two vector fields in $T(S^2)$. We find that

$$[V_x, V_y] = -V_z,$$

where $V_z = \partial_\phi$.

Frobenius' theorem

Suppose that in some region of a d-dimensional manifold M we are given $n < d$ linearly independent tangent-vector fields X_i. Such a set is called a *distribution* by differential geometers. (The concept has nothing to do with probability, or with objects like "$\delta(x)$" which are also called "distributions.") At each point x, the span $\langle X_i(x)\rangle$ of the field vectors forms a subspace of the tangent space TM_x, and we can picture this subspace as a fragment of an n-dimensional surface passing through x. It is possible that these surface fragments fit together to make a stack of smooth surfaces — called a *foliation* — that fill out the d-dimensional space, and have the given X_i as their tangent vectors.

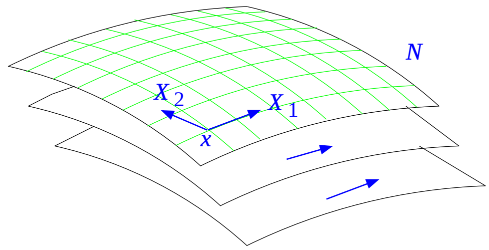

Figure 11.4: *A local foliation.*

If this is the case then starting from x and taking steps only along the X_i we find ourselves restricted to the n-surface, or n-*submanifold*, N passing though the original point x.

Alternatively, the surface fragments may form such an incoherent jumble that starting from x and moving only along the X_i we can find our way to any point in the neighbourhood of x. It is also possible that some intermediate case applies, so that moving along the X_i restricts us to an m-surface, where $d > m > n$. The Lie bracket provides us with the appropriate tool with which to investigate these possibilities.

First a definition: If there are functions ${c_{ij}}^k(x)$ such that

$$[X_i, X_j] = {c_{ij}}^k(x) X_k, \tag{11.25}$$

i.e. the Lie brackets close within the set $\{X_i\}$ at each point x, then the distribution is said to be *involutive.* and the vector fields are said to be "in involution" with each other. When our given distribution is involutive, then the first case holds, and, at least locally, there is a foliation by n-submanifolds N. A formal statement of this is:

Theorem (Frobenius): A smooth (C^∞) involutive distribution is completely integrable: locally, there are co-ordinates $x^\mu, \mu = 1, \dots, d$ such that $X_i = \sum_{\mu=1}^n X_i^\mu \partial_\mu$, and the surfaces N through each point are in the form $x^\mu =$ const. for $\mu = n+1, \dots, d$. Conversely, if such co-ordinates exist then the distribution is involutive.

A half-proof: If such co-ordinates exist then it is obvious that the Lie bracket of any pair of vectors in the form $X_i = \sum_{\mu=1}^n X_i^\mu \partial_\mu$ can also be expanded in terms of the first n basis vectors. A logically equivalent statement exploits the geometric interpretation of the Lie bracket: If the Lie brackets of the fields X_i do *not* close within the n-dimensional span of the X_i, then a sequence of back-and-forth manœvres along the X_i allows us to escape into a new direction, and so the X_i *cannot* be tangent to an n-surface. Establishing the converse — that closure implies the existence of the foliation — is rather more technical, and we will not attempt it.

Involutive and non-involutive distributions appear in classical mechanics under the guise of *holonomic* and *anholonomic* constraints. In mechanics, constraints are not usually given as a list of the directions (vector fields) in which we are free to move, but instead as a list of restrictions imposed on the permitted motion. In a d-dimensional mechanical system we might have set of m independent constraints of the form $\omega^i_\mu(q)\dot{q}^\mu = 0$, $i = 1, \ldots, m$. Such restrictions are most naturally expressed in terms of the covector fields

$$\omega^i = \sum_{\mu=1}^{d} \omega^i_\mu(q) dq^\mu, \quad i = 1 \leq i \leq m. \tag{11.26}$$

We can write the constraints as the m conditions $\omega^i(\dot{q}) = 0$ that must be satisfied if $\dot{q} \equiv \dot{q}^\mu \partial_\mu$ is to be an allowed motion. The list of constraints is known a *Pfaffian* system of equations. These equations indirectly determine an $n = d - m$ dimensional distribution of permitted motions. The Pfaffian system is said to be *integrable* if this distribution is involutive, and hence integrable. In this case there is a set of m functions $g^i(q)$ and an invertible m-by-m matrix ${f^i}_j(q)$ such that

$$\omega^i = \sum_{j=1}^{m} {f^i}_j(q) dg^j. \tag{11.27}$$

The functions $g^i(q)$ can, for example, be taken to be the co-ordinate functions x^μ, $\mu = n + 1, \ldots, d$, that label the foliating surfaces N in the statement of Frobenius' theorem. The system of integrable constraints $\omega^i(\dot{q}) = 0$ thus restricts us to the surfaces $g^i(q) = \mathit{constant}$.

For example, consider a particle moving in three dimensions. If we are told that the velocity vector is constrained by $\omega(\dot{q}) = 0$, where

$$\omega = x\, dx + y\, dy + z\, dz \tag{11.28}$$

we realize that the particle is being forced to move on a sphere passing through the initial point. In spherical co-ordinates the associated distribution is the set $\{\partial_\theta, \partial_\phi\}$, which is clearly involutive because $[\partial_\theta, \partial_\phi] = 0$. The functions $f(x, y, z)$ and $g(x, y, z)$ from the previous paragraph can be taken to be $r = \sqrt{x^2+y^2+z^2}$, and the constraint covector written as $\omega = f\,dg = r\,dr$.

The foliation is the family of nested spheres whose centre is the origin. (The foliation is not global because it becomes singular at $r = 0$.) Constraints like this, which restrict the motion to a surface, are said to be *holonomic*.

Suppose, on the other hand, we have a ball rolling on a table. Here, we have a five-dimensional configuration manifold $M = \mathbb{R}^2 \times S^3$, parametrized by the centre of mass $(x, y) \in \mathbb{R}^2$ of the ball and the three Euler angles $(\theta, \phi, \psi) \in S^3$ defining its orientation. Three no-slip rolling conditions

$$\begin{aligned} \dot{x} &= \dot{\psi}\sin\theta\sin\phi + \dot{\theta}\cos\phi, \\ \dot{y} &= -\dot{\psi}\sin\theta\cos\phi + \dot{\theta}\sin\phi, \\ 0 &= \dot{\psi}\cos\theta + \dot{\phi}, \end{aligned} \tag{11.29}$$

(see exercise 11.17) link the rate of change of the Euler angles to the velocity of the centre of mass. At each point in this five-dimensional manifold we are free to roll the ball in two directions, and so we might expect that the reachable configurations constitute a two-dimensional surface embedded in the full five-dimensional space. The two vector fields

$$\begin{aligned} \mathbf{roll_x} &= \partial_x - \sin\phi\cot\theta\,\partial_\phi + \cos\phi\,\partial_\theta + \operatorname{cosec}\theta\sin\phi\,\partial_\psi, \\ \mathbf{roll_y} &= \partial_y + \cos\phi\cot\theta\,\partial_\phi + \sin\phi\,\partial_\theta - \operatorname{cosec}\theta\cos\phi\,\partial_\psi, \end{aligned} \tag{11.30}$$

describing the permitted x- and y-direction rolling motion are not in involution, however. By calculating enough Lie brackets we eventually obtain five linearly independent velocity vector fields, and starting from one configuration we can reach any other. The no-slip rolling condition is said to be *non-integrable*, or *anholonomic*. Such systems are tricky to deal with in Lagrangian dynamics.

The following exercise provides a familiar example of the utility of non-holonomic constraints:

Exercise 11.1: *Parallel Parking using Lie Brackets*. The configuration space of a car is four dimensional, and parameterized by co-ordinates (x, y, θ, ϕ), as shown in figure 11.5. Define the following vector fields:

a) (front wheel) **drive** $= \cos\phi(\cos\theta\,\partial_x + \sin\theta\,\partial_y) + \sin\phi\,\partial_\theta$.
b) **steer** $= \partial_\phi$.
c) (front wheel) **skid** $= -\sin\phi(\cos\theta\,\partial_x + \sin\theta\,\partial_y) + \cos\phi\,\partial_\theta$.

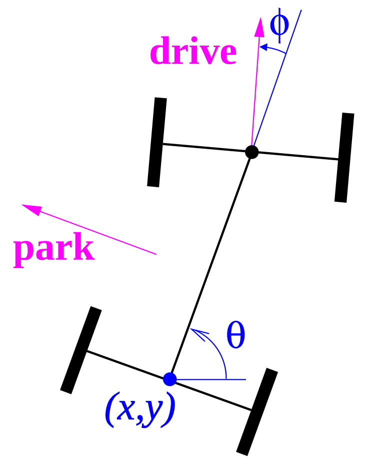

Figure 11.5: Co-ordinates for car parking

d) **park** $= -\sin\theta\,\partial_x + \cos\theta\,\partial_y$.

Explain why these are apt names for the vector fields, and compute the six Lie brackets:

$$[\mathbf{steer}, \mathbf{drive}], \quad [\mathbf{steer}, \mathbf{skid}], \quad [\mathbf{skid}, \mathbf{drive}],$$
$$[\mathbf{park}, \mathbf{drive}], \quad [\mathbf{park}, \mathbf{park}], \quad [\mathbf{park}, \mathbf{skid}].$$

The driver can use only the operations $(\pm)$ **drive** and $(\pm)$ **steer** to manœvre the car. Use the geometric interpretation of the Lie bracket to explain how a suitable sequence of motions (forward, reverse, and turning the steering wheel) can be used to manoeuvre a car sideways into a parking space.

