

Geometry of Mathematical Physics III (Epiphany 2022-23)

Stefano Cremonesi

`stefano.cremonesi@durham.ac.uk`

Department of Mathematical Sciences, Durham University

April 23, 2023

Contents

5	Abelian gauge theories	2
5.1	Electromagnetism as a $U(1)$ gauge theory	2
5.1.1	Maxwell's equations and relativity	3
5.1.2	Maxwell's equations: variational principle	6
5.1.3	Gauge Symmetry	8
5.2	$U(1)$ global symmetry	10
5.3	$U(1)$ gauge symmetry	11
5.4	Gauge redundancy and gauge fixing	17
5.5	$U(1)$ Wilson line and Wilson loop	21
6	Applications of abelian gauge theories	26
6.1	Global vortices	26
6.2	Derrick's theorem	31
6.3	Gauged vortices	33
6.3.1	Bogomol'nyi vortices	37
6.4	The Dirac monopole (à la Wu and Yang)	38
7	Non-abelian gauge theories	44
7.1	Compact Lie algebras	44
7.2	Non-abelian gauge theories: fields	50
7.3	Non-abelian gauge theories: action and EoM	55
7.4	* Non-abelian Wilson lines and Wilson loops	59
8	Applications of non-abelian gauge theories	60
8.1	The 't Hooft-Polyakov monopole	60
8.2	Instantons	65
9	Bundles, connections, curvature and sections*	66
9.1	The tangent bundle	67
9.2	Fibre bundles	74
9.3	Connection, holonomy and curvature	79

Chapter 5

Abelian gauge theories

This term we will learn how to formulate gauge theories, a special subset of field theories which describe most forces in modern physics. For example, the Standard Model of elementary particles is a gauge theory based on the group $G = SU(3) \times SU(2) \times U(1)$, and accounts for the strong, weak and electromagnetic interaction.

In this chapter we will start by looking at abelian gauge theories, the formulation of which is based on an abelian Lie group, called the gauge group. The abelian restriction will allow us to acquaint us with the key concepts in gauge theory without complicating the underlying mathematics too much. As we will see in the next chapter, despite their simplicity abelian gauge theories already exhibit interesting field configurations, which make use of the geometric and topological properties of abelian gauge theory. We will encounter magnetic monopoles, vortices and other interesting effects. In the second half of the term we will generalize all of this to non-abelian gauge theories, which have even richer mathematical and physical properties.

5.1 Electromagnetism as a $U(1)$ gauge theory

We will soon delve into the abstract idea that underlies abelian gauge theories, starting from a field theory with a $U(1)$ global symmetry and promoting the constant $U(1)$ parameter to a local function of spacetime. But before we do that, let us take a fresh look at Maxwell's theory of electromagnetism, and describe it as a relativistic field theory that can be based on a gauge symmetry principle. Excellent references for some foundational material are [Landau and Lifshitz, 2013], [Griffiths, 2014] and [Jackson, 2021]. The coupling of electromagnetism to field theories, which we will study later, is a standard topic in nearly all books on quantum field theory, see *e.g.* [Peskin, 1995].

5.1.1 Maxwell's equations and relativity

The Maxwell equations describing which electric (\mathbf{E}) and magnetic fields (\mathbf{B}) are induced by the electric charge density ρ and current \mathbf{j} are (in natural units)

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \rho, & \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} &= \mathbf{j}, \\ \nabla \cdot \mathbf{B} &= 0, & \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0. \end{aligned} \quad (5.1)$$

We call the equations in the first line the inhomogeneous Maxwell equations, since they have sources for the electric and magnetic fields in the right-hand side, and the equations in the second line the homogeneous Maxwell equations, since they don't.

The behaviour of Maxwell equations under Lorentz transformations can be worked out as follows. Starting from an inertial frame with a charge distribution ρ at rest, we can perform a boost

$$\Lambda = \begin{pmatrix} \cosh \lambda & \sinh \lambda & 0 & 0 \\ \sinh \lambda & \cosh \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5.2)$$

to another inertial frame moving at a relative speed $\tanh \lambda$, in which there is now also a non-zero current \mathbf{j} . As resting charges only source electric fields and steady currents source magnetic fields, this implies that Lorentz transformations will also mix up electric and magnetic fields.

In order to understand how to write the Maxwell equations in a manifestly Lorentz invariant way, and how the electric and magnetic field transform under Lorentz transformations, let us first focus on the sources appearing in the right-hand side of the inhomogeneous Maxwell equations. The charge density ρ and the current \mathbf{j} can be repackaged into a Lorentz 4-vector J^μ , such that $J^0 = \rho$ and $J^i = j^i$. The **continuity equation** (or **local conservation law**)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 \quad (5.3)$$

can then be written as

$$\partial_\mu J^\mu = 0. \quad (5.4)$$

Since J^μ is a Lorentz vector, a Lorentz transformation acts as

$$J^\mu(x) \mapsto J'^\mu(x) = \Lambda^\mu{}_\nu J^\nu(\Lambda^{-1}x), \quad (5.5)$$

which indeed leaves the continuity equation invariant.¹

¹Recall that $\partial_0 = \frac{\partial}{\partial t}$ and $\partial_i = \frac{\partial}{\partial x^i} = (\nabla)_i$, and that the Lorentz transformation $x^\mu \mapsto x'^\mu = \Lambda^\mu{}_\nu x^\nu$ of the spacetime coordinates implies the following Lorentz transformation of the derivatives:

$$\partial_\mu \mapsto \partial'_\mu = \Lambda_\mu{}^\rho \partial_\rho = (\Lambda^{-1})^\rho{}_\mu \partial_\rho.$$

REMARK:

In the following I may use the shorthand notation $J^\mu \mapsto \Lambda^\mu{}_\nu J^\nu$ for the transformation law (5.5), with the understanding that if the object in question is a field then the argument must transform appropriately.

The transformation property of J^μ and the assumption of Lorentz symmetry (or ‘Lorentz invariance’) requires that the inhomogeneous Maxwell equations in the first line of (5.1) be the temporal and spatial components of a Lorentz 4-vector equation respectively. The similarity between the two rows of (5.1) suggests that the same should be true of the homogeneous Maxwell equations in the second line of (5.1).

