Monopoles and the mathematics of gauge theories

Jack M^{c}Kenna

March 24, 2016

Abstract

We model the Dirac monopole to address the asymmetry of Maxwell's equations. Following observations on the gauge-invariance of Maxwell's electromagnetism and its role in the U(1) gauge symmetry of a field theory, we develop the theory of fibre bundles and connections on principal bundles to better understand gauge fields, and observe the geometrical role of the gauge potential and gauge field in the connection and curvature on a principal bundle. We return to the Dirac monopole with a more sophisticated model, and generalise to non-Abelian gauge theories. After investigating spontaneous symmetry breaking, Goldstone's theorem and the Brout-Englert-Higgs mechanism, we investigate the 't Hooft-Polyakov monopole. We observe that its magnetic charge is topological in origin. We relate this to the Dirac monopole by a gauge transform which moves the monopole source to the Higgs field. Finally, we derive the Bogomol'nyi bound and the Bogomol'nyi equation, and observe the Prasad-Sommerfeld solution in a particular case.

Contents

Chapter 1

Introduction to gauges, fields and monopoles in electromagnetism

We begin by observing some features of electromagnetism, both of the electromagnetic field itself and of charged particles which feel it, that hint at a deeper geometrical structure to the theory. We develop a naïve model of the Dirac monopole, and examine how the use of gauges on different patches of space might help remove its problematic singularity. We mention the physical necessity of implementing field theories, and the role of electromagnetism in a field theory with a simple gauge symmetry. We will later develop the theory of fibre bundles and connections on principal bundles, which will allow us to properly investigate electromagnetism as a fully-fledged gauge theory. The tools that we develop will naturally be general enough to apply to other gauge theories that play a central role in modern theoretical physics. As we develop our geometric theory we will consistently return to the Dirac monopole in increasing sophistication and develop further models from it.

Unless otherwise specified, we use units wherein $c = 1$ $(= \mu_0 = \epsilon_0)$ throughout. We will take $\hbar = 1$ after this first chapter.

1.1 Maxwell's equations and gauge invariance

The starting point of the modern study of electromagnetism is, of course, Maxwell's equations; experimentally observed laws describing the electric and magnetic fields. Including the effects of charged particles, they are:

$$\nabla \cdot \mathbf{B} = 0 \tag{1.1}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \tag{1.2}$$

$$\nabla \cdot \mathbf{E} = \rho \tag{1.3}$$

$$\frac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{B} = -\mathbf{j} \tag{1.4}$$

In a vacuum, the current terms ρ and $\mathbf{j}$ are set to 0. Basic vector calculus then tells us that we can obtain two potentials, a vector potential $\mathbf{A}$ for the magnetic field and an associated scalar potential ϕ for the electric field, satisfying

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad \mathbf{E} = \frac{\partial \mathbf{A}}{\partial t} - \nabla \phi \tag{1.5}$$

It is readily observed that the system is unchanged under a *gauge transformation*

$$\mathbf{A} \mapsto \mathbf{A} - \nabla \chi, \quad \phi \mapsto \phi + \frac{\partial \chi}{\partial t} \tag{1.6}$$

Maxwell's equations are not Galilean invariant, but they are Lorentz invariant. In view of this, we can recast them in the language of 4-vectors (tensors) whereby they take on an even more elegant form. We work in Minkowski space ($\eta_{\alpha\beta} = \mathrm{diag}(1, -1, -1, -1)$) and define the 4-potential to be $A_\mu = (\phi, \mathbf{A})$. Now the gauge transformation takes the form

$$A_\mu \mapsto A_\mu + \partial_\mu \chi \tag{1.7}$$

We define the *field strength tensor*, which captures all components of the electric and magnetic fields.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix} \tag{1.8}$$

Note in particular that, owing to its symmetry considerations, the field strength tensor is invariant under change of gauge!

Maxwell's equations are reduced to two properties of the field strength tensor. The homogeneous equations 1.1 and 1.2 become

$$\partial_\xi F_{\mu\nu} + \partial_\mu F_{\nu\xi} + \partial_\nu F_{\xi\mu} = 0 \tag{1.9}$$

which is now simply a statement of a symmetry property of the field strength. The inhomogeneous equations 1.3 and 1.4 become

$$\partial_\nu F^{\mu\nu} = j^\mu \tag{1.10}$$

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

From this standpoint, we may consider applying the Lagrangian formulation of mechanics to the electromagnetic field. We find that a suitable Lagrangian is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + A_\mu j^\mu \tag{1.11}$$

Varying the resulting action with respect to A_μ reproduces the dynamical equation 1.10. In the absence of matter, a gauge transform leaves the action unchanged.

1.2 Magnetic monopoles

Despite the notable beauty and utility of Maxwell's equations, there is an asymmetry in them: $\nabla \cdot \mathbf{B} = 0$ implies that there are no isolated magnetic charges. Dirac [2] introduced a point magnetic charge, the *magnetic monopole*.

1.2.1 The Dirac monopole

We consider a monopole of strength g sitting at the origin, modifying 1.1 to

$$\nabla \cdot \mathbf{B} = 4\pi g\, \delta^3(\mathbf{r}) \tag{1.12}$$

Potential theory tells us that the magnetic field generated is

$$\mathbf{B} = g\mathbf{r}/r^3 \tag{1.13}$$

Now the flux Φ obtained by integrating $\mathbf{B}$ over a sphere S (of radius R) is

$$\Phi = \int_S \mathbf{B} \cdot \mathrm{d}\mathbf{S} = 4\pi g \tag{1.14}$$

From hereon in, however, the monopole is less well-behaved, particularly if we try to find its vector potential. We might try a potential $\mathbf{A}^N$ defined by

$$A_x^N = \frac{-gy}{r(r+z)} \quad A_y^N = \frac{gx}{r(r+z)} \quad A_z^N = 0 \tag{1.15}$$

This potential satisfies

$$\nabla \times \mathbf{A}^N = g\mathbf{r}/r^3 + 4\pi g \delta(x)\delta(y)\theta(-z) \tag{1.16}$$

Hence $\nabla \times \mathbf{A}^N = \mathbf{B}$ except on the negative z-axis – but in particular, it's a suitable potential in the upper half of the universe (excepting the origin). The singularity along the z-axis is is called the *Dirac string* and is a consequence of the (poor) geometry we have imposed.

We further define the potential $\mathbf{A}^S$ by

$$A_x^S = \frac{gy}{r(r-z)} \quad A_y^S = \frac{-gx}{r(r-z)} \quad A_z^S = 0 \tag{1.17}$$

and now have $\nabla \times \mathbf{A}^S = \mathbf{B}$ except along the positive z-axis. We note that our choice of gauge moves the Dirac string around. This is the best we could hope for. The presence of a singularity is an essential requirement of our non-zero flux 1.14, because if there were a single vector field $\mathbf{A}$ satisfying $\mathbf{B} = \nabla \times \mathbf{A}$ without a singularity, by Stokes' theorem the flux would be zero. We can only avoid this problem by abandoning the notion of a *globally* defined vector potential.

1.2.2 The Wu-Yang monopole

Wu and Yang [10] observed that following this treatment, the structure of the Dirac monopole is well modelled by the theory of *fibre bundles*. We will develop this theory and demonstrate its application to the magnetic monopole, but we can outline the Wu-Yang monopole without it. Essentially, we employ two vector potentials to describe a monopole in a way which (hopefully) eliminates the singularity. For example, we adopt $\mathbf{A}^N$ in the northern hemisphere and $\mathbf{A}^S$ in the southern hemisphere of a sphere surrounding the monopole. Together, these vector potentials give the field $\mathbf{B} = g\mathbf{r}/r^3$, non-singular everywhere on the sphere. On the sphere's equator, i.e. the boundary between the hemispheres, our potentials should be related by a gauge transformation, $\mathbf{A}^N - \mathbf{A}^S = \nabla\Lambda$. Converting to polar co-ordinates (r, θ, ϕ) we see that

$$\mathbf{A}^N - \mathbf{A}^S = \frac{2g}{r \sin\theta} \hat{\mathbf{e}}_\phi = \nabla(2g\phi) \tag{1.18}$$

Our gauge transformation connecting $\mathbf{A}^N$ and $\mathbf{A}^S$ is simply

$$\Lambda = 2g\phi \tag{1.19}$$

Λ is not well-defined at $\theta = 0, \pi$ but we only perform the gauge transform in a neighbourhood of $\theta = \pi/2$ so the singularities are happily absent from the theory.

The total flux is now

$$\Phi = \int_S \nabla \times \mathbf{A} \cdot \mathrm{d}\mathbf{S} = \int_{U_N} \nabla \times \mathbf{A}^N \cdot \mathrm{d}\mathbf{S} + \int_{U_S} \nabla \times \mathbf{A}^S \cdot \mathrm{d}\mathbf{S} \tag{1.20}$$

where U_N, U_S are the northern and southern hemispheres, respectively. By Stokes' theorem,

$$\begin{aligned}\Phi &= \int_{\text{equator}} \mathbf{A}^N \cdot \mathrm{d}\mathbf{s} - \int_{\text{equator}} \mathbf{A}^S \cdot \mathrm{d}\mathbf{s} = \int_{\text{equator}} (\mathbf{A}^N - \mathbf{A}^S) \cdot \mathrm{d}\mathbf{s} \\ &= \int_{\text{equator}} \nabla(2g\phi) \cdot \mathrm{d}\mathbf{s} = 4g\pi\end{aligned}$$

which agrees with our result 1.14 for the Dirac monopole. This perspective of using changes of gauge to "patch together" local regions of space(time) will be fundamental to our treatment of gauge theories.

1.2.3 Charge quantisation

We will not, in general, be considering quantum systems, but we note that charge quantisation is apparent already if we quantise the mechanics of a charged particle in a monopole field. Charged particles in an electromagnetic field are experimentally observed to obey the *Lorenz force law*, $\mathbf{F} = q(\mathbf{E} + \dot{\mathbf{x}} \times \mathbf{B})$. A suitable (non-relativistic) Lagrangian is

$$\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m \dot{\mathbf{x}}^2 + q(\mathbf{A} \cdot \dot{\mathbf{x}} - \phi) \tag{1.21}$$

The Euler-Lagrange equation then reproduces the Lorenz force law. It is readily seen that a change of gauge introduces only a total derivative to the Lagrangian so the action is left unchanged.

Switching to the Hamiltonian formalism and quantising, we obtain an operator strongly reminiscent of the Schrödinger operator:

$$\hat{\mathcal{H}} = \frac{1}{2m}(-i\hbar\nabla - q\mathbf{A})^2 + q\phi \tag{1.22}$$

This motivates the definition of what is in the literature called the *covariant derivative*, $\mathcal{D}$:

$$\mathcal{D} = -i\hbar\nabla - q\mathbf{A} \tag{1.23}$$

We will encounter the covariant derivative again and again, and will investigate it rigorously. Though here we have introduced it for what may seem like solely physical reasons, its role is rooted in the geometry that we are posing by way of gauge invariance.

Now consider a point particle of electric charge e and mass m moving in the field of a magnetic monopole of charge g. The Schrödinger equation of the particle is

$$\frac{1}{2m}\mathcal{D}^2\psi(\mathbf{x}) = E\psi(\mathbf{x}) \tag{1.24}$$

Under a gauge transform $\mathbf{A} \mapsto \mathbf{A} + \nabla\Lambda$, the wavefunction transforms as $\psi \mapsto \exp(\mathrm{i}e\Lambda/\hbar)\psi$ – an early hint of U(1) playing a role as a symmetry group. We use two gauges on U^N and U^S related by $\mathbf{A}^N - \mathbf{A}^S = \nabla(2g\phi)$. Hence, our wavefunctions on these hemispheres ψ^N and ψ^S respectively are related by the phase change

$$\psi^N(\mathbf{x}) = \exp\left(\frac{\mathrm{i}e\Lambda}{\hbar}\right)\psi^S(\mathbf{x}) \tag{1.25}$$

Around the equator $\theta = \pi/2$ we can consider the behaviour of the wavefunction as ϕ goes from 0 to 2π. We certainly want the wavefunction to be single-valued, which imposes the *Dirac quantisation condition* for the magnetic charge:

$$g = \frac{\hbar n}{2e}, \quad n \in \mathbb{Z} \tag{1.26}$$

This remarkable result tells us that if the magnetic monopole exists, the magnetic charge takes on discrete values – and similarly, the existence of a single monopole anywhere in the universe not only explains but enforces quantisation of electric charge.

1.3 Symmetries and gauge fields in classical field theory

The theory of fields is introduced to physics in order to mediate "action at a distance". Owing to relativistic causality, we know that some change in a physical system can not realistically affect all spatially separated points of the system "at once"; the change must propagate from its source at the speed of causality. We assume, as implicit in section 1.1, that the reader is already familiar with the notion of the electromagnetic fields (or field) as functions on spacetime. Changes in the field propagate as EM waves at speed c. Further, conservation laws become local in field theories: charge cannot disappear at one location and reappear at another without a current propagating through the field in between.

However, it is not *a priori* apparent why the electromagnetic field has a special role in gauge theories. In this section we quickly recap classical field theories, and how the notion of a local or gauge symmetry necessitates the introduction of Maxwell's electromagnetism. Full details of the principle of stationary action for fields can be found in Ryder [7].

1.3.1 The Lagrangian formalism and Noether's theorem

The Lagrangian formalism is well known for particle mechanics and can be extended to field theories in a straightforward way. We begin with a real scalar field, $\phi(x^\mu)$. We will apply a variational principle to an action of the form:

$$\mathcal{S}[\phi, x^\mu] = \int \mathcal{L}(\phi, \partial_\mu\phi, x^\mu)\,\mathrm{d}^4x. \tag{1.27}$$

Here $\mathcal{L}$ is the *Lagrangian density.* Properly, the Lagrangian associated with it is

$$L = \int \mathcal{L}\,\mathrm{d}^3x \tag{1.28}$$

but it is common to refer to $\mathcal{L}$ simply as the Lagrangian. It is typically assumed that $\mathcal{L}$ depends only on ϕ and its first derivatives, for simplicity. We are assuming here for completeness that the Lagrangian may also depend on the co-ordinate x^μ, which is sometimes the case in interacting systems, though this will not be true for the systems we consider. For example, in relativistic field theories we often expect that a scalar field should obey the *Klein-Gordon equation*,

$$(\Box + m^2)\phi = 0, \tag{1.29}$$

and a suitable Lagrangian is

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{m^2}{2}\phi^2. \tag{1.30}$$

It is a fact that we will take on trust from quantum field theory that, upon quantisation of a field theory, quadratic terms in the Lagrangian ($\sim -m^2\phi^2$) will yield particles of mass m. Henceforth, and in particular later when we consider the Brout-Englert-Higgs mechanism, we will simply read off the mass of a classical field by looking at or for such quadratic terms.

By varying the action 1.27 we obtain the condition for stationary action:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu}\left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right] = 0 \tag{1.31}$$

This is the *Euler-Lagrange equation* for ϕ, notably similar to its equivalent in particle mechanics.

Noether's theorem also applies to the field theory: if the action is invariant under some group of transformations on ϕ and x^μ, then there exist (one or more) conserved quantities — combinations of the fields and their derivatives invariant under the transform. Returning to the variation in the action, we may rewrite the surface term in particular to obtain

$$\begin{aligned} \delta\mathcal{S} = &\int_R \left\{\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu\left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right]\right\} \delta\phi \, \mathrm{d}^4x \\ &+ \int_{\partial R} \left\{\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}[\delta\phi + (\partial_\nu \phi)\delta x^\mu] - \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\partial_\nu \phi - \delta^\mu_\nu \mathcal{L}\right] \delta x^\nu\right\} \mathrm{d}\sigma_\mu \end{aligned} \tag{1.32}$$

The first square bracket in the surface integral is the total variation $\Delta\phi$. The second defines the *energy-momentum tensor* for the field:

$$\theta^\mu_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\partial_\nu \phi - \delta^\mu_\nu \mathcal{L}. \tag{1.33}$$

So to rewrite 1.32,

$$\delta\mathcal{S} = \int_R \left\{\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu\left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right]\right\} \delta\phi \, \mathrm{d}^4x + \int_{\partial R} \left\{\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\Delta\phi - \theta^\mu_\nu \, \delta x^\nu\right\} \mathrm{d}\sigma_\mu \tag{1.34}$$

Suppose the action is invariant under a group of transformations on x^μ and ϕ, which for infinitesimal transformations take the form

$$\Delta x^\mu = X^\mu_\nu \delta\omega^\nu, \quad \Delta\phi = \Phi_\mu \delta\omega^\mu \tag{1.35}$$

characterised by the infinitesimal parameter $\delta\omega^\mu$. Our symmetry groups will always be Lie groups, and here the X^μ_ν are the generators of the corresponding Lie algebra; as written here, this is a four-parameter group of transformations but we could extend the expression to encompass more indices, or a multiplet of scalar fields.

Assuming that the transformed ϕ still obeys the Euler-Lagrange equation 1.31, the principle of stationary action enforces

$$\int_{\partial R} \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\Phi_\nu - \theta^\mu_\sigma X^\sigma_\nu\right] \delta\omega^\nu \, \mathrm{d}\sigma_\mu = 0$$
$$\Rightarrow \int_{\partial R} J^\mu_\nu \, \mathrm{d}\sigma_\mu := \int_{\partial R} \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\Phi_\nu - \theta^\mu_\sigma X^\sigma_\nu\right] \mathrm{d}\sigma_\mu = 0$$

By Stokes' theorem,

$$\int_R \partial_\mu J^\mu_\nu \, \mathrm{d}^4 x = 0 \tag{1.36}$$

Since R is arbitrary, we have our conservation equation,

$$\partial_\mu J^\mu_\nu = 0. \tag{1.37}$$

Hence, J^μ_ν is a conserved (i.e. divergence-free) Noether current. There is an associated conserved (i.e. time-independent) Noether charge:

$$Q_\nu = \int_\sigma J^\mu_\nu \, \mathrm{d}\sigma_\mu \tag{1.38}$$

where σ_μ is a spacelike hypersurface. If we choose this to be $t = \text{constant}$, we have

$$Q_\nu = \int_V J^0_\nu \, \mathrm{d}^3 x \tag{1.39}$$

integrating over the 3-volume V. We observe conservation of Q_ν by integrating over V:

$$\int_V \partial_0 J^0_\nu \, \mathrm{d}^3 x + \int_V \partial_i J^i_\nu \, \mathrm{d}^3 x = 0 \tag{1.40}$$

The second term becomes a surface integral by Stokes' theorem, which vanishes as we takes the boundary surface tending to infinity, leaving

$$\frac{\mathrm{d}}{\mathrm{d}t} \int J^0_\nu \, \mathrm{d}^3 x = \frac{\mathrm{d}Q_\nu}{\mathrm{d}t} = 0 \tag{1.41}$$

which is Noether's theorem in the field theory.

Typical symmetries in the co-ordinate are symmetries of Minkowski space: translations in either space or time, and rotations (i.e. both spatial rotations and Lorentz boosts). In particular, if we consider spacetime translations, we have $\Phi_\mu = 0$ and $X^\mu_\nu = \delta^\mu_\nu$, so

$$J^\mu_\nu = -\theta^\mu_\nu. \tag{1.42}$$

We define the conserved *4-momentum* to be

$$P_\nu = \int_V \theta^0_\nu \, \mathrm{d}^3 x. \tag{1.43}$$

Its time component is

$$P_0 = \int_V \theta^0_0 \, \mathrm{d}^3 x = \int_V \left[\frac{\partial \mathcal{L}}{\partial \dot{\phi}}\dot{\phi} - \mathcal{L}\right] \mathrm{d}^3 x. \tag{1.44}$$

By analogy with the Hamiltonian formalism, we consider this to be the energy of the field, and define the *Hamiltonian density* of the field to be

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} \dot{\phi} - \mathcal{L}. \tag{1.45}$$

The spacetime translation and rotation symmetries are the maximal symmetries of Minkowski space, and hence by considering these we obtain all the Noether charges owing to symmetries of the co-ordinate space: the 4-momentum as above, and an analogous angular momentum tensor by considering rotations. Any further conserved quantities that we know exist — such as electric charge — must be due to symmetries of the field, rather than of the co-ordinate space. Hence, the scalar field must have more than one component, and to consider it further we will promote it from having one component to two.

1.3.2 The complex scalar field and the gauge field

When the scalar field has two components ϕ_1, ϕ_2, we may equally consider it to be a complex scalar field by defining

$$\begin{aligned} \phi &= (\phi_1 + \imath\phi_2)/\sqrt{2} \\ \phi^* &= (\phi_1 - \imath\phi_2)/\sqrt{2} \end{aligned}$$

where we regard ϕ and ϕ^* as independent fields. Since the action should be real, we take the Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)(\partial^\mu \phi^*) - m^2 \phi^* \phi. \tag{1.46}$$

The Euler-Lagrange equations produce two Klein-Gordon equations:

$$\begin{aligned} (\Box + m^2)\phi &= 0 \\ (\Box + m^2)\phi^* &= 0 \end{aligned}$$

The Lagrangian is clearly invariant under a U(1) symmetry, i.e. explicitly under a transformation

$$\phi \mapsto \mathrm{e}^{-\mathrm{i}\Lambda}\phi, \quad \phi^* \mapsto \mathrm{e}^{\mathrm{i}\Lambda}\phi^* \tag{1.47}$$

where here Λ is a real constant. The infinitesimal forms and generators are easily obtained:

$$\begin{aligned} \delta\phi &= -\mathrm{i}\Lambda\phi, & \delta\phi^* &= \mathrm{i}\Lambda\phi^* \\ \Rightarrow \delta(\partial_\mu\phi) &= -\mathrm{i}\Lambda\partial_\mu\phi, & \delta(\partial_\mu\phi^*) &= \mathrm{i}\Lambda\partial_\mu\phi^* \end{aligned}$$

$$\Rightarrow \Phi = -\mathrm{i}\phi, \quad \Phi^* = \mathrm{i}\phi^* \quad (X = 0)$$

The conserved current is

$$\begin{aligned} J^\mu &= \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}(-\mathrm{i}\phi) + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^*)}(\mathrm{i}\phi^*) \\ &= \mathrm{i}(\phi^* \partial^\mu \phi - \phi \partial^\mu \phi^*) \end{aligned} \tag{1.48}$$

and $\partial_\mu J^\mu = 0$. The conserved charge is

$$Q = \int J^0 \, \mathrm{d}V = \mathrm{i} \int \left(\phi^* \frac{\partial \phi}{\partial t} - \phi \frac{\partial \phi^*}{\partial t} \right) \mathrm{d}V. \tag{1.49}$$

We'd like the physical meaning of this real quantity to be the electric charge, but there are several issues with this interpretation. There is no mention of the electron charge e; this is clearly not a quantum property of the field theory as there is no mention of $\hbar$, and indeed the charge is not quantised. Finally, it's clear that the charge and current vanish when $\phi = \phi^*$, i.e. in the case that ϕ is real.

We note also that in deriving these quantities, we considered symmetry under a global transformation. However, as observed earlier, we cannot physically perform this internal rotation equally at all points in spacetime, as this would violate causality. We abandon the requirement that Λ is constant, and "promote" the symmetry to be a local or *gauge* symmetry by now taking $\Lambda = \Lambda(x^\mu)$.