11.2.2 Lie derivative

Another derivative that we can define is the *Lie derivative* along a vector field X. It is defined by its action on a scalar function f as

$$\mathcal{L}_X f \stackrel{\text{def}}{=} Xf, \tag{11.31}$$

on a vector field by

$$\mathcal{L}_X Y \stackrel{\text{def}}{=} [X, Y], \tag{11.32}$$

and on anything else by requiring it to be a *derivation*, meaning that it obeys Leibniz' rule. For example, let us compute the Lie derivative of a covector

F. We first introduce an arbitrary vector field Y and plug it into F to get the scalar function $F(Y)$. Leibniz' rule is then the statement that

$$\mathcal{L}_X F(Y) = (\mathcal{L}_X F)(Y) + F(\mathcal{L}_X Y). \tag{11.33}$$

Since $F(Y)$ is a function and Y is a vector, both of whose derivatives we know how to compute, we know the first and third of the three terms in this equation. From $\mathcal{L}_X F(Y) = XF(Y)$ and $F(\mathcal{L}_X Y) = F([X,Y])$, we have

$$XF(Y) = (\mathcal{L}_X F)(Y) + F([X,Y]), \tag{11.34}$$

and so

$$(\mathcal{L}_X F)(Y) = XF(Y) - F([X,Y]). \tag{11.35}$$

In components, this becomes

$$\begin{aligned}(\mathcal{L}_X F)(Y) &= X^\nu \partial_\nu (F_\mu Y^\mu) - F_\nu (X^\mu \partial_\mu Y^\nu - Y^\mu \partial_\mu X^\nu) \\ &= (X^\nu \partial_\nu F_\mu + F_\nu \partial_\mu X^\nu) Y^\mu. \end{aligned} \tag{11.36}$$

Note how all the derivatives of Y^μ have cancelled, so $\mathcal{L}_X F(\ )$ depends only on the local value of Y. The Lie derivative of F is therefore still a covector field. This is true in general: the Lie derivative does not change the tensor character of the objects on which it acts. Dropping the passive spectator field Y^ν, we have a formula for $\mathcal{L}_X F$ in components:

$$(\mathcal{L}_X F)_\mu = X^\nu \partial_\nu F_\mu + F_\nu \partial_\mu X^\nu. \tag{11.37}$$

Another example is provided by the Lie derivative of a type $(0,2)$ tensor, such as a metric tensor. This is

$$(\mathcal{L}_X g)_{\mu\nu} = X^\alpha \partial_\alpha g_{\mu\nu} + g_{\mu\alpha} \partial_\nu X^\alpha + g_{\alpha\nu} \partial_\mu X^\alpha. \tag{11.38}$$

The Lie derivative of a metric measures the extent to which the displacement $x^\alpha \to x^\alpha + \epsilon X^\alpha(x)$ deforms the geometry. If we write the metric as

$$g(\ ,\ ) = g_{\mu\nu}(x)\, dx^\mu \otimes dx^\nu, \tag{11.39}$$

we can understand both this geometric interpretation and the origin of the three terms appearing in the Lie derivative. We simply make the displacement $x^\alpha \to x^\alpha + \epsilon X^\alpha$ in the coefficients $g_{\mu\nu}(x)$ and in the two dx^α. In the latter we write

$$d(x^\alpha + \epsilon X^\alpha) = dx^\alpha + \epsilon \frac{\partial X^\alpha}{\partial x^\beta} dx^\beta. \tag{11.40}$$

Then we see that

$$\begin{aligned} g_{\mu\nu}(x)\, dx^\mu \otimes dx^\nu &\to [g_{\mu\nu}(x) + \epsilon(X^\alpha \partial_\alpha g_{\mu\nu} + g_{\mu\alpha}\partial_\nu X^\alpha + g_{\alpha\nu}\partial_\mu X^\alpha)]\, dx^\mu \otimes dx^\nu \\ &= [g_{\mu\nu} + \epsilon(\mathcal{L}_X g)_{\mu\nu}]\, dx^\mu \otimes dx^\nu. \end{aligned} \tag{11.41}$$

A displacement field X that does not change distances between points, *i.e.* one that gives rise to an *isometry*, must therefore satisfy $\mathcal{L}_X g = 0$. Such an X is said to be a *Killing field* after Wilhelm Killing who introduced them in his study of non-euclidean geometries.

The geometric interpretation of the Lie derivative of a vector field is as follows: In order to compute the X directional derivative of a vector field Y, we need to be able to subtract the vector $Y(x)$ from the vector $Y(x + \epsilon X)$, divide by ϵ, and take the limit $\epsilon \to 0$. To do this we have somehow to get the vector $Y(x)$ from the point x, where it normally resides, to the new point $x + \epsilon X$, so both vectors are elements of the same vector space. The Lie derivative achieves this by carrying the old vector to the new point along the field X.

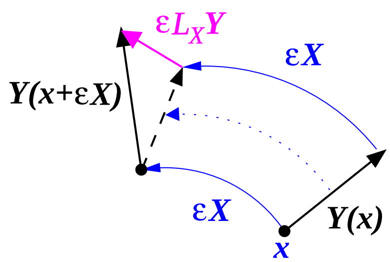

Figure 11.6: Computing the Lie derivative of a vector.

Imagine the vector Y as drawn in ink in a flowing fluid whose velocity field is X. Initially the tail of Y is at x and its head is at $x + Y$. After flowing for a time ϵ, its tail is at $x + \epsilon X$ — *i.e* exactly where the tail of $Y(x + \epsilon X)$ lies. Where the head of transported vector ends up depends how the flow has stretched and rotated the ink, but it is this distorted vector that is subtracted from $Y(x + \epsilon X)$ to get $\epsilon \mathcal{L}_X Y = \epsilon[X, Y]$.

Exercise 11.2: The metric on the unit sphere equipped with polar co-ordinates is

$$g(\ ,\ ) = d\theta \otimes d\theta + \sin^2\theta\, d\phi \otimes d\phi.$$

Consider

$$V_x = -\sin\phi\, \partial_\theta - \cot\theta \cos\phi\, \partial_\phi,$$

which is the vector field of a rigid rotation about the x axis. Compute the Lie derivative $\mathcal{L}_{V_x} g$, and show that it is zero.

Exercise 11.3: Suppose we have an unstrained block of material in real space. A co-ordinate system ξ^1, ξ^2, ξ^3, is attached to the material of the body. The point with co-ordinate ξ is located at $(x^1(\xi), x^2(\xi), x^3(\xi))$ where x^1, x^2, x^3 are the usual $\mathbf{R^3}$ Cartesian co-ordinates.

a) Show that the induced metric in the ξ co-ordinate system is

$$g_{\mu\nu}(\xi) = \sum_{a=1}^{3} \frac{\partial x^a}{\partial \xi^\mu}\frac{\partial x^a}{\partial \xi^\nu}.$$

b) The body is now deformed by an infinitesimal strain vector field $\eta(\xi)$. The atom with co-ordinate ξ^μ is moved to what was $\xi^\mu + \eta^\mu(\xi)$, or equivalently, the atom initially at Cartesian co-ordinate $x^a(\xi)$ is moved to $x^a + \eta^\mu \partial x^a/\partial\xi^\mu$. Show that the new induced metric is

$$g_{\mu\nu} + \delta g_{\mu\nu} = g_{\mu\nu} + \mathcal{L}_\eta g_{\mu\nu}.$$

c) Define the *strain tensor* to be 1/2 of the Lie derivative of the metric with respect to the deformation. If the original ξ co-ordinate system coincided with the Cartesian one, show that this definition reduces to the familiar form

$$e_{ab} = \frac{1}{2}\left(\frac{\partial \eta_a}{\partial x^b} + \frac{\partial \eta_b}{\partial x^a}\right),$$

all tensors being Cartesian.

d) Part c) gave us the geometric definitition of *infinitesimal strain*. If the body is deformed substantially, the *Cauchy-Green finite strain tensor* is defined as

$$E_{\mu\nu}(\xi) = \frac{1}{2}\left(g_{\mu\nu} - g^{(0)}_{\mu\nu}\right),$$

where $g^{(0)}_{\mu\nu}$ is the metric in the undeformed body and $g_{\mu\nu}$ the metric in the deformed body. Explain why this is a reasonable definition.

11.3 Exterior calculus

11.3.1 Differential forms

The objects we introduced in section 11.1, the dx^μ, are called one-forms, or differential one-forms. They are fields living in the cotangent bundle T^*M

of M. More precisely, they are *sections* of the cotangent bundle. Sections of the bundle whose fibre above $x \in M$ is the p-th skew-symmetric tensor power $\bigwedge^p(T^*M_x)$ of the cotangent space are known as p-forms.

For example,

$$A = A_\mu dx^\mu = A_1 dx^1 + A_2 dx^2 + A_3 dx^3, \tag{11.42}$$

is a 1-form,

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = F_{12} dx^1 \wedge dx^2 + F_{23} dx^2 \wedge dx^3 + F_{31} dx^3 \wedge dx^1, \tag{11.43}$$

is a 2-form, and

$$\Omega = \frac{1}{3!} \Omega_{\mu\nu\sigma}\, dx^\mu \wedge dx^\nu \wedge dx^\sigma = \Omega_{123}\, dx^1 \wedge dx^2 \wedge dx^3, \tag{11.44}$$

is a 3-form. All the coefficients are skew-symmetric tensors, so, for example,

$$\Omega_{\mu\nu\sigma} = \Omega_{\nu\sigma\mu} = \Omega_{\sigma\mu\nu} = -\Omega_{\nu\mu\sigma} = -\Omega_{\mu\sigma\nu} = -\Omega_{\sigma\nu\mu}. \tag{11.45}$$

In each example we have explicitly written out all the independent terms for the case of three dimensions. Note how the $p!$ disappears when we do this and keep only distinct components. In d dimensions the space of p-forms is $d!/p!(d-p)!$ dimensional, and all p-forms with $p > d$ vanish identically.