Let’s now focus on the left-hand side of the inhomogeneous Maxwell equations, which is equal to the current 4-vector J^μ . Spacetime derivatives appear linearly, so we need a ∂_ν on the left-hand side, with the ν index suitably contracted with a tensor linear in the electric and magnetic field, in such a way that a μ index stays free (that is, uncontracted). The simplest option is that the left-hand side is $\partial^\mu X$ for a scalar field X , but an equation of the form $\partial^\mu X = J^\mu$ is immediately ruled out by counting degrees of freedom: it cannot account for the electric and magnetic fields \mathbf{E} and \mathbf{B} and hence reproduce the left-hand side of the inhomogeneous Maxwell equation. In order to match the upper index of J^μ on the right-hand side, the derivative ∂_ν must therefore act on a second rank Lorentz tensor $F^{\mu\nu}$,² which is linear in the electric and magnetic field, with the ν index contracted so that only the μ index remains free.

The electric and magnetic field \mathbf{E} and \mathbf{B} have $3+3 = 6$ components in total, whereas a second rank tensor³ has $4 \cdot 4 = 16$ components, so there still appears to be a mismatch of degrees of freedom. This is fixed by requiring that $F^{\mu\nu}$ be antisymmetric, that is $F^{\mu\nu} = -F^{\nu\mu}$: then it has $\frac{4 \cdot 3}{2} = 6$ components, corresponding to the reducible representation $(1, 0) \oplus (0, 1)$ of the Lorentz group.

To summarize, we are led to write the inhomogeneous Maxwell equations as

$$\partial_\nu F^{\mu\nu} = J^\mu \quad (5.6)$$

for a second rank antisymmetric tensor $F^{\mu\nu} = -F^{\nu\mu}$ which is linear in \mathbf{E} and \mathbf{B} . Comparing with the first line of (5.1) determines

$$[F^{\mu\nu}] = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}. \quad (5.7)$$

²Recall that by definition a Lorentz tensor with two indices transforms as

$$F^{\mu\nu}(x) \mapsto F'^{\mu\nu}(x) = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}(\Lambda^{-1}x)$$

under a Lorentz transformation.

³Vocabulary: a tensor with n indices is called an n -th rank tensor or equivalently a rank- n tensor.

Lowering indices to $F_{\mu\nu} = \eta_{\mu\rho}\eta_{\nu\sigma}F^{\rho\sigma}$, we have

$$[F_{\mu\nu}] = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}. \quad (5.8)$$

In other words for $i = 1, 2, 3$ we have

$$F_{i0} = -F_{0i} = E_i, \quad F_{ij} = \epsilon_{ijk}B_k. \quad (5.9)$$

$F_{\mu\nu}$ used to be called the Faraday tensor, and is now most commonly called the **field strength** tensor, because its components encode the strength of the electric and magnetic fields.

By a similar logic, it is not hard to see that the homogeneous Maxwell equations in the second line of (5.1) can also be written covariantly – that is, in Lorentz tensor notation – as

$$\epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0, \quad (5.10)$$

where $\epsilon^{\mu\nu\rho\sigma}$ is the completely antisymmetric tensor with four indices, normalized such that $\epsilon^{0123} = 1$.

REMARKS:

1. In practice this means that one gets a relative minus sign when swapping any two indices. *E.g.* $\epsilon^{3201} = -1$ as one needs to swap indices an odd number of times to arrive there from ϵ^{0123} . One way to see that is

$$\epsilon^{3201} = -\epsilon^{3021} = \epsilon^{1023} = -\epsilon^{0123}.$$

2. A fancier mathematical way of saying the same thing is: for any permutation σ of $0, 1, 2, 3$ we set $\epsilon^{\sigma(0),\sigma(1),\sigma(2),\sigma(3)} = \text{sign}(\sigma)$, where $\text{sign}(\sigma)$ is the signature of σ . The signature of a permutation σ is defined to be $+1$ (respectively -1) if the permutation is even (resp. odd), which means that $(\sigma(0), \sigma(1), \sigma(2), \sigma(3))$ is obtained from $(0, 1, 2, 3)$ by an even (resp. odd) number of transpositions (or swaps).
3. Note that in a situation with four indices the ‘cyclical’ vs. ‘anti-cyclical’ method useful for ϵ_{ijk} does not work anymore.
4. If we lower all four indices using the Minkowski metric, one of them is temporal and three of them are spatial, so we pick up a minus sign:

$$\epsilon_{0123} = \eta_{00}\eta_{11}\eta_{22}\eta_{33}\epsilon^{0123} = -1. \quad (5.11)$$

In summary, using the field strength $F_{\mu\nu}$ and the 4-current J^μ we can write the **Maxwell equations** as

$$\partial_\nu F^{\mu\nu} = J^\mu, \quad \epsilon^{\mu\nu\rho\sigma} \partial_\nu F_{\rho\sigma} = 0. \quad (5.12)$$

Proof. [Ex 1]

The inhomogeneous Maxwell equations imply the local conservation equation for the electromagnetic current J^μ :

$$\partial_\mu J^\mu = 0. \quad (5.13)$$

Proof. Using the inhomogeneous Maxwell equations we find

$$\partial_\mu J^\mu = \partial_\mu \partial_\nu F^{\mu\nu} = 0 \quad (5.14)$$

The first equality are just Maxwell's equations and the second equality follows from the anti-symmetry of the field strength $F^{\mu\nu} = -F^{\nu\mu}$, along with the commutativity of partial derivatives $\partial_\nu \partial_\mu = \partial_\mu \partial_\nu$.⁴ We have

$$\partial_\mu \partial_\nu F^{\mu\nu} = -\partial_\mu \partial_\nu F^{\nu\mu} = -\partial_\nu \partial_\mu F^{\nu\mu} = -\partial_\mu \partial_\nu F^{\mu\nu} \quad (5.15)$$

where we have relabelled (ν, μ) as (μ, ν) in the last step. As we see, this expression is equal to minus itself, so it must be zero.

5.1.2 Maxwell's equations: variational principle

How can we write down a Lorentz invariant Lagrangian density that will give us (5.12) as its Euler-Lagrange equations (or equations of motion, or EoM)? You can try playing around but you will soon realise that using $F^{\mu\nu}$ as the dynamical field(s) will not allow you to recover Maxwell's equations.