Now the infinitesimal change in ϕ is $\delta\phi = -\mathrm{i}\Lambda\phi$, and similarly $\delta\phi^* = \mathrm{i}\Lambda\phi^*$. The change in the derivative $\partial_\mu\phi$ is, to first order,

$$\begin{aligned} \partial_\mu\phi &\mapsto \partial_\mu\phi - \mathrm{i}(\partial_\mu\Lambda)\phi - \mathrm{i}\Lambda(\partial_\mu\phi) \\ \Rightarrow \delta(\partial_\mu\phi) &= -\mathrm{i}(\partial_\mu\Lambda)\phi - \mathrm{i}\Lambda(\partial_\mu\phi) \end{aligned} \tag{1.50}$$

and similarly $\delta(\partial_\mu\phi^*) = \mathrm{i}(\partial_\mu\Lambda)\phi^* + \mathrm{i}\Lambda(\partial_\mu\phi^*)$. There is an extra term in $\partial_\mu\Lambda$ occurring in the transformation of the derivatives of the fields $\partial_\mu\phi$: they no longer transform covariantly with ϕ. Indeed, even the action is no longer invariant! The change in the Lagrangian is

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial\phi^*}\delta\phi^* + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi) + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^*)}\delta(\partial_\mu\phi^*). \tag{1.51}$$

Using the Euler-Lagrange equations, we can reduce this to

$$\delta\mathcal{L} = J^\mu\partial_\mu\Lambda \tag{1.52}$$

where the current J^μ is as given at 1.48. To make the Lagrangian invariant, we must introduce more terms to it. First, we introduce a 4-vector A_μ coupled to the current J^μ, so that an additional term in the Lagrangian is

$$\mathcal{L}_1 = -eJ^\mu A_\mu \tag{1.53}$$

We demand that under gauge transforms, A_μ transforms like

$$A_\mu \mapsto A_\mu + \frac{1}{e}\partial_\mu\Lambda. \tag{1.54}$$

Hence,

$$\delta\mathcal{L}_1 = -e(\delta J^\mu)A_\mu - J^\mu\partial_\mu\Lambda \tag{1.55}$$

The last term will cancel out our original problematic term, but we've introduced yet another non-vanishing term. We note that

$$\delta J^\mu = 2\phi^*\phi\,\partial^\mu\Lambda \tag{1.56}$$

$$\Rightarrow \delta\mathcal{L} + \delta\mathcal{L}_1 = -2eA_\mu(\partial^\mu\Lambda)\phi^*\phi \tag{1.57}$$

To compensate for this, we add a second term to the Lagrangian:

$$\mathcal{L}_2 = e^2A_\mu A^\mu\phi^*\phi \tag{1.58}$$

$$\Rightarrow \delta\mathcal{L}_2 = 2eA_\mu(\partial^\mu\Lambda)\phi^*\phi \tag{1.59}$$

Now, as desired, $\delta\mathcal{L} + \delta\mathcal{L}_1 + \delta\mathcal{L}_3 = 0$, i.e. $\mathcal{L} + \mathcal{L}_1 + \mathcal{L}_2$ is invariant. However, we must presume that this vector A_μ contributes to the Lagrangian by itself. We add a final gauge-invariant term by defining a tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.60}$$

and an associated scalar Lagrangian to be

$$\mathcal{L}_3 = -\tfrac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{1.61}$$

Our final, total Lagrangian for the original scalar field ϕ coupled to the new *gauge field* is:

$$\begin{aligned} \mathcal{L}_{\text{tot}} &= \mathcal{L} + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \\ &= (\partial_\mu \phi + \mathrm{i}eA_\mu \phi)(\partial^\mu \phi^* - \mathrm{i}eA^\mu \phi^*) - m^2 \phi^* \phi - \tfrac{1}{4} F^{\mu\nu} F_{\mu\nu} \end{aligned} \tag{1.62}$$

Remarkably, a structure fundamentally similar to the electromagnetic field has arisen naturally simply by demanding a gauge symmetry in our scalar field. The gauge potential A_μ couples to the current J^μ with coupling strength given by e, which we now happily associate with the charge of the ϕ field. We draw several conclusions.

1. The derivative of the scalar field $\partial_\mu \phi$ is no longer a covariant quantity when we impose a gauge symmetry. However, if (as in the previous chapter) we define the *covariant derivative* operator $\mathcal{D}_\mu$ by

$$\mathcal{D}_\mu \phi = (\partial_\mu + \mathrm{i}eA_\mu)\phi, \tag{1.63}$$

then the covariant derivative of the field $\mathcal{D}_\mu \phi$ is a covariant quantity. Nonetheless, in this sense, "covariance" is a physical notion. If the reader has some experience with differential geometry, they are aware that a covariant derivative is typically defined on a manifold by specifying a *connection*: a means of comparing vectors in different tangent spaces and hence defining parallel transport. But is is not intrinsic to the manifold; we are (mostly[1]) free to specify the connection. Here, we see that the covariant derivative is affected by our vector potential, and hence by our choice of gauge. When we look at connections on principal bundles, we will find an explanation of the geometrical role that the gauge potentials play in our theory.

2. In fact, this vector potential and field-strength tensor are not just *similar* to the electromagnetism of Maxwell's theory, but are precisely equivalent to it. By our definition of $F_{\mu\nu}$, the homogeneous Maxwell equations will necessarily hold when we identify its components with the electric and magnetic fields in the standard way. Electromagnetism as a gauge field is a required element of the theory in order to guarantee local U(1) invariance. Note that since we have an action already specified, by varying A_μ we can obtain the inhomogeneous Maxwell equations. The Euler-Lagrange equations for this variational principle are

$$\frac{\partial \mathcal{L}}{\partial A_\mu} - \frac{\partial}{\partial x^\nu}\left(\frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)}\right) = 0. \tag{1.64}$$

The dynamical equation that we obtain is

$$\partial_\nu F^{\mu\nu} = -e\mathcal{J}^\mu \tag{1.65}$$

where the *covariant current* is

$$\mathcal{J}^\mu = \mathrm{i}(\phi^* \mathcal{D}^\mu \phi - \phi \mathcal{D}^\mu \phi^*). \tag{1.66}$$

[1] Though other properties of the manifold, such as a metric, may naturally influence our choice of connection.

The antisymmetry of $F^{\mu\nu}$ shows that $\partial_\mu \mathcal{J}^\mu = 0$, so once we have introduced the electromagnetic field, it is the covariant current which plays the role of the Noether current rather than the previously derived expression for the current J^μ.

3. We may observe that the gauge field that we have derived is massless: clearly there is no term in the Lagrangian like $\mathcal{L}_\mathrm{M} = -M^2 A_\mu A^\mu$. Indeed, such a term would not be gauge-invariant. We conclude further that gauge invariance requires a massless gauge field.

4. The electron charge e appears as a coupling constant, demonstrating the dual role of electric charge: it is both a conserved quantity and the strength with which a particle interacts with the electromagnetic field. This dynamical aspect of charge is a consequence of the gauge principle that the symmetry is local.

In the following chapters, we will investigate the geometrical nature of the gauge field by outlining the theory of fibre bundles and connections on principal bundles, where we will identity the connection with the gauge potential and the associated curvature with the field strength. We will generalise to larger gauge symmetries (e.g. SU(2) in Yang-Mills theory) and examine the role of the associated massless gauge fields coupling to particles. Finally, we will study the implications of spontaneous breaking of such gauge symmetries to sophisticated models of monopoles such as the 't Hooft-Polyakov monopole and the Bogomol'nyi bound.

Chapter 2

Fibre bundles

We develop the theory of fibre bundles in this and the following chapter, in particular looking at vector bundles and sections of vector bundles. We examine principal bundles and the association between a vector bundle and a principal bundle. We define connections on principal bundles and the notions of curvature, parallel transport and covariant derivatives. We also mention frame bundles to illustrate how our gauge theory exhibits similarity to the connection in general relativity.

We assume the reader is familiar with the basic theory of the calculus on manifolds, connections on a manifold, and Lie groups and Lie algebras. We follow Nakahara's [4] development of the theory, and many examples are done in more depth in his text.

2.1 Motivation and definition

A fibre bundle is a manifold that locally looks like the (Cartesian) product of two manifolds, but globally may exhibit a non-trivial (and hence more interesting!) structure. A cylindrical hairbrush provides an intuitive example of a trivial bundle where the "base space" is the cylinder and the "fibres" are line segments; it is *trivial* in the sense that as a topological space it is simply the formal product space of those two manifolds. The reader has in fact already been introduced to specific examples of fibre bundles: namely, the tangent bundle (and cotangent bundle) on a manifold. A perspective on fibre bundles is that we wish to generalise the notion of a tangent bundle on a manifold to fibres that are not essentially $\mathbb{R}^n$, and whose elements do not carry that essential structure of tangent vectors as derivations on the manifold; but, in a sense, the broad picture will be the same.

Definition 2.1.1. A *fibre bundle* (E, π, M, F, G) or $\pi : E \to M$ or $\quad E \xrightarrow{\pi} M \quad$ is

1. A differentiable manifold E, the *total space*
2. A differentiable manifold M, the *base space*
3. A differentiable manifold F, the *(typical) fibre*
4. A surjection $\pi : E \to M$, the *projection*, satisfying $\pi^{-1}(p) = F_p \cong F$. F_p is called the fibre at p.
5. A Lie group G, the *structure group*, acting on F from the left

6. A set of open coverings $\{U_i\}$ of M, each equipped with a diffeomorphism $\phi_i : U_i \times F \to \pi^{-1}(U_i)$ such that $\pi\phi_i(p,f) = p$. ϕ or its inverse are typically called the *local trivialisation*, as ϕ^{-1} : $\pi^{-1}(U_i) \to U_i \times F$.

7. Finally, we demand suitable transitions between patches by appropriately defining the *transition function*: writing $\phi_i(p,f) = \phi_{i,p}(f)$, $\phi_{i,p} : F \to F_p$ is itself a diffeomorphism; so on overlapping patches of the bundle $U_i \cap U_j \neq \emptyset$ we require that it is $t_{ij}(p) = \phi_{i,p}^{-1}\phi_{j,p} : F \to F \in G$. (Note indeed it acts on F, as elements of G must.)

(Strictly, the reader may note that this depends on our choice of $\{U_i, \phi_i\}$. We have more accurately defined here a *co-ordinate bundle*, and fibre bundles are properly equivalence classes of co-ordinate bundles. Generally we will be happy to consider a specific set of co-ordinate systems on a given bundle.)

Essentially, given a chart U_i on M, π^{-1} is a direct product diffeomorphic (under ϕ^{-1}) to $U_i \times F$. Note that on overlapping patches $U_i \cap U_j \neq \emptyset$ we have two such maps ϕ_i, ϕ_j, and for $u \in E$ with $\pi(u) = p \in U_i \cap U_j$, there are *two* elements of F assigned to u:

$$\phi_i^{-1}(u) = (p, f_i)$$
$$\phi_j^{-1}(u) = (p, f_j).$$

Then $\exists\ t_{ij} \to U_i \cap U_j \to G$ specifying the transition at p:

$$f_i = t_{ij}(p) f_j. \tag{2.1}$$

We see that the consistency conditions on $t_{ij}(p)$, i.e. the conditions we need to glue the bundle together, are:

$$\begin{aligned} &t_{ii}(p) = \mathrm{id} && p \in U_i \\ &t_{ij}(p) = t_{ji}(p)^{-1} && p \in U_i \cap U_j \\ &t_{ij}(p) t_{jk}(p) = t_{ik}(p) && p \in U_i \cap U_j \cap U_k. \end{aligned}$$

The *trivial bundle* has all $t_{ij}(p) = \mathrm{id}$.

Given $E \xrightarrow{\pi} M$, the set of transition functions is clearly not unique. Let $\{U_i\}$ cover M with $\{\phi_i\}, \{\widetilde{\phi_j}\}$ local trivialisations generating the same bundle. The transition functions are

$$\begin{aligned} t_{ij}(p) &= \phi_{i,p}^{-1}\phi_{j,p} \\ \widetilde{t}_{ij}(p) &= \widetilde{\phi}_{i,p}^{-1}\widetilde{\phi}_{j,p}. \end{aligned} \tag{2.2}$$

Consider $g_i(p) : F \to F$ at each $p \in U_i$ given by

$$g_i(p) = \phi_{i,p}^{-1}\widetilde{\phi}_{i,p}. \tag{2.3}$$

We require $g_i(p)$ to be a homeomorphism in G for these local trivilisations to describe the same fibre bundle. Combining 2.2 and 2.3 we obtain

$$\widetilde{t}_{ij}(p) = g_i(p)^{-1} t_{ij}(p) g_j(p). \tag{2.4}$$

In physical applications, the t_{ij} are realised as the gauge transformations for gluing local patches together; the g_i are the choice of gauge in the patch. On a trivial bundle we may take $t_{ij} = \mathrm{id}$, so a general transition function is $\widetilde{t_{ij}}(p) = g_i(p)^{-1} g_j(p)$.

Definition 2.1.2. A *section* or *cross section* $s : M \to E$ is a smooth map satisfying $\pi s = \mathrm{id}_M$.

Note that $s(p) = s|_p \in F_p = \pi^{-1}(p)$; i.e. a section smoothly applies an element in the fibre to every point on the base space. The set of sections on M is denotes $\Gamma(M, E)$. For $U \subset M$ we may think of *local sections* defined only on U, the set of which is denoted $\Gamma(U, E)$. For example, $\Gamma(M, TM)$ is the set $\mathfrak{X}(M)$ of vector fields on M. Not all fibre bundle admit global sections!

2.1.1 Example: the Möbius strip

The Möbius strip is the unique non-trivial bundle $E \xrightarrow{\pi} S^1$ with fibre $F = [-1, 1]$ — note that strictly F is not a manifold but this suffices as an example. Let $U_1 = (0, 2\pi)$, $U_2 = (-\pi, \pi)$ cover S^1, and let $U_1 \cap U_2 = A \cup B$, where

$$A = (0, \pi) \quad B = (\pi, 2\pi).$$

Without loss of generality, local trivialisations (on A) are

$$\begin{gathered}\phi_1^{-1}(u) = (\theta, t) \quad \phi_2^{-1} = (\theta, t) \\ \theta \in A, \qquad t \in F.\end{gathered}$$

On B, we have (up to equivalence) two choices:

1. $\phi_1^{-1}(u) = (\theta, t), \quad \phi_2^{-1}(u) = (\theta, t);$
2. $\phi_1^{-1}(u) = (\theta, t), \quad \phi_2^{-1}(u) = (\theta, -t).$

In each case, $t_{12}(\theta)|_A : t \mapsto t$, i.e. the identity map. But in case (1), $t_{12}(\theta)|_B : t \mapsto t$ again and E is the trivial bundle, the cylinder $M \times F = S^1 \times [-1, 1]$ – whereas in case (2), $t_{12}(\theta)|_B : t \mapsto -t$ and we obtain the Möbius strip. Hence we see that the cylinder has $G = \{e\}$, while the Möbius strip has $G = \{e, g|g^2 = e\} \cong \mathbb{Z}_2$. (Again, strictly, here G is not a Lie group.)

2.2 Relations between bundles

2.2.1 Reconstructing the fibre bundle

The definition of the fibre bundle is dense, and it is natural to wonder exactly how much of it we need before the fibre bundle is entirely determined. The answer is that given M, $\{U_i\}$, $t_{ij}(p)$, F and G, we can uniquely determine π, E and ϕ_i and reconstruct the bundle.

Define $X = \bigcup_i U_i \times F$ and introduce an equivalence relation $\sim$ between $(p, f) \in U_i \times F$ and $(q, f') \in U_j \times F$ by

$$(p, f) \sim (q, f') \Leftrightarrow p = q, f' = t_{ij}(p)f.$$

Then $E = X/\sim$. If $[(p, f)] \in E$, we define the projection

$$\pi : [(p, f)] \mapsto p,$$

and local trivialisation $\phi_i : U_i \times F \to \pi^{-1}(U_i)$ by

$$\phi_i : (p, f) \mapsto [(p, f)].$$

We may use this to construct new fibre bundles: given (E, π, M, F, G) we can use M, $t_{ij}(p)$, G and a new fibre F' to construct a new bundle $E' \xrightarrow{\pi'} M$.

2.2.2 Bundle maps

Definition 2.2.1. Let $E \xrightarrow{\pi} M$ and $E' \xrightarrow{\pi'} M'$ be fibre bundles. A smooth map $\bar{f} : E' \to E$ is a *bundle map* if it maps each fibre F'_p of E' onto F_q of E. Then $\bar{f}$ naturally induces a smooth map $f : M' \to M$ with $f(p) = q$, and the following diagram commutes:

$$\begin{array}{ccc} E' & \xrightarrow{\bar{f}} & E \\ {\scriptstyle \pi'}\downarrow & & \downarrow{\scriptstyle \pi} \\ M' & \xrightarrow{f} & M \end{array} .$$

Note that not every smooth map $\bar{f} : E' \to E$ is a bundle map: it could map two points in the same fibre in E' to points on different fibres in E.

Earlier we stated that formally fibre bundles are equivalence classes of co-ordinate bundles. This can be made rigorous by saying that two fibres bundles are equivalent if there exists a bundle map with $f = \mathrm{id}_M$ and $\bar{f}$ a diffeomorphism; see Nakahara [4].

2.2.3 Pullback bundles

Let $E \xrightarrow{\pi} M$ be a fibre bundle with typical fibre F. Given a map $f : N \to M$, we can define a new fibre bundle over N also with typical fibre F in a straightforward way.

Definition 2.2.2. The *pullback* of E by f is

$$f^*E = \{(p, u) \in N \times E | f(p) = \pi(u)\}.$$

The fibre of F_p of f^*E is just a copy of $F_{f(p)}$. We define

$$\begin{aligned} f^*E \xrightarrow{\pi_1} N \text{ by } \pi_1 : (p, u) \mapsto p \\ \text{and } \pi_2 : f^*E \to E : (p, u) \mapsto u \quad ; \end{aligned}$$

then f^*E is a fibre bundle with the bundle map

$$\begin{array}{ccc} f^*E & \xrightarrow{\pi_2} & E \\ {\scriptstyle \pi'_1}\downarrow & & \downarrow{\scriptstyle \pi} \\ N & \xrightarrow{f} & M \end{array} .$$

Note that if $N = M$ and $f = \mathrm{id}_M$, f^*E and E are equivalent.

Let $\{U_i\}$ cover M with local trivialisations $\{\phi_i\}$. $\{f^{-1}(U_i)\}$ defines a covering on f^* that is locally trivial, i.e. admitting local trivialisations. Take $u \in E$ with $\pi(u) = f(p) \in U_i$, for some $p \in N$. Then if $\phi_i^{-1}(u) = (f(p), f_I)$, we must have for $\{\psi_i\}$ the local trivialisations of f^*E $\psi_i^{-1}(p, u) = (p, f_i)$. The transition function t_{ij} at $f(p) \in U_i \cap U_j$ maps f_j to $f_i = t_{ij}(f(p))f_j$; the corresponding transition function t^*_{ij} of f^*E at $p \in f^{-1}(U_i) \cap f^{-1}(U_j)$ must also map f_j to $f_i = t^*_{ij}(p)f_j$, i.e.:

$$t^*_{ij}(p) = t_{ij}(f(p)) \tag{2.5}$$

So the transition function of the pullback is the pullback of the transition function.

2.2.4 Homotopy

Let $f, g : M' \to M$. These maps are said to be *homotopic* if there exists a smooth map $F : M' \times [0,1] \to M$ such that $F(p,0) = f(p)$ and $F(p,1) = g(p) \forall p \in M'$.

Theorem 2.2.3 (Homotopy axiom)**.** Let $E \xrightarrow{\pi} M$ be a fibre bundle with typical fibre F, and let f and g be homotopic maps from N to M. Then f^*E and g^*E are equivalent as bundles.

Proof. See Steenrod [8]. □

Let M be a manifold which is contractible to a single point. Then there exists a homotopy $F : M \times [0,1] \to M$ such that

$$F(p,0) = p, \quad F(p,1) = p_0 \text{ for fixed } p_0 \in M.$$

Let $E \xrightarrow{\pi} M$ be a fibre bundle over M and consider pullback bundles h_0^*E, h_1^*E, where $h_t(p) = F(p,t)$. h_1^*E is defined over a single point p_0, so is the trivial bundle; on the other hand, $h_0^*E = E$ as h_0^* is the identity. However, according to the homotopy axiom, $h_1^*E = h_0^*E = E$, so E is a trivial bundle. That is,

Corollary 2.2.4. Let $E \xrightarrow{\pi} M$ be a fibre bundle. E is trivial if M is contractible to a point.

For example, the tangent bundle $T\mathbb{R}^m$ is trivial.

2.3 Vector bundles

2.3.1 Definition and examples

Definition 2.3.1. A *vector bundle* $E \xrightarrow{\pi} M$ is a fibre bundle whose fibre is a vector space.

Let $F = \mathbb{R}^k$ and M be an m-dimensional manifold. k is called the *fibre dimension* and we often abuse notation to write $k = \dim E$, although the total space E is $(m+k)$-dimensional.

The transition functions invertible map between vector spaces, so necessarily

$$t_{ij}(p) \in \mathrm{GL}_k(\mathbb{R}) \quad \forall p \in M. \tag{2.6}$$

(If $F = \mathbb{C}^k$, instead $t_{ij}(p) \in \mathrm{GL}_k(\mathbb{C}) \quad \forall p \in M$.)

Example: Tangent bundles as vector bundles

As noted, the prototype of vector bundles (and fibre bundles in general) is the tangent bundle. For m-dimensional M, TM has typical fibre $\mathbb{R}^m$. Consider $u \in TM$ with $\pi(u) = p \in U_i \cap U_j$, $\{U_i\}$ covering M. Let $x^\mu = \phi_i(p)$ ($y^\mu = \phi_j(p)$) be co-ordinate systems of U_i (U_j). The vector V corresponding to u is expressed

$$V = V^\mu \frac{\partial}{\partial x^\mu}\bigg|_p = \widetilde{V}^\mu . \frac{\partial}{\partial y^\mu}\bigg|_p$$

The local trivialisations are therefore

$$\phi_i^{-1}(u) = (p, \{V^\mu\}), \quad \phi_j^{-1}(u) = (p, \{\widetilde{V}^\mu\}).$$

The fibre co-ordinates are related by

$$V^\mu = G^\mu{}_\nu(p)\widetilde{V}^\nu,$$

where

$$\{G^\mu{}_\nu(p)\} = \left\{ \left(\frac{\partial x^\mu}{\partial y^\nu}\right)_p \right\} \in \mathrm{GL}_m(\mathbb{R}).$$

Hence, a tangent bundle is explicitly $(TM, \pi, M, \mathbb{R}^m, \mathrm{GL}_m(\mathbb{R}))$ and sections of TM are the vector fields on M:

$$\Gamma(M, TM) = \mathfrak{X}(M).$$

2.3.2 Line bundles

A vector bundle whose fibre is one-dimensional ($F = \mathbb{R}$ or $\mathbb{C}$) is a *line bundle*. For example, both the trivial line bundle (infinite cylinder) $S^1 \times \mathbb{R}$ and the infinite Möbius strip[1] are line bundles on S^1. Line bundles have the Abelian structure group $\mathrm{GL}_1(\mathbb{R})$ $(\mathrm{GL}_1(\mathbb{C})) = \mathbb{R}(\mathbb{C}) \setminus \{0\}$.

In physics, the trivial line bundle $L = \mathbb{R}^3 \times \mathbb{C}$ is associated with non-relativistic quantum mechanics on $\mathbb{R}^3$ – a wavefunction ψ is a section of L. When we introduce a monopole at the origin, the wavefunction is defined on $\mathbb{R}^3 \setminus \{0\}$ and we take a complex line bundle on $\mathbb{R}^3 \setminus \{0\}$. If we only look at the wavefunction on a surrounding sphere S^2 we take a complex line bundle over S^2. (Note that S^2 is a deformation retract of $\mathbb{R}^3 \setminus \{0\}$, i.e. there is a homotopy between the identity map on $\mathbb{R}^3 \setminus \{0\}$ and its retraction to the sphere S^2.)