As with the wedge products in chapter one, we regard a p-form as a p-linear skew-symetric function with p slots into which we can drop vectors to get a number. For example the basis two-forms give

$$dx^\mu \wedge dx^\nu(\partial_\alpha, \partial_\beta) = \delta^\mu_\alpha \delta^\nu_\beta - \delta^\mu_\beta \delta^\nu_\alpha. \tag{11.46}$$

The analogous expression for a p-form would have $p!$ terms. We can define an algebra of differential forms by "wedging" them together in the obvious way, so that the product of a p-form with a q-form is a $(p+q)$-form. The wedge product is associative and distributive but not, of course, commutative. Instead, if a is a p-form and b a q-form, then

$$a \wedge b = (-1)^{pq}\, b \wedge a. \tag{11.47}$$

Actually it is customary in this game to suppress the "$\wedge$" and simply write $F = \frac{1}{2} F_{\mu\nu}\, dx^\mu dx^\nu$, it being assumed that you know that $dx^\mu dx^\nu = -dx^\nu dx^\mu$ — what else could it be?

11.3.2 The exterior derivative

These p-forms may seem rather complicated, so it is perhaps surprising that all the vector calculus (div, grad, curl, the divergence theorem and Stokes' theorem, *etc.*) that you have learned in the past reduce, in terms of them, to two simple formulæ! Indeed Élie Cartan's calculus of p-forms is slowly supplanting traditional vector calculus, much as Willard Gibbs' and Oliver Heaviside's vector calculus supplanted the tedious component-by-component formulæ you find in Maxwell's *Treatise on Electricity and Magnetism*.

The basic tool is the *exterior derivative* "d", which we now define axiomatically:

i) If f is a function (0-form), then df coincides with the previous definition, *i.e.* $df(X) = Xf$ for any vector field X.

ii) d is an *anti-derivation*: If a is a p-form and b a q-form then

$$d(a \wedge b) = da \wedge b + (-1)^p a \wedge db. \tag{11.48}$$

iii) *Poincaré's lemma*: $d^2 = 0$, meaning that $d(da) = 0$ for any p-form a.

iv) d is linear. That $d(\alpha a) = \alpha da$, for constant α follows already from i) and ii), so the new fact is that $d(a + b) = da + db$.

It is not immediately obvious that axioms i), ii) and iii) are compatible with one another. If we use axiom i), ii) and $d(dx^i) = 0$ to compute the d of $\Omega = \frac{1}{p!}\Omega_{i_1,\ldots,i_p} dx^{i_1} \cdots dx^{i_p}$, we find

$$\begin{aligned} d\Omega &= \frac{1}{p!}(d\Omega_{i_1,\ldots,i_p})\, dx^{i_1} \cdots dx^{i_p} \\ &= \frac{1}{p!}\partial_k \Omega_{i_1,\ldots,i_p}\, dx^k dx^{i_1} \cdots dx^{i_p}. \end{aligned} \tag{11.49}$$

Now compute

$$d(d\Omega) = \frac{1}{p!}\left(\partial_l \partial_k \Omega_{i_1,\ldots,i_p}\right) dx^l dx^k dx^{i_1} \cdots dx^{i_p}. \tag{11.50}$$

Fortunately this is zero because $\partial_l \partial_k \Omega = \partial_k \partial_l \Omega$, while $dx^l dx^k = -dx^k dx^l$.

As another example let $A = A_1 dx^1 + A_2 dx^2 + A_3 dx^3$, then

$$\begin{aligned} dA &= \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\right) dx^1 dx^2 + \left(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\right) dx^3 dx^1 + \left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\right) dx^2 dx^3 \\ &= \frac{1}{2} F_{\mu\nu} dx^\mu dx^\nu, \end{aligned} \tag{11.51}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{11.52}$$

You will recognize the components of curl $\mathbf{A}$ hiding in here.

Again, if $F = F_{12}dx^1dx^2 + F_{23}dx^2dx^3 + F_{31}dx^3dx^1$ then

$$dF = \left(\frac{\partial F_{23}}{\partial x^1} + \frac{\partial F_{31}}{\partial x^2} + \frac{\partial F_{12}}{\partial x^3}\right) dx^1dx^2dx^3. \tag{11.53}$$

This looks like a divergence.

The axiom $d^2 = 0$ encompasses both "curl grad $= 0$" and "div curl $= 0$", together with an infinite number of higher-dimensional analogues. The familiar "curl $=\nabla\times$", meanwhile, is only defined in three dimensional space.

The exterior derivative takes p-forms to $(p+1)$-forms *i.e.* skew-symmetric type $(0,p)$ tensors to skew-symmetric $(0,p+1)$ tensors. How does "d" get around the fact that the derivative of a tensor is not a tensor? Well, if you apply the transformation law for A_μ, and the chain rule to $\frac{\partial}{\partial x^\mu}$ to find the transformation law for $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, you will see why: all the derivatives of the $\frac{\partial z^\nu}{\partial x^\mu}$ cancel, and $F_{\mu\nu}$ is a *bona-fide* tensor of type $(0,2)$. This sort of cancellation is why skew-symmetric objects are useful, and symmetric ones less so.

Exercise 11.4: Use axiom ii) to compute $d(d(a\wedge b))$ and confirm that it is zero.

Closed and exact forms

The Poincaré lemma, $d^2 = 0$, leads to some important terminology:

i) A p-form ω is said to be *closed* if $d\omega = 0$.

ii) A p-form ω is said to *exact* if $\omega = d\eta$ for some $(p-1)$-form η.

An exact form is necessarily closed, but a closed form is not necessarily exact. The question of when closed $\Rightarrow$ exact is one involving the global topology of the space in which the forms are defined, and will be subject of chapter 13.

Cartan's formulæ

It is sometimes useful to have expressions for the action of d coupled with the evaluation of the subsequent $(p+1)$ forms.

If f, η, ω, are $0,1,2$-forms, respectively, then $df, d\eta, d\omega$, are $1,2,3$-forms. When we plug in the appropriate number of vector fields X, Y, Z, then, after some labour, we will find

$$\begin{aligned} df(X) &= Xf. &(11.54)\\ d\eta(X,Y) &= X\eta(Y) - Y\eta(X) - \eta([X,Y]). &(11.55)\\ d\omega(X,Y,Z) &= X\omega(Y,Z) + Y\omega(Z,X) + Z\omega(X,Y)\\ &\quad -\omega([X,Y],Z) - \omega([Y,Z],X) - \omega([Z,X],Y). &(11.56)\end{aligned}$$

These formulæ, and their higher-p analogues, express d in terms of geometric objects, and so make it clear that the exterior derivative is itself a geometric object, independent of any particular co-ordinate choice.

Let us demonstate the correctness of the second formula. With $\eta = \eta_\mu dx^\mu$, the left-hand side, $d\eta(X,Y)$, is equal to

$$\partial_\mu \eta_\nu \, dx^\mu dx^\nu(X,Y) = \partial_\mu \eta_\nu (X^\mu Y^\nu - X^\nu Y^\mu). \tag{11.57}$$

The right hand side is equal to

$$X^\mu \partial_\mu(\eta_\nu Y^\nu) - Y^\mu \partial_\mu(\eta_\nu X^\nu) - \eta_\nu(X^\mu \partial_\mu Y^\nu - Y^\mu \partial_\mu X^\nu). \tag{11.58}$$

On using the product rule for the derivatives in the first two terms, we find that all derivatives of the components of X and Y cancel, and are left with exactly those terms appearing on left.

Exercise 11.5: Let ω^i, $i = 1,\ldots,r$, be a linearly independent set of one-forms defining a Pfaffian system (see sec. 11.2.1) in d dimensions.

i) Use Cartan's formulæ to show that the corresponding $(d-r)$-dimensional distribution is involutive if and only if there is an r-by-r matrix of 1-forms ${\theta^i}_j$ such that

$$d\omega^i = \sum_{j=1}^r {\theta^i}_j \wedge \omega^j.$$

ii) Show that the conditions in part i) are satisfied if there are r functions g^i and an invertible r-by-r matrix of functions ${f^i}_j$ such that

$$\omega^i = \sum_{j=1}^r {f^i}_j dg^i.$$

In this case foliation surfaces are given by the conditions $g^i(x) = \text{const.}$, $i = 1,\ldots,r$.

It is also possible, but considerably harder, to show that i) $\Rightarrow$ ii). Doing so would constitute a proof of Frobenius' theorem.

Exercise 11.6: Let ω be a closed two-form, and let $\mathrm{Null}(\omega)$ be the space of vector fields X such that $\omega(X,\ \ ) = 0$. Use the Cartan formulæ to show that if $X, Y \in \mathrm{Null}(\omega)$, then $[X, Y] \in \mathrm{Null}(\omega)$.

Lie derivative of forms

Given a p-form ω and a vector field X, we can form a $(p-1)$-form called $i_X\omega$ by writing

$$i_X\omega(\underbrace{\ldots\ldots}_{p-1\,\text{slots}}) = \omega(\overbrace{X, \underbrace{\ldots\ldots}_{p-1\,\text{slots}}}^{p\,\text{slots}}). \tag{11.59}$$

Acting on a 0-form, i_X is defined to be 0. This procedure is called the *interior multiplication* by X. It is simply a contraction

$$\omega_{j_i j_2\ldots j_p} \to \omega_{k j_2 \ldots j_p} X^k, \tag{11.60}$$

but it is convenient to have a special symbol for this operation. It is perhaps surprising that i_X turns out to be an anti-derivation, just as is d. If η and ω are p and q forms respectively, then

$$i_X(\eta \wedge \omega) = (i_X\eta) \wedge \omega + (-1)^p \eta \wedge (i_X\omega), \tag{11.61}$$

even though i_X involves no differentiation. For example, if $X = X^\mu \partial_\mu$, then

$$\begin{aligned} i_X(dx^\mu \wedge dx^\nu) &= dx^\mu \wedge dx^\nu (X^\alpha \partial_\alpha,\ \ ), \\ &= X^\mu dx^\nu - dx^\mu X^\nu, \\ &= (i_X dx^\mu) \wedge (dx^\nu) - dx^\mu \wedge (i_X dx^\nu). \end{aligned} \tag{11.62}$$

One reason for introducing i_X is that there is a nice (and profound) formula for the Lie derivative of a p-form in terms of i_X. The formula is called the *infinitesimal homotopy relation*. It reads

$$\mathcal{L}_X\omega = (d\, i_X + i_X d)\omega. \tag{11.63}$$

This formula is proved by verifying that it is true for functions and one-forms, and then showing that it is a derivation – in other words that it satisfies Leibniz' rule. From the derivation property of the Lie derivative, we immediately deduce that that the formula works for any p-form.