Let us hence try something else. The second equation of (5.12) implies that we can write

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (5.16)$$

in any star-shaped open subset in $\mathbb{R}^{4,5}$. We say that (5.16) holds *locally*. Conversely, (5.16) implies

$$\epsilon^{\mu\nu\rho\sigma} \partial_\nu F_{\rho\sigma} = \epsilon^{\mu\nu\rho\sigma} \partial_\nu (\partial_\rho A_\sigma - \partial_\sigma A_\rho) = \epsilon^{\mu\nu\rho\sigma} \partial_\nu \partial_\rho A_\sigma - \epsilon^{\mu\nu\rho\sigma} \partial_\nu \partial_\sigma A_\rho = 0 - 0 = 0 \quad (5.17)$$

⁴We assume that all fields are smooth functions, hence they have continuous second partial derivatives and Schwarz/Clairaut's theorem applies. It turns out that this assumption is false for generic field configurations in quantum field theory, but we are only doing classical field theory here, and we'll leave that story for another day.

⁵This is known as the Poincaré lemma, which is a generalization of the fact that $\nabla \times \mathbf{F} = 0$ implies $\mathbf{F} = \nabla \phi$ locally (see AMV). An open set U is called star-shaped (or a star domain) if there exists a point $p \in U$ such that for any $q \in U$, the line segment from p to q is contained in U .

by using that each of the two terms is symmetric with respect to swapping the order of the derivatives but is contracted with an epsilon tensor, which is antisymmetric in all indices. The second equation of (5.12) is hence automatic (it is called the **Bianchi identity**) and we need only worry about the first one.

In the theory of electromagnetism, A^μ is called the electromagnetic 4-vector potential: its time component $A^0 = \phi$ is the electric ‘scalar potential’, and its space components $A^i = A_i$ are the components of the magnetic ‘vector potential’ \mathbf{A} . (In this pre-relativistic terminology, ‘scalar’ and ‘vector’ refer to spatial rotations, not to Lorentz transformations). Using (5.16) and (5.9), we recover the relations between electromagnetic fields and electromagnetic potentials from the theory of electromagnetism:

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}. \quad (5.18)$$

Proof. [Ex 2]

We now declare that A_μ is the dynamical field, which also enables us to include J^μ as a source in the action. Maxwell’s equations follow from the action⁶

$$S[A_\mu] = \int d^4x \left(-\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + A_\mu J^\mu \right). \quad (5.19)$$

Proof. We work out the Euler-Lagrange equations

$$\frac{\partial\mathcal{L}}{\partial A_\mu} - \partial_\nu \frac{\partial\mathcal{L}}{\partial(\partial_\nu A_\mu)} = 0 \quad (5.20)$$

for the Lagrangian density

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + A_\mu J^\mu. \quad (5.21)$$

For the first term we have

$$\frac{\partial\mathcal{L}}{\partial A_\mu} = \frac{\partial}{\partial A_\mu} (A_\nu J^\nu) = \frac{\partial A_\nu}{\partial A_\mu} J^\nu = \delta_\nu^\mu J^\nu = J^\mu. \quad (5.22)$$

Remember: repeated indices are summed over and are dummy. You should never use the same letter for different indices, or you will get wrong results: this is the reason why I relabelled

⁶The overall minus sign is there to ensure that the Hamiltonian of the electromagnetic field is positive definite. More about this later.

the dummy index as ν here. For the second term we have

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)} &= \frac{\partial F_{\alpha\beta}}{\partial(\partial_\nu A_\mu)} \frac{\partial}{\partial F_{\alpha\beta}} \left(-\frac{1}{4} F^{\rho\sigma} F_{\rho\sigma} \right) \\
&= -\frac{1}{4} \cdot 2F^{\alpha\beta} \frac{\partial}{\partial(\partial_\nu A_\mu)} (\partial_\alpha A_\beta - \partial_\beta A_\alpha) \\
&= -\frac{1}{2} F^{\alpha\beta} (\delta_\alpha^\nu \delta_\beta^\mu - \delta_\beta^\nu \delta_\alpha^\mu) \\
&= -\frac{1}{2} (F^{\nu\mu} - F^{\mu\nu}) = F^{\mu\nu} .
\end{aligned} \tag{5.23}$$

In deriving (5.23) we used the chain rule in the first line. In the second line we used the definition (5.16) of the field strength $F_{\alpha\beta}$ in terms of derivatives of A_μ , and the identity

$$\frac{\partial}{\partial X_{a_1 \dots a_n}} (X^{b_1 \dots b_n} X_{b_1 \dots b_n}) = 2X^{a_1 a_2 \dots a_n} , \tag{5.24}$$

which works for any tensor X [Ex 3]. In the third line of (5.23) we just calculated derivatives, and in the final equality in the fourth line we used the antisymmetry of the field strength.

The Euler-Lagrange equations then give

$$J^\mu - \partial_\nu F^{\mu\nu} = 0 , \tag{5.25}$$

which reproduce the inhomogeneous Maxwell's equations.

REMARK:

It is also possible to derive the action (5.19) (without the source term) by using the Lorentz force to show that the energy stored in the electromagnetic fields (which equals the Hamiltonian) is $\frac{1}{2} \int d^3x (\mathbf{E}^2 + \mathbf{B}^2)$, and then finding the associated Lagrangian.

5.1.3 Gauge Symmetry

The technical trick we have used has an interesting consequence: the physical fields that we can measure are the electric and magnetic field \mathbf{E} and \mathbf{B} , *i.e.* the components of the field strength tensor $F_{\mu\nu}$, not the dynamical field A_μ that we use to define the action and obtain equations of motion. In fact, A_μ is not uniquely defined: we are free to shift $A_\mu(x)$ by a derivative of an arbitrary smooth function $\alpha(x)$

$$A_\mu(x) \mapsto A_\mu(x) + \partial_\mu \alpha(x) \tag{5.26}$$

without altering the physical fields which appear in the Maxwell equations and which can be measured:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \mapsto \partial_\mu A_\nu - \partial_\nu A_\mu + \partial_\mu \partial_\nu \alpha - \partial_\nu \partial_\mu \alpha = F_{\mu\nu} . \tag{5.27}$$

A symmetry for which the parameters of the transformation depend on space-time is called a **gauge symmetry**.⁷ Equation (5.26) is called the **gauge transformation** of A_μ . The field A_μ is then called the **gauge field** (or the **gauge connection**). Gauge field configurations which differ by a gauge transformations are considered physically equivalent, since they give rise to the same physically observable electric and magnetic fields.

You should contrast gauge symmetries with the symmetries you studied so far: their parameters did not depend on space-time in any way. They are called **global symmetries**, and they relate physically inequivalent (though isomorphic) configurations.