2.3.3 Frames

On a tangent bundle TM, each fibre has a natural basis $\{\partial/\partial x^\mu\}$ endowed by the co-ordinate system x^μ on a chart U_i. If M has a metric we can also use the orthonormal basis $\{\mathbf{e}_a\}$. Each member (of either basis) is a vector field and the sets form linearly independent vector fields over U_i. It is always possible to choose m linearly independent tangent vectors over U_i, but not necessarily throughout M. By definition,

$$\frac{\partial}{\partial x^\mu} = \begin{array}{ccccccc} (0, & \ldots, & 0, & 1, & 0, & \ldots, & 0) \\ & & & \mu & & & \end{array} ; \tag{2.7}$$

$$\mathbf{e}_a = \begin{array}{ccccccc} (0, & \ldots, & 0, & 1, & 0, & \ldots, & 0) \\ & & & \alpha & & & \end{array} \tag{2.8}$$

These define a local *frame* over U_i.

Let $E \xrightarrow{\pi} M$ be a vector bundle with fibre $\mathbb{R}^k$ (or $\mathbb{C}^k$). On a chart U_i, $\pi^{-1}(U_i)$ is trivial:

$$\pi^{-1}(U_i) = U_i \times \mathbb{R}^k,$$

and we can choose k linearly independent sections $\{e_1(p), \ldots, e_k(p)\}$ over U_i defining a *frame* over U_i. Given a frame over U_i we have a natural map:

$$F_p \to F : V = V^\alpha e_\alpha(p) \mapsto \{V^\alpha\}_\alpha \in F. \tag{2.9}$$

The local trivialisation is

$$\phi_i^{-1}(V) = (p, \{V^\alpha\}_\alpha),$$

[1] Allowing $G = \mathbb{Z}_2$ again here.

so by definition we have

$$\phi_i(p, \{0, \quad \ldots, \quad 0, \quad \underset{\alpha}{1}, \quad 0, \quad \ldots, \quad 0\}) = e_\alpha(p).$$

Let $U_i \cap U_j \neq \emptyset$ and consider a change of frame. On U_i we have frame $\{e_1(p), \ldots, e_k(p)\}$ and on U_j we have $\{\tilde{e}_1(p), \ldots, \tilde{e}_k(p)$, $p \in U_i \cap U_j$. A vector $\tilde{e}_\beta(p)$ is expressed as

$$\tilde{e}_\beta(p) = e_\alpha(p) G(p)^\alpha{}_\beta \tag{2.10}$$

for $G(p)^\alpha{}_\beta \in \mathrm{GL}_k(\mathbb{R})$. A vector $V \in \pi^{-1}(p)$ can be expressed

$$V = V^\alpha e_\alpha(p) = \widetilde{V}^\alpha \widetilde{e}_\alpha(p) \tag{2.11}$$

Combining 2.10 and 2.11 we find

$$\widetilde{V}^\beta(p) = G^{-1}(p)^\beta{}_\alpha V^\alpha(p) \tag{2.12}$$

where

$$G^{-1}(p)^\beta{}_\alpha G(p)^\alpha{}_\gamma = G(p)^\beta{}_\alpha G^{-1}(p)^\alpha{}_\gamma = \delta^\beta{}_\gamma \tag{2.13}$$

– that is, the transition function $t_{ij}(p)$ is the matrix $G^{-1}(p)$.

The set of frames at each point form fibres of the *frame bundle*, which we will develop further.

2.3.4 Sections of vector bundles

Let s, s' be sections (as defined at 2.1.2) of a vector bundle $E \xrightarrow{\pi} M$. We define vector addition and scalar multiplication pointwise:

$$(s + s')(p) = s(p) + s'(p), \quad (fs)(p) = f(p)s(p) \tag{2.14}$$

. for $p \in M, f \in \mathfrak{F}(M)$

The null vector is left invariant under $\mathrm{GL}_k(\mathbb{F})$ and plays a distinguished role: any vector bundle E admits a global section, the null section $s_0 \in \Gamma(M, E)$ such that $\phi_i^{-1}(s_0, p) = (p, 0)$ in any local trivialisation.

Vector bundles naturally carry, or can be endowed with, additional structures that we need not discuss: for example, similarly to the cotangent bundle, any vector bundle $E \xrightarrow{\pi} M$ with typical fibre F gives rise to the dual bundle $E' \xrightarrow{\pi} M$ whose typical fibre is F', the dual space to F. Vector bundles can be combined in manners similar to the tensor sum or tensor product in a generally intuitive way. We can extend from the tangent and cotangent bundles in this manner to spaces of forms as sections of the bundle of totally antisymmetric tensors. An example of an additional structure that might be placed on a vector bundle is a fibre metric, an inner product defined pointwise between (possibly local) sections.

2.4 Principal bundles

Definition 2.4.1. A *principal bundle* has typical fibre F identical to its structure group G. We denote a principal bundle by $P \xrightarrow{\pi} M = P(M, G)$, and may call it a *G-bundle (over M)*.

Vector bundles are not the only bundles that carry a nice structure. Another privileged choice of typical fibre is the structure group of the bundle; as a Lie group, G is itself a manifold. We already know that elements of G, as transition functions, act on the fibre from the left, but now we are free to also define another action from the right.

Let $\phi_i : U_i \times G \to \pi^{-1}(U_i)$ be the local trivialisation

$$\phi_i^{-1}(u) = (p, g_i), \quad u \in \pi^{-1}(U_i), \quad p = \pi(u).$$

The right action of G on $\pi^{-1}(U_i)$ is defined by

$$\phi_i^{-1}(ua) = (p, g_i a) \equiv ua = \phi_i(p, g_i a) \quad (\forall a \in G\ \forall u \in \pi^{-1}(U_i)). \tag{2.15}$$

As the left and right actions commute, this definition is independent of the local trivialisation. Moreover, if $p \in U_i \cap U_j$,

$$ua = \phi_j(p, g_j a) = \phi_j(p, t_{ji} g_i a) = \phi_i(p, g_i a).$$

Note that this action $P \times G \to P : (u, a) \mapsto ua$ does not map a point out of its original fibre: $\pi(ua) = \pi(u)$.

The right action is transitive (possessing only a single orbit) as G acts on itself transitively, and $F_p = \pi^{-1}(p)$ is diffeomorphic to G. Hence, if $\pi(u) = p$ the fibre is $\pi^{-1}(p) = \{ua | a \in G\}$. The action is also free: that is, if $ua = u$ for some $u \in P$, $a = e$. Writing $u = \phi_i(p, g_i)$, the condition is $\phi_i(p, g_i a) = \phi_i(p, g_i)a = ua = u = \phi_i(p, g_i)$, but ϕ_i is bijective so $a = e$.

Given a section $s_i(p)$ over U_i we define a preferred local trivialisation $\phi_i : U_i \times G \to \pi^{-1}(U_i)$ as follows. For $u \in \pi^{-1}(U_i)$, $p \in U_i$, take the unique element $g_u \in G$ with $u = s_i(p)g_u$. Then define

$$\phi_i^{-1}(u) = (p, g_u) \tag{2.16}$$

whence

$$s_i(p) = \phi_i(p, e) \text{ in the patch } (U_i, \phi_i). \tag{2.17}$$

This is the *canonical local trivialisation.*

By definition, $\phi_i(p, g) = \phi_i(p, e)g = s_i(p)g$. If $p \in U_i \cap U_j$, sections $s_i(p)$, $s_j(p)$ are related by the transition function $t_{ij}(p)$ under

$$\begin{aligned} s_i(p) &= \phi_i(p, e) = \phi_j(p, t_{ji}(p)e) = \phi_j(p, t_{ji}(p)) \\ &= \phi_j(p, e)t_{ji}(p) = s_j(p)t_{ji}(p). \end{aligned} \tag{2.18}$$

2.4.1 Example: the U(1)-bundle on S^2 and the monopole

The bundle $P(S^2, \mathrm{U}(1))$ is the setting in which we can more formally investigate the magnetic monopole. Let $\{U_N, U_S\}$ be an open cover of S^2 comprising the northern and southern hemispheres respectively. Parametrising by the standard polar angles, we have

$$\begin{aligned} U_N &= \{(\theta, \phi) | 0 \le \theta \le \pi/2 + \epsilon, 0 \le \phi < 2\pi\}, \\ U_S &= \{(\theta, \phi) | \pi/2 - \epsilon \le \theta \le 0, 0 \le \phi < 2\pi\}. \end{aligned}$$

The intersection on which we glue together our bundle, or (as we begin to consider equivalent) perform our gauge transform, is essentially just the equator. Let ϕ_N and ϕ_S be the local trivialisations:

$$\phi_N^{-1}(u) = (p, \exp(\mathrm{i}\alpha_N)), \quad \phi_S^{-1}(u) = (p, \exp(\mathrm{i}\alpha_S)) \tag{2.19}$$

where $p = \pi(u)$. Consider a transition function of the form $\mathrm{e}^{\mathrm{i}n\phi}$, where as we have seen $n \in \mathbb{Z}$ to ensure that $t_{NS}(p)$ remains well-defined all the way around the equator. (As we have a map between the equator S^1 and U(1), the reader might note that this integer $n \in \mathbb{Z}$ characterises the *homotopy group* of U(1).) The fibre co-ordinates α_N and α_S are related on the equator by

$$\mathrm{e}^{\mathrm{i}\alpha_N} = \mathrm{e}^{\mathrm{i}n\phi}\mathrm{e}^{\mathrm{i}\alpha_S}. \tag{2.20}$$

If $n = 0$, the transition function is the unit element of U(1), which is to say that we have a trivial bundle $P = S^2 \times S^1$. When $n \neq 0$, the $U(1)$-bundle P_n is twisted (non-trivial). To quote Nakahara [4], "It is remarkable that the topological structure of a fibre bundle is characterised by an integer".

As we will discover in our further development of gauge theories, a U(1)-bundle is particularly nice because the group is Abelian. Hence, the right action and left action are equivalent. Under the action of $g = \mathrm{e}^{\mathrm{i}\lambda}$, we have

$$\phi_N^{-1}(ug) = (p, \mathrm{e}^{\mathrm{i}(\alpha_N+\lambda)}), \tag{2.21}$$

$$\phi_S^{-1}(ug) = (p, \mathrm{e}^{\mathrm{i}(\alpha_S+\lambda)}). \tag{2.22}$$

The physical perspective, which we hope to generalise, is that this action corresponds to the U(1)-gauge transformation.

2.4.2 Associated bundles

There is a firm relationship between vector bundles and principal bundles, which is an expression of the role of gauge fields in field theories with gauge symmetry. Given a principal bundle $P(M, G)$ we may construct an *associated fibre bundle* as follows. Suppose we have an action of G on a manifold F from the left. Then we can define an action of $g \in G$ on $P \times F$ by

$$(u, f) \mapsto (ug, g^{-1}f) \tag{2.23}$$

where $u \in P$ and $f \in F$. Then the associated fibre bundle (E, π, M, G, F, P) is an equivalence class $P \times F/G$ where we identify points of the form (u, f) and $(ug, g^{-1}f)$.

If we take F to be a k-dimensional vector space V, the action of G is given by a k-dimensional representation ρ. As above, the *associated vector bundle* $P \times_\rho V$ is defined by identifying points (u, v) and $(ug, \rho(g)^{-1}v)$ of $P \times V$. Of course, we must check that this is still a fibre bundle. The fibre bundle structure is as follows: the projection $\pi : E \to M$ is defined by $\pi_E(u, v) = \pi(u)$, which is well-defined since $\pi(u) = \pi(ug)$ implies $\pi_E(ug, \rho(g)^{-1}v) = \pi_E(u, v)$. The local trivialisation is $\psi_i : U_i \times V \to \pi_E^{-1}(U_i)$ and the transition function is $\rho(t_{ij}(p)$ where $t_{ij}(p)$ is that of P.

Conversely, a vector bundle naturally induces a principal bundle associated with it. For $E \xrightarrow{\pi} M$ a vector bundle with $\dim E = k$, the *associated principal bundle* $P(E) = P(M, G)$ is the principal bundle over M with the same transition functions, where the structure group (and now fibre) is $\mathrm{GL}_k(\mathbb{F})$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, where $V = \mathbb{F}^k$). Recall that we have a prescription for reconstructing the bundle in 2.2.1.

Example: the frame bundle

For an m-dimensional manifold M, the principal bundle associated to the tangent bundle TM is the *frame bundle*

$$LM \equiv \bigcup_{p \in M} L_p M \tag{2.24}$$

where L_pM is the set of frames at p. Given co-ordinates x^μ on a chart U_i, T_pM has the natural basis $\{\partial/\partial x^\mu\}$ on U_i. As defined in 2.3.3, a frame $u = \{X_1, \ldots, X_m\}$ at p is expressed

$$X_\alpha = X^\mu{}_\alpha \left.\frac{\partial}{\partial x^\mu}\right|_p \quad 1 \leq \alpha \leq m \tag{2.25}$$

where $(X^\mu{}_\alpha)$ is an element of $\mathrm{GL}_m(\mathbb{R})$ such that the $\{X_\alpha\}$ form a linearly independent set. We define the local trivialisation $\phi_i : U_i \times \mathrm{GL}_m(\mathbb{R}) \to \pi^{-1}(U_i)$ by $\phi_i^{-1}(u) = (p, (X^\mu{}_\alpha))$. The bundle structure is defined in the following steps:

1. Projection: For a frame $u = (X_1, \ldots, X_m)$ at p, the projection is defined $\pi_L : LM \to M : u \mapsto p$.
2. Action: An element $a = (a^i{}_j) \in \mathrm{GL}_m(\mathbb{R})$ acts on the frame u by $(u, a) \mapsto ua$, where ua is a frame at p defined by
$$Y_\beta = X_\alpha a^\alpha{}_\beta. \tag{2.26}$$
Conversely, given two frames $\{X_\alpha\}$ and $\{Y_\beta\}$ at p there exists an element of $\mathrm{GL}_m(\mathbb{R})$ such that 2.26 is satisfied. Hence, $\mathrm{GL}_m(\mathbb{R})$ acts on LM transitively.
3. Transition functions: For overlapping charts U_i, U_j with co-ordinate systems x^μ, y^μ respectively, for $p \in U_i \cap U_j$ we have
$$X_\alpha = X^\mu{}_\alpha \left.\frac{\partial}{\partial x^\mu}\right|_p = \widetilde{X}^\mu{}_\alpha \left.\frac{\partial}{\partial y^\mu}\right|_p. \tag{2.27}$$
Since $X^\mu{}_\alpha = (\partial x^\mu/\partial y^\nu)_p \widetilde{X}^\nu{}_\alpha$, the transition function is
$$t^L_{ij}(p) = \left(\left(\frac{\partial x^\mu}{\partial y^\nu}\right)_p\right) \in \mathrm{GL}_m(\mathbb{R}). \tag{2.28}$$

We have, as claimed, constructed from TM a principal frame bundle LM with the same transition functions. It should not be surprising that this bundle has physical relevance. General relativity can, with some caveats, be viewed in the context of fibre bundles. Local Lorentz transforms are chosen to correspond to the right action while general co-ordinate transforms naturally correspond to the left action, and the frame bundle becomes their natural setting.

2.4.3 Triviality of bundles

Recall that a fibre bundle is trivial if it is the direct product of the base space and the fibre. We now prove a condition for a principal bundle to be trivial.

Theorem 2.4.2. A principal bundle is trivial if and only if it admits a global section.

Proof. Let (P, π, M, G) be a principal bundle over M and let $s \in \Gamma(M, P)$ be a global section. For $a \in G$, $s(p)a$ is in the fibre at p. The right action is transitive and free, so any element $u \in P$ can be written uniquely as $s(p)a$ for some $p \in M$ and $a \in G$.

Define the map
$$\Phi : P \to M \times G : s(p)a \mapsto (p, a). \tag{2.29}$$
Φ is a homeomorphism. Hence P is homeomorphic to a trivial bundle $M \times G$ – i.e., it is trivial.

Conversely, suppose $P \cong M \times G$. Let $\phi : M \times G \to P$ be a (global) trivialisation. Taking a fixed element $g \in G$, the section $s_g : M \to P$ defined by $s_g(p) = \phi(p, g)$ is a global section. □

The same cannot be true for vector bundles, however, as we have already noted that every vector bundle admits the null section as a global section. Nonetheless, given a vector bundle E, it shares the same transition functions as its associated principal bundle $P(E)$, meaning it admits the same global structure.

Corollary 2.4.3. A vector bundle E is trivial if and only if its associated principal bundle $P(E)$ admits a global non-vanishing section.

Chapter 3

Connections on principal bundles

A connection on a manifold provides us with a way of comparing tangent vectors in different tangent spaces, and hence define a well-behaved derivative of a vector field (i.e. physically "covariant"). We need some way to *parallel transport* (without change) vectors across a manifold. There are infinitely many choices of connection on an arbitrary manifold, though some are preferred owing to additional structure – e.g. the Levi-Civita connection for a manifold with a metric.

A connection on a principal bundle plays much the same role. Geometrically, it will allow us to decide what tangent vectors to the bundle are "horizontal" and hence define a notion of parallel transport. In turn, we construct a covariant derivative, not only on the principal bundle, but also on the associated vector bundle to covariantly differentiate vector fields. Functionally, the connection on a principal bundle will be a one-form which takes values in the Lie algebra of the structure group. We assume familiarity with Lie groups and algebras, but will briefly recap (without proof) some essential knowledge; proofs can be find in Nakahara [4] or elsewhere.

3.1 A review of Lie groups and algebras

A Lie group G is a group which is also a smooth manifold, where the group operation is compatible with the smooth structure. Given elements $a, g \in G$ we define the left- and right-translation actions $L_a, R_a : G \to G$ by

$$L_a g = ag \tag{3.1}$$

$$R_a g = ga \tag{3.2}$$

These diffeomorphisms on G induce push-forwards between tangent spaces of G: $L_{a*} : T_g G \to T_{ag} G$ and $R_{a*} : T_g G \to T_{ga} G$. There are a special class of vector fields on G which are invariant under the (left) action.

Definition 3.1.1. A vector field X on a Lie group G is a *left-invariant vector field* if $L_{a*}X|_g = X|_{ag}$.

We denote the set of left-invariant vector fields on G by $\mathfrak{g}$. An element of $\mathfrak{g}$ is entirely determined by its value at the identity and we have a natural isomorphism between $\mathfrak{g}$ and $T_e G$. Hence a tangent vector at the identity $V \in T_e G$ uniquely generates a left-invariant vector field $X_V \in \mathfrak{g}$ across the entire manifold. As $\mathfrak{g}$ is a set of vector fields, it is a subset of $\mathfrak{X}(G)$ and the Lie bracket is defined on $\mathfrak{g}$. In fact $\mathfrak{g}$ is closed under the Lie bracket. For a matrix Lie group, the push-forward(s) of elements of T_eG have particularly nice expressions:

$$L_{g*}V = gV \tag{3.3}$$

$$[X_V, X_W]\,|_g = L_{g*}\,[V, W] = g\,[V, W] \tag{3.4}$$

where the multiplication is standard vector and matrix multiplication. More generally, we have an action of the group on its algebra via the push-forward. We now define the Lie algebra as the set of left-invariant vector fields with this closed Lie bracket.

Definition 3.1.2. The *Lie algebra* of a Lie group G is the set of left-invariant vector fields $\mathfrak{g}$ equipped with the Lie bracket $[\ ,\ ] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. In general, we use lower-case gothic script to denote the Lie algebra of a given Lie group, e.g. $\mathfrak{so}(n)$ is the Lie algebra of $\mathrm{SO}(n)$.

Moreover, the action of the Lie bracket can be expressed as $[T_\alpha, T_\beta] = f_{\alpha\beta}{}^{\gamma} T_\gamma$, where $\{T_\alpha\}$ is a set of generators of $\mathfrak{g}$ and $f_{\alpha\beta}{}^{\gamma}$ are called the structure constants.

Any vector field $X \in \mathfrak{X}(M)$ generates a flow in M. Left-invariant vector fields are associated with a special class of flow on the Lie group.

Definition 3.1.3. A curve $\phi : \mathbb{R} \to G$ is a *one-parameter subgroup* of G if it satisfies

$$\phi(t)\phi(s) = \phi(t + s). \tag{3.5}$$

This implies that $\phi(0) = e$ and $\phi(t)^{-1} = \phi(-t)$. The subgroup which the curve traces is Abelian, even if the group is not. We note the similarities with the exponential function on $\mathbb{R}$. As with any curve, given a one-parameter subgroup $\phi : \mathbb{R} \to G$ there exists a vector field X such that on the curve,

$$\frac{\mathrm{d}\phi^\mu(t)}{\mathrm{d}t} = X^\mu(\phi(t)). \tag{3.6}$$

In fact, for a one-parameter subgroup X is left-invariant ($X \in \mathfrak{g}$) and hence defined uniquely; i.e. there is a bijection between the one-parameter subgroups of G and elements of $\mathfrak{g}$. We are motivated to define the exponential map.

Definition 3.1.4. Let G be a Lie group and $V \in T_eG$. The *exponential map* $\exp : T_eG \to G$ is defined:

$$\exp(tV) = \phi_V(t) \tag{3.7}$$

where $\phi_V(t)$ is the one-parameter subgroup generated by $X_V|g$.

For matrix Lie groups, this coincides with matrix exponentiation. We find that, for $g \in G, A \in \mathfrak{g}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} g\exp(tA)|_{t=0} = L_{g*}A = X_A|_g \quad [\, = gA\,] \tag{3.8}$$

where the last equality holds in the sense we defined earlier at 3.3 for matrix Lie groups. In turn, the curve $g\exp(tA)$ defines a map $\sigma_t : G \to G$ by $\sigma_t(g) = g\exp(tA)$ which is also expressed as a right translation,

$$\sigma_t = R_{\exp(tA)}. \tag{3.9}$$

We generalise this to the case where a Lie group acts on another manifold.

Definition 3.1.5. Let G be a Lie group and M be a manifold. The *action* of G on M is a differentiable map $\sigma : G \times M \to M$ which satisfies

1. $\sigma(e,p) = p$ for any $p \in M$
2. $\sigma(g_1, \sigma(g_2, p)) = \sigma(g_1 g_2, p)$

We often denote $gp = \sigma(g,p)$.

Clearly a flow is an action of $\mathbb{R}$ on M. A periodic flow can be regarded as an action of e.g. U(1) or SO(2) on M and we could explicitly construct a new flow whose group is U(1). For a general action of a Lie group G on M, elements of the Lie algebra $\mathfrak{g}$ naturally induce vector fields in M also. For $V \in \mathfrak{g}$, define a flow in M by

$$\sigma(t,x) = \exp(tV)x. \tag{3.10}$$

This is a one-parameter group of transformations and defines the *induced vector field*, $V^\#$:

$$V^\#\big|_x = \frac{\mathrm{d}}{\mathrm{d}t}\exp(tV)x\bigg|_{t=0}. \tag{3.11}$$

Equivalently, we have defined a map $\# : \mathfrak{g} \to \mathfrak{X}(M)$ by $V \mapsto V^\#$.

A Lie group acts on itself in a special way.

Definition 3.1.6. For $a \in G$, we define the *adjoint representation* of G to be the homomorphism

$$\mathrm{ad}_a : G \to G : g \mapsto aga^{-1} \tag{3.12}$$

We note that $\mathrm{ad}_a e = e$ so the push-forward ad_{a*} will map the tangent space at the identity to itself. Equivalently, we have induced an action on the algebra:

Definition 3.1.7. The *adjoint map* of a Lie group G is the action $\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g}$ by

$$\mathrm{Ad}_a \equiv \mathrm{ad}_{a*}|_{TeG} \tag{3.13}$$

It follows that $\mathrm{Ad}_a \mathrm{Ad}_b = \mathrm{Ad}_{ab}$ and $\mathrm{Ad}_{a^{-1}} = \mathrm{Ad}_a{}^{-1}$. For a matrix Lie group G, the adjoint representation and map also become straightforward matrix operations. Take $g \in G, V \in \mathfrak{g}$ and let $\sigma_V = \exp(tV)$. Then we have:

$$\mathrm{ad}_g \sigma_V(t) = g\exp(tV)g^{-1} = \exp(tgVg^{-1}) \tag{3.14}$$

$$\begin{aligned} \mathrm{Ad}_g V &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\mathrm{ad}_g \exp(tV)\right]\bigg|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\exp(tgVg^{-1})\bigg|_{t=0} = gVg^{-1}. \end{aligned} \tag{3.15}$$

These expressions generalise for a more general Lie group via the push-forward of the adjoint representation, or the left- and right-translations, as before; note that the left- and right-translations commute, so their push-forwards do also.