That the formula is true for functions should be obvious: Since $i_X f = 0$ by definition, we have

$$(d\, i_X + i_X d)f = i_X df = df(X) = Xf = \mathcal{L}_X f. \tag{11.64}$$

To show that the formula works for one forms, we evaluate

$$\begin{aligned}(d\, i_X + i_X d)(f_\nu\, dx^\nu) &= d(f_\nu X^\nu) + i_X(\partial_\mu f_\nu\, dx^\mu dx^\nu)\\ &= \partial_\mu(f_\nu X^\nu)dx^\mu + \partial_\mu f_\nu(X^\mu dx^\nu - X^\nu dx^\mu)\\ &= (X^\nu \partial_\nu f_\mu + f_\nu \partial_\mu X^\nu)dx^\mu. \end{aligned} \tag{11.65}$$

In going from the second to the third line, we have interchanged the dummy labels $\mu \leftrightarrow \nu$ in the term containing dx^ν. We recognize that the 1-form in the last line is indeed $\mathcal{L}_X f$.

To show that $di_X + i_X d$ is a derivation we must apply $d\, i_X + i_X d$ to $a \wedge b$ and use the anti-derivation property of i_x and d. This is straightforward once we recall that d takes a p-form to a $(p+1)$-form while i_X takes a p-form to a $(p-1)$-form.

Exercise 11.7: Let

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

Use the anti-derivation property of i_X to show that

$$i_X \omega = \frac{1}{(p-1)!}\, \omega_{\alpha i_2 \dots i_p} X^\alpha dx^{i_2} \cdots dx^{i_p},$$

and so verify the equivalence of (11.59) and (11.60).

Exercise 11.8: Use the infinitesimal homotopy relation to show that $\mathcal{L}$ and d commute, *i.e.* for ω a p-form, we have

$$d\,(\mathcal{L}_X \omega) = \mathcal{L}_X(d\omega).$$

11.4 Physical applications

11.4.1 Maxwell's equations

In relativistic[3] four-dimensional tensor notation the two source-free Maxwell's equations

$$\begin{aligned} \operatorname{curl}\mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\ \operatorname{div}\mathbf{B} &= 0, \end{aligned} \tag{11.66}$$

reduce to the single equation

$$\frac{\partial F_{\mu\nu}}{\partial x^\lambda} + \frac{\partial F_{\nu\lambda}}{\partial x^\mu} + \frac{\partial F_{\lambda\mu}}{\partial x^\nu} = 0. \tag{11.67}$$

where

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{11.68}$$

The "F" is traditional, for Michael Faraday. In form language, the relativistic equation becomes the even more compact expression $dF = 0$, where

$$\begin{aligned} F &\equiv \frac{1}{2} F_{\mu\nu} dx^\mu dx^\nu \\ &= B_x dydz + B_y dzdx + B_z dxdy + E_x dxdt + E_y dydt + E_z dzdt, \end{aligned} \tag{11.69}$$

is a Minkowski-space 2-form.

Exercise 11.9: Verify that the source-free Maxwell equations are indeed equivalent to $dF = 0$.

The equation $dF = 0$ is automatically satisfied if we introduce a 4-vector 1-form potential $A = -\phi dt + A_x dx + A_y dy + A_z dz$ and set $F = dA$.

The two Maxwell equations with sources

$$\begin{aligned} \operatorname{div}\mathbf{D} &= \rho, \\ \operatorname{curl}\mathbf{H} &= \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \end{aligned} \tag{11.70}$$

[3] In this section we will use units in which $c = \epsilon_0 = \mu_0 = 1$. We take the Minkowski metric to be $g_{\mu\nu} = \operatorname{diag}(-1, 1, 1, 1)$ where $x^0 = t$, $x^1 = x$, *etc.*

reduce in 4-tensor notation to the single equation

$$\partial_\mu F^{\mu\nu} = J^\nu. \tag{11.71}$$

Here $J^\mu = (\rho, \mathbf{j})$ is the current 4-vector.

This source equation takes a little more work to express in form language, but it can be done. We need a new concept: the *Hodge "star" dual* of a form. In d dimensions the "$\star$" map takes a p-form to a $(d-p)$-form. It depends on both the metric and the *orientation*. The latter means a canonical choice of the order in which to write our basis forms, with orderings that differ by an even permutation being counted as the same. The full d-dimensional definition involves the Levi-Civita duality operation of chapter 10 , combined with the use of the metric tensor to raise indices. Recall that $\sqrt{g} = \sqrt{\det g_{\mu\nu}}$. (In Minkowski-signature metrics we should replace $\sqrt{g}$ by $\sqrt{-g}$.) We define "$\star$" to be a linear map

$$\star : \bigwedge^p (T^*M) \to \bigwedge^{(d-p)} (T^*M) \tag{11.72}$$

such that

$$\star\, dx^{i_1} \dots dx^{i_p} \stackrel{\text{def}}{=} \frac{1}{(d-p)!} \sqrt{g} g^{i_1 j_1} \dots g^{i_p j_p} \epsilon_{j_1 \cdots j_p j_{p+1} \cdots j_d} dx^{j_{p+1}} \dots dx^{j_d}. \tag{11.73}$$

Although this definition looks a trifle involved, computations involving it are not so intimidating. The trick is to work, whenever possible, with oriented orthonormal frames. If we are in euclidean space and $\{\mathbf{e}^{*i_1}, \mathbf{e}^{*i_2}, \dots, \mathbf{e}^{*i_d}\}$ is an ordering of the orthonormal basis for $(T^*M)_x$ whose orientation is equivalent to $\{\mathbf{e}^{*1}, \mathbf{e}^{*2}, \dots, \mathbf{e}^{*d}\}$ then

$$\star\,(\mathbf{e}^{*i_1} \wedge \mathbf{e}^{*i_2} \wedge \cdots \wedge \mathbf{e}^{*i_p}) = \mathbf{e}^{*i_{p+1}} \wedge \mathbf{e}^{*i_{p+2}} \wedge \cdots \wedge \mathbf{e}^{*i_d}. \tag{11.74}$$

For example, in three dimensions, and with x, y, z, our usual Cartesian co-ordinates, we have

$$\begin{aligned} \star\, dx &= dydz, \\ \star\, dy &= dzdx, \\ \star\, dz &= dxdy. \end{aligned} \tag{11.75}$$

An analogous method works for Minkowski-signature $(-,+,+,+)$ metrics, except that now we must include a minus sign for each negatively normed

dt factor in the form being "starred." Taking $\{dt, dx, dy, dz\}$ as our oriented basis, we therefore find[4]

$$\begin{aligned}
\star\, dxdy &= -dzdt,\\
\star\, dydz &= -dxdt,\\
\star\, dzdx &= -dydt,\\
\star\, dxdt &= dydz,\\
\star\, dydt &= dzdx,\\
\star\, dzdt &= dxdy.
\end{aligned} \tag{11.76}$$

For example, the first of these equations is derived by observing that $(dxdy)(-dzdt) = dtdxdydz$, and that there is no "dt" in the product $dxdy$. The fourth follows from observing that that $(dxdt)(-dydx) = dtdxdydz$, but there is a negative-normed "dt" in the product $dxdt$.

The $\star$ map is constructed so that if

$$\alpha = \frac{1}{p!}\alpha_{i_1 i_2 \ldots i_p} dx^{i_1} dx^{i_2} \cdots dx^{i_p}, \tag{11.77}$$

and

$$\beta = \frac{1}{p!}\beta_{i_1 i_2 \ldots i_p} dx^{i_1} dx^{i_2} \cdots dx^{i_p}, \tag{11.78}$$

then

$$\alpha \wedge (\star\beta) = \beta \wedge (\star\alpha) = \langle \alpha, \beta \rangle\, \sigma, \tag{11.79}$$

where the inner product $\langle \alpha, \beta \rangle$ is defined to be the invariant

$$\langle \alpha, \beta \rangle = \frac{1}{p!} g^{i_1 j_1} g^{i_2 j_2} \cdots g^{i_p j_p} \alpha_{i_1 i_2 \ldots i_p} \beta_{j_1 j_2 \ldots j_p}, \tag{11.80}$$

and σ is the *volume form*

$$\sigma = \sqrt{g}\, dx^1 dx^2 \cdots dx^d. \tag{11.81}$$

In future we will write $\alpha \star \beta$ for $\alpha \wedge (\star\beta)$. Bear in mind that the "$\star$" in this expression is acting β and is not some new kind of binary operation.

We now apply these ideas to Maxwell. From the field-strength 2-form

$$F = B_x dydz + B_y dzdx + B_z dxdy + E_x dxdt + E_y dydt + E_z dzdt, \tag{11.82}$$

[4] See for example: Misner, Thorn and Wheeler, *Gravitation*, (MTW) page 108.

we get a dual 2-form

$$\star F = -B_x dxdt - B_y dydt - B_z dzdt + E_x dydz + E_y dzdx + E_z dxdy. \quad (11.83)$$

We can check that we have correctly computed the Hodge star of F by taking the wedge product, for which we find

$$F \star F = \frac{1}{2}(F_{\mu\nu}F^{\mu\nu})\sigma = (B_x^2 + B_y^2 + B_z^2 - E_x^2 - E_y^2 - E_z^2)dtdxdydz. \quad (11.84)$$

Observe that the expression $B^2 - E^2$ is a Lorentz scalar. Similarly, from the current 1-form

$$J \equiv J_\mu dx^\mu = -\rho\, dt + j_x dx + j_y dy + j_z dz, \quad (11.85)$$

we derive the dual current 3-form

$$\star J = \rho\, dxdydz - j_x dtdydz - j_y dtdzdx - j_z dtdxdy, \quad (11.86)$$

and check that

$$J \star J = (J_\mu J^\mu)\sigma = (-\rho^2 + j_x^2 + j_y^2 + j_z^2)dtdxdydz. \quad (11.87)$$

Observe that

$$d \star J = \left(\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j}\right) dtdxdydz = 0, \quad (11.88)$$

expresses the charge conservation law.