Performing a gauge transformation (5.26) has the following effect on the action (5.19):

$$\begin{aligned} S[A_\mu] \mapsto S[A_\mu + \partial_\mu \alpha] &= \int d^4x \left(-\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + A_\mu J^\mu + (\partial_\mu \alpha) J^\mu \right) \\ &= S[A_\mu] + \int d^4x (\partial_\mu \alpha) J^\mu \end{aligned} \quad (5.28)$$

At first sight the action does not seem to be invariant under a gauge transformation, since

$$\delta_\alpha S[A_\mu] \equiv S[A_\mu + \partial_\mu \alpha] - S[A_\mu] = \int d^4x (\partial_\mu \alpha) J^\mu \quad (5.29)$$

does not seem to vanish. But this is too fast: we can perform a partial integration of the extra term and discard the boundary term⁸ to write the gauge variation of the action as

$$\delta_\alpha S[A_\mu] = - \int d^4x \alpha (\partial_\mu J^\mu) = 0, \quad (5.30)$$

which vanishes thanks to the conservation of the current J^μ that couples to the electromagnetic gauge field A_μ .

REMARKS:

1. We can write

$$A_\mu \rightarrow A_\mu + \partial_\mu \alpha = e^{i\alpha} (A_\mu + i\partial_\mu) e^{-i\alpha}, \quad (5.31)$$

so we can think about our gauge transformations as being related to the group $G = U(1)$, but now its parameter α depends on where we are in space-time. $G = U(1)$ is called the **gauge group**. The field A_μ transforms in the adjoint representation, except for the derivative term. This rewriting may look silly since the adjoint representation of $G = U(1)$ is trivial, but we will see later that this form generalizes to other gauge groups in a natural way. We will also understand the rôle and meaning of the extra derivative term.

⁷As I will stress later, this is a misnomer: a gauge ‘symmetry’ is not really a symmetry of a physical system. Rather, it is a redundancy in our description of the system.

⁸We assume that the fields obey boundary conditions such that this holds, e.g. that they vanish fast enough at infinity, or that they obey (along with the gauge parameter) periodic boundary conditions.

2. You have encountered field theories with $U(1)$ global symmetries and conserved currents before, for example the Schroedinger action, problem 24 from the exercises, and problem 2 from the 4th problem class. Can we use the currents found there to couple them to electromagnetism? If so, can we identify the $U(1)$ global symmetry of these field theories with the $U(1)$ gauge symmetry found above?

The answer to the previous question is yes, and we will learn how to do this systematically next. But first, let us briefly remind ourselves of the concept of $U(1)$ global symmetry and set notation for what follows.

5.2 $U(1)$ global symmetry

Consider (for simplicity) a complex scalar field $\phi(x)$.⁹ The action¹⁰

$$\begin{aligned} S_0[\phi, \bar{\phi}] &= \int d^4x \mathcal{L}_0(\phi, \bar{\phi}, \partial_\mu \phi, \partial_\mu \bar{\phi}), \\ \mathcal{L}_0 &= -|\partial_\mu \phi|^2 - V(\phi, \bar{\phi}) = -|\partial_\mu \phi|^2 - U(|\phi|^2) \\ &= |\dot{\phi}|^2 - |\nabla \phi|^2 - U(|\phi|^2) \end{aligned} \quad (5.32)$$

is invariant under **global** $G = U(1)$ transformations

$$g : \phi(x) \mapsto e^{i\alpha} \phi(x)$$

where $\alpha \sim \alpha + 2\pi$ is a **constant** parameter, and $g = e^{i\alpha} \in U(1)$ is a constant group element. The requirement of $U(1)$ invariance restricts the scalar potential $V(\phi, \bar{\phi})$ to only depend on the invariant $|\phi|^2$. Because the scalar field ϕ is multiplied by a single power of the $U(1)$ group element $g = e^{i\alpha}$, we say that it has charge 1.

REMARKS:

1. The continuous $U(1)$ symmetry ensures the existence of a **conserved current**

$$\begin{aligned} j^\mu &= -i(\bar{\phi} \partial^\mu \phi - \phi \partial^\mu \bar{\phi}) \\ \partial_\mu j^\mu &= 0 \end{aligned} \quad (5.33)$$

⁹Recall that mathematically, this is a map from Minkowski space-time $\mathbb{R}^{1,3}$ to \mathbb{C} , which associates a complex number to each point in space-time:

$$\begin{aligned} \phi: \mathbb{R}^{1,3} &\rightarrow \mathbb{C} \\ x^\mu &\mapsto \phi(x) \end{aligned}$$

Greek indices μ, ν, \dots are space-time indices running from 0 to 3. (Roman indices i, j, \dots are spatial indices running from 1 to 3. Index 0 is for time.)

Unless we explicitly state otherwise, we will typically assume that all fields are smooth.

¹⁰ $|\partial_\mu \phi|^2$ is a short-hand notation for $\partial_\mu \bar{\phi} \partial^\mu \phi$, where Einstein summation convention (repeated indices are summed over) is understood. Recalling that we work with Minkowski metric $[\eta_{\mu\nu}] = (-1, +1, +1, +1)$, this means that $|\partial_\mu \phi|^2 = -|\partial_0 \phi|^2 + |\partial_i \phi|^2 = -|\dot{\phi}|^2 + |\nabla \phi|^2$.

and of a **conserved charge**

$$Q = \int d^3x j^0$$

$$\frac{d}{dt}Q = 0 \tag{5.34}$$

by Noether's theorem.

2. A **global symmetry** relates **physically distinct configurations**.

*** EXERCISE:**

[Ex 6] Consider a field theory with action (5.32) and scalar potential

$$V(\phi, \bar{\phi}) = \lambda(|\phi|^2 - a^2)^2,$$

with parameters $\lambda, a > 0$, see figure 5.1.

1. Calculate the energy (or “Hamiltonian”)

$$E = \int d^3x (|\partial_0\phi|^2 + |\partial_i\phi|^2 + V(\phi, \bar{\phi}))$$

$$= \int d^3x (|\dot{\phi}|^2 + |\nabla\phi|^2 + V(\phi, \bar{\phi})) .$$

You may use the relation between the Lagrangian and Hamiltonian densities, or calculate the Noether charge associated to invariance under time translations $t \mapsto t + c$.

2. Show that the configurations of least energy (“vacua”, or “ground states”) parametrize a circle in field space.

3. Show that different vacua are related by global $U(1)$ transformations.

5.3 $U(1)$ gauge symmetry

To make the global symmetry local, or a gauge symmetry, we promote the constant parameter α to a function of spacetime $\alpha(x)$. For subtle reasons that we might return to later, the parameter $\alpha(x)$ of a gauge transformation should approach 0 (sufficiently fast) at infinity.