3.2 Preliminary definitions

Let u be an element of a principal bundle $P(M,G)$ and let G_p be the fibre at $p = \pi(u)$. As the bundle is itself a manifold, we can consider its tangent space at u, T_uP. However, we are also free to consider the tangent space of the base manifold M, T_pM. We will frequently switch between these spaces so it is important to keep the distinction clear! Geometrically, to define the connection, we want to separate the tangent space T_uP into "vertical and horizontal" subspaces. We will see that this abstract notion is equivalent to another definition of the connection as a Lie-algebra-valued form.

Definition 3.2.1. The *vertical subspace* V_uP is the subspace of T_uP which is tangent to G_p at u.

We construct V_uP as follows: take an element of the Lie algebra $A \in \mathfrak{g}$. By the right action $R_{\exp(tA)}u = u\exp(tA)$, we define a curve in P through u. As we established in the previous chapter, this curve lies within the fibre G_p. Now define a vector $A^\# \in T_uP$ by its action on a smooth function $f : P \to \mathbb{R}$ as follows:

$$A^\# f(u) = \frac{\mathrm{d}}{\mathrm{d}t} f(u\exp(tA))\Big|_{t=0}. \tag{3.16}$$

This vector must be tangent to G_p at u, so it is in the vertical subspace. In this manner, we define a vector at each point of P and generate a vector field $A^\#$, the *fundamental vector field*. We now have a vector space isomorphism $\# : \mathfrak{g} \to V_uP : A \mapsto A^\#$. Further, $\#$ preserves the Lie algebra structure: $\left[A^\#, B^\#\right] = [A,B]^\#$ so this is a Lie algebra isomorphism.

Given the vertical subspace, we now define the horizontal subspace H_uP as a complement to V_uP in T_uP. There are many complements so we uniquely specify which by defining a connection on P.

Definition 3.2.2. Let $P(M,G)$ be a principal bundle. A *connection* on P is a unique separation of the tangent space T_uP into the vertical subspace V_uP and the horizontal subspace H_uP such that

1. $T_uP = V_uP \oplus H_uP$;
2. A smooth vector field X on P is separated into smooth vector fields $X^V \in V_uP$ and $X^H \in H_uP$ with $X = X^V + X^H$;
3. $H_{ug}P = R_{g*}H_uP$ for arbitrary $u \in P, g \in G$.

The third condition tells us that horizontal subspaces on the same fibre $H_uP, H_{ug}P$ are related by the linear map R_{g*}. This is enough to specify all the horizontal subspaces on the same fibre once we have one of them. Since we will be defining parallel transport as "movement through horizontal subspaces", this condition ensures that if a point u is parallel transported, so are its constant multiples ug.

This geometrical definition of the connection is intuitive, but difficult to work with. The more technical definition that follows is more useful mechanically, and will be shown to be equivalent to the geometrical definition.

3.3 The connection one-form

Definition 3.3.1. A *connection one-form* $\omega \in \mathfrak{g} \otimes T^*P$ is a projection of T_uP onto the vertical component $V_uP \cong \mathfrak{g}$: that is, ω satisfies the conditions

1. $\omega(A^\#) = A$ for $A \in \mathfrak{g}$;
2. $R_g^*\omega = \mathrm{Ad}_{g^{-1}}\omega$.

We now define the horizontal subspace H_uP as the kernel of ω,

$$H_uP = \{X \in T_uP | \omega(X) = 0\}. \tag{3.17}$$

Proposition 3.3.2. The horizontal subspaces as defined in 3.17 satisfy

$$R_{g*}H_uP = H_{ug}P \tag{3.18}$$

– i.e., our definitions 3.2.2 and 3.3.1 are equivalent.

Proof. For a fixed point $u \in P$, define the horizontal subspace H_uP as a kernel as in 3.17. Take $X \in H_uP$ and thence $R_{g*}X$. Now,

$$\omega(R_{g*}X) = R_g^*\omega(X) = g^{-1}\omega(X)g = 0 \tag{3.19}$$

as $\omega(X) = 0$ by definition. Hence, $R_{g*}X \in H_uP$, so $R_{g*}H_uP \subset H_{ug}P$. But R_{g*} is an invertible linear map, so every $Y \in H_{ug}P$ is expressed as the push-forward of some $X \in H_uP$, so $H_{ug}P \subset R_{g*}H_uP$ also. □

The Lie-algebra-valued one-form definition of the connection is frequently called the *Ehresmann connection.*

3.3.1 The local connection form and the gauge potential

Let $\{U_i\}$ be an open covering of M and let σ_i be a local section on each U_i. We introduce a Lie-algebra-valued one-form $\mathcal{A}_i$ on U_i by

$$\mathcal{A}_i = \sigma_i^*\omega \in \mathfrak{g} \otimes \Omega^1(U_i). \tag{3.20}$$

(The choice of letter for the local connection form is deliberate!)

Conversely, given a Lie-algebra-valued one-form $\mathcal{A}_i$ and a local section σ_i on U_i, there is a connection one-form ω with $\mathcal{A}_i = \sigma_i^*\omega$.

Proposition 3.3.3. The connection one-form we want is

$$\omega_i = g_i^{-1}\pi^*\mathcal{A}_i g_i + g_i^{-1}\mathrm{d}_P g_i \tag{3.21}$$

where d_P is the exterior derivative on P, and g_i is given by the canonical local trivialisation:

$$\phi_i^{-1}(u) = (p, g_i) \text{ for } u = \sigma_i(p)g_i \tag{3.22}$$

Proof. We first show that $\sigma_i^*\omega_i = \mathcal{A}_i$. For $X \in T_pM$, we have

$$\begin{aligned}\sigma_i^*\omega_i(X) &= \omega_i(\sigma_{i*}X)\\ &= \pi^*\mathcal{A}_i(\sigma_{i*}X) + \mathrm{d}_P g_i(\sigma_{i*}X)\end{aligned}$$

(where we have noted that $g_i = e$ at $u = \sigma_i$)

$$= \mathcal{A}_i(\pi_*\sigma_{i*}) + \mathrm{d}_P g_i(\sigma_{i*}X).$$

But $\pi_*\sigma_{i*} = \mathrm{id}_{T_pM}$, and $\mathrm{d}_P g_i(\sigma_{i*}X) = 0$ since $g = e$ *along* $\sigma_{i*}X$ by definition of our canonical local trivialisation!

$$\Rightarrow \sigma_i^*\omega_i(X) = \mathcal{A}_i(X).$$

It remains to show that this one-form is a connection one-form according to our earlier definition.

Let $X = A^\# \in V_uP, A \in \mathfrak{g}$. Hence, $\pi_*X = 0$. Now,

$$\begin{aligned}\omega_i(A^\#) &= g_i^{-1}\mathrm{d}_P g_i(A^\#)\\ &= g_i(u)^{-1}\frac{\mathrm{d}g(u\exp(tA))}{\mathrm{d}t}\bigg|_{t=0}\\ &= g_i(u)^{-1}g_i(u)\frac{\mathrm{d}\exp(tA)}{\mathrm{d}t}\bigg|_{t=0}\\ &= A.\end{aligned}$$

Finally, take $X \in T_u P, h \in G$. We have

$$R_h^* \omega_i X = \omega_i(R_{h*}X) = g_{i\,uh}^{-1} \mathcal{A}_i(\pi_* R_{h*} X) g_{i\,uh} + g_{i\,uh}^{-1} \mathrm{d}_P g_{i\,uh}(R_{h*}X).$$

Now $g_{i\,uh} = g_{i\,u} h$ and $\pi_* R_{h*} X = \pi_* X$ (certainly $\pi R_h = \pi$), so

$$\begin{aligned} R_h^* \omega_i X &= h^{-1} g_{i\,u}^{-1} \mathcal{A}_i(\pi_* X) g_{i\,u} h + h^{-1} g_{i\,u}^{-1} \mathrm{d}_P g_{i\,u}(X) h \\ &= h^{-1} \omega_i(X) h \end{aligned}$$

where in the second term we have noted that, taking $\gamma(t)$ is a curve through $u = \gamma(0)$ with tangent vector X at u,

$$\begin{aligned} g_{i\,uh}^{-1} \mathrm{d}_P g_{i\,uh}(R_{h*}X) &= g_{i\,uh}^{-1} \frac{\mathrm{d}}{\mathrm{d}t} g_{i\,\gamma(t)h} \bigg|_{t=0} \\ &= h^{-1} g_{i\,u}^{-1} \frac{\mathrm{d}}{\mathrm{d}t} g_{i\,\gamma(t)} \bigg|_{t=0} h \\ &= h^{-1} g_{i\,u}^{-1} \mathrm{d}_P g_{i\,u}(X) h. \end{aligned}$$

□

We've defined a connection one-form ω_i on each patch $(\pi^{-1}(U_i))$ of P. For ω to be uniquely defined, i.e. for the separation $T_u P = H_u P \oplus V_u P$ to be unique, we must have $\omega_i = \omega_j$ on $\pi^{-1}(U_i \cap U_j)$. A unique one-form is then defined on P by

$$\omega|_{\pi^{-1}(U_i)} = \omega_i.$$

Before we can discuss the compatibility condition on such a global connection, we need a technical lemma.

Lemma 3.3.4. Let $P(M, G)$ be a principal bundle and σ_i, σ_j be local sections over U_i, U_j respectively such that $U_i \cap U_j \neq \emptyset$. For $X \in T_p M, p \in U_i \cap U_j$, $\sigma_{i*}X$ and $\sigma_{j*}X$ satisfy:

$$\sigma_{j*}X = R_{t_{ij}*}(\sigma_{i*}X) + (t_{ij}^{-1} \mathrm{d}t_{ij}(X))^{\#}. \tag{3.23}$$

Proof. Consider a curve $\gamma : [0, 1] \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$. Since $\sigma_i(p)$ and $\sigma_j(p)$ are related by the transition function by $\sigma_j(p) = \sigma_i(p) t_{ij}(p)$, we have

$$\begin{aligned} \sigma_{j*}X &= \frac{\mathrm{d}}{\mathrm{d}t} \sigma_j(\gamma(t)) \bigg|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \left[\sigma_i(\gamma(t)) t_{ij}(\gamma(t)) \right] \bigg|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \sigma_i(\gamma(t)) \bigg|_{t=0} \cdot t_{ij}(p) + \sigma_i(p) \frac{\mathrm{d}}{\mathrm{d}t} t_{ij}(\gamma(t)) \bigg|_{t=0} \\ &= R_{t_{ij}*}(\sigma_{i*}X) + \sigma_j(p) t_{ij}(p)^{-1} \frac{\mathrm{d}}{\mathrm{d}t} t_{ij}(\gamma(t)) \bigg|_{t=0} \end{aligned}$$

because, by assumption, G is a matrix group with $R_{g*}X = Xg$ (or this is the definition of that multiplication). Further,

$$\begin{aligned} t_{ij}(p)^{-1} \mathrm{d}t_{ij}(X) &= t_{ij}(p)^{-1} \frac{\mathrm{d}}{\mathrm{d}t} t_{ij}(\gamma(t)) \bigg|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \left[t_{ij}(p)^{-1} t_{ij}(\gamma(t)) \right] \bigg|_{t=0} \in T_e G \cong \mathfrak{g}. \end{aligned}$$

Checking 3.16, we see therefore that the second term is by definition $(t_{ij}^{-1}\mathrm{d}t_{ij}(X))^{\#}$ at $\sigma_j(p)$. $\square$

Now we simply apply our connection one-form to 3.23:

$$\begin{aligned}\sigma_j^*\omega(X) &= R_{t_{ij}}^*\omega(\sigma_{i*}X) + t_{ij}^{-1}\mathrm{d}t_{ij}(X) \\ &= t_{ij}^{-1}\omega(\sigma_{i*}X)t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}(X).\end{aligned} \tag{3.24}$$

This is true for all $X \in T_pM$, so we have our compatibility condition:

$$\mathcal{A}_j = t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}. \tag{3.25}$$

(This bears some similarity to how Christoffel symbols transform in differential geometry or general relativity. This might not be too surprising after our previous remarks on frame bundles. Loosely, general relativity can be interpreted as a gauge theory of the Poincaré group, but there are several problems; for example, the Poincaré group is not compact, and gauge symmetries are usually internal symmetries rather than symmetries of spacetime.)

Conversely, given an open covering $\{U_i\}$, local sections $\{\sigma_i\}$ and compatible (i.e., obeying 3.25) local forms $\{\mathcal{A}_i\}$, we may construct the $\mathfrak{g}$-valued one-form ω over P. Since a non-trivial one-form does not admit a global section, the pullback $\mathcal{A}_i = \sigma_i^*\omega$ exists locally but not globally. In gauge theories, $\mathcal{A}_i$ is identified with the *gauge potential* or *Yang-Mills potential*.

The connection one-form ω is defined over $P(M,G)$ – in fact, there are many connection one-formes on $P(M,G)$, which all share the same global information about the structure of the bundle. However, any gauge potential $\mathcal{A}_i$ is associated with the trivial bundle $\pi^{-1}(U_i)$ (for example, see our comments in the proof of Proposition 3.3.3), and it can't have any global information about the structure of P. It is ω, or the combination of the $\{\mathcal{A}_i\}$ with the compatibility condition 3.25, that carry the global information about P's structure.

3.3.2 Example: gauge potentials on a U(1)-bundle

We wish to consider the case where on some region – such as, say, the equator of S^2 – we have two local connections. Let $P(M,G)$ be a principal bundle, U a chart of M, and σ_1, σ_2 local sections on U. We can write $\sigma_2(p) = \sigma_1(p)g(p)$. Then by 3.25, the local forms satisfy

$$\mathcal{A}_2 = g^{-1}\mathcal{A}_1 g + g^{-1}\mathrm{d}g. \tag{3.26}$$

In components, this is

$$\mathcal{A}_{2\mu} = g(p)^{-1}\mathcal{A}_{1\mu}g(p) + g(p)^{-1}\partial_\mu g(p) \tag{3.27}$$

which is evidently a gauge transform as seen at 1.7!

If we now take $G = \mathrm{U}(1)$, overlapping charts U_i, U_j and local connection forms $\mathcal{A}_i, \mathcal{A}_j$ respectively, the transition function $t_{ij} : U_i \cap U_j \to \mathrm{U}(1)$ takes the form

$$t_{ij}(p) = \exp\left[\mathrm{i}\chi(p)\right], \quad \chi(p) \in \mathbb{R}. \tag{3.28}$$

Now $\mathcal{A}_i$ and $\mathcal{A}_j$ are related by

$$\begin{aligned}\mathcal{A}_j(p) &= t_{ij}(p)^{-1}\mathcal{A}_i(p)t_{ij}(p) + t_{ij}(p)^{-1}\mathrm{d}t_{ij}(p) \\ &= \mathcal{A}_i(p) + \mathrm{i}\,\mathrm{d}\chi(p)\end{aligned} \tag{3.29}$$

or componentwise

$$\mathcal{A}_{j\mu} = \mathcal{A}_{i\mu} + \mathrm{i}\partial_\mu\chi. \tag{3.30}$$

This is an extremely familiar expression! Our connection $\mathcal{A}_\mu$ differs from the standard electromagnetic 4-potential A_μ just by the Lie algebra factor:

$$\mathcal{A}_\mu = \mathrm{i}A_\mu. \tag{3.31}$$

3.3.3 Horizontal lift and parallel transport

Having used the connection one-form to separate the bundle's tangent space into its horizontal and vertical subspaces, there is an extremely natural and intuitive way of defining parallel transport on the bundle: a vector is *parallel transported* if it is, in a sense, moving through the horizontal subspace at every point on the path. That path is now through the bundle, and we associate it to the projected path in the manifold by the notion of *horizontal lift*.

Definition 3.3.5. Let $P(M, G)$ be a G-bundle, $\gamma : [0, 1] \to M$ a curve in the base space. $\bar{\gamma} : [0, 1] \to P$ is a *horizontal lift* of γ if:

- $\pi\bar{\gamma} = \gamma$.
- The tangent vector to $\bar{\gamma}(t)$ always belongs to $H_{\bar{\gamma}(t)}P$.

Let $\widetilde{X}$ be a tangent vector to $\bar{\gamma}$. Then by definition, $\omega(\widetilde{X}) = 0$. This is an ODE, so at least locally there is a unique horizontal lift. We prove that we can do even better:

Theorem 3.3.6. Let $\gamma : [0, 1] \to M$ and $u_0 \in \pi^{-1}(\gamma(0))$. Then there exists a *unique* horizontal lift $\bar{\gamma} : [0, 1] \to P$ with $\bar{\gamma}(0) = u_0$.

Proof. We will construct $\bar{\gamma}$. Let U_i be a chart containing γ and take a section σ_i over U_i. A horizontal lift must be written as $\bar{\gamma}(t) = \sigma_i(\gamma(t))g_i(t)$ (where we write $g_i(t) = g_i(\gamma(t)) \in G$, &c.). WLOG take a section such that $\sigma_i(\gamma(0)) = \bar{\gamma}(0)$ — that is, $g_i(0) = e$. Let X be tangent to $\gamma(t)$ at $\gamma(0)$: then $\widetilde{X} = \sigma_{i*}X$ is tangent to $\bar{\gamma}$ at $u_0 = \bar{\gamma}(0)$. But $\widetilde{X}$ is horizontal: $\omega(\widetilde{X}) = 0$. □

Essentially by Lemma 3.3.4, we have

$$\widetilde{X} = g_i(t)^{-1}\sigma_{i*}Xg_i(t) + [g_i(t)^{-1}\,\mathrm{d}g_i(X)]^{\#}. \tag{3.32}$$

Applying ω to this, we obtain

$$\begin{aligned} 0 = \omega(\widetilde{X}) &= g_i(t)^{-1}\omega(\sigma_{i*}X)g_i(t) + g_i(t)^{-1}\frac{\mathrm{d}g_i(t)}{\mathrm{d}t} \\ \Rightarrow \frac{\mathrm{d}g_i(t)}{\mathrm{d}t} &= -\omega(\sigma_{i*X})g_i(t) \end{aligned} \tag{3.33}$$

which has a unique solution. Locally, $\omega(\sigma_{i*}X) = \sigma_i^*\omega(X) = \mathcal{A}_i(X)$ and the above becomes

$$\frac{\mathrm{d}g_i(t)}{\mathrm{d}t} = -\mathcal{A}_i(X)g_i(t). \tag{3.34}$$

Formally, the solution satisfying $g_i(0) = e$ is

$$\begin{aligned} g_i(\gamma(t)) &= \mathcal{P}\,\exp\left(-\int_0^t \mathcal{A}_{i\mu}\frac{\mathrm{d}x^\mu}{\mathrm{d}t}\,\mathrm{d}t\right) \\ &= \mathcal{P}\,\exp\left(-\int_{\gamma(0)}^{\gamma(t)} \mathcal{A}_{i\mu}\,\mathrm{d}x^\mu\right) \end{aligned}$$

where $\mathcal{P}$ is a path-ordering operator for the exponential — note that in general, we may be working over a non-Abelian Lie algebra $\mathfrak{g}$! Now the horizontal lift is fully written as $\bar{\gamma}(t) = \sigma_i(\gamma(t))g_i(\gamma(t))$.

Corollary 3.3.7. The horizontal lift is well-defined: if $\bar{\gamma}'$ is another lift of γ with $\bar{\gamma}'(0) = \bar{\gamma}(0)g$, then $\bar{\gamma}'(t) = \bar{\gamma}(t)g \,\forall t \in [0,1]$. Hence, choosing a different $\bar{\gamma}(0) = u_0'$ does not essentially change the lift.

This follows immediately from the right-invariance of the horizontal subspace: $R_{g*}H_uP = H_{ug}P$.

Let $\gamma : [0,1] \to M$, $u_0 \in \pi^{-1}(\gamma(0))$. We've established there is a unique horizontal lift $\bar{\gamma}(t)$ that starts at u_0, so there is certainly a unique point $u_1 = \bar{\gamma}(1) \in \pi^{-1}(\gamma(1))$. So we define u_1 to be the *parallel transport of u_0 along $\bar{\gamma}$*.

This defines a map between fibres

$$\Gamma(\bar{\gamma}) : \pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(1)) \tag{3.35}$$

such that $\Gamma(\bar{\gamma})(u_0) = u_1$. In local form, we have

$$u_1 = \sigma_i(1)\mathcal{P}\exp\left(-\int_{\gamma(0)}^{\gamma(1)} \mathcal{A}_{i\mu}\mathrm{d}x^\mu\right). \tag{3.36}$$

Corollary 3.3.7 ensures $\Gamma(\bar{\gamma})$ commutes with the right action R_g:

$$R_g\Gamma(\bar{\gamma})(u_0) = u_1g \qquad\qquad \Gamma(\bar{\gamma})R_g(u_0) = \Gamma(\bar{\gamma})(u_0g) \tag{3.37}$$

Observe that $\bar{\gamma}(t)g$ is a horizontal lift through u_0g, u_1g. By uniqueness, $u_1g = \Gamma(\bar{\gamma})(u_0g)$ so indeed $\Gamma(\bar{\gamma})$ commutes with the right action.

In fact we have an equivalence relation across the whole bundle: $u \sim v$ iff $u, v \in P$ are on the same horizontal lift.

3.3.4 Holonomy

Consider two curves in the base space which share endpoints: $\alpha, \beta : [0,1] \to M$ with $\alpha(0) = \beta(0) = p_0$ and $\alpha(1) = \beta(1) = p_1$. Take horizontal lifts $\bar{\alpha}, \bar{\beta}$ with $\bar{\alpha}(0) = \bar{\beta}(0) = u_0$. In general, there is no reason for $\bar{\alpha}(1)$ and $\bar{\beta}(1)$ to be equal.

Similarly, if we consider a loop in the base space $\gamma : [0,1] \to M$ at $p = \gamma(0) = \gamma(1)$, in general $\bar{\gamma}(0)$ and $\bar{\gamma}(1)$ are unequal. This defines a transformation on the fibre:

$$\tau_\gamma : \pi^{-1}(p) \to \pi^{-1}(p)$$

compatible with the right action of the group

$$\tau_\gamma(ug) = \tau_\gamma(u)g. \tag{3.38}$$

Note that τ_γ depends on our choices of γ and ω.

Take $u \in P$ with $\pi(u) = p$ and consider the set of loops C_pM at p:

$$C_pM = \{\gamma : [0,1] \to M \mid \gamma(0) = \gamma(1) = p\}. \tag{3.39}$$

The subset of elements of G

$$\Phi_u = \{g \in G \mid \tau_\gamma(u) = ug, \gamma \in C_pM\} \tag{3.40}$$

is in fact a subgroup of the structure group G, called the *holonomy group* at u. Note that if $\alpha, \beta, \gamma = \alpha * \beta$ are loops at p, where $\alpha * \beta(t)$ is the loop that traverses α once and then β once as t ranges over $[0, 1]$, we have:

$$\begin{aligned}
&\tau_\gamma = \tau_\beta \tau_\alpha \\
\Rightarrow &\tau_\gamma(u) = \tau_\beta \tau_\alpha(u) = \tau_\beta(u g_\alpha) = \tau_\beta(u) g_\alpha = u g_\beta g_\alpha \\
\Rightarrow &g_\gamma = g_\beta g_\alpha.
\end{aligned}$$

The constant loop $c : [0, 1] \to \{p\}$ defines the identity transformation $\tau_c : u \mapsto u$. The inverse loop $\gamma^{-1}(t) = \gamma(1 - t)$ induces $\tau_{\gamma^{-1}} = \tau_\gamma^{-1}$, whence $g_{\gamma^{-1}} = g_\gamma^{-1}$.