Writing out the terms explicitly shows that the source-containing Maxwell equations reduce to $d\star F = \star J$. All four Maxwell equations are therefore very compactly expressed as

$$\boxed{dF = 0, \quad d \star F = \star J.}$$

Observe that current conservation $d\star J = 0$ follows from the second Maxwell equation as a consequence of $d^2 = 0$.

Exercise 11.10: Show that for a p-form ω in d euclidean dimensions we have

$$\star\star\omega = (-1)^{p(d-p)}\omega.$$

Show, further, that for a Minkowski metric an additional minus sign has to be inserted. (For example, $\star\star F = -F$, even though $(-1)^{2(4-2)} = +1$.)

11.4.2 Hamilton's equations

Hamiltonian dynamics takes place in *phase space*, a manifold with co-ordinates $(q^1, \ldots, q^n, p^1, \ldots, p^n)$. Since momentum is a naturally covariant vector,[5] phase space is usually the *co-tangent bundle* T^*M of the configuration manifold M. We are writing the indices on the p's upstairs though, because we are considering them as co-ordinates in T^*M.

We expect that you are familiar with Hamilton's equation in their q, p setting. Here, we shall describe them as they appear in a modern book on Mechanics, such as Abrahams and Marsden's *Foundations of Mechanics*, or V. I. Arnold's *Mathematical Methods of Classical Mechanics*.

Phase space is an example of a *symplectic manifold*, a manifold equipped with a *symplectic form* — a closed, non-degenerate, 2-form field

$$\omega = \frac{1}{2}\omega_{ij} dx^i dx^j. \tag{11.89}$$

Recall that the word *closed* means that $d\omega = 0$. *Non-degenerate* means that for any point x the statement that $\omega(X, Y) = 0$ for all vectors $Y \in TM_x$ implies that $X = 0$ at that point (or equivalently that for all x the matrix $\omega_{ij}(x)$ has an inverse $\omega^{ij}(x)$).

Given a *Hamiltonian* function H on our symplectic manifold, we define a velocity vector-field v_H by solving

$$dH = -i_{v_H}\omega = -\omega(v_H, \ ) \tag{11.90}$$

for v_H. If the symplectic form is $\omega = dp^1 dq^1 + dp^2 dq^2 + \cdots + dp^n dq^n$, this is nothing but a fancy form of Hamilton's equations. To see this, we write

$$dH = \frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial p^i} dp^i \tag{11.91}$$

and use the customary notation $(\dot{q}^i, \dot{p}^i)$ for the velocity-in-phase-space components, so that

$$v_H = \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}^i \frac{\partial}{\partial p^i}. \tag{11.92}$$

Now we work out

$$\begin{aligned} i_{v_H}\omega &= dp^i dq^i (\dot{q}^j \partial_{q^j} + \dot{p}^j \partial_{p^j}, \quad) \\ &= \dot{p}^i dq^i - \dot{q}^i dp^i, \end{aligned} \tag{11.93}$$

[5] To convince yourself of this, remember that in quantum mechanics $\hat{p}_\mu = -i\hbar \frac{\partial}{\partial x^\mu}$, and the gradient of a function is a covector.

so, comparing coefficients of dp^i and dq^i on the two sides of $dH = -i_{v_H}\omega$, we read off

$$\dot{q}^i = \frac{\partial H}{\partial p^i}, \quad \dot{p}^i = -\frac{\partial H}{\partial q^i}. \tag{11.94}$$

Darboux' theorem, which we will not try to prove, says that for any point x we can always find co-ordinates p, q, valid in some neigbourhood of x, such that $\omega = dp^1 dq^1 + dp^2 dq^2 + \cdots dp^n dq^n$. Given this fact, it is not unreasonable to think that there is little to gained by using the abstract differential-form language. In simple cases this is so, and the traditional methods are fine. It may be, however, that the neigbourhood of x where the Darboux co-ordinates work is not the entire phase space, and we need to cover the space with overlapping p, q co-ordinate charts. Then, what is a p in one chart will usually be a combination of p's and q's in another. In this case, the traditional form of Hamilton's equations loses its appeal in comparison to the co-ordinate-free $dH = -i_{v_H}\omega$.

Given two functions H_1, H_2 we can define their *Poisson bracket* $\{H_1, H_2\}$. Its importance lies in Dirac's observation that the passage from classical mechanics to quantum mechanics is accomplished by replacing the Poisson bracket of two quantities, A and B, with the commutator of the corresponding operators $\hat{A}$, and $\hat{B}$:

$$i[\hat{A}, \hat{B}] \quad \longleftrightarrow \quad \hbar\{A, B\} + O\left(\hbar^2\right). \tag{11.95}$$

We define the Poisson bracket by[6]

$$\{H_1, H_2\} \stackrel{\text{def}}{=} \left.\frac{dH_2}{dt}\right|_{H_1} = v_{H_1} H_2. \tag{11.96}$$

Now, $v_{H_1} H_2 = dH_2(v_{H_1})$, and Hamilton's equations say that $dH_2(v_{H_1}) = \omega(v_{H_1}, v_{H_2})$. Thus,

$$\{H_1, H_2\} = \omega(v_{H_1}, v_{H_2}). \tag{11.97}$$

The skew symmetry of $\omega(v_{H_1}, v_{H_2})$ shows that despite the asymmetrical appearance of the definition we have skew symmetry: $\{H_1, H_2\} = -\{H_2, H_1\}$.

Moreover, since

$$v_{H_1}(H_2 H_3) = (v_{H_1} H_2) H_3 + H_2(v_{H_1} H_3), \tag{11.98}$$

[6]Our definition differs in sign from the traditional one, but has the advantage of minimizing the number of minus signs in subsequent equations.

the Poisson bracket is a derivation:

$$\{H_1, H_2H_3\} = \{H_1, H_2\}H_3 + H_2\{H_1, H_3\}. \tag{11.99}$$

Neither the skew symmetry nor the derivation property require the condition that $d\omega = 0$. What does need ω to be closed is the *Jacobi identity*:

$$\{\{H_1, H_2\}, H_3\} + \{\{H_2, H_3\}, H_1\} + \{\{H_3, H_1\}, H_2\} = 0. \tag{11.100}$$

We establish Jacobi by using Cartan's formula in the form

$$\begin{aligned} d\omega(v_{H_1}, v_{H_2}, v_{H_3}) &= v_{H_1}\omega(v_{H_2}, v_{H_3}) + v_{H_2}\omega(v_{H_3}, v_{H_1}) + v_{H_3}\omega(v_{H_1}, v_{H_2}) \\ &\quad -\omega([v_{H_1}, v_{H_2}], v_{H_3}) - \omega([v_{H_2}, v_{H_3}], v_{H_1}) - \omega([v_{H_3}, v_{H_1}], v_{H_2}). \end{aligned} \tag{11.101}$$

It is relatively straight-forward to interpret each term in the first of these lines as Poisson brackets. For example,

$$v_{H_1}\omega(v_{H_2}, v_{H_3}) = v_{H_1}\{H_2, H_3\} = \{H_1, \{H_2, H_3\}\}. \tag{11.102}$$

Relating the terms in the second line to Poisson brackets requires a little more effort. We proceed as follows:

$$\begin{aligned} \omega([v_{H_1}, v_{H_2}], v_{H_3}) &= -\omega(v_{H_3}, [v_{H_1}, v_{H_2}]) \\ &= dH_3([v_{H_1}, v_{H_2}]) \\ &= [v_{H_1}, v_{H_2}]H_3 \\ &= v_{H_1}(v_{H_2}H_3) - v_{H_2}(v_{H_1}H_3) \\ &= \{H_1, \{H_2, H_3\}\} - \{H_2, \{H_1, H_3\}\} \\ &= \{H_1, \{H_2, H_3\}\} + \{H_2, \{H_3, H_1\}\}. \end{aligned} \tag{11.103}$$

Adding everything togther now shows that

$$\begin{aligned} 0 &= d\omega(v_{H_1}, v_{H_2}, v_{H_3}) \\ &= -\{\{H_1, H_2\}, H_3\} - \{\{H_2, H_3\}, H_1\} - \{\{H_3, H_1\}, H_2\} \end{aligned} \tag{11.104}$$

If we rearrange the Jacobi identity as

$$\{H_1, \{H_2, H_3\}\} - \{H_2, \{H_1, H_3\}\} = \{\{H_1, H_2\}, H_3\}, \tag{11.105}$$

we see that it is equivalent to

$$[v_{H_1}, v_{H_2}] = v_{\{H_1,H_2\}}.$$

The algebra of Poisson brackets is therefore *homomorphic* to the algebra of the Lie brackets. The correspondence is not an *isomorphism*, however: the assignment $H \mapsto v_H$ fails to be one-to-one because constant functions map to the zero vector field.

Exercise 11.11: Use the infinitesimal homotopy relation, to show that $\mathcal{L}_{v_H}\omega = 0$, where v_H is the vector field corresponding to H. Suppose now that the phase space is $2n$ dimensional. Show that in local Darboux co-ordinates the $2n$-form $\omega^n/n!$ is, up to a sign, the phase-space volume element $d^n p\, d^n q$. Show that $\mathcal{L}_{v_H}\omega^n/n! = 0$ and that this result is *Liouville's theorem* on the conservation of phase-space volume.