If we try to write a kinetic term for ϕ , we immediately seem to run into trouble. Under a $U(1)$ gauge transformation

$$\partial_\mu\phi \mapsto \partial_\mu\phi' \equiv \partial_\mu(e^{i\alpha}\phi) = e^{i\alpha} (\partial_\mu\phi + i(\partial_\mu\alpha)\phi) \tag{5.35}$$

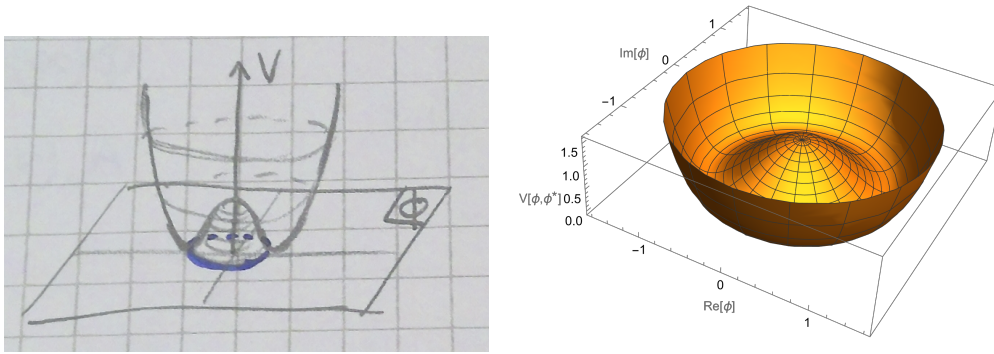


Figure 5.1: The scalar potential $V(\phi, \bar{\phi}) = \lambda(|\phi|^2 - a^2)^2$.

since now α depends on spacetime. Therefore the naive kinetic term $-\partial_\mu \phi^2$ is not invariant under a $U(1)$ gauge transformation. We say that it is not gauge invariant.

This is a serious problem. But there is a way to fix it: we replace the derivative $\partial_\mu \phi$ by the so called **gauge covariant derivative**

$$D_\mu \phi := \partial_\mu \phi - iA_\mu \phi \quad (5.36)$$

which includes a new field A_μ (the **gauge field**), whose purpose is to transform under gauge transformations precisely in such a way to cancel the unwanted second term in (5.35). This happens if under a $U(1)$ gauge transformation

$$A_\mu \mapsto A'_\mu = A_\mu + \partial_\mu \alpha, \quad (5.37)$$

because then

$$\begin{aligned} D_\mu \phi &= (\partial_\mu \phi - iA_\mu \phi) \mapsto D'_\mu \phi' \equiv (\partial_\mu \phi' - iA'_\mu \phi') \\ &= e^{i\alpha} (\partial_\mu \phi + i(\partial_\mu \alpha)\phi - iA_\mu \phi - i(\partial_\mu \alpha)\phi) \\ &= e^{i\alpha} (\partial_\mu \phi - iA_\mu \phi) = e^{i\alpha} D_\mu \phi, \end{aligned} \quad (5.38)$$

using (5.35) and (5.37). Replacing derivatives ∂_μ by gauge covariant derivatives D_μ makes the gauge kinetic term of ϕ invariant under $U(1)$ gauge transformations.

Note that (5.37) mimics precisely the gauge transformation (5.26) of the 4-vector potential in the theory of electromagnetism. Having introduced a new $U(1)$ gauge field A_μ , we now need to write a gauge invariant kinetic term for it. But we know how to do it: we just write the Maxwell Lagrangian from the theory of electromagnetism.

Putting everything together, we find that the action

$$\begin{aligned}
 S[\phi, \bar{\phi}, A_\mu] &= \int d^4x \mathcal{L}(\phi, \bar{\phi}, A_\nu, \partial_\mu \phi, \partial_\mu \bar{\phi}, \partial_\mu A_\nu), \\
 \mathcal{L} &= \mathcal{L}_0(\phi, \bar{\phi}, D_\mu \phi, \overline{D_\mu \phi}) + \mathcal{L}_{\text{Maxwell}}(\partial_\mu A_\nu) \\
 &= -\overline{D_\mu \phi} D^\mu \phi - U(|\phi|^2) - \frac{1}{4g^2} F_{\mu\nu} F^{\mu\nu},
 \end{aligned} \tag{5.39}$$

where A_μ is a real **gauge field** (or mathematically, a “gauge connection”) and

$$\begin{aligned}
 D_\mu \phi &:= (\partial_\mu - iA_\mu)\phi && \text{covariant derivative of } \phi \\
 F_{\mu\nu} &:= \partial_\mu A_\nu - \partial_\nu A_\mu && \text{field strength of } A_\mu,
 \end{aligned} \tag{5.40}$$

is invariant under $G = U(1)$ **gauge transformations**

$$\begin{aligned}
 \phi(x) &\mapsto e^{i\alpha(x)}\phi(x) \\
 A_\mu(x) &\mapsto A_\mu(x) + \partial_\mu \alpha(x).
 \end{aligned} \tag{5.41}$$

REMARKS:

1. To linear order in the gauge field A_μ

$$\mathcal{L} = \mathcal{L}_0 + j^\mu A_\mu + \dots \tag{5.42}$$

The scalar field is coupled (via covariant derivatives) to the gauge field A_μ , and not to the field strength $F_{\mu\nu}$. To leading order, the gauge field A_μ couples directly to the conserved current j^μ of the theory with $U(1)$ global symmetry, which is built out of the scalar field. This type of coupling is called the **minimal coupling**.

A common alternative normalization to the one we use is obtained by rescaling the gauge field by one power of the gauge coupling: $A_\mu \rightarrow gA_\mu$. In that normalization the Lagrangian density is

$$\begin{aligned}
 \mathcal{L} &= -((\partial^\mu + igA^\mu)\bar{\phi})(\partial_\mu - igA_\mu)\phi - U(|\phi|^2) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
 &= \mathcal{L}_0 + gj^\mu A_\mu + \dots
 \end{aligned}$$

where the ellipses denote terms quadratic in the gauge field. This alternative normalization makes it clear that the **gauge coupling** g controls the strength of the coupling between the conserved current j^μ of the theory with $U(1)$ global symmetry and the gauge field A_μ . In the following we will typically stick to the convention in which the gauge coupling g appears in front of the kinetic term for the gauge field, rather than inside gauge covariant derivatives.