Physically, we will see that in the presence of a non-zero gauge field, motion around a closed path can have a physical effect on our system. We will define curvature next and see its role in this phenomenon, by analogy to curvature on manifolds.

3.4 Curvature

3.4.1 The covariant derivative and the curvature two-form

We generalise the exterior derivative on a manifold $\mathrm{d} : \Omega^r(M) \to \Omega^{r+1}(M)$ on real-valued r-forms

$$\eta \in \Omega^r(M) : \underbrace{TM \otimes \cdots \otimes TM}_{\mathrm{r}} \to \mathbb{R} \tag{3.41}$$

to vector-valued r-forms on the bundle,

$$\phi \in \Omega^r(P) \otimes V : TP \otimes \cdots \otimes TP \to V \tag{3.42}$$

where $\dim V = k$. A general such vector-valued form can be expressed as

$$\phi = \sum_{\alpha=1}^{k} \phi^\alpha \otimes \mathbf{e}_\alpha \tag{3.43}$$

where $\{\mathbf{e}_\alpha\}$ is a basis of V and $\phi^\alpha \in \Omega^r(P)$.

Our connection ω on separates $T_u P$ into $H_u P \oplus V_u P$ and so separates a given vector $X \in T_u P$ into $X^H + X^V$.

Definition 3.4.1. Let $\phi \in \Omega^r(P) \otimes V, X_1, \ldots, X_{r+1} \in T_u P$. The *covariant derivative* $\mathrm{d}\phi$ of ϕ is defined by:

$$\mathrm{d}\phi(X_1, \ldots, X_{r+1}) = \mathrm{d}_P \phi(X_1^H, \ldots, X_{r+1}^H) \tag{3.44}$$

where $\mathrm{d}_P \phi \equiv \mathrm{d}_P \phi^\alpha \otimes \mathbf{e}_\alpha$ (with summation).

Definition 3.4.2. The *curvature two-form* Ω is the covariant derivative of the connection one-form ω:

$$\Omega = \mathrm{d}\omega \in \Omega^2(P) \otimes \mathfrak{g} \tag{3.45}$$

It is easy to show that the curvature satisfies $R_a^* \Omega = a^{-1} \Omega a$; compare with the definition (3.3.1) of the connection.

Since the forms we are working with are $\mathfrak{g}$-valued, we use the structure of the Lie algebra in our treatment. Consider a $\mathfrak{g}$-valued p-form $\zeta = \zeta^\alpha \otimes T_\alpha$ and a $\mathfrak{g}$-valued q-form $\eta = \eta^\alpha \otimes T_\alpha$, with $\zeta^\alpha \in \Omega^p(M), \eta^\alpha \in \Omega^q(M)$ and $\{T_\alpha\}$ a basis of $\mathfrak{g}$. The commutator of ζ and η is

$$\begin{aligned}
[\zeta, \eta] &= \zeta \wedge \eta - (-1)^{pq} \eta \wedge \zeta \\
&= T_\alpha T_\beta \zeta^\alpha \wedge \eta^\beta - (-1)^{pq} T_\beta T_\alpha \eta^\beta \wedge \zeta^\alpha \\
&= [T_\alpha, T_\beta] \otimes \zeta^\alpha \wedge \eta^\beta \\
&= f_{\alpha\beta}{}^\gamma T_\gamma \otimes \zeta^\alpha \wedge \eta^\beta.
\end{aligned}$$

If $\zeta = \eta$, we have

$$[\zeta, \zeta] = 2\zeta \wedge \zeta = f_{\alpha\beta}{}^\gamma T_\gamma \otimes \zeta^\alpha \wedge \zeta^\beta. \tag{3.46}$$

Lemma 3.4.3. If $X \in H_uP, Y \in V_uP$, then $[X, Y] \in H_uP$.

Proof. We use the equivalence of the commutator and the Lie derivative. Let $g(t)$ be a flow (at least locally) generating Y. Then:

$$[Y, X] = \lim_{t\to 0} t^{-1} \left(R_{g(t)*}X - X \right) \tag{3.47}$$

A connection satisfies $R_{g*}H_uP = H_{ug}P$, so $R_{g(t)*}X$ is horizontal, so $[Y, X]$ is. □

With this structure we can prove a useful expression for the curvature.

Theorem 3.4.4 (Cartan's structure equation)**.** Let $X, Y \in T_uP$. Then Ω and ω satisfy

$$\Omega(X, Y) = \mathrm{d}_P\omega(X, Y) + [\omega(X), \omega(Y)], \tag{3.48}$$

which we may write

$$\Omega = \mathrm{d}_P\omega + \omega \wedge \omega. \tag{3.49}$$

Sketch Proof. We omit a full proof; see Nakahara [4] for details. The sketch is to consider the three cases:

1. $X, Y \in H_uP$;
2. $X \in H_uP$, $Y \in V_uP$;
3. $X, Y \in V_uP$;

and use the structure outlined above and the definition of the connection and covariant derivative to show that 3.48 is satisfied in each case. By linearity and antisymmetry of Ω, these cases are sufficient to extend the theorem to any vectors $X, Y \in T_uP$.

To obtain 3.49 from 3.48, note that

$$\begin{aligned}
[\omega, \omega](X, Y) &= [T_\alpha, T_\beta]\omega^\alpha \wedge \omega^\beta(X, Y) \\
&= [T_\alpha, T_\beta][\omega^\alpha(X)\omega^\beta(Y) - \omega^\beta(X)\omega^\alpha(Y)] \\
&= [\omega(X), \omega(Y)] - [\omega(Y), \omega(X)] \\
&= 2[\omega(X), \omega(Y)].
\end{aligned}$$

Hence $\Omega(X, Y) = (\mathrm{d}_P\omega + \frac{1}{2}[\omega, \omega])(X, Y) = (\mathrm{d}_P\omega + \omega \wedge \omega)(X, Y)$. □

3.4.2 The geometrical interpretation of curvature

The curvature tensor on a manifold expresses the failure of parallel transport to "close" or commute — a vector parallel transported around a loop, or equivalently parallel transported to a different point by two different paths, may not be invariant. There is a very similar interpretation for the curvature on a principal bundle.

We first observe that $\Omega(X,Y)$ gives the *vertical* component of the Lie bracket $[X,Y]$ of *horizontal* vectors $X, Y \in H_uP$.

$$\begin{aligned} &\omega(X) = \omega(Y) = 0 \\ \Rightarrow\ &\mathrm{d}_P\omega(X,Y) = X\omega(Y) - Y\omega(X) - \omega([X,Y]). \\ &\qquad\qquad\quad = -\omega([X,Y]) \end{aligned}$$

Further, $X^H = X$, $Y^H = Y$;

$$\Rightarrow \Omega(X,Y) = \mathrm{d}_P\omega(X,Y) = -\omega([X,Y]).$$

Consider co-ordinates $\{x^\mu\}$ on a chart U. Let $V = \frac{\partial}{\partial x^1}$, $W = \frac{\partial}{\partial x^2}$. Take a parallelogram γ whose corners are $O = \{0\}$, $P = \{\epsilon, 0, \ldots, 0\}$, $Q = \{\epsilon, \delta, \ldots, 0\}$ and $R = \{0, \delta, \ldots, 0\}$, and let $\bar{\gamma}$ be a horizontal lift of γ. Let $X, Y \in H_uP$ such that $\pi_*X = \epsilon V, \pi_*Y = \delta W$. Then

$$\pi_*([X,Y]^H) = \epsilon\delta[V,W] = \epsilon\delta\left[\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}\right] = 0; \tag{3.50}$$

i.e. $[X,Y]$ is vertical, and $\bar{\gamma}$ fails to close. The failure is proportional to the vertical vector $[X,Y]$ connecting initial and final points on the same fibre, and the curvature measures the distance:

$$\Omega(X,Y) = -\omega([X,Y]) = A \text{ where } A \in \mathfrak{g},\ A^\# = [X,Y]. \tag{3.51}$$

Note that in this specific case, we have the option of "integrating" the locally constant[1] curvature over the area $(\epsilon\delta)$ enclosed by γ, or performing the path integral of the connection at (3.36) over γ, to obtain the distance on the fibre between the initial and final points. This suggests that there is something like a non-Abelian Stokes theorem, at least for Lie-algebra-valued one- and two-forms, and surfaces and curves; however, going further than this comment is beyond the scope of this treatment.

We conclude our discussion of the algebra by just mentioning the Ambrose-Singer theorem, which expresses the holonomy group in terms of the curvature.

Theorem 3.4.5 (The Ambrose-Singer theorem)**.** Let $P(M,G)$ be a principal G-bundle over a connected manifold M. Then the Lie algebra $\mathfrak{h}$ of Φ_{u_0} (for some $u_0 \in P$) agrees with the subalgebra $\mathfrak{g}' < \mathfrak{g}$ spanned by

$$\{\Omega_u(X,Y) \mid X, Y \in H_uP\}$$

where u is a point on the same horizontal lift as u_0.

[1] i.e., ϵ and δ are small enough that to effective order the curvature is constant.

3.4.3 Local form of curvature

The local form $\mathcal{F}$ of the curvature Ω is defined by

$$\mathcal{F} = \sigma^*\Omega. \tag{3.52}$$

In terms of the gauge potential $\mathcal{A}$ we have

$$\mathcal{F} = \mathrm{d}\mathcal{A} + \mathcal{A} \wedge \mathcal{A} \tag{3.53}$$

(where d is the exterior derivative on M): note that $\mathcal{A} = \sigma^*\omega$, $\sigma^*\mathrm{d}_P\omega = \mathrm{d}\sigma^*\omega$, and $\sigma^*(\zeta \wedge \eta) = \sigma^*\zeta \wedge \sigma^* \wedge \eta$, so by Cartan (3.4.4),

$$\begin{aligned}\mathcal{F} &= \sigma^*(\mathrm{d}_P\omega + \omega \wedge \omega)\\ &= \mathrm{d}\sigma^*\omega + \sigma^*\omega \wedge \sigma^*\omega\\ &= \mathrm{d}\mathcal{A} + \mathcal{A} \wedge \mathcal{A}.\end{aligned}$$

Its action on vectors of TM is

$$\mathcal{F}(X,Y) = \mathrm{d}\mathcal{A}(X,Y) + [\mathcal{A}(X), \mathcal{A}(Y)]. \tag{3.54}$$

Componentwise, on a chart U whose co-ordinates are x^μ $[= \varphi^\mu(p)]$, let $\mathcal{A} = \mathcal{A}_\mu \mathrm{d}x^\mu$ be the gauge potential ($\mathcal{A}_\mu \in \mathfrak{g}$). Writing $\mathcal{F} = \frac{1}{2}\mathcal{F}_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$, we compute directly

$$\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu] \tag{3.55}$$

As a result we identify $\mathcal{F}$ with the *(Yang-Mills) field strength* — and call it that, to avoid confusion with the curvature Ω. We note the similarity to electromagnetism, an Abelian gauge theory, and that the commutator plays a role in non-Abelian theories.

Since $\mathcal{A}_\mu$ and $\mathcal{F}_{\mu\nu}$ are $\mathfrak{g}$-valued, we may write them in terms of a basis $\{T_\alpha\}$ of $\mathfrak{g}$:

$$\mathcal{A}_\mu = {A_\mu}^\alpha T_\alpha \qquad\qquad \mathcal{F}_{\mu\nu} = {F_{\mu\nu}}^\alpha T_\alpha \tag{3.56}$$

Since $[T_\alpha, T_\beta] = {f_{\alpha\beta}}^\gamma T_\gamma$, we obtain

$${F_{\mu\nu}}^\alpha = \partial_\mu {A_\nu}^\alpha - \partial_\nu {A_\mu}^\alpha + {f_{\beta\gamma}}^\alpha {A_\mu}^\beta {A_\nu}^\gamma \tag{3.57}$$

Theorem 3.4.6. On overlapping charts U_i, U_j of M with local field strengths $\mathcal{F}_i, \mathcal{F}_j$, the compatibility condition is

$$\mathcal{F}_j = \mathrm{Ad}_{t_{ij}^{-1}}\mathcal{F}_i = t_{ij}^{-1}\mathcal{F}_i\, t_{ij} \tag{3.58}$$

Proof. Note that $\mathrm{d}t^{-1} = -t^{-1}\,\mathrm{d}t\, t^{-1}$ (as $\mathrm{d}(tt^{-1}) = 0$). So:

$$\begin{aligned}\mathcal{F}_i &= \mathrm{d}\mathcal{A}_i + \mathcal{A}_i \wedge \mathcal{A}_i\\ \mathcal{F}_j &= \mathrm{d}\mathcal{A}_j + \mathcal{A}_j \wedge \mathcal{A}_j\\ \mathcal{A}_j &= t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}\end{aligned}$$

$$\begin{aligned}\mathcal{F}_j =&\mathrm{d}\left(t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}\right) + \left(t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}\right) \wedge \left(t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}t_{ij}\right)\\ =& \left[-t_{ij}^{-1}\,\mathrm{d}t_{ij} \wedge t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\mathrm{d}\mathcal{A}_i t_{ij} - t_{ij}^{-1}\mathcal{A}_i \wedge \mathrm{d}t_{ij} - t_{ij}^{-1}\,\mathrm{d}t_{ij}\, t_{ij}^{-1} \wedge \mathrm{d}t_{ij}\right]\\ &+ \left[t_{ij}^{-1}\mathcal{A}_i \wedge \mathcal{A}_i t_+ t_{ij}^{-1}\mathcal{A}_i \wedge \mathrm{d}t_{ij} + t_{ij}^{-1}\,\mathrm{d}t_{ij} \wedge t_{ij}^{-1}\mathcal{A}_i t_{ij} + t_{ij}^{-1}\,\mathrm{d}t_{ij}\, t_{ij}^{-1} \wedge \mathrm{d}t_{ij}\right]\\ =&t_{ij}^{-1}\left(\mathrm{d}\mathcal{A}_i + \mathcal{A}_i \wedge \mathcal{A}_i\right)t_{ij}\\ =&t_{ij}^{-1}\mathcal{F}_i\, t_{ij}\end{aligned}$$

$\square$

$\mathcal{A}$ is called a *pure gauge* if, locally, $\mathcal{A} = g^{-1}\mathrm{d}g$, in which case it is easily seen that $\mathcal{F} = 0$. Less easily, the converse is also true: if the field strength vanishes on a patch, the gauge potential can be written as a pure gauge there.

3.4.4 The Bianchi identity

Since ω and Ω are $\mathfrak{g}$-forms, we may also expand them over the basis $\{T_\alpha\}$ of $\mathfrak{g}$ as

$$\omega = \omega^\alpha T_\alpha, \qquad \Omega = \Omega^\alpha T_\alpha.$$

Hence,

$$\Omega^\alpha = \mathrm{d}_P\omega^\alpha + f_{\beta\gamma}{}^\alpha \omega^\beta \wedge \omega^\gamma. \tag{3.59}$$

Taking the exterior derivative of this expression, we have

$$\mathrm{d}_P\Omega^\alpha = f_{\beta\gamma}{}^\alpha \mathrm{d}_P\omega^\beta \wedge \omega^\gamma + f_{\beta\gamma}{}^\alpha \omega^\beta \wedge \mathrm{d}_P\omega^\gamma. \tag{3.60}$$

Recall that $\omega(X) = 0$ for $X \in H_uP$, so:

$$\mathrm{d}\Omega(X, Y, Z) = \mathrm{d}_P\Omega(X^H, Y^H, Z^H) = 0 \quad \forall X, Y, Z \in T_uP; \tag{3.61}$$

i.e. we have established the *Bianchi identity*:

$$\mathrm{d}\Omega = 0. \tag{3.62}$$

The local form of this identity is obtained by noting that $\sigma^*\mathrm{d}_P\Omega = \mathrm{d}\cdot\sigma^*\Omega = \mathrm{d}\mathcal{F}$, so,

$$\begin{aligned}\sigma^*(\mathrm{d}_P\omega \wedge \omega - \omega \wedge \mathrm{d}_P\omega) &= \mathrm{d}\sigma^*\omega \wedge \sigma^*\omega - \sigma^*\omega \wedge \mathrm{d}\sigma^*\omega \\ &= \mathrm{d}\mathcal{A} \wedge \mathcal{A} - \mathcal{A} \wedge \mathrm{d}\mathcal{A} \\ &= \mathcal{F} \wedge \mathcal{A} - \mathcal{A} \wedge \mathcal{F}\end{aligned} \tag{3.63}$$

Hence we have established the identity

$$\mathcal{D}\mathcal{F} = \mathrm{d}\mathcal{F} + [\mathcal{A}, \mathcal{F}] = 0 \tag{3.64}$$

where the operator $\mathcal{D}$ acts on $\eta \in \Omega^p(M) \otimes \mathfrak{g}$ by

$$\mathcal{D}\eta = \mathrm{d}\eta + [\mathcal{A}, \eta]. \tag{3.65}$$

3.5 The covariant derivative on associated vector bundles

Recall that given a principal bundle, we define an associated vector bundle. With our connection one-form ω on a principal bundle $P(M, G)$, we can naturally define the covariant derivative on the associated vector bundle. We often want to differentiate sections on this associated vector bundle but find that the unmodified derivative will break some symmetry (e.g. the gauge symmetry) of the system, as seen at 1.23. We saw that covariant differentiation was essential in constructing gauge-invariant actions.

Let $P(M, G)$ have projection π_P. Take a chart U_i of M and a section σ_i over U_i. The canonical trivialisation is $\phi_i(p, e) = \sigma_i(p)$. Let $\bar{\gamma}$ be a horizontal lift of a curve $\gamma : [0, 1] \to U_i$. Denote $\gamma(0) = p_0, \bar{\gamma}(0) = u_0$. Associated with P is the vector bundle $E = P \times_\rho V$ with projection π_E. Let

$X \in T_{p_0}M$ be a tangent vector to γ at p_0. Let $s \in \Gamma(M, E)$ be a section (i.e. vector field) of M. Write elements of E as equivalence classes:

$$[(u, v)] = \{(ug, \rho(g)^{-1}v \mid u \in P, v \in V, g \in G\}. \tag{3.66}$$

Taking a representative of the equivalence class is the same as fixing the gauge: we could write

$$s(p) = [(\sigma_i(p), \xi(p))],$$

but if we worked with this, we would be dependent on our choice of local section σ_i. For example, we might try to define parallel transport of a vector in E along a curve γ by saying that it is parallel transported if $\xi(\gamma(t))$ remains constant, but this is clearly dependent on the choice of local section. Instead, we'll say a vector is parallel transported if it is constant with respect to a horizontal lift $\bar{\gamma}$ of γ — and we will discover it does not matter which lift we choose.

Definition 3.5.1. A vector $s(p)$ is *parallel transported* along a path γ if it satisfies $s(\gamma(t)) = [(\bar{\gamma}(t), \eta(\gamma(t)))]$ and η is constant along γ.

This definition is intrinsic in that it does not depend on the choice of lift: note that if η is constant along γ, so is its constant multiple $a^{-1}g$. So in the above definition, $\bar{\gamma}$ is an arbitrary horizontal lift of γ.

We can now define the covariant derivative of the section along the path:

Definition 3.5.2. Let $s(p)$ be a section of E. Along a curve $\gamma : [0, 1] \to M$ we have $s(t) = [(\bar{\gamma}(t), \eta(t))]$, $\bar{\gamma}(t)$ an arbitrary horizontal lift. The *covariant derivative* of the section $s(t)$ along $\gamma(t)$ at $p_0 = \gamma(0)$ is defined by:

$$\nabla_X s = \left[\left(\bar{\gamma}(0), \; \frac{\mathrm{d}}{\mathrm{d}t}\, \eta(\gamma(t))\Big|_{t=0} \right)\right] \tag{3.67}$$

where X is the tangent vector to $\gamma(t)$ at p_0.

For this to be intrinsic, it shouldn't depend on the choice of lift. It is straightforward to check that if we take $\bar{\gamma}' = \bar{\gamma}a$ as a different representative lift, $\nabla_X s$ is unchanged. Hence $\nabla_X s$ depends only on X and $s \in \Gamma(M, E)$ but not on the choice of lift - equivalently, it depends on γ and ω, but not on the local trivialisation or anything else.

We've defined the covariant derivative at a point $p_0 = \gamma(0)$. But for X a vector field, we equally have a map

$$\nabla_X : \Gamma(M, E) \to \Gamma(M, E) : s \mapsto \nabla_X s. \tag{3.68}$$

Given $x \in \mathcal{X}(M)$ with $X(p) = X_p \in T_pM$, there is a curve $\gamma(t)$ with $\gamma(0) = p, \dot{\gamma}(0) = X$. Any lift $\bar{\gamma}$ lets us compute $\nabla_X s|_p = \nabla_{X_p} s$.

We can also define:

$$\nabla : \Gamma(M, E) \to \Gamma(M, E) \otimes \Omega^1(M) : s \mapsto \nabla s \tag{3.69}$$

where $\nabla s(X) = \nabla_X s$ for $s \in \Gamma(M, E)$. This map ∇ then satisfies the usual properties of a connection or covariant derivative on a manifold.