The classical mechanics of spin

It is sometimes said in books on quantum mechanics that the spin of an electron, or other elementary particle, is a purely quantum concept and cannot be described by classical mechanics. This statement is false, but spin *is* the simplest system in which traditional physicist's methods become ugly and it helps to use the modern symplectic language. A "spin" **S** can be regarded as a fixed length vector that can point in any direction in $\mathbb{R}^3$. We will take it to be of unit length so that its components are

$$\begin{aligned} S_x &= \sin\theta\cos\phi, \\ S_y &= \sin\theta\sin\phi, \\ S_z &= \cos\theta, \end{aligned} \tag{11.106}$$

where θ and ϕ are polar co-ordinates on the two-sphere S^2.

The surface of the sphere turns out to be both the configuration space and the phase space. In particular the phase space for a spin is *not* the cotangent bundle of the configuration space. This has to be so: we learned from Niels Bohr that a $2n$-dimensional phase space contains roughly one quantum state for every $\hbar^n$ of phase-space volume. A cotangent bundle always has infinite volume, so its corresponding Hilbert space is necessarily infinite dimensional. A quantum spin, however, has a *finite-dimensional* Hilbert space so its classical phase space must have a finite total volume.

This finite-volume phase space seems un-natural in the traditional view of mechanics, but it fits comfortably into the modern symplectic picture.

We want to treat all points on the sphere alike, and so it is natural to take the symplectic 2-form to be proportional to the element of area. Suppose that $\omega = \sin\theta\, d\theta d\phi$. We could write $\omega = -d\cos\theta\, d\phi$ and regard ϕ as "q" and $-\cos\theta$ as "p" (Darboux' theorem in action!), but this identification is singular at the north and south poles of the sphere, and, besides, it obscures the spherical symmetry of the problem, which is manifest when we think of ω as d(area).

Let us take our hamiltonian to be $H = BS_x$, corresponding to an applied magnetic field in the x direction, and see what Hamilton's equations give for the motion. First we take the exterior derivative

$$d(BS_x) = B(\cos\theta\cos\phi d\theta - \sin\theta\sin\phi d\phi). \tag{11.107}$$

This is to be set equal to

$$-\omega(v_{BS_x},\ \ ) = v^\theta(-\sin\theta)d\phi + v^\phi\sin\theta d\theta. \tag{11.108}$$

Comparing coefficients of $d\theta$ and $d\phi$, we get

$$v_{(BS_x)} = v^\theta\partial_\theta + v^\phi\partial_\phi = B(\sin\phi\partial_\theta + \cos\phi\cot\theta\partial_\phi), \tag{11.109}$$

i.e. B times the velocity vector for a rotation about the x axis. This velocity field therefore describes a steady Larmor precession of the spin about the applied field. This is exactly the motion predicted by quantum mechanics. Similarly, setting $B = 1$, we find

$$\begin{aligned} v_{S_y} &= -\cos\phi\partial_\theta + \sin\phi\cot\theta\partial_\phi, \\ v_{S_z} &= -\partial_\phi. \end{aligned} \tag{11.110}$$

From these velocity fields we can compute the Poisson brackets:

$$\begin{aligned} \{S_x, S_y\} &= \omega(v_{S_x}, v_{S_y}) \\ &= \sin\theta d\theta d\phi(\sin\phi\partial_\theta + \cos\phi\cot\theta\partial_\phi, -\cos\phi\partial_\theta + \sin\phi\cot\theta\partial_\phi) \\ &= \sin\theta(\sin^2\phi\cot\theta + \cos^2\phi\cot\theta) \\ &= \cos\theta = S_z. \end{aligned}$$

Repeating the exercise leads to

$$\begin{aligned} \{S_x, S_y\} &= S_z, \\ \{S_y, S_z\} &= S_x, \\ \{S_z, S_x\} &= S_y. \end{aligned} \tag{11.111}$$

These Poisson brackets for our classical "spin" are to be compared with the commutation relations $[\hat{S}_x, \hat{S}_y] = i\hbar\hat{S}_z$ *etc.* for the quantum spin operators $\hat{S}_i$.

11.5 Covariant derivatives

Covariant derivatives are a general class of derivatives that act on sections of a vector or tensor bundle over a manifold. We will begin by considering derivatives on the tangent bundle, and in the exercises indicate how the idea generalizes to other bundles.

11.5.1 Connections

The Lie and exterior derivatives require no structure beyond that which comes for free with our manifold. Another type of derivative that can act on tangent-space vectors and tensors is the *covariant derivative* $\nabla_X \equiv X^\mu \nabla_\mu$. This requires an additional mathematical object called an *affine connection*.

The covariant derivative is defined by:

i) Its action on scalar functions as

$$\nabla_X f = Xf. \tag{11.112}$$

ii) Its action a basis set of tangent-vector fields $\mathbf{e}_a(x) = e_a^\mu(x)\partial_\mu$ (a local frame, or *vielbein*[7]) by introducing a set of functions ${\omega^i}_{jk}(x)$ and setting

$$\nabla_{\mathbf{e}_k}\mathbf{e}_j = \mathbf{e}_i {\omega^i}_{jk}. \tag{11.113}$$

ii) Extending this definition to any other type of tensor by requiring ∇_X to be a derivation.

iii) Requiring that the result of applying ∇_X to a tensor is a tensor of the same type.

The set of functions ${\omega^i}_{jk}(x)$ is the *connection*. In any local co-ordinate chart we can choose them at will, and different choices define different covariant derivatives. (There may be global compatibility constraints, however, which appear when we assemble the charts into an atlas.)

[7]In practice *viel*, "many", is replaced by the appropriate German numeral: *ein-, zwei-, drei-, vier-, fünf-, ...*, indicating the dimension. The word *bein* means "leg."

Warning: Despite having the appearance of one, $\omega^i{}_{jk}$ is **not** a tensor. It transforms inhomogeneously under a change of frame or co-ordinates — see equation (11.132).

We can, of course, take as our basis vectors the co-ordinate vectors $\mathbf{e}_\mu \equiv \partial_\mu$. When we do this it is traditional to use the symbol Γ for the co-ordinate frame connection instead of ω. Thus,

$$\nabla_\mu \mathbf{e}_\nu \equiv \nabla_{\mathbf{e}_\mu} \mathbf{e}_\nu = \mathbf{e}_\lambda \Gamma^\lambda{}_{\nu\mu}. \tag{11.114}$$

The numbers $\Gamma^\lambda{}_{\nu\mu}$ are often called *Christoffel symbols*.

As an example consider the covariant derivative of a vector $f^\nu \mathbf{e}_\nu$. Using the derivation property we have

$$\begin{aligned} \nabla_\mu (f^\nu \mathbf{e}_\nu) &= (\partial_\mu f^\nu)\mathbf{e}_\nu + f^\nu \nabla_\mu \mathbf{e}_\nu \\ &= (\partial_\mu f^\nu)\mathbf{e}_\nu + f^\nu \mathbf{e}_\lambda \Gamma^\lambda{}_{\nu\mu} \\ &= \mathbf{e}_\nu \left\{ \partial_\mu f^\nu + f^\lambda \Gamma^\nu{}_{\lambda\mu} \right\}. \end{aligned} \tag{11.115}$$

In the first line we have used the defining property that $\nabla_{\mathbf{e}_\mu}$ acts on the functions f^ν as ∂_μ, and in the last line we interchanged the dummy indices ν and λ. We often abuse the notation by writing only the components, and set

$$\nabla_\mu f^\nu = \partial_\mu f^\nu + f^\lambda \Gamma^\nu{}_{\lambda\mu}. \tag{11.116}$$

Similarly, acting on the components of a mixed tensor, we would write

$$\nabla_\mu A^\alpha{}_{\beta\gamma} = \partial_\mu A^\alpha{}_{\beta\gamma} + \Gamma^\alpha{}_{\lambda\mu} A^\lambda{}_{\beta\gamma} - \Gamma^\lambda{}_{\beta\mu} A^\alpha{}_{\lambda\gamma} - \Gamma^\lambda{}_{\gamma\mu} A^\alpha{}_{\beta\lambda}. \tag{11.117}$$

When we use this notation, we are no longer regarding the tensor components as "functions."

Observe that the plus and minus signs in (11.117) are required so that, for example, the covariant derivative of the scalar function $f_\alpha g^\alpha$ is

$$\begin{aligned} \nabla_\mu (f_\alpha g^\alpha) &= \partial_\mu (f_\alpha g^\alpha) \\ &= (\partial_\mu f_\alpha) g^\alpha + f_\alpha (\partial_\mu g^\alpha) \\ &= \left(\partial_\mu f_\alpha - f_\lambda \Gamma^\lambda{}_{\alpha\mu}\right) g^\alpha + f_\alpha \left(\partial_\mu g^\alpha + g^\lambda \Gamma^\alpha{}_{\lambda\mu}\right) \\ &= (\nabla_\mu f_\alpha) g^\alpha + f_\alpha (\nabla_\mu g^\alpha), \end{aligned} \tag{11.118}$$

and so satisfies the derivation property.