2. The group of gauge transformations

$$\mathcal{G} = \mathcal{U}(1) := \left\{ \begin{array}{ll} g : \mathbb{R}^{1,3} & \rightarrow G = U(1) \\ x^\mu & \mapsto g(x) = e^{i\alpha(x)} \end{array} \right\} \quad (5.43)$$

is infinite-dimensional, since it associates independent transformations $g(x)$ for the fields at different points x^μ , and there are infinitely many points in space-time. We use calligraphic letters to distinguish the gauge group from the associated finite-dimensional (for $G = U(1)$, one-dimensional) Lie group. Later on, once we have familiarized ourselves with this distinction, we will typically drop this notation and simply use G for the gauge group, with a common abuse of notation.

3. A “**gauge symmetry**” relates **physically equivalent configurations**, which are to be **identified**. The term “gauge symmetry” is therefore a misnomer: it is **not a symmetry**, but rather a **redundancy** in our description of the theory.

The identification of field configurations which differ by a gauge transformation¹¹ leads to non-trivial topological properties of gauge fields, which in turn ensure the existence of topological solitons and instantons, non-trivial gauge field configurations which are stable for topological reasons. We will study these configurations in later chapters.

From now on we omit writing the dependence on the space-time coordinate x . It is understood that all fields and all gauge transformation parameters depend on x .

4. Under a $U(1)$ gauge transformation (5.41),

$$\begin{aligned} D_\mu \phi &\mapsto e^{i\alpha} D_\mu \phi, \\ F_{\mu\nu} &\mapsto F_{\mu\nu} \end{aligned} \quad (5.44)$$

We say that the covariant derivative $D_\mu \phi$ of ϕ is **gauge covariant**, because it transforms in a representation of G for all x (the same representation of ϕ , namely the charge 1 representation here), and that the field strength $F_{\mu\nu}$ is **gauge invariant**, because it does not change under a gauge transformation (in fancy language, it transforms in the trivial, or “singlet”, representation).

5. It is useful to think of the covariant derivative $D_\mu = \partial_\mu - iA_\mu$ as a **differential operator**, which acts on everything to its right. The partial derivative ∂_μ acts by differentiating all that appears to its right, while the gauge field A_μ , like all functions of x , acts by multiplying all that appears to its right. Requiring that under a $U(1)$ gauge transformation

$$D_\mu \equiv \partial_\mu - iA_\mu \mapsto D'_\mu \equiv \partial_\mu - iA'_\mu = e^{i\alpha} D_\mu e^{-i\alpha}, \quad (5.45)$$

¹¹See section 2.6 of [Manton and Sutcliffe, 2004] if you want to read more about this.

so that

$$D_\mu \phi \mapsto e^{i\alpha} D_\mu e^{-i\alpha} e^{i\alpha} \phi = e^{i\alpha} D_\mu \phi \quad (5.46)$$

as desired, implies the gauge transformation of the gauge field

$$A_\mu \mapsto A'_\mu = A_\mu + \partial_\mu \alpha \quad (5.47)$$

and vice versa.

Proof. We have already proven the implication (5.45) \Leftarrow (5.47) in (5.38). For the opposite implication (5.45) \Rightarrow (5.47), we expand (5.45) and act with ∂_μ on everything to its right. There are two options: either ∂_μ acts on $e^{-i\alpha}$, which produces the function $(\partial_\mu e^{-i\alpha}) = -i e^{-i\alpha} (\partial_\mu \alpha)$, or ∂_μ goes through $e^{-i\alpha}$, which produces the differential operator $e^{-i\alpha} \partial_\mu$.¹² Then we find

$$\begin{aligned} D_\mu \equiv \partial_\mu - iA_\mu &\mapsto D'_\mu \equiv \partial_\mu - iA'_\mu = e^{i\alpha} (\partial_\mu - iA_\mu) e^{-i\alpha} \\ &= e^{i\alpha} e^{-i\alpha} (-i\partial_\mu \alpha) + e^{i\alpha} e^{-i\alpha} \partial_\mu - i e^{i\alpha} e^{-i\alpha} A_\mu \\ &= \partial_\mu - i(A_\mu + \partial_\mu \alpha), \end{aligned}$$

which comparing the initial expression and the final result implies

$$A_\mu \mapsto A'_\mu = A_\mu + \partial_\mu \alpha .$$

Furthermore, defining the commutator $[X, Y] := XY - YX$, we have

$$[D_\mu, D_\nu] = -iF_{\mu\nu}, \quad (5.48)$$

so the field strength controls the non-commutativity of covariant derivatives.

Proof. [Ex 7]

6. The gauge field A_μ is only defined **locally**, namely **in a patch**, which we take to be such that the Poincaré lemma applies. As we saw in the gauge theory formulation of electromagnetism, the Bianchi identity $\epsilon^{\mu\nu\rho\sigma} \partial_\nu F_{\rho\sigma} = 0$ implies $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ only if the Poincaré lemma applies.

What this means is the following. Consider two patches $U^{(1)}$ and $U^{(2)}$ with a non-trivial overlap $U^{(1)} \cap U^{(2)} \neq \emptyset$. Then the gauge fields $A_\mu^{(1)}$ and $A_\mu^{(2)}$ defined in the two patches are related by a gauge transformation

$$A_\mu^{(1)} = A_\mu^{(2)} + \partial_\mu \alpha^{(12)}$$

¹²If you are confused by these statements and manipulations, act with the differential operator on any smooth test function $f(x)$. If X and Y are two differential operators, then $X = Y$ iff $Xf = Yf$ for all smooth test functions. Similarly $X \mapsto Y$ iff $Xf \mapsto Yf$ for all smooth test functions.

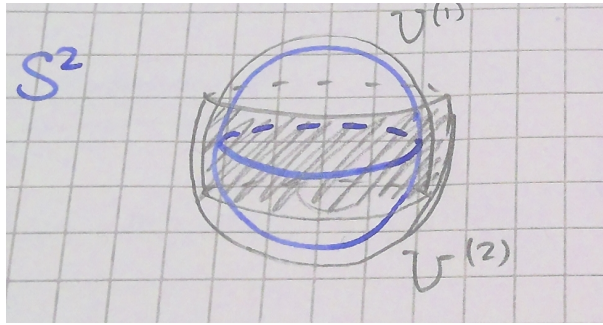


Figure 5.2: Two patches which cover a 2-sphere S^2 , and their overlap.

on the overlap $U^{(1)} \cap U^{(2)}$, so that the field strengths agree: $F_{\mu\nu}^{(1)} = F_{\mu\nu}^{(2)}$.¹³ Mathematically, the gauge transformation parameter $\alpha^{(12)}$ that relates the gauge fields in the two patches is called a “transition function”. Charged fields are also defined locally, in patches. For consistency, they also transform by a gauge transformation when we switch to another patch.