3.5.1 Local co-ordinates for the covariant derivative

Let $P(M, G)$ be a G-bundle, $E = P \times_\rho V$ the associated vector bundle. Take a local section $\sigma_i \in \Gamma(U_i, P)$ with canonical local trivialisation $\sigma_i(p) = \phi(p, e)$. Let $\gamma : [0, 1] \to U_i \subset M$ and $\bar{\gamma}$ be a horizontal lift. Write

$$\bar{\gamma}(t) = \sigma_i(t) g_i(t), \quad g_i(t) = g_i(\gamma(t)) \in G.$$

Now take a section of E, $\mathbf{e}_\alpha(p) = [(\sigma_i(p), {\mathbf{e}_\alpha}^0)]$ where ${\mathbf{e}_\alpha}^0$ is the fixed α^{th} basis vector of V; $({\mathbf{e}_\alpha}^0)^\beta = {\delta_\alpha}^\beta$. Then, as a function of the path parameter, the section is

$$\begin{aligned}\mathbf{e}_\alpha(t) &= [(\bar{\gamma}(t) g_i(t)^{-1}, {\mathbf{e}_\alpha}^0)] \\ &= [(\bar{\gamma}(t), g_i(t)^{-1} {\mathbf{e}_\alpha}^0)],\end{aligned} \tag{3.70}$$

so we see that $g_i(t)^{-1}$ acts on ${\mathbf{e}_\alpha}^0$ to account for the change of basis along γ. The covariant derivative of $\mathbf{e}_\alpha$ is

$$\begin{aligned}\nabla_X \mathbf{e}_\alpha &= \left[\left(\bar{\gamma}(0),\ \frac{\mathrm{d}}{\mathrm{d}t}\left[g_i(t)^{-1}{\mathbf{e}_\alpha}^0\right]\Big|_{t=0}\right)\right] \\ &= \left[\left(\bar{\gamma}(0),\ -g_i(t)^{-1}\,\frac{\mathrm{d}g_i(t)}{\mathrm{d}t}\,g_i(t)^{-1}{\mathbf{e}_\alpha}^0\Big|_{t=0}\right)\right] \\ &= [(\bar{\gamma}(0) g_i(0)^{-1}, \mathcal{A}_i(X){\mathbf{e}_\alpha}^0)],\end{aligned} \tag{3.71}$$

where we have made use of 3.34. Now the local expression is simply

$$\nabla_X \mathbf{e}_\alpha = [(\sigma_i(0), \mathcal{A}_i(X){\mathbf{e}_\alpha}^0)]. \tag{3.72}$$

$\mathcal{A}_i(X){\mathbf{e}_\alpha}^0$ is a vector, taking value in the Lie algebra, but how does it act? Write $\mathcal{A}_i = \mathcal{A}_{i\mu}\mathrm{d}x^\mu = {\mathcal{A}_{i\mu}}^\alpha{}_\beta\mathrm{d}x^\mu$, where ${\mathcal{A}_{i\mu}}^\alpha{}_\beta = {\mathcal{A}_{i\mu}}^\gamma (T_\gamma)^\alpha{}_\beta$ — that is, we expand $\mathcal{A}_{i\mu}$ over the basis $\{T_\gamma\}$ of the Lie algebra, and the α and β indices refer to the action of elements of the algebra. Hence,

$$\begin{aligned}\mathcal{A}_i(X){\mathbf{e}_\alpha}^0 &= \frac{\mathrm{d}x^\mu}{\mathrm{d}t}{\mathbf{e}_\beta}^0 {\mathcal{A}_{i\mu}}^\beta{}_\gamma {\delta_\alpha}^\gamma \\ &= \frac{\mathrm{d}x^\mu}{\mathrm{d}t}{\mathcal{A}_{i\mu}}^\beta{}_\alpha {\mathbf{e}_\beta}^0\end{aligned} \tag{3.73}$$

$$\begin{aligned}\Rightarrow \nabla_X \mathbf{e}_\alpha &= \left[\left(\sigma_i(0),\ \frac{\mathrm{d}x^\mu}{\mathrm{d}t}\ {\mathcal{A}_{i\mu}}^\beta{}_\alpha {\mathbf{e}_\beta}^0\right)\right] \\ &= \frac{\mathrm{d}x^\mu}{\mathrm{d}t}{\mathcal{A}_{i\mu}}^\beta{}_\alpha \mathbf{e}_\beta.\end{aligned} \tag{3.74}$$

More generally,

$$\nabla \mathbf{e}_\alpha = {\mathcal{A}_i}^\beta{}_\alpha \mathbf{e}_\beta, \tag{3.75}$$

and for co-ordinate curves,

$$\nabla_\mu \mathbf{e}_\alpha \equiv \nabla_{\frac{\partial}{\partial x^\mu}} \mathbf{e}_\alpha = {\mathcal{A}_{i\mu}}^\beta{}_\alpha \mathbf{e}_\beta. \tag{3.76}$$

It is straightforward to extend this to general sections of the vector bundle. Let $s(p) = [(\sigma_i(p), \xi_i(p))] = {\xi_i}^\alpha(p)\mathbf{e}_\alpha$, where $\xi_i(p) = {\xi_i}^\alpha(p){\mathbf{e}_\alpha}^0$. Then:

$$\nabla_X s = \left[\left(\sigma_i(0),\ \frac{\mathrm{d}\xi_i}{\mathrm{d}t} + \mathcal{A}_i(X)\xi_i\Big|_{t=0}\right)\right] = \frac{\mathrm{d}x^\mu}{\mathrm{d}t}\left(\frac{\partial {\xi_i}^\alpha}{\partial x^\mu} + {\mathcal{A}_{i\mu}}^\alpha{}_\beta {\xi_i}^\beta\right)\mathbf{e}_\alpha. \tag{3.77}$$

By the compatibility condition for $\mathcal{A}_j$ we find that ∇_X is independent of the local trivialisation. We have essentially answered the question we posed at 1.23: up to some fiddling with constants, we have rederived the expression for what we naïvely named the covariant derivative earlier.

Example: connection coefficients and the frame bundle

Let FM be the frame bundle over M, TM the associated vector bundle:

$$FM = P(M, \mathrm{GL}_m(\mathbb{R})) \quad TM = FM \times_\rho \mathbb{R}^m, \quad m = \dim M$$

where ρ is the matrix representation of $\mathrm{GL}_m(\mathbb{R})$. The Lie algebra $\mathfrak{gl}_m(\mathbb{R})$ is just $m \times m$ real matrices. Write. $\mathcal{A}_i = \Gamma^\alpha_{\mu\beta} \mathrm{d}x^\mu$, so

$$\nabla_\mu \mathbf{e}_\alpha = [(\sigma_i(0), \Gamma_\mu {\mathbf{e}_\alpha}^0)] = \Gamma^\beta_{\mu\alpha} \mathbf{e}_\beta.$$

For a general vector field or section

$$s(p) = [(\sigma_i(p), X_i(p))] = {X_i}^\alpha(p)\mathbf{e}_\alpha,$$

we have

$$\nabla_\mu s = \left(\frac{\partial}{\partial x^\mu} {X_i}^\alpha + \Gamma^\alpha_{\mu\beta} X^\beta \right) \mathbf{e}_\alpha.$$

It's clear now that the role of the indices on the connection coefficient are quite distinct: μ is the $\Omega^1(M)$ index, while α and β are the $\mathfrak{gl}_m(\mathbb{R})$ indices.

Example: a $\mathrm{U}(1)$ gauge field coupled to a complex scalar field ϕ

Let P be a $\mathrm{U}(1)$-bundle, $E = P \times_\rho \mathbb{C}$, $\rho : \mathrm{U}(1) \hookrightarrow \mathbb{C}$ naturally. Take a connection ω with $\mathcal{A}_i = \sigma_i^* \omega = \mathcal{A}_{i\mu} \mathrm{d}x^\mu$, where

$$\mathcal{A}_{i\mu} = \mathcal{A}_i \frac{\partial}{\partial x^\mu}$$

is the Maxwell potential. Take a curve $\gamma : [0, 1] \to M$ with $\dot{\gamma}(0) = X$. Given σ_i, a horizontal lift is

$$\bar{\gamma}(t) = \sigma_i(t) \mathrm{e}^{\mathrm{i}\varphi(t)}.$$

Take $1 \in \mathbb{C}$ to be a basis vector, so the basic section is

$$e = [(\sigma_i(p), 1)].$$

Let $\phi(p) = [(\sigma_i(p), \Phi(p))] = \Phi(p)e$ be a section of E: this is the complex scalar field. With respect to $\bar{\gamma}(t)$,

$$\phi(t) = \Phi(t)[(\bar{\gamma}(t), g(t)^{-1})]$$

where obviously $g(t) = \mathrm{e}^{\mathrm{i}\varphi(t)}$.

The covariant derivative is:

$$\begin{aligned}
\nabla_X \phi &= \frac{\mathrm{d}\Phi}{\mathrm{d}t}[(\bar{\gamma}(0), g(0)^{-1})] + \Phi(0)[(\bar{\gamma}(0), g(0)^{-1}\mathcal{A}_i(X))] \\
&= \left(\frac{\mathrm{d}\Phi}{\mathrm{d}t} + \mathcal{A}_{i\mu} \Phi \frac{\mathrm{d}x^\mu}{\mathrm{d}t} \right) e \\
&= X^\mu \left(\frac{\partial \Phi}{\partial x^\mu} + \mathcal{A}_{i\mu} \Phi \right) e.
\end{aligned}$$

Chapter 4

Gauge theories, symmetry breaking and monopoles

4.1 Gauge theories

Having developed the mathematics of connections on principal bundles, we henceforth physically identity the gauge potential with the local form of the connection, and the Yang-Mills field strength with the local form of the curvature. We have a geometric picture of the physical gauge theory.

4.1.1 Electromagnetism as a $\mathrm{U}(1)$ gauge theory

Maxwell's electromagnetism is captured by a U(1)-bundle over $\mathbb{R}^4$. Since the gauge group is Abelian and one-dimensional, we drop the algebra indices α and β and take ${f_{\alpha\beta}}^{\gamma} = 0$. By Corollary 2.2.4, the bundle is trivial — we need only one local trivialisation. Take $\mathcal{A} = \mathcal{A}_\mu \mathrm{d}x^\mu$ where $\mathcal{A}_\mu = \mathrm{i}A_\mu$, that is, differing from the usual Maxwell potential just by the Lie algebra factor. The field strength is

$$\begin{aligned} \mathcal{F} &= \mathrm{d}\mathcal{A} \\ \Rightarrow \mathcal{F}_{\mu\nu} &= \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu \end{aligned}$$

where $\mathcal{F} = \frac{1}{2}\mathcal{F}_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$. The Bianchi identity is

$$\mathrm{d}\mathcal{F} = \mathcal{F} \wedge \mathcal{A} - \mathcal{A} \wedge \mathcal{F} = 0 \tag{4.1}$$

which tells us $\mathcal{F}$ is exact — which is unsurprising as $\mathcal{F} = \mathrm{d}\mathcal{A}$ is closed. In components this gives us the symmetry

$$\partial_\rho \mathcal{F}_{\mu\nu} + \partial_\mu \mathcal{F}_{\nu\rho} + \partial_\nu \mathcal{F}_{\rho\mu} = 0. \tag{4.2}$$

If we define $*\mathcal{F}_{\mu\nu} = \frac{1}{2}\mathcal{F}^{\rho\sigma}\epsilon_{\rho\sigma\mu\nu}$ to be the Hodge dual of $\mathcal{F}_{\mu\nu}$, we see that in components the Bianchi identity can be written

$$\partial_\mu * \mathcal{F}^{\mu\nu} = 0. \tag{4.3}$$

Once again, we identify the components $\mathcal{F}_{\mu\nu} = \mathrm{i}F_{\mu\nu}$ with the electromagnetic field strength tensor. Then reading off the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ as

$$E_i = F_{0i}, \quad B_i = -\tfrac{1}{2}\epsilon_{ijk}F_{jk} \quad \text{(spatial indices)} \tag{4.4}$$

we obtain the homogeneous Maxwell equations

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \qquad\qquad \nabla \cdot \mathbf{B} = 0.$$

The dynamical equations require us to specify an action. A gauge-invariant Lagrangian is

$$\mathcal{L}_{\mathrm{M}} = -\frac{1}{4}\mathcal{F} \wedge *\mathcal{F} \tag{4.5}$$

which generates the *Maxwell action* $\mathcal{S}_{\mathrm{M}}[\mathcal{A}]$. In the absence of current, this is

$$\mathcal{S}_{\mathrm{M}}[\mathcal{A}] = -\frac{1}{4}\int_{\mathbb{R}^4} \mathcal{F} \wedge *\mathcal{F} = \frac{1}{4}\int_{\mathbb{R}^4} \mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}\mathrm{d}^4x = -\frac{1}{4}\int_{\mathbb{R}^4} F_{\mu\nu}F^{\mu\nu}\mathrm{d}^4x. \tag{4.6}$$

Varying $\mathcal{S}_{\mathrm{M}}[\mathcal{A}]$ with respect to $\mathcal{A}_\mu$ gives the equation of motion

$$\partial_\mu \mathcal{F}^{\mu\nu} = 0, \tag{4.7}$$

which is equivalent to the remaining Maxwell equations in the vacuum

$$\nabla \cdot \mathbf{E} = 0 \qquad\qquad \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = 0. \tag{4.8}$$

4.1.2 The Dirac monopole revisited

The choice of $\mathbb{R}^4$ as the base space for the U(1) theory above made the principal bundle trivial, and Poincaré's lemma ensured $\mathcal{F}$ was globally exact. More interesting things occur with a more interesting choice of base space.

For the Dirac monopole, we assume time-independence and use $\mathbb{R}^3\backslash\{0\}$, which we have noted is homotopically equivalent to S^2. The principal bundle is $P(S^2, \mathrm{U}(1))$. As in section 2.4.1, we cover S^2 with two charts:

$$U_{\mathrm{N}} = \{(\theta,\phi)|0 \leq \theta \leq \pi/2 + \epsilon, 0 \leq \phi < 2\pi\},$$
$$U_{\mathrm{S}} = \{(\theta,\phi)|\pi/2 - \epsilon \leq \theta \leq 0, 0 \leq \phi < 2\pi\}.$$

Let ω be a connection one-form on P. Take local sections $\sigma_{\mathrm{N}}, \sigma_{\mathrm{S}}$ on $U_{\mathrm{N}}, U_{\mathrm{S}}$ respectively. The local gauge potentials are

$$\mathcal{A}_{\mathrm{N}} = \sigma_{\mathrm{N}}^*\omega \qquad\qquad \mathcal{A}_{\mathrm{S}} = \sigma_{\mathrm{S}}^*\omega.$$

Take $\mathcal{A}_{\mathrm{N}}$ and $\mathcal{A}_{\mathrm{S}}$ to be of Wu-Yang form as in section 1.2.2:

$$\mathcal{A}_{\mathrm{N}} = \mathrm{i}g(1-\cos\theta)\,\mathrm{d}\theta \qquad\qquad \mathcal{A}_{\mathrm{S}} = -\mathrm{i}g(1+\cos\theta)\,\mathrm{d}\theta$$

where g is the strength of the monopole. Let t_{NS} be the transition function, defined on the equator $U_{\mathrm{N}} \cap U_{\mathrm{S}}$. We saw in section 2.4.1 that, as a map from S^1 to U(1), it is classified by the first homotopy group of the circle $\pi_1(\mathrm{U}(1)) = \mathbb{Z}$. Write

$$t_{\mathrm{NS}}(\phi) = \exp[\mathrm{i}\varphi(\phi)], \quad \varphi \in \mathbb{R}. \tag{4.9}$$

On $U_{\mathrm{N}} \cap U_{\mathrm{S}}$ the gauge potentials are related by

$$\mathcal{A}_{\mathrm{N}} = t_{\mathrm{NS}}^{-1}\mathcal{A}_{\mathrm{S}}t_{\mathrm{NS}} + t_{\mathrm{NS}}^{-1}\,\mathrm{d}t_{\mathrm{NS}} = \mathcal{A}_{\mathrm{S}} + \mathrm{i}\,\mathrm{d}\varphi. \tag{4.10}$$

Hence,

$$\mathrm{d}\varphi = -\mathrm{i}(\mathcal{A}_{\mathrm{N}} - \mathcal{A}_{\mathrm{S}}) = 2g\,\mathrm{d}\phi. \tag{4.11}$$

Integrating over S^1 we see that the change in φ is

$$\Delta\varphi = \int_{S^1} \mathrm{d}\varphi = \int_0^{2\pi} 2g\,\mathrm{d}\phi = 4\pi g. \tag{4.12}$$

For unique definition of the transition function, this must be a multiple of 2π, so

$$\Delta\varphi/2\pi = 2g \in \mathbb{Z} \tag{4.13}$$

which reproduces the quantisation condition of the Dirac monopole observed in section 1.2.3. The integer $2g$ represents the homotopy class of the bundle. We could also have obtained this, as at 1.2.2, by obtaining the total flux. Taking $F_{\mathrm{N}} = \mathrm{d}A_{\mathrm{N}}, F_{\mathrm{S}} = \mathrm{d}A_{\mathrm{S}}$, the total flux Φ is

$$\begin{aligned}\Phi &= \int_{S^2} \mathbf{B}\cdot\mathrm{d}\mathbf{S} = \int_{U_{\mathrm{N}}} \mathrm{d}A_{\mathrm{N}} + \int_{U_{\mathrm{S}}} \mathrm{d}A_{\mathrm{S}} \\ &= \int_{S^1} A_{\mathrm{N}} - \int_{S^1} A_{\mathrm{S}} = 2g\int_0^{2\pi} \mathrm{d}\phi = 4\pi g \end{aligned} \tag{4.14}$$

4.1.3 Non-Abelian gauge theories

The Abelian nature of U(1) simplified the gauge theory somewhat. However, non-Abelian gauge theories play an essential role in particle physics. The outline of the gauge theory is essentially the same, but now we must consider the commutator and representation of any fields. As developed in the previous chapter, the connection and curvature (hence also $\mathcal{A}$ and $\mathcal{F}$) are always in the adjoint representation.

Let G be a compact semi-simple Lie group such as SO(N) or SU(N). The Lie algebra $\mathfrak{g}$ has generators $\{T_\alpha\}$ satisfying the commutation relations

$$[T_\alpha, T_\beta] = f_{\alpha\beta}{}^{\gamma} T_\gamma \tag{4.15}$$

for structure constants $f_{\alpha\beta}{}^{\gamma}$; an element g of G near the identity can be expressed

$$g = \exp(\theta^\alpha T_\alpha). \tag{4.16}$$

Take a connection one-form ω, with local expression

$$\mathcal{A} = \mathcal{A}_\mu\,\mathrm{d}x^\mu = A_\mu{}^\alpha T_\alpha\,\mathrm{d}x^\mu. \tag{4.17}$$

The covariant derivative is

$$\mathcal{D}_\mu = \partial_\mu + \mathcal{A}_u \tag{4.18}$$

$$= \mathbb{I}\,\partial_\mu + T_\alpha A_\mu{}^\alpha \tag{4.19}$$

where $\mathbb{I}$ is the identity matrix for the representation. The Yang-Mills field strength is

$$\mathcal{F}_{\mu\nu} = \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu] \tag{4.20}$$

with components over the algebra basis

$$F_{\mu\nu}{}^\alpha = \partial_\mu A_\nu{}^\alpha - \partial_\nu A_\mu{}^\alpha + f_{\beta\gamma}{}^\alpha A_\mu{}^\beta A_\nu{}^\gamma. \tag{4.21}$$

We note that the field is self-interacting owing to the presence of the A^2 term, even in the absence of a source.

Under a gauge transform with transition function $g = g(p)$ our forms transform as expected:

$$\mathcal{A}'_\mu = g^{-1}\mathcal{A}_\mu g + g^{-1}\partial_\mu g$$
$$\mathcal{F}'_{\mu\nu} = g^{-1}\mathcal{F}^{\mu\nu} g$$

The Bianchi identity, $\mathcal{D}\mathcal{F} = 0$, in components is

$$\mathcal{D}_\rho \mathcal{F}_{\mu\nu} + \mathcal{D}_\mu \mathcal{F}_{\nu\rho} + \mathcal{D}_\nu \mathcal{F}_{\rho\mu} = 0. \tag{4.22}$$

Once again we can define the dual field strength $*\mathcal{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\mathcal{F}_{\rho\sigma}$ and the Bianchi identity is expressed

$$\mathcal{D}_\mu * \mathcal{F}^{\mu\nu} = \partial_\mu * \mathcal{F}^{\mu\nu} + [\mathcal{A}_\mu, *\mathcal{F}^{\mu\nu}] = 0. \tag{4.23}$$

For the dynamical equations of motion, a gauge-invariant action has Lagrangian

$$\mathcal{L}_{\rm YM} = -\frac{1}{2}\operatorname{tr}(\mathcal{F} \wedge *\mathcal{F}) = -\frac{1}{2}\operatorname{tr}(\mathcal{F}^{\mu\nu}\mathcal{F}_{\mu\nu}) \tag{4.24}$$

where the trace is taken with respect to the group matrix in the representation. At this point we choose our algebra basis to satisfy $\operatorname{tr}(T_\alpha T_\beta) = \frac{1}{2}\delta_{\alpha\beta}$, so that in components

$$\mathcal{L}_{\rm YM} = -\frac{1}{2}F^{\mu\nu\alpha}F_{\mu\nu}{}^{\beta}\operatorname{tr}(T_\alpha T_\beta) = -\frac{1}{4}F^{\mu\nu\alpha}F_{\mu\nu}{}^{\alpha}. \tag{4.25}$$

The equation of motion ($\mathcal{D} * \mathcal{F} = 0$) derived from the action with this Lagrangian is

$$\mathcal{D}_\mu \mathcal{F}^{\mu\nu} = \partial_\mu \mathcal{F}^{\mu\nu} + [\mathcal{A}_\mu, \mathcal{F}^{\mu\nu}] = 0 \tag{4.26}$$

or in components

$$\partial_\mu F^{\mu\nu\alpha} + f_{\beta\gamma}{}^{\alpha} A_\mu{}^{\beta} F^{\mu\nu\gamma} = 0. \tag{4.27}$$

We observe key physical differences from the Abelian case. As mentioned, the field is self-interacting, and so acts as a source for itself (note $\partial_\mu F^{\mu\nu a} \neq 0$), even in the absence of charge. The electromagnetic field itself did not carry any charge, but in e.g. Yang-Mills theory, the gauge field is "charged". This is similar to general relativity, where the gravitational field carries its own mass-energy, and couples to itself — i.e. self-gravitates.

There are further implications in the quantum theory of gauge fields, as mentioned in Ryder [7]. In the Abelian case, there is no third-order A coupling in the Lagrangian. A third-order coupling which is exhibited is $\phi^*\phi A$ which will tell us that the charged field emits gauge bosons, while the electromagnetic field, not being charged, does not emit photons. However, in the non-Abelian case, there is a third-order A coupling in the $[\mathcal{A}_\mu, \mathcal{F}_{\mu\nu}]$ terms, so the gauge field itself may emit gauge bosons.

Finally, we might consider the physical implications of the Bianchi identity in each case. In Abelian theories, interpreting the components of the field-strength tensor tells us $\nabla \cdot \mathbf{B} = 0$ so there are no magnetic monopoles. However, in the non-Abelian case, defining similarly the magnetic induction $\mathcal{B}_i = -\frac{1}{2}\epsilon_{ijk}\mathcal{F}_{jk}$, a vector in co-ordinate and internal space, we have $\nabla \cdot \mathcal{B} \neq 0$. Hence, there exist magnetic monopoles with the same charge as the field. In particular, if we introduce a scalar field in the same representation of the gauge group, $\boldsymbol{\phi}$, and define the "ordinary" magnetic field to be $\mathbf{B} = \boldsymbol{\phi} \cdot \mathcal{B}$, then explicitly $\nabla \cdot \mathbf{B} \neq 0$ and the theory has ordinary (internal scalar) magnetic monopoles. This aspect of the theory was noticed by 't Hooft and Polyakov, and we return to it later.

4.2 Spontaneous symmetry breaking

The process of spontaneous symmetry breaking is important in many physical systems — for example, spin alignment in ferromagnetism — but we can picture it simply. Consider a pen balanced on end on a table. While upright, the system has obvious rotational symmetry. However, being balanced on end is not the *ground state* of the system, and is unstable. In fact the ground state is *degenerate*; the pen could fall over in any direction — but once it does, a particular phase is singled out and the earlier symmetry is broken. The direction in which it fell is random, and related to the other degenerate possibilities by rotations.

A similar phenomenon is observed in ferromagnetism. Atoms in a ferromagnet interact through a rotationally-invariant spin-spin interaction, but their (degenerate) ground state is one in which all the spins within a domain are aligned, which is no longer rotationally symmetric. This spontaneous magnetisation is random, and the degenerate ground states are reached by rotation again. In this case, the symmetry-breaking depends on temperature; at high temperatures the ground state becomes symmetric once more as the atoms will be randomly oriented.

The essential aspects of the phenomenon are the same in each case:

1. A parameter of the system assumes a critical value, beyond which point
2. the symmetric configuration is no longer stable, and
3. the ground state is degenerate.

The phenomenon of spontaneous symmetry breaking was introduced to condensed matter physics around 1960 by Anderson. Shortly afterwards, Brout, Englert and Higgs pointed out near-simultaneously[1] that the implications of spontaneous symmetry breaking to gauge theories were quite different to those of other theories. We develop this, following Ryder [7].