Parallel transport

We have defined the covariant derivative *via* its formal calculus properties. It has, however, a geometrical interpretation. As with the Lie derivative, in order to compute the derivative along X of the vector field Y, we have to somehow carry the vector $Y(x)$ from the tangent space TM_x to the tangent space $TM_{x+\epsilon X}$, where we can subtract it from $Y(x+\epsilon X)$. The Lie derivative carries Y along with the X flow. The covariant derivative implicitly carries Y by "parallel transport". If $\gamma : s \mapsto x^\mu(s)$ is a parameterized curve with tangent vector $X^\mu \partial_\mu$, where

$$X^\mu = \frac{dx^\mu}{ds}, \tag{11.119}$$

then we say that the vector field $Y(x^\mu(s))$ is *parallel transported* along the curve γ if

$$\nabla_X Y = 0, \tag{11.120}$$

at each point $x^\mu(s)$. Thus, a vector that in the vielbein frame $\mathbf{e}_i$ at x has components Y^i will, after being parallel transported to $x + \epsilon X$, end up components

$$Y^i - \epsilon \omega^i{}_{jk} Y^j X^k. \tag{11.121}$$

In a co-ordinate frame, after parallel transport through an infinitesimal displacement δx^μ, the vector $Y^\nu \partial_\nu$ will have components

$$Y^\nu \to Y^\nu - \Gamma^\nu{}_{\lambda\mu} Y^\lambda \delta x^\mu, \tag{11.122}$$

and so

$$\begin{aligned} \delta x^\mu \nabla_\mu Y^\nu &= Y^\nu(x^\mu + \delta x^\mu) - \{Y^\nu(x) - \Gamma^\nu{}_{\lambda\mu} Y^\lambda \delta x^\mu\} \\ &= \delta x^\mu \{\partial_\mu Y^\nu + \Gamma^\nu{}_{\lambda\mu} Y^\lambda\}. \end{aligned} \tag{11.123}$$

Curvature and torsion

As we said earlier, the connection $\omega^i{}_{jk}(x)$ is not itself a tensor. Two important quantities which *are* tensors, are associated with ∇_X:

i) The *torsion*

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{11.124}$$

The quantity $T(X,Y)$ is a vector depending linearly on X, Y, so T at the point x is a map $TM_x \times TM_x \to TM_x$, and so a tensor of type (1,2). In a co-ordinate frame it has components

$$T^\lambda{}_{\mu\nu} = \Gamma^\lambda{}_{\mu\nu} - \Gamma^\lambda{}_{\nu\mu}. \tag{11.125}$$

ii) The *Riemann curvature tensor*

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_Z Z - \nabla_{[X,Y]} Z. \tag{11.126}$$

The quantity $R(X,Y)Z$ is also a vector, so $R(X,Y)$ is a linear map $TM_x \to TM_x$, and thus R itself is a tensor of type (1,3). Written out in a co-ordinate frame, we have

$$R^\alpha{}_{\beta\mu\nu} = \partial_\mu \Gamma^\alpha{}_{\beta\nu} - \partial_\nu \Gamma^\alpha{}_{\beta\mu} + \Gamma^\alpha{}_{\lambda\mu}\Gamma^\lambda{}_{\beta\nu} - \Gamma^\alpha{}_{\lambda\nu}\Gamma^\lambda{}_{\beta\mu}. \tag{11.127}$$

If our manifold comes equipped with a metric tensor $g_{\mu\nu}$ (and is thus a *Riemann manifold*), and if we require both that $T = 0$ and $\nabla_\mu g_{\alpha\beta} = 0$, then the connection is uniquely determined, and is called the *Riemann*, or *Levi-Civita*, connection. In a co-ordinate frame it is given by

$$\Gamma^\alpha{}_{\mu\nu} = \frac{1}{2} g^{\alpha\lambda} \left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right). \tag{11.128}$$

This is the connection that appears in General Relativity.

The curvature tensor measures the degree of path dependence in parallel transport: if $Y^\nu(x)$ is parallel transported along a path $\gamma : s \mapsto x^\mu(s)$ from a to b, and if we deform γ so that $x^\mu(s) \to x^\mu(s) + \delta x^\mu(s)$ while keeping the endpoints a, b fixed, then

$$\delta Y^\alpha(b) = -\int_a^b R^\alpha{}_{\beta\mu\nu}(x) Y^\beta(x) \delta x^\mu \, dx^\nu. \tag{11.129}$$

If $R^\alpha{}_{\beta\mu\nu} \equiv 0$ then the effect of parallel transport from a to b will be independent of the route taken.

The geometric interpretation of $T_{\mu\nu}$ is less transparent. On a two-dimensional surface a connection is torsion free when the tangent space "rolls without slipping" along the curve γ.

Exercise 11.12: *Metric compatibility*. Show that the Riemann connection

$$\Gamma^\alpha{}_{\mu\nu} = \frac{1}{2} g^{\alpha\lambda} \left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right).$$

follows from the torsion-free condition $\Gamma^\alpha{}_{\mu\nu} = \Gamma^\alpha{}_{\nu\mu}$ together with the *metric compatibility condition*

$$\nabla_\mu g_{\alpha\beta} \equiv \partial_\mu \, g_{\alpha\beta} - \Gamma^\nu{}_{\alpha\mu} \, g_{\nu\beta} - \Gamma^\nu{}_{\alpha\mu} \, g_{\alpha\nu} = 0.$$

Show that "metric compatibility" means that that the operation of raising or lowering indices commutes with covariant derivation.

Exercise 11.13: *Geodesic equation.* Let $\gamma : s \mapsto x^\mu(s)$ be a parametrized path from a to b. Show that the Euler-Lagrange equation that follows from minimizing the distance functional

$$S(\gamma) = \int_a^b \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}\, ds,$$

where the dots denote differentiation with respect to the parameter s, is

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0.$$

Here $\Gamma^\mu{}_{\alpha\beta}$ is the Riemann connection (11.128).

Exercise 11.14: Show that if A^μ is a vector field then, for the Riemann connection,

$$\nabla_\mu A^\mu = \frac{1}{\sqrt{g}}\frac{\partial \sqrt{g} A^\mu}{\partial x^\mu}.$$

In other words, show that

$$\Gamma^\alpha{}_{\alpha\mu} = \frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}}{\partial x^\mu}.$$

Deduce that the Laplacian acting on a scalar field ϕ can be defined by setting either

$$\nabla^2\phi = g_{\mu\nu}\nabla_\mu\nabla_\nu\phi,$$

or

$$\nabla^2\phi = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\mu}\left(\sqrt{g}g^{\mu\nu}\frac{\partial\phi}{\partial x^\nu}\right),$$

the two definitions being equivalent.

11.5.2 Cartan's form viewpoint

Let $\mathbf{e}^{*j}(x) = e^{*j}_{\ \mu}(x)dx^\mu$ be the basis of one-forms dual to the vielbein frame $\mathbf{e}_i(x) = e_i^\mu(x)\partial_\mu$. Since

$$\delta^i_j = \mathbf{e}^{*i}(\mathbf{e}_j) = e^{*j}_{\ \mu}e_i^\mu, \tag{11.130}$$

the matrices $e^{*j}_{\ \mu}$ and e_i^μ are inverses of one-another. We can use them to change from roman vielbein indices to greek co-ordinate frame indices. For example:

$$g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j) = e_i^\mu g_{\mu\nu} e_j^\nu, \tag{11.131}$$

and

$$\omega^i{}_{jk} = e^{*i}{}_\nu(\partial_\mu e^\nu_j)e^\mu_k + e^{*i}{}_\lambda e^\nu_j e^\mu_k \Gamma^\lambda{}_{\nu\mu}. \tag{11.132}$$

Cartan regards the connection as being a matrix $\mathbf{\Omega}$ of one-forms with matrix entries $\omega^i{}_j = \omega^i{}_{j\mu} dx^\mu$. In this language equation (11.113) becomes

$$\nabla_X \mathbf{e}_j = \mathbf{e}_i \omega^i{}_j(X). \tag{11.133}$$

Cartan's viewpoint separates off the index μ, which refers to the direction $\delta x^\mu \propto X^\mu$ in which we are differentiating, from the matrix indices i and j that act on the components of the vector or tensor being differentiated. This separation becomes very natural when the vector space spanned by the $\mathbf{e}_i(x)$ is no longer the tangent space, but some other "internal" vector space attached to the point x. Such internal spaces are common in physics, an important example being the "colour space" of gauge field theories. Physicists, following Hermann Weyl, call a connection on an internal space a "gauge potential." To mathematicians it is simply a connection on the vector bundle that has the internal spaces as its fibres.

Cartan also regards the torsion $\mathbf{T}$ and curvature $\mathbf{R}$ as forms; in this case vector- and matrix-valued two-forms, respectively, with entries

$$T^i = \frac{1}{2} T^i_{\mu\nu} dx^\mu dx^\nu, \tag{11.134}$$

$$R^i{}_k = \frac{1}{2} R^i{}_{k\mu\nu} dx^\mu dx^\nu. \tag{11.135}$$

In his form language the equations defining the torsion and curvature become *Cartan's structure equations*:

$$d\mathbf{e}^{*i} + \omega^i{}_j \wedge \mathbf{e}^{*j} = T^i, \tag{11.136}$$

and

$$d\omega^i{}_k + \omega^i{}_j \wedge \omega^j{}_k = R^i{}_k. \tag{11.137}$$

The last equation can be written more compactly as

$$d\mathbf{\Omega} + \mathbf{\Omega} \wedge \mathbf{\Omega} = \mathbf{R}. \tag{11.138}$$

From this, by taking the exterior derivative, we obtain the *Bianchi identity*

$$d\mathbf{R} - \mathbf{R} \wedge \mathbf{\Omega} + \mathbf{\Omega} \wedge \mathbf{R} = 0. \tag{11.139}$$

On a Riemann manifold, we can take the vielbein frame $\mathbf{e}_i$ to be orthonormal. In this case the roman-index metric $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$ becomes δ_{ij}. There is then no distinction between covariant and contravariant roman indices, and the connection and curvature forms, $\mathbf{\Omega}$, $\mathbf{R}$, being infinitesimal rotations, become skew symmetric matrices:

$$\omega_{ij} = -\omega_{ji}, \quad R_{ij} = -R_{ji}. \tag{11.140}$$

11.6 Further exercises and problems

Exercise 11.15: Consider the vector fields $X = y\partial_x$, $Y = \partial_y$ in $\mathbb{R}^2$. Find the flows associated with these fields, and use them to verify the statements made in section 11.2.1 about the geometric interpretation of the Lie bracket.

Exercise 11.16: Show that the pair of vector fields $L_z = x\partial_y - y\partial_x$ and $L_y = z\partial_x - x\partial_z$ in $\mathbb{R}^3$ is in involution wherever they are both non-zero. Show further that the general solution of the system of partial differential equations

$$\begin{aligned}(x\partial_y - y\partial_x)f &= 0,\\ (x\partial_z - z\partial_x)f &= 0,\end{aligned}$$

in $\mathbb{R}^3$ is $f(x, y, z) = F(x^2 + y^2 + z^2)$, where F is an arbitrary function.