This local definition of A_μ is responsible for most of the topological and geometric properties of gauge theories. To give you an appetizer, consider a space-time of the form $\mathbb{R} \times (\mathbb{R}^3 \setminus p)$, where the first factor of \mathbb{R} is parametrized by time, and the second factor is space, which is flat Euclidean space \mathbb{R}^3 except that we excise the point p (we could equally excise a 3-ball).¹⁴ It turns out that this space-time is not contractible to a point, but only to a 2-sphere surrounding the point p . (Perhaps you can figure it in your mind. If not, just trust me for now.) Last term, when you learned about stereographic projections, you saw that a 2-sphere can be covered by two patches, see figure 5.2. For instance, we can take patch $U^{(1)}$ to cover everything north of the southern tropic, and patch $U^{(2)}$ to cover everything south of the northern tropic. The two patches overlap in the region between the two tropics near the equator, so we need to specify how the gauge field in the northern patch and the gauge field in the southern patch are related in this region where both are defined. As we will see, this freedom allows us to define a magnetic monopole, namely a pointlike magnetic charge, sitting at point p . This is very surprising, because Maxwell’s equations allow electric charge densities but not magnetic charge densities in the right-hand sides. As we will see later, we can by-pass this limitation by exploiting the topology of the gauge field.

*** EXERCISE:**

¹³Naively you might want to impose the simpler identification $A_\mu^{(1)} = A_\mu^{(2)}$, but taking into account that gauge fields are only defined modulo gauge transformations, one is led to the more general (and mathematically correct) identification in the main text. It took physicists several decades to appreciate this point.

¹⁴I use the symbol \setminus to denote set difference. If you are used to the ordinary $-$ sign to denote set difference as well, please let me know and I’ll change my notation accordingly.

[Ex 8] So far I have assumed for simplicity that the complex scalar field ϕ has charge 1. Go through this chapter and work out how all formulae change if ϕ has charge $q \in \mathbb{Z}$ rather than charge 1.

5.4 Gauge redundancy and gauge fixing

A good reference for this topic is section 6 of David Tong's QFT lecture notes [Tong, 2006].

Let us start from the equations of motion (EoM) of the theory of **scalar electrodynamics**, which is described by the action (5.39). We recall here the Lagrangian density

$$\mathcal{L} = -|D_\mu \phi|^2 - V(\bar{\phi}, \phi) - \frac{1}{4g^2} F_{\mu\nu}^2,$$

where $F_{\mu\nu}^2 \equiv F_{\mu\nu} F^{\mu\nu}$ etc, and the scalar potential takes the form $V(\bar{\phi}, \phi) = U(|\phi|^2)$ to ensure gauge invariance. Then the Euler-Lagrange equations are

$$\begin{aligned} 1) \quad D_\mu D^\mu \phi &= -\frac{\partial V}{\partial \bar{\phi}} \equiv -U'(|\phi|^2)\phi \\ 2) \quad \partial_\nu F^{\mu\nu} &= g^2 J^\mu \end{aligned} \tag{5.49}$$

where

$$J_\mu = -i(\bar{\phi} D_\mu \phi - \phi D_\mu \bar{\phi}) = j_\mu - 2A_\mu |\phi|^2 \tag{5.50}$$

is a conserved current. The EoM for $\bar{\phi}$ is the complex conjugate of the EoM for ϕ , so I will not write it explicitly. Note that upon gauging the global $U(1)$ symmetry, the conserved current j_μ (5.34) of the scalar field theory with global $U(1)$ symmetry gets a correction term, due to the presence of the gauge field A_μ in the covariant derivatives.

Proof. **[Ex 10]**

Let us now consider the transformation properties of the EoM (5.2) under a $U(1)$ gauge transformation (5.41). The equations transform as

$$\begin{aligned} 1) &\mapsto e^{i\alpha} 1) && \text{(gauge covariant)} \\ 2) &\mapsto 2) && \text{(gauge invariant)} \end{aligned} \tag{5.51}$$

Therefore, if a field configuration (ϕ, A_μ) solves the EoM (5.49), then any gauge transformed field configuration $(\phi' = e^{i\alpha}\phi, A'_\mu = A_\mu + \partial_\mu \alpha)$ also solves the EoM (5.49): the EoM only determine (ϕ, A_μ) up to a gauge transformation.

Given some initial data $(\phi^{(0)}, A_\mu^{(0)})$ specifying the field configuration at an initial time t_0 , we cannot uniquely determine the field configuration (ϕ, A_μ) at a later time $t > t_0$. Indeed

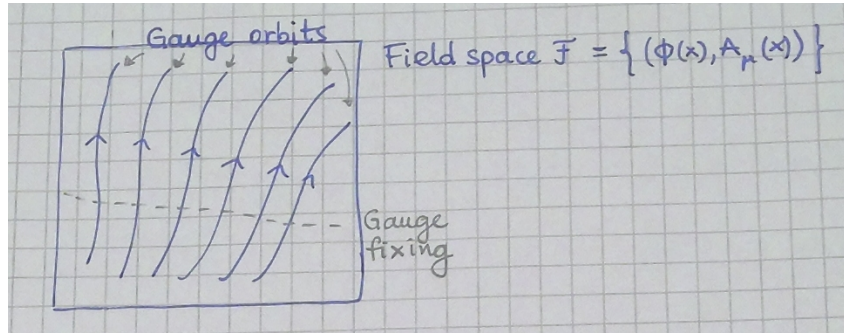


Figure 5.3: The space of all field configurations decomposes into the disjoint union of gauge orbits, each represents a single physical configuration. A complete gauge fixing selects a single representative for each orbit.

$(\phi' = e^{i\alpha}\phi, A'_\mu = A_\mu + \partial_\mu\alpha)$ is as good a solution of the EoM as (ϕ, A_μ) , and obeys the same initial condition provided that the gauge parameter α obeys the conditions $\alpha(t_0, \vec{x}) = 0 \pmod{2\pi}$ and $\partial_\mu\alpha(t_0, \vec{x}) = 0$ at the initial time t_0 .