We turn to field theory and look for a system in which a symmetry of the Lagrangian is not shared by the ground state solution, i.e. the vacuum. We want a Lagrangian that has a symmetry to begin with, so we start with complex ϕ^4 theory and a Lagrangian of the form

$$\begin{aligned}\mathcal{L} &= (\partial_\mu \phi)(\partial^\mu \phi^*) - m^2 \phi^* \phi - \lambda(\phi^* \phi)^2 && (4.28)\\ &= (\partial_\mu \phi)(\partial^\mu \phi^*) - V(\phi, \phi^*). && (4.29)\end{aligned}$$

The λ term is a self-interaction. In the usual scalar field theory, the m^2 term is associated with particles of mass m, but here we will regard m^2 as a parameter only, and in particular we will consider the implications of allowing it to be negative. We note that $\mathcal{L}$ is invariant under the *global* transformation

$$\phi \mapsto \exp(\mathrm{i}\Lambda)\phi \quad (\Lambda \text{ constant}). \tag{4.30}$$

The ground state is obtained by minimising the potential V. We have

$$\frac{\partial V}{\partial \phi} = m^2 \phi^* + 2\lambda \phi^* (\phi^* \phi) \tag{4.31}$$

so if we had, as usual, $m^2 > 0$, the minimum occurs at $\phi^* = \phi = 0$. However, when we let $m^2 < 0$ there is now a local maximum at $\phi = 0$ and a minimum along

$$|\phi|^2 = -\frac{m^2}{2\lambda} = a^2, \tag{4.32}$$

[1] As did Guralnik, Hagen and Kibble. Higgs is said to prefer to call it the ABEGHHK'tH mechanism, for Anderson, Brout, Englert, Guralnik, Hagen, Higgs, Kibble and 't Hooft.

i.e. at $|\phi| = a$. We will not consider quantising our fields, but it is useful to import the terminology and refer to this as the *vacuum expectation value* of the field. These vacua are degenerate. If we consider a decomposition of ϕ into fields ϕ_1 and ϕ_2 satisfying $\phi = \phi_1 + \mathrm{i}\phi_2$, they lie along the circle $|\phi| = a$, and are related to each other by rotations.

We might work in polar co-ordinates henceforth, setting

$$\phi(x) = \rho(x)\mathrm{e}^{\mathrm{i}\theta(x)} \tag{4.33}$$

so that our complex field is expressed in two real scalar fields. We *choose* out of the degenerate vacua one in particular to be the physical one: we take

$$\phi = a \quad (a \in \mathbb{R}) \tag{4.34}$$

$$\Rightarrow \rho = a, \quad \theta = 0. \tag{4.35}$$

We shift to new fields

$$\phi(x) = [\rho'(x) + a]\mathrm{e}^{\mathrm{i}\theta(x)} \tag{4.36}$$

so that now ρ' and θ both have vanishing vacuum expectation values. We consider these to be our physical fields, and want to rewrite the Lagrangian so it is expressed in terms of these fields. From 4.29 and 4.32 the potential is now

$$V = \lambda(\phi^*\phi - a^2)^2 - \lambda a^4$$

and the kinetic term is

$$(\partial_\mu\phi)(\partial^\mu\phi^*) = (\partial_\mu\rho')(\partial^\mu\rho') + (\rho' + a)^2(\partial_\mu\theta)(\partial^\mu\theta).$$

Inspecting our Lagrangian $\mathcal{L} = (\partial_\mu\phi)(\partial^\mu\phi^*) - V(\phi, \phi^*)$, we see that there is a term in ρ'^2, so ρ' has a mass given by

$$m^2_{\rho'} = 4\lambda a^2 \tag{4.37}$$

— but now there is no term in θ^2, so θ is a *massless field*. As a result of the spontaneous symmetry breaking, what would have been two *massive* fields (the real components of ϕ) have become one massive and one massless field. Excitations away from the vacuum state cost energy to move in the radial direction, but are unopposed in the angular direction. Quantisation will associate a massless particle with the θ field, called the *Goldstone boson*.

This result is general: spontaneous symmetry breaking of a continuous symmetry causes the existence of a massless particle, the Goldstone particle. This statement is Goldstone's theorem, which we will prove in the following section.

For future reference, we can also perform the same analysis of the ground state using a "Cartesian" rather than polar decomposition of ϕ. In place of 4.33, we write

$$\phi(x) = a + \frac{\phi_1(x) + \mathrm{i}\phi_2(x)}{\sqrt{2}} \tag{4.38}$$

so that ϕ_1 and ϕ_2 now have vacuum expectation value zero. After some algebra, the (non-constant) terms of the Lagrangian are

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi_1)^2 + \tfrac{1}{2}(\partial_\mu\phi_2) - 2\lambda a^2\phi_1^2 - \sqrt{2}\lambda\phi_1(\phi_1^2 + \phi_2^2) - \frac{\lambda}{4}(\phi_1^2 + \phi_2^2). \tag{4.39}$$

Hence, of our chosen physical fields, one (ϕ_2) is again massless, whereas one is massive with mass $(m_{\phi_1})^2 = 4\lambda a^2$ — entirely matching our analysis in polar co-ordinates.

4.3 Goldstone's theorem

In the previous section, we note that the real component fields of ϕ formed a two-dimensional representation of U(1), and that the Lagrangian had U(1) symmetry. One of the representative fields had a non-vanishing vacuum expectation value, and in the end there was one massless particle, which we called the Goldstone boson, and one massive one. We wish to generalise this argument: if the Lagrangian $\mathcal{L}$ is invariant under some symmetry group G, how many Goldstone bosons will there be?

4.3.1 Non-Abelian global symmetries

We first consider the case of a more general global (not gauge!) symmetry, by choosing a specific non-Abelian example. We will consider SO(3), and now take our scalar field ϕ_i ($i = 1, 2, 3$) to live in the fundamental representation of SO(3)[2] — that is, it is a *Lorentz*-scalar but carries an internal SO(3) symmetry. In the literature, objects in such a representation are sometimes called *isovectors* and its orientation in the representation space is referred to as *isospin*. Physically, isospin is a "quantum number" associated with the strong force which obeys a global (not gauge) SU(2) symmetry. We will generally refer to it as internal symmetry, internal space, &c. instead.

The Lagrangian is

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_i)(\partial^\mu \phi_i) - \frac{m^2}{2}\phi_i\phi_i - \lambda(\phi_i\phi_i)^2 \tag{4.40}$$

(with summation convention), and is invariant under internal rotations, which generate the symmetry group $G = \mathrm{SO}(3)$:

$$\begin{aligned} G : \phi_i \mapsto\ & \mathrm{e}^{\mathrm{i}Q_k\alpha_k}\phi_i\mathrm{e}^{-\mathrm{i}Q_k\alpha_k} \\ &= (\mathrm{e}^{\mathrm{i}T_k\alpha_k})_{ij}\phi_j = U_{ij}\phi_j = [U(g)\phi]_i. \end{aligned} \tag{4.41}$$

Here the α_i are rotation angles in internal space, the Q_i are generators of the group, and the T_i are matrices obeying the Lie algebra of the group, necessarily of the same dimension as the representation (so three-dimensional in this case). The matrix $U(g)$ corresponds to the particular group element g, and is unitary if the T_i are Hermitian — this is not an essential requirement here, but would be necessary to consider the quantum case.

Once again we look for minima of the potential,

$$V = \frac{m^2}{2}\phi_i\phi_i + \lambda(\phi_i\phi_i)^2 \tag{4.42}$$

which occurs at $\phi_i = 0$ for $m^2 > 0$. When $m^2 < 0$, the minimum occurs for

$$|\phi_0| = (\phi_1^2 + \phi_2^2 + \phi_3^2)^{\frac{1}{2}} = \left(\frac{-m^2}{4\lambda}\right) \equiv a \tag{4.43}$$

Once again we see that we have degenerate vacua and are free to choose one of them as physical. We take

$$\phi_0 = a\hat{\mathbf{e}}_3 \tag{4.44}$$

so the vacuum value of ϕ, ϕ_0, points in the $\hat{\mathbf{e}}_3$ direction in internal space. It's clear that now ϕ_0 is not invariant under the entire symmetry group G: rotations about other axes in internal space will

[2] Or the adjoint representation of SU(2), though we should be aware of the double cover of SU(2) over SO(3).

not preserve it. There is a subgroup $H < G$ under which it is invariant: namely, rotations about this axis,

$$H = \left\{ h \in G \mid U(h) = \mathrm{e}^{\mathrm{i}T_3\alpha_3} \right\}. \tag{4.45}$$

However, by construction, V certainly was invariant under the entirety of G, and it's this that will give rise to Goldstone bosons. We wish to determine how many there are.

As in the previous section, we rewrite our Lagrangian in terms of physical fields with non-zero vacuum expectation values. Letting

$$\phi_3 = \chi + a \tag{4.46}$$

the potential is rewritten in terms of ϕ_1, ϕ_2 and χ, using 4.43, as

$$V = \lambda[(\phi_i\phi_i - a^2)^2 - a^4]. \tag{4.47}$$

Originally, each component ϕ_i field was massive. Now, χ is the only field with a quadratic term, and its characteristic mass is $m_\chi^2 = 8a^2\lambda$. The ϕ_1 and ϕ_2 fields are massless, and give rise to Goldstone bosons after spontaneous symmetry breaking.

4.3.2 Generalising to Goldstone's theorem

We can understand our treatment of the vacuum, i.e. the minimum of the potential, in a very general way. We will expand $V(\phi)$ about its minimum value according to Taylor's theorem. At the minimum,

$$\left.\frac{\partial V}{\partial \phi_a}\right|_{\phi=\phi_0} = 0, \tag{4.48}$$

so we expand as

$$V(\phi) = V(\phi_0) + \frac{1}{2}\left(\frac{\partial^2 V}{\partial\phi_i\partial\phi_j}\right)_{\phi=\phi_0} \chi_i\chi_j + O(\chi^3), \tag{4.49}$$

where $\chi_i(x) = \phi_i(x) - \phi_0$. Since ϕ_0 is a minimum, the *mass matrix* is non-negative:

$$M_{ij} = \left(\frac{\partial^2 V}{\partial\phi_i\partial\phi_j}\right)_{\phi=\phi_0} \geq 0. \tag{4.50}$$

For what fields is it zero — i.e., for what fields do we have Goldstone bosons? We use the invariance of V under G.

$$V(\phi_0) = V(U(g)\phi_0) = V(\phi_0) + \frac{1}{2}\left(\frac{\partial^2 V}{\partial\phi_i\partial\phi_j}\right)_{\phi=\phi_0} \delta\phi_i\delta\phi_j + O(\delta\phi^3), \tag{4.51}$$

where the $\delta\phi_i$ are variations in the fields ϕ_i due to the group transformation. Invariance demands that the second term in the above is zero:

$$\left(\frac{\partial^2 V}{\partial\phi_i\partial\phi_j}\right)_{\phi=\phi_0} \delta\phi_i\delta\phi_j = 0. \tag{4.52}$$

What are the variations? It's clear from our earlier discussion that they depend on whether or not the group element is in the subgroup of symmetries of ϕ_0. If $g \in H$, $U(g)\phi_0 = \phi_0$, so $\delta\phi_i = 0$, or equivalently,

$$\delta\phi = \left(\frac{\partial U}{\partial\alpha_3}\right)_{\alpha_3=0} \delta\alpha_3 = 0, \tag{4.53}$$

and condition 4.52 is automatically satisfied. On the other hand, if $g \notin H$, we have

$$\delta\phi_m = \left[\left(\frac{\partial U}{\partial \alpha_i}_{\alpha_i=0}\right)\phi_0\right]_m \delta\alpha_i \neq 0, \tag{4.54}$$

so for $M_{ij}[U'(0)\phi_0] = 0$ we must have zero mass in the field $U'(0)\phi_0$. This is the statement of Goldstone's theorem: a field whose mass is not *required* to be zero obeys 4.52; and the number of such fields is dim $\mathfrak{h}$. (In our previous example, $H = \mathrm{SO}(2) \cong \mathrm{U}(1)$.) Elements $g \notin H$ form cosets G/H; hence, the number of Goldstone bosons is dim G/H.

4.4 Spontaneous symmetry breaking for gauge symmetries

4.4.1 The Abelian case

We return to the case of U(1) symmetry, and now promote it to a gauge symmetry:

$$\phi \mapsto \mathrm{e}^{\mathrm{i}\Lambda(x)}\phi \tag{4.55}$$

Coupling the complex scalar field ϕ to the gauge field (with coupling constant given by the electron charge e) is done by replacing the co-ordinate derivative with the covariant derivative and introducing the gauge field term. This yields the Lagrangian

$$\mathcal{L} = (\partial_\mu + \mathrm{i}eA_\mu)\phi(\partial^\mu - \mathrm{i}eA^\mu)\phi^* - m^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{4.56}$$

As before, when we allow $m^2 < 0$, the minimum (vacuum) is attained at, in the *absence* of the gauge field,

$$|\phi| = a \equiv \left(\frac{-m^2}{2\lambda}\right)^{\frac{1}{2}}. \tag{4.57}$$

Following the "Cartesian" prescription, we set

$$\phi(x) = a + \frac{\phi_1(x) + \mathrm{i}\phi_2(x)}{\sqrt{2}}, \tag{4.58}$$

where, a priori, we think that our physical fields are ϕ_1 and ϕ_2. Rewriting the Lagrangian as before, we find that

$$\begin{aligned}\mathcal{L} = &-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + e^2a^2A_\mu A^\mu + \tfrac{1}{2}(\partial_\mu\phi_1)^2 + \tfrac{1}{2}(\partial_\mu\phi_2)^2 \\ &- 2\lambda a^2\phi_1^2 + \sqrt{2}eaA^\mu\partial_\mu\phi_2 + \text{cubic} + \text{quartic terms.}\end{aligned}$$

There are several remarks to be made at this point. Clearly, ϕ_1 is still massive. However, rather oddly, it seems that the photon has *become massive*. Further, the final term we've written suggests that ϕ_2 and A_μ *mix* under propagation. We regard this behaviour as unphysical and reassure ourselves by noticing that we can gauge the errant ϕ_2 field away: for small enough Λ, we approximate 4.55 by the linear term, and obtain from 4.58

$$\left.\begin{aligned}\phi_1' &= \phi_1 - \Lambda\phi_2 \\ \phi_2' &= \phi_2 + \Lambda\phi_1 + \sqrt{2}\Lambda a\end{aligned}\right\}. \tag{4.59}$$

We observe that, like A_μ, ϕ_2 undergoes an inhomogeneous transformation (a rotation and a translation in the (ϕ_1, ϕ_2) plane), so we regard it as not having any direct physical interpretation. We choose our gauge such that $\phi_2 = 0$, which is called the *physical* or *unitary gauge*; in particular, the mixed term disappears, and the Lagrangian becomes

$$\begin{aligned}\mathcal{L} = &-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + e^2a^2A_\mu A^\mu + \tfrac{1}{2}(\partial_\mu\phi_1)^2 \\ &- 2\lambda a^2\phi_1^2 + \text{coupling terms.}\end{aligned} \tag{4.60}$$

It now contains only two fields, the photon and ϕ_1, and they are both massive. Recall that in spontaneous breaking of the global symmetry, the field ϕ_2 field became massless and was associated with the Goldstone boson, but here it has totally disappeared. Moreover, the photon, which was massless when introduced as the gauge field, has become massive owing to the broken symmetry. This is the *Brout-Englert-Higgs phenomenon.* It is summarised in the Abelian model by saying that spontaneous breaking of a gauge symmetry results not in a massless Goldstone boson, but the total disappearance of that field.

- *Goldstone mode* (spontaneous breaking of global U(1) symmetry):
  2 massive scalar fields $\to$ 1 massive scalar field + 1 massless scalar field

- *BEH mode* (spontaneous breaking of gauge U(1) symmetry):
  $\left.\begin{array}{l}\text{2 massive scalar fields} \\ \text{+ 1 photon}\end{array}\right\} \to \left\{\begin{array}{l}\text{1 massive scalar field} \\ \text{+ 1 massive photon}\end{array}\right.$

We note that the degrees of freedom are preserved in each case. This is trivial in the Goldstone case as all fields have one degree of freedom. In the Higgs case, the massless photon has two degrees of freedom: Ryder [7] presents a polarisation argument; alternatively it starts with four, but gauge freedom fixes one, and another is fixed through the fact that the gauge field Lagrangian $\mathcal{L}_\mathrm{M} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ has no kinetic term in A_0, and hence the A_0 field is *not dynamical* and is fixed by initial data in the other fields. On the other hand, the massive photon has three degrees of freedom, because ([7]) it has a physical transverse polarisation state. We therefore say that the photon has *eaten* a degree of freedom from the scalar field, and, in doing so, acquires mass.

4.4.2 The non-Abelian Brout-Englert-Higgs phenomenon

We illustrate spontaneous breaking of a non-Abelian gauge symmetry by returning to the example of O(3). The structure constants for the algebra $\mathfrak{o}(3)$ are given by the Levi-Civita symbol ϵ_{ijk}, and we take the coupling constant to be g. With triplet scalar Higgs field ϕ_i, the Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(\mathcal{D}_\mu\phi_i)(\mathcal{D}^\mu\phi_i) - \frac{m^2}{2}\phi_i\phi_i - \lambda(\phi_i\phi_i)^2 - \frac{1}{4}F_{\mu\nu}{}^iF^{\mu\nu i}, \tag{4.61}$$

where

$$\mathcal{D}_\mu\phi_i = \partial_\mu\phi_i + g\epsilon_{ijk}A_\mu{}^j\phi_k, \tag{4.62}$$

$$F_{\mu\nu}{}^i = \partial_\mu A_\nu{}^i - \partial_\nu A_\mu{}^i + g\epsilon^{ijk}A_\mu{}^jA_\nu{}^k. \tag{4.63}$$

When $m^2 < 0$, the potential is minimised in the absence of the gauge field at

$$|\phi_0| = a \equiv \left(\frac{-m^2}{2\lambda}\right)^{\frac{1}{2}}. \tag{4.64}$$

Once again, we choose

$$\phi_0 = a\hat{\mathbf{e}}_3. \tag{4.65}$$

The physical fields are ϕ_1, ϕ_2 and $\chi = \phi_3 - a$. Relevant terms in the Lagrangian are

$$\begin{aligned}\mathcal{L} =& \tfrac{1}{2}[(\partial_\mu\phi_1)^2 + (\partial_\mu\phi_2)^2 + (\partial_\mu\chi)^2] + ag[(\partial_\mu\phi_1)A^{\mu 2} - (\partial_\mu\phi_2)A^{\mu 1}] \\ &+ \tfrac{a^2g^2}{2}[({A_\mu}^1)^2 + ({A_\mu}^2)^2] - \tfrac{1}{4}(\partial_\mu {A_\nu}^i - \partial_\nu {A_\mu}^i)^2 \\ &- 4a^2\lambda\chi^2 + \text{cubic} + \text{quartic terms}.\end{aligned} \tag{4.66}$$

We again observe mixing of the ϕ_1 and ϕ_2 fields with A_μ. We gauge so that at *every* point, $\boldsymbol{\phi}(x) = \hat{\mathbf{e}}_3\phi_3 = \hat{\mathbf{e}}_3(\chi + a)$. This eliminates ϕ_1 and ϕ_2, leaving

$$\left.\begin{aligned}\mathcal{D}_\mu\phi_1 &= g(a+\chi){A_\mu}^2 \\ \mathcal{D}_\mu\phi_2 &= -g(a+\chi){A_\mu}^1 \\ \mathcal{D}_\mu\phi_3 &= \partial_\mu\chi\end{aligned}\right\} \tag{4.67}$$

$$\Rightarrow (\mathcal{D}_\mu\phi_i)^2 = a^2g^2[({A_\mu}^1)^2 + ({A_\mu}^2)^2] + (\partial_\mu\chi). \tag{4.68}$$

The Lagrangian is reduced to

$$\begin{aligned}\mathcal{L} =& -\tfrac{1}{4}(\partial_\mu {A_\nu}^i - \partial_\nu {A_\mu}^i)^2 - \tfrac{1}{2}a^2g^2[({A_\mu}^1)^2 + ({A_\mu}^2)^2] \\ &+ \tfrac{1}{2}(\partial_\mu\chi)^2 - 4a^2\lambda\chi^2 + \text{cubic} + \text{quartic terms}.\end{aligned} \tag{4.69}$$

We have one massive scalar field (χ), two massive vector fields (${A_\mu}^1$ and ${A_\mu}^2$), and one massless vector field (${A_\mu}^3$). Once again, the Goldstone bosons have disappeared and the gauge bosons have acquired mass. To summarise the spontaneous breaking of non-Abelian symmetry:

- *Goldstone mode* (global O(3) symmetry):
  3 massive scalar fields → 1 massive scalar field + 2 massless scalar fields
- *BEH mode* (gauge O(3) symmetry):

$$\left.\begin{array}{l}\text{3 massive scalar fields} \\ \text{+ 3 massless vector fields}\end{array}\right\} \rightarrow \left\{\begin{array}{l}\text{1 massive scalar field} \\ \text{+ 2 massive vector fields} \\ \text{+ 1 massless vector field}\end{array}\right.$$

Degrees of freedom are once again conserved; following comments in the Abelian case, we see that $3 + 3 \times 2 = 9 = 1 + 2 \times 3 + 2$. Again, the vector fields that acquire mass have each eaten one degree of freedom from the massive scalar fields they eliminate.

One massive vector field remains because the subgroup that remains unbroken is $H = \mathrm{U}(1)$, which has one generator. By Goldstone's theorem, the number of massless vector fields remaining is dim H, while the number of eating fields (or missing Goldstone bosons) is dim G/H. Hence, the total number of gauge bosons remains at dim G — which we might expect, as the gauge field is in the regular representation of the internal symmetry group. In fact, the scalar field which survives does so because we chose the fields ϕ_i at the outset to be an internal triplet.

4.5 Monopoles in non-Abelian gauge theories

4.5.1 The 't Hooft-Polyakov monopole

In a non-Abelian gauge theory, with the introduction of spontaneous symmetry breaking, the field equations yield a solution corresponding to magnetic charge, and magnetic monopoles exist.

Nonetheless, the matter and gauge fields carry electric charge only: the origin of magnetic charge is topological.

Consider again an O(3) symmetry group with triplet gauge field ${F_{\mu\nu}}^a$ coupled (with coupling constant the electron charge e) to a triplet scalar Higgs field ϕ^a. The Lagrangian is

$$\mathcal{L} = -\frac{1}{4}{F_{\mu\nu}}^a F^{\mu\nu a} + \frac{1}{2}(\mathcal{D}_\mu\phi^a)(\mathcal{D}^\mu\phi^a) - \frac{m^2}{2}\phi^a\phi^a - \lambda(\phi^a\phi^a)^2, \tag{4.70}$$

where the field and covariant derivative are

$$\begin{aligned} {F_{\mu\nu}}^a &= \partial_\mu {A_\nu}^a - \partial_\nu {A_\mu}^a + e\epsilon^{abc}{A_\mu}^b{A_\nu}^c, &(4.71)\\ \mathcal{D}_\mu\phi^a &= \partial_\mu\phi^a + e\epsilon^{abc}{A_\mu}^b\phi^c. &(4.72)\end{aligned}$$

Following our discussion of the BEH mechanism, we will primarily be interested in potential functions which have minima at $|\phi|^2 = \phi^a\phi^a = a^2 \neq 0$, so we choose

$$V(|\phi|^2) = \lambda(|\phi|^2 - F^2)^2, \quad F^2 = -\frac{m^2}{4\lambda}. \tag{4.73}$$

Now the Lagrangian

$$\mathcal{L} = -\frac{1}{4}{F_{\mu\nu}}^a F^{\mu\nu a} + \frac{1}{2}(\mathcal{D}_\mu\phi^a)(\mathcal{D}^\mu\phi^a) - V(|\phi|^2) \tag{4.74}$$

matches the previous at 4.70 up to a constant term. We note that we could always perform a gauge transform such that the time component of the potential is zero. In such a gauge, $\mathcal{D}_0\phi^a = \dot{\phi}^a$ and ${F_{0i}}^a = \dot{A}_i{}^a$. Then the energy of this solution is obtained (recall 1.44) from the integral

$$E = \int \mathrm{d}^3x \left[\frac{1}{2}|\dot{\phi}|^2 + \frac{1}{2}(\mathcal{D}_i\phi^a)(\mathcal{D}^i\phi^a) + \frac{1}{2}\dot{A}_i{}^a\dot{A}^{ia} + \frac{1}{4}{F_{ij}}^a F^{ija} + V(|\phi|^2)\right]. \tag{4.75}$$

For finite energy as $r = |\mathbf{x}| \to \infty$ we must have

$$V(|\phi|^2) \to 0, \quad |\mathcal{D}_i\phi^a|^2 = o(r^{-3}). \tag{4.76}$$

A general treatment of these conditions can be found in Fry's notes [3] but we consider static solutions where the gauge potential takes the non-trivial asymptotic form

$$A_i{}^a = -\epsilon_{iab}\frac{r^b}{er^2} \quad (r\to\infty), \qquad {A_0}^a = 0. \tag{4.77}$$

(Note that the solution being static assumes more than that we can simply gauge ${A_0}^a$ away.) We further assume the scalar field has the asymptotic form

$$\phi^a = F\frac{r^a}{r} \quad (r\to\infty) \tag{4.78}$$

In a manner of speaking, the field is radial; the scalar field, for example, in the x-direction in space only has internal 1-component, and so on. Polyakov [5] called this the "hedgehog solution".