Exercise 11.17: In the rolling conditions (11.29) we are using the "Y" convention for Euler angles. In this convention θ and ϕ are the usual spherical polar co-ordinate angles with respect to the space-fixed xyz axes. They specify the direction of the body-fixed Z axis about which we make the final ψ rotation — see figure 11.7.

a) Show that (11.29) are indeed the no-slip rolling conditions

$$\begin{aligned}\dot{x} &= \omega_y,\\ \dot{y} &= -\omega_x,\\ 0 &= \omega_z,\end{aligned}$$

where $(\omega_x, \omega_y, \omega_z)$ are the components of the ball's angular velocity in the xyz space-fixed frame.

b) Solve the three constraints in (11.29) so as to obtain the vector fields $\mathbf{roll_x}$, $\mathbf{roll_y}$ of (11.30).

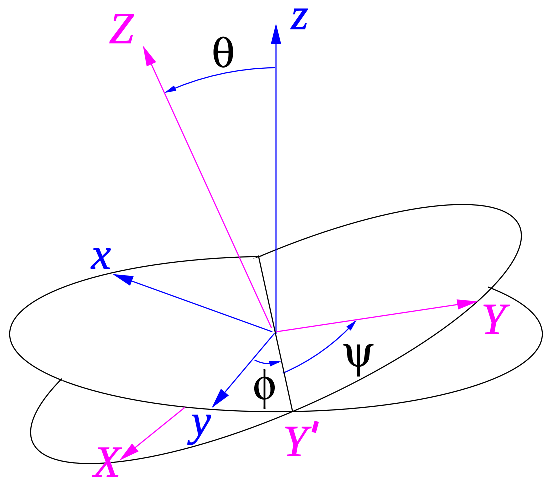

Figure 11.7: *The "Y" convention for Euler angles. The XYZ axes are fixed to the ball, and the the xyz axes are fixed in space. We first rotate the ball through an angle ϕ about the z axis, thus taking $y \to Y'$, then through θ about Y', and finally through ψ about Z, so taking $Y' \to Y$.*

c) Show that

$$[\mathbf{roll_x}, \mathbf{roll_y}] = -\mathbf{spin_z},$$

where $\mathbf{spin_z} \equiv \partial_\phi$, corresponds to a rotation about a vertical axis through the point of contact. This is a new motion, being forbidden by the $\omega_z = 0$ condition.

d) Show that

$$\begin{aligned}
[\mathbf{spin_z}, \mathbf{roll_x}] &= \mathbf{spin_x}, \\
[\mathbf{spin_z}, \mathbf{roll_y}] &= \mathbf{spin_y},
\end{aligned}$$

where the new vector fields

$$\begin{aligned}
\mathbf{spin_x} &\equiv -(\mathbf{roll_y} - \partial_y), \\
\mathbf{spin_y} &\equiv \quad (\mathbf{roll_x} - \partial_x),
\end{aligned}$$

correspond to rotations of the ball about the space-fixed x and y axes through its centre, and with the centre of mass held fixed.

We have generated five independent vector fields from the original two. Therefore, by sufficient rolling to-and-fro, we can position the ball anywhere on the table, and in any orientation.

Exercise 11.18: The semi-classical dynamics of charge $-e$ electrons in a magnetic solid are governed by the equations[8]

$$\begin{aligned}\dot{\mathbf{r}} &= \frac{\partial \epsilon(\mathbf{k})}{\partial \mathbf{k}} - \dot{\mathbf{k}} \times \mathbf{\Omega},\\ \dot{\mathbf{k}} &= -\frac{\partial V}{\partial \mathbf{r}} - e\dot{\mathbf{r}} \times \mathbf{B}.\end{aligned}$$

Here $\mathbf{k}$ is the Bloch momentum of the electron, $\mathbf{r}$ is its position, $\epsilon(\mathbf{k})$ its band energy (in the extended-zone scheme), and $\mathbf{B}(\mathbf{r})$ is the external magnetic field. The components Ω_i of the *Berry curvature* $\mathbf{\Omega}(\mathbf{k})$ are given in terms of the periodic part $|u(\mathbf{k})\rangle$ of the Bloch wavefunctions of the band by

$$\Omega_i = i\epsilon_{ijk}\frac{1}{2}\left(\left\langle \frac{\partial u}{\partial k_j}\middle| \frac{\partial u}{\partial k_k}\right\rangle - \left\langle \frac{\partial u}{\partial k_k}\middle| \frac{\partial u}{\partial k_j}\right\rangle\right).$$

The only property of $\mathbf{\Omega}(\mathbf{k})$ needed for the present problem, however, is that $\mathrm{div}_{\mathbf{k}}\mathbf{\Omega} = 0$.

a) Show that these equations are Hamiltonian, with

$$H(\mathbf{r}, \mathbf{k}) = \epsilon(\mathbf{k}) + V(\mathbf{r})$$

and with

$$\omega = dk_i dx_i - \frac{e}{2}\epsilon_{ijk}B_i(\mathbf{r})dx_j dx_k + \frac{1}{2}\epsilon_{ijk}\Omega_i(\mathbf{k})dk_j dk_k.$$

as the symplectic form.[9]

b) Confirm that the ω defined in part b) is closed, and that the Poisson brackets are given by

$$\begin{aligned}\{x_i, x_j\} &= -\frac{\epsilon_{ijk}\Omega_k}{(1 + e\mathbf{B}\cdot\mathbf{\Omega})},\\ \{x_i, k_j\} &= -\frac{\delta_{ij} + eB_i\Omega_j}{(1 + e\mathbf{B}\cdot\mathbf{\Omega})},\\ \{k_i, k_j\} &= \frac{\epsilon_{ijk}eB_k}{(1 + e\mathbf{B}\cdot\mathbf{\Omega})}.\end{aligned}$$

c) Show that the conserved phase-space volume $\omega^3/3!$ is equal to

$$(1 + e\mathbf{B}\cdot\mathbf{\Omega})d^3k d^3x,$$

instead of the naïvely expected $d^3k d^3x$.

[8] M. C. Chang, Q. Niu, *Phys. Rev. Lett.* **75** (1995) 1348.

[9] C. Duval, Z. Horváth, P. A. Horváthy, L. Martina, P. C. Stichel, *Modern Physics Letters B* **20** (2006) 373.

The following pair of exercises show that Cartan's expression for the curvature tensor remains valid for covariant differentiation in "internal" spaces. There is, however, no natural concept analogous to the torsion tensor for internal spaces.

Exercise 11.19: *Non-abelian gauge fields as matrix-valued forms.* In a non-abelian Yang-Mills gauge theory, such as QCD, the vector potential

$$A = A_\mu dx^\mu$$

is matrix-valued, meaning that the components A_μ are matrices which do not necessarily commute with each other. (These matrices are elements of the Lie algebra of the gauge group, but we won't need this fact here.) The matrix-valued curvature, or field-strength, 2-form F is defined by

$$F = dA + A^2 = \frac{1}{2} F_{\mu\nu} dx^\mu dx^\nu.$$

Here a combined matrix and wedge product is to be understood:

$$(A^2)^a{}_b \equiv A^a{}_c \wedge A^c{}_b = A^a{}_{c\mu} A^c{}_{b\nu}\, dx^\mu dx^\nu.$$

i) Show that $A^2 = \frac{1}{2}[A_\mu, A_\nu] dx^\mu dx^\nu$, and hence show that

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu].$$

ii) Define the *gauge-covariant derivatives*

$$\nabla_\mu = \partial_\mu + A_\mu,$$

and show that the commutator $[\nabla_\mu, \nabla_\nu]$ of two of these is equal to $F_{\mu\nu}$. Show further that if X, Y are two vector fields with Lie bracket $[X, Y]$ and $\nabla_X \equiv X^\mu \nabla_\mu$, then

$$F(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}.$$

iii) Show that F obeys the Bianchi identity

$$dF - FA + AF = 0.$$

Again wedge and matrix products are to be understood. This equation is the non-abelian version of the source-free Maxwell equation $dF = 0$.

iv) Show that, in any number of dimensions, the Bianchi identity implies that the 4-form $\mathrm{tr}\,(F^2)$ is closed, *i.e.* that $d\,\mathrm{tr}\,(F^2) = 0$. Similarly show that the $2n$-form $\mathrm{tr}\,(F^n)$ is closed. (Here the "tr" means a trace over the roman matrix indices, and not over the greek space-time indices.)

v) Show that,

$$\mathrm{tr}\,(F^2) = d\left\{\mathrm{tr}\,\left(AdA + \frac{2}{3}A^3\right)\right\}.$$

The 3-form $\mathrm{tr}\,(AdA + \frac{2}{3}A^3)$ is called a *Chern-Simons* form.

Exercise 11.20: *Gauge transformations*. Here we consider how the matrix-valued vector potential transforms when we make a change of gauge. In other words, we seek the non-abelian version of $A_\mu \to A_\mu + \partial_\mu \phi$.

i) Let g be an invertable matrix, and δg a matrix describing a small change in g. Show that the corresponding change in the inverse matrix is given by $\delta(g^{-1}) = -g^{-1}(\delta g)g^{-1}$.

ii) Show that under the *gauge transformation*

$$A \to A^g \equiv g^{-1}Ag + g^{-1}dg,$$

we have $F \to g^{-1}Fg$. (Hint: The labour is minimized by exploiting the covariant derivative identity in part ii) of the previous exercise).

iii) Deduce that $\mathrm{tr}\,(F^n)$ is *gauge invariant*.

iv) Show that a necessary condition for the matrix-valued gauge field A to be "pure gauge", *i.e.* for there to be a position dependent matrix $g(x)$ such that $A = g^{-1}dg$, is that $F = 0$, where F is the curvature two-form of the previous exercise. (If we are working in a simply connected region, then $F = 0$ is also a *sufficient* condition for there to be a g such that $A = g^{-1}dg$, but this is a little harder to prove.)

In a gauge theory based on a Lie group G, the matrices g will be elements of the group, or, more generally, they will form a matrix representation of the group.