't Hooft [9] has shown that regular solutions with these asymptotic forms exist. The equations of motion for ϕ are seen from 4.70 to be

$$4\lambda(F - |\phi|^2)\phi^a = \mathcal{D}_\mu(\mathcal{D}^\mu\phi^a). \tag{4.79}$$

The asymptotic form of ϕ (4.78) implies $|\phi| = F$, so the LHS of the above vanishes at infinity. We observe also that $\mathcal{D}_\mu \phi^a$ vanishes at infinity: for spatial i,

$$\begin{aligned}
\mathcal{D}_i \phi^a &= F\partial_i \left(\frac{r^a}{r}\right) + e\epsilon^{abc} A_i{}^b F \frac{r^c}{r} \\
&= F\left(\frac{\delta^{ia}}{r} - \frac{r^i r^a}{r^3}\right) - \epsilon^{abc}\epsilon_{ibm} F \frac{r^m r^c}{r^3} \\
&= 0.
\end{aligned} \tag{4.80}$$

Hence, at infinity, ϕ is covariantly constant and must take on its vacuum value, but now with the topologically non-trivial boundary condition 4.78, rather than the usual assumptions we made previously (cf. 4.65) that the vacuum was $\phi^{1,2} = 0, \phi^3 \neq 0$. However, we will shortly see that $F_{\mu\nu}{}^a$ will *not* be zero at infinity: in fact, there will be a radial magnetic field.

We wish to generalise the definition of the electromagnetic field $F_{\mu\nu}$ so that in the usual case (of vacuum behaviour) of the scalar field ϕ, we get the standard electromagnetic field. Define:

$$F_{\mu\nu} = \frac{1}{|\phi|}\phi^a F_{\mu\nu}{}^a - \frac{1}{e|\phi|^3}\epsilon_{abc}\phi^a(\mathcal{D}_\mu\phi^b)(\mathcal{D}_\nu\phi^c). \tag{4.81}$$

If we now set

$$\begin{array}{lcl} A_\mu{}^{1,2} & = & 0 \\ A_\mu{}^3 & = & A_\mu \neq 0 \\ \phi^{1,2} & = & 0 \\ \phi^3 & = & F \neq 0 \end{array}, \tag{4.82}$$

this clearly reduces to the usual electromagnetic field, with the only contribution from the $a = 3$ case in the first term. If we now generalise the gauge potential similarly, according to

$$A_\mu = \frac{1}{|\phi|}\phi^a A_\mu{}^a, \tag{4.83}$$

we obtain the expression

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \frac{1}{e|\phi|^3}\epsilon_{abc}\phi^a(\mathcal{D}_\mu\phi^b)(\mathcal{D}_\nu\phi^c). \tag{4.84}$$

This is similar to, if more complicated than, our usual definition of the electromagnetic field - but we note that it reduces to what we want in the usual case where ϕ is fixed in internal space. Now inserting the asymptotic behaviour from 4.77 and 4.78, we see immediately (by symmetry) that $A_\mu = 0$ — which tells us that all of the field is contributed by the Higgs field! Furthermore, we observe after a little more work that

$$F_{0i} = 0, \qquad F_{ij} = -\frac{1}{er^3}\epsilon_{ijk}r^k. \tag{4.85}$$

This corresponds precisely to a radial magnetic field,

$$B_k = \frac{r^k}{er^3}. \tag{4.86}$$

Integrating over the sphere at infinity, we find the flux to be

$$\Phi = \frac{4\pi}{e}. \tag{4.87}$$

As the total solid angle of the sphere at infinity is 4π, we see that the magnetic charge g is such that $eg = 1$. By comparison with the Dirac monopole, this is twice the Dirac unit — cf. section 4.1.2.

We have established that the configuration of gauge and scalar fields with the specified asymptotic behaviour carried a magnetic charge: viewed from infinity, there is a radial magnetic field. 't Hooft showed that this field is everywhere non-singular, and therefore has finite energy. He estimates the monopole mass to be $\sim \frac{4\pi}{e^2} M_W \approx 137 M_W$, where M_W is a typical vector boson mass — this is extremely heavy and would make the monopole difficult to detect.

Nonetheless, at this point, the origin of the magnetic charge is still a little unclear. How do fields carrying only electric charge arrange to simulate magnetic charge? We define the magnetic current as

$$K^\mu = \partial_\nu * F^{\mu\nu} = \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} \tag{4.88}$$

by extension of the homogeneous Maxwell equations. Then, from 4.84, we have

$$K^\mu = -\frac{1}{2e}\epsilon^{\mu\nu\rho\sigma}\epsilon_{abc}\partial_\nu\hat{\phi}^a\partial_\rho\hat{\phi}^b\partial_\sigma\hat{\phi}^c, \tag{4.89}$$

where

$$\hat{\phi}^a = \frac{1}{|\phi|}\phi^a. \tag{4.90}$$

Hence, the magnetic current depends only on the Higgs field, as observed at 4.85. Moreover, the current is identically conserved:

$$\partial_\mu K^\mu = 0. \tag{4.91}$$

This conservation does not follow from a symmetry of the Lagrangian — it's not a Noether current. We might anticipate instead that it depends simply on the non-trivial boundary conditions. The conserved charge is

$$\begin{aligned} M &= \frac{1}{4\pi}\int K^0\,\mathrm{d}^3x \\ &= -\frac{1}{8\pi e}\oint_{S^2}\epsilon_{ijk}\epsilon_{abc}\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c(\mathrm{d}^2S)_i, \end{aligned} \tag{4.92}$$

where S^2 is the sphere at infinity, the boundary of the static field configuration ϕ. ϕ must be single-valued, so as $(\mathrm{d}^2S)_i$ covers the sphere once, the internal space of ϕ will be covered an integral number of times, say d. (For example, we could relax our definition of $\hat{\phi}$ to $(\cos d\phi\,\sin\theta, \sin d\phi\,\sin\theta, \cos\theta)$.) Using the asymptotic behaviour of ϕ, the integral is $8\pi d$, so ultimately

$$M = \frac{d}{e}, \quad d \in \mathbb{Z}. \tag{4.93}$$

$\hat{\phi}$ is the unit vector in the field space so it describes a sphere[3] S^2 in field space. The boundary describes a mapping of S^2 in co-ordinate space to the $\hat{\phi}$ manifold, and d is the *Brouwer degree* of the mapping, which is necessarily integral — it displays the topological nature of the 't Hooft-Polyakov monopole.

In 't Hooft's model, the non-Abelian gauge group is SO(3), with electromagnetism represented by the Abelian subgroup U(1). Now the existence of magnetic charge depends on the $\hat{\phi}$ manifold, so the gauge theory must be spontaneously broken. We saw that it is in fact the vacuum manifold; here, $G = \mathrm{SO}(3)$ and $H = \mathrm{U}(1)$ and the vacuum manifold is invariant under H. Its space is G/H, the set of transformations not related by one in H.

[3]With opposite points identified.

The existence of magnetic monopoles now requires non-trivial mapping of G/H onto S^2, the boundary in co-ordinate space. This is the *second homotopy group* of G/H, $\pi_2(G/H)$ and monopoles exist where this is non-trivial. We quote some facts from algebraic topology, as mentioned in Ryder [7]:

1. $\pi_2(G/H) \cong \ker\phi$, where $\phi : \pi_1(H) \hookrightarrow \pi_1(G)$ is the natural inclusion. Note that every closed path in H is also closed in G.
2. The first homotopy group π_1 is trivial if the group is simply connected, $\cong \mathbb{Z}_2$ if the group is doubly connected, &c.
3. SO(3) is doubly connected (consider what happens when we identify opposite points on the sphere), while U(1) is infinitely connected; i.e. $\pi_1(G) = \mathbb{Z}_2, \pi_1(H) = \mathbb{Z}$.

Hence, $\pi_2(G/H)$ is the additive group of even integers, and the monopole charge is twice the Dirac quantum.

4.5.2 The relationship between the 't Hooft-Polyakov and Dirac monopoles

In fact there is a firm relationship between the 't Hooft-Polyakov monopole and the Dirac monopole despite their apparent differences. Consider a Dirac monopole with a string singularity along the negative z-axis, with vector potential

$$A_r = a_\theta = 0, \quad A_\phi = \frac{g}{r}\frac{1-\cos\theta}{\sin\theta}. \tag{4.94}$$

We insert this into the SU(2) theory, and align the vector potential in the third internal direction:

$$\begin{aligned} &A_\mu = A_\mu^a T_a \\ &A_0 = a_r = a_\theta = 0, \quad A_\phi = T_3\left(-\frac{g}{r}\right)\left(\frac{1-\cos\theta}{\sin\theta}\right) \end{aligned} \tag{4.95}$$

Introduce a scalar field ϕ with vacuum expectation value F in the same alignment:

$$\phi = T_3 F. \tag{4.96}$$

We will transform A_μ, ϕ by a space-dependent internal gauge transform. A general SU(2) gauge transform may be parametrised by the Euler angles (α, β, γ):

$$S = \mathrm{e}^{(\mathrm{i}/2)\alpha T_3}\mathrm{e}^{(\mathrm{i}/2)\beta T_2}\mathrm{e}^{(\mathrm{i}/2)\gamma T_3} \tag{4.97}$$

$$= \begin{pmatrix} \cos\beta/2\,\mathrm{e}^{\mathrm{i}(\alpha+\gamma/2)} & \sin\beta/2\,\mathrm{e}^{\mathrm{i}(-\gamma+\alpha/2)} \\ -\sin\beta/2\,\mathrm{e}^{\mathrm{i}(\gamma-\alpha/2)} & \cos\beta/2\,\mathrm{e}^{\mathrm{i}(\alpha+\gamma/2)} \end{pmatrix}. \tag{4.98}$$

Set $\gamma = -\alpha = \phi, \beta = -\theta$ to obtain

$$S = \begin{pmatrix} \cos\theta/2 & -\mathrm{e}^{-\mathrm{i}\phi}\sin\theta/2 \\ e^{\mathrm{i}\theta}\sin\theta/2 & \cos\theta/2 \end{pmatrix} \tag{4.99}$$

$$\Rightarrow S^{-1} = \begin{pmatrix} \cos\theta/2 & \mathrm{e}^{-\mathrm{i}\phi}\sin\theta/2 \\ -^{\mathrm{i}\theta}\sin\theta/2 & \cos\theta/2 \end{pmatrix}. \tag{4.100}$$

The gauge transform is

$$A'_\mu = S A_\mu S^{-1} + \frac{2\mathrm{i}}{e}\, S\partial_\mu S^{-1}, \tag{4.101}$$

and we evaluate

$$
\begin{aligned}
\partial_r S^{-1} &= 0 \\
\partial_\theta S^{-1} &= \frac{1}{2r}\begin{pmatrix} -\sin\theta/2 & \mathrm{e}^{-\mathrm{i}\phi}\cos\theta/2 \\ -\mathrm{e}^{\mathrm{i}\phi}\cos\theta/2 & -\sin\theta/2 \end{pmatrix} \\
\partial_\phi S^{-1} &= \frac{-\mathrm{i}}{r\sin\theta}\begin{pmatrix} 0 & \mathrm{e}^{-\mathrm{i}\phi}\sin\theta/2 \\ \mathrm{e}^{\mathrm{i}\phi}\sin\theta/2 & 0 \end{pmatrix}.
\end{aligned}
$$

From our quantisation condition, $g = 1/e$, and in the new gauge our potential is

$$
\begin{aligned}
A'_0 = A'_r &= 0 \\
A'_\theta &= \frac{1}{er}\left(T_1\sin\phi - T_2\cos\phi\right) && (4.102) \\
A'_\phi &= \frac{1}{er}\left(T_1\cos\theta\cos\phi + T_2\cos\theta\sin\phi - T_3\sin\theta\right). && (4.103)
\end{aligned}
$$

We can now evaluate the Cartesian components of A: for example, note that

$$
\begin{aligned}
A'_x &= a'_r\cos\phi\sin\theta + A'_\theta\cos\phi\cos\theta - A'_\phi\sin\phi \\
&= \frac{1}{er}\left[T_2\left(\frac{-z}{r}\right) + T_3\left(\frac{y}{r}\right)\right].
\end{aligned}
$$

The other components are similar, and we recover exactly the hedgehog solution from the beginning of the section (cf. 4.77). What does the Higgs field look like in the new gauge? We have

$$
\begin{aligned}
\phi' &= S\phi S^{-1} \\
&= F\begin{pmatrix} \cos\theta & \mathrm{e}^{-\mathrm{i}\phi}\sin\theta \\ -\mathrm{e}^{\mathrm{i}\phi}\sin\theta & -\cos\theta \end{pmatrix} \\
&= F(\sin\theta\cos\phi\, T_1 + \sin\theta\cos\phi\, T_2 + \cos\theta\, T_3)
\end{aligned}
$$

which says precisely that

$$
\phi'^a = F\frac{r^a}{r}. \tag{4.104}
$$

which is again the field in the hedgehog solution (4.78). Hence, after our gauge transform, the Dirac string has disappeared, and now the monopole source lives in the Higgs field, as in the 't Hooft-Polyakov monopole. The gauge transform on

$$
F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \frac{1}{e|\phi|^3}\epsilon_{abc}\phi^a(\partial_\mu\phi^b)(\partial_\nu\phi^c)
$$

has transferred responsibility for the monopole from the Dirac term to the topological Higgs term.

4.5.3 The Bogomol'nyi bound and the Prasad-Sommerfeld solution

Rewriting the energy of the time-independent 't Hooft-Polyakov monopole from 4.75 in slightly abbrieviated form,

$$
E = \int \mathrm{d}^3x\left[\frac{1}{4}(F_{ij}{}^a)^2 + \frac{1}{2}(\mathcal{D}_i\phi^a)^2 + \lambda(|\phi|^2 - F^2)^2\right]. \tag{4.105}
$$

Following Bogomol'nyi [1], we note that

$$(F_{ij}{}^a - \epsilon_{ijk}\mathcal{D}_k\phi^a)^2 = (F_{ij}{}^a) - 2\epsilon_{ijk}(\mathcal{D}_k\phi^a)F_{ij}{}^a + 2(\mathcal{D}i\phi^a)^2. \tag{4.106}$$

Hence,

$$E = \int \mathrm{d}^3x \left\{ \frac{1}{4}(F_{ij}{}^a - \epsilon_{ijk}\mathcal{D}_k\phi^a)^2 + \left[\frac{\epsilon_{ijk}(\mathcal{D}_k\phi^a)F_{ij}{}^a}{2}\right] + \lambda(|\phi|^2 - F^2)^2 \right\}. \tag{4.107}$$

In fact the term in square brackets is a total divergence. Let

$$S_k = \frac{1}{2}\epsilon_{ijk}F_{ij}{}^a\phi^a \tag{4.108}$$

$$\Rightarrow \partial_k S_k = \frac{1}{2}\epsilon_{ijk}\,[(\partial_k F_{ij}{}^a)\phi^a + F_{ij}{}^a\partial_k\phi^a]. \tag{4.109}$$

Recall that by the Bianchi identity,

$$\begin{aligned} \mathcal{D}_k F_{ij}{}^a &= 0 = \partial_k F_{ij}{}^a + \epsilon^{abc}A_k{}^b F_{ij}{}^c \\ \Rightarrow \partial_k F_{ij}{}^a\phi^a &= -\epsilon^{abc}A_k{}^b F_{ij}{}^c\phi^a. \end{aligned}$$

Hence,

$$\begin{aligned} \partial_k S_k &= \frac{1}{2}\epsilon_{ijk}\,[-\epsilon^{abc}A_k{}^b F_{ij}{}^c\phi^a + F_{ij}{}^c\partial_k\phi^a] \\ &= \frac{1}{2}\epsilon_{ijk}\,[\epsilon^{cba}A_k{}^b\phi^a F_{ij}{}^c + F_{ij}{}^c\partial_k\phi^a] \\ &= \frac{1}{2}\epsilon_{ijk}(\mathcal{D}_k\phi^c)F_{ij}{}^c \text{ as desired.} \end{aligned} \tag{4.110}$$

This term becomes a surface integral of S_i — but we immediately see that in fact S_i is the integrand in 4.92, and this term will be proportional to the charge d on the monopole. In fact,

$$E_d = \frac{4\pi M_W}{e^2}d + \int \mathrm{d}^3x \left[\frac{1}{4}(F_{ij}{}^a - \epsilon_{ijk}\mathcal{D}_k\phi^a)^2 + \lambda(|\phi|^2 - F^2)^2\right] \tag{4.111}$$

In the Bogomol'nyi bound, $\lambda/e^2 \to 0$ and we see

$$E_d \geq \frac{4\pi M_W}{e^2}d \tag{4.112}$$

This holds more generally under the assumption that $V(\phi)$ is nonnegative and zero only at the Higgs vacuum. The minimal energy is attained for fields that satisfy the *Bogomol'nyi equation*,

$$F_{ij}{}^a = \epsilon_{ijk}\mathcal{D}_k\phi^a, \tag{4.113}$$

which is alternately expressed in co-ordinate-free notation as $\mathcal{F} = *\mathcal{D}\phi$.

In the case $d = 1$, we seek solutions satisfying the 't Hooft-Polyakov ansatz:

$$\begin{array}{rcl} \phi^a &=& F\frac{r^a}{r}s(r) \\ A_i{}^a &=& -\epsilon_{iab}\frac{r^b}{er^2}v(r). \end{array} \tag{4.114}$$

Substituting 4.114 into 4.113, a lengthy but straightforward calculation derives a sufficient condition on v and s. Switching to the dimensionless variable $\xi = eFr$, it is

$$\frac{\mathrm{d}v}{\mathrm{d}\xi} = s(1-v); \quad \frac{\mathrm{d}s}{\mathrm{d}\xi} = \frac{v(2-v)}{\xi^2}; \tag{4.115}$$

with boundary conditions $v \to 1, s \to 1$ as $r \to \infty$. An exact solution was observed by Prasad and Sommerfeld [6]:

$$s(\xi) = \coth\xi - \frac{1}{\xi}; \quad v(\xi) = 1 - \xi \operatorname{csch}\xi' \tag{4.116}$$

This is the Prasad-Sommerfeld solution, or the Bogomol'nyi-Prasad-Sommerfeld monopole, with energy

$$E_1 = \frac{4\pi M_W}{e^2}. \tag{4.117}$$

We note finally that the Bogomol'nyi equation

$$\mathcal{F} = *\mathcal{D}\phi \tag{4.118}$$

together with the Bianchi identity

$$\mathcal{D}\mathcal{F} = 0 \tag{4.119}$$

together imply the *Yang-Mills-Higgs equations*, which are the equations of motion when $\lambda = 0$:

$$\mathcal{D}\mathcal{F} = \mathcal{D} * \mathcal{D}\phi = 0; \tag{4.120}$$

$$\mathcal{D} * \mathcal{F} = \mathcal{D}\mathcal{D}\phi = [\mathcal{D}, \mathcal{D}]\phi = [\mathcal{F}, \phi] = [*\mathcal{D}\phi, \phi]. \tag{4.121}$$

The first is immediate, while the second and third equalities in the second equation can be established by, e.g., switching to co-ordinates. Hence, we might expect that the Bogomol'nyi equation (with the Bianchi identity) gives rise to an interesting field theory of monopoles. Indeed, the general theory of solutions to Bogomol'nyi's equation leads to the rich and beautiful theory of monopole moduli spaces, which we unfortunately do not have time to describe here.

Chapter 5

Conclusions

We began by observing simple consequences of gauge invariance in Maxwell's equations, particularly when we adopted a field theory to account for physical relativistic causality. We also noted the asymmetry between electricity and electromagnetism in Maxwell's equations, and hypothesised the Dirac monopole to model isolated magnetic charge. Using gauge-invariance of Maxwell's theory, we were able to choose gauges on different patches of space to move the Dirac string singularity, and remove it as long as we were comfortable with this patchwork of different gauges. We developed the theory of fibre bundles to allow us to smoothly glue together these gauge choices on different regions. In defining connections on principal bundles, we saw the geometrical role played by the physical gauge potential as the connection, and by the gauge field as the curvature, in this theory. We noted the physical utility of gauge theories not only in electromagnetism but in its relevance to general relativity. We returned to the Dirac and Wu-Yang monopoles as U(1) gauge theories with the benefit of this geometry. In fact, we were able to go further, and generalise to non-Abelian gauge theories, which had key physical differences, including the possibility of regular magnetic monopoles. Returning to field theories, we investigated spontaneous symmetry breaking and derived Goldstone's theorem. We observed the Brout-Englert-Higgs effect of spontaneous symmetry breaking in non-Abelian gauge theories, and the phenomenon of vector fields "eating" scalar fields to acquire mass. We illustrated the role of the BEH mechanism in the 't Hooft-Polyakov monopole in the SU(2) gauge theory, where the matter and gauge fields carry only electric charge and the magnetic charge is topological in origin. Nonetheless, we were able to relate the Dirac monopole to the 't Hooft-Polyakov monopole by way of a gauge transformation which removed the Dirac string and transferred the monopole source to the Higgs field. Finally, we investigated the Bogomol'nyi bound on the energy of the 't Hooft-Polyakov monopole, derived the Bogomol'nyi equation, and noted the Bogomol'nyi-Prasad-Sommerfeld solution for the lowest-charged such monopoles. We hope the reader is now equipped to explore the rich theory of monopoles further, e.g. in arbitrary gauge groups, or the theory of monopole moduli spaces in the context of Bogomol'nyi's equation.

Bibliography

[1] Evgeny Bogomol'nyi, *Soviet Journal of Nuclear Physics*, 1976, **24** 449; *Yadernaya Fizika*, 1976, **24** 861

[2] Paul Adrien Maurice Dirac, *Proceedings of the Royal Society*, 1931, A **133** 60

[3] James N. Fry, *'t Hooft-Polyakov Magnetic Monopoles*, University of Florida Department of Physics, Fall 2003. Retrieved from `http://www.phys.ufl.edu/~fry/7097/monopole.pdf` on March 22$^{\text{nd}}$ 2016

[4] Mikio Nakahara, *Geometry, Topology and Physics*, IOP Publishing, 1990

[5] Alexander Markovich Polyakov, *JETP Letters*, 1974, **20** 194; *Soviet Physics JETP*, 1976, **41** 988

[6] Manoj Prasad and Charles Sommerfeld, *Physical Review Letters*, 1975, **35** 760

[7] Lewis H. Ryder, *Quantum Field Theory*, Cambridge University Press, 1986

[8] Norman Steenrod, *The Topology of Fibre Bundles*, Princeton University Press, 1951

[9] Gerard 't Hooft, *Nuclear Physics*, 1974, **B79** 276

[10] Tai Tsun Wu and Chen Ning Yang, *Physical Review*, 1975, D **12** 3845