We appear to be in trouble: we would like the EoM to define a **well-posed initial value problem** and determine **uniquely** physically observable fields at later times. This is not the case if we regard field configurations which differ by a gauge transformation as physically inequivalent. If instead we declare field configurations which differ by a gauge transformation to be physically equivalent, then the issue disappears and the initial value problem is well-posed. We will therefore **identify field configurations related by a gauge transformation**,

$$(\phi, A_\mu) \sim (\phi' = e^{i\alpha}\phi, A'_\mu = A_\mu + \partial_\mu\alpha). \quad (5.52)$$

Physically observable quantities must then be **gauge invariant**, such as for example the field strength $F_{\mu\nu}$, the magnitude of the scalar field $|\phi|^2$, or the conserved current J_μ . This explains remark 3 in the previous section.

The picture to keep in mind for gauge theories is that field space $\mathcal{F} = \{\phi(x), A_\mu(x)\}$ is foliated¹⁵ by **gauge orbits** traced by the action of the gauge group

$$\mathcal{G} \cdot (\phi(x), A_\mu(x)) = \{(e^{i\alpha(x)}\phi(x), A_\mu(x) + \partial_\mu\alpha(x) \mid \alpha(x) \sim \alpha(x) + 2\pi\}.$$

In down to earth terms, a gauge orbit simply consists of all the field configurations which are related by a gauge transformation.

Then the identification (5.52) of field configurations related by gauge transformations states

¹⁵Foliation is a mathematical term, from 'folia', Latin for 'leaf'. You can look up the technical definition if you are interested. For our purposes, you can take it to mean that field space is a union of disjoint orbits of the gauge group.

the correspondence¹⁶

$$\text{Physical configuration} \longleftrightarrow \text{Gauge orbit} .$$

Rather than working with the redundant description of field space \mathcal{F} subject to the gauge symmetry \mathcal{G} , it is often useful to “**fix a gauge**” (or pick a gauge, that is, picking a single **representative** for each gauge orbit). Any representative does the job – after all any two representatives of a given gauge orbit are physically equivalent – but we need to ensure that the gauge fixing cuts each orbit once and only once, as in figure 5.3. If that is not the case, and there is some leftover gauge symmetry that is not fixed, we refer to the gauge fixing as partial or incomplete, and further conditions must be specified in order to have a complete gauge fixing. The topic of gauge fixing is rather technical, and plays an important role in the quantization of gauge theories. Here we will content ourselves with giving a few standard examples of (partial) gauge fixing, which may be useful later on.

EXAMPLES:

1. **Lorenz gauge:**

This gauge is defined by imposing the constraint

$$\partial_\mu A^\mu = 0 \tag{5.53}$$

on the gauge field 4-vector A_μ . This can always be achieved. Indeed, if we are given a representative A_μ which does not obey the Lorenz gauge condition (5.53), then we can find another representative $A'_\mu = A_\mu + \partial_\mu \alpha$ in the same gauge orbit which obeys the Lorenz gauge constraint

$$0 = \partial_\mu A'^\mu = \partial_\mu A^\mu + \partial_\mu \partial^\mu \alpha \tag{5.54}$$

by picking α to be a solution of the inhomogeneous equation

$$\partial_\mu \partial^\mu \alpha = -\partial_\mu A^\mu , \tag{5.55}$$

which exists.¹⁷

Let us discuss pros and cons of the Lorenz gauge. The main advantage of the Lorenz gauge is that the constraint (5.53) is **Lorentz invariant**.¹⁸ The main disadvantage of the

¹⁶If you are formally minded, you would say that the physical configuration space \mathcal{C} is the quotient of the field space \mathcal{F} by the gauge group \mathcal{G} ,

$$\mathcal{C} = \mathcal{F}/\mathcal{G} ,$$

namely the set of equivalence classes of field configurations under the equivalence relation (5.52).

¹⁷Here the right-hand side $-\partial_\mu A^\mu$ is given and acts as a source in a relativistic Poisson equation for α . Solutions can be found by the method of Green’s functions.

¹⁸The Lorenz gauge is due to the Danish physicist Ludvig Lorenz, not to be confused with the more famous Dutch physicist Hendrik Lorentz, who is responsible for the Lorentz transformations which leave the laws of special relativity invariant, as well as for introducing the Lorentz force which acts on relativistic particles moving in a magnetic field. Click on the names of the physicists to see who is who.

Lorenz gauge is that it only fixes the gauge partially. Indeed, if we are in Lorenz gauge we are free to perform gauge transformations with parameters α such that $\partial_\mu \partial^\mu \alpha = 0$ and we will remain in the Lorenz gauge. (This corresponds to adding a solution of the homogeneous equation in (5.55).)

2. Coulomb gauge (or radiation gauge):

This gauge is defined by imposing the constraint

$$\nabla \cdot \vec{A} = 0 \quad (5.56)$$

on the vector potential \vec{A} , which is the spatial part of the 4-vector A_μ . This can always be achieved, by a similar reasoning to above.

Compared to the Lorenz gauge, the Coulomb gauge has the clear drawback of not being Lorentz covariant. So this gauge fixing spoils the manifest relativistic symmetry of the formalism, which is not ideal. (The physics of the system remains Lorentz invariant, because gauge transformations are unphysical, they are just a redundancy in our description.) Another drawback, in common with the Lorenz gauge, is that the Coulomb gauge constraint (5.56) only fixes the gauge partially. The argument is the same as for the Lorenz gauge, except that we are using spatial indices only instead of full space-time indices.

On the other hand, a pro of the Coulomb gauge is that the temporal component A_0 of the gauge potential (aka the ‘electric scalar potential’ in electromagnetism) is determined by the charge density $\rho = J^0$ as in electrostatics:

$$A_0(t, \vec{x}) \propto \int d^3x' \frac{\rho(t, \vec{x}')}{|\vec{x} - \vec{x}'|}. \quad (5.57)$$

So if the charge density $\rho = 0$, for instance for ‘pure electromagnetism’, in which there is no charged matter ϕ , we have

$$A_0 = 0$$

in Coulomb gauge. On the other hand, if there are charged fields and hence $\rho \neq 0$, then $A_0 \neq 0$.

* EXERCISE:

Determine the proportionality factor in (5.57). [Hint: use $\nabla^2 \frac{1}{4\pi|\vec{x}|} = \delta^{(3)}(\vec{x})$.]

REMARK: [Ex 11]

It is often possible to fix a gauge where¹⁹

$$A_0 = 0 .$$

In this gauge the **energy** (or more precisely the Hamiltonian) of **scalar electrodynamics** (5.39) is

$$E = \int d^3x \left[|\partial_0\phi|^2 + |(\nabla - i\vec{A})\phi|^2 + U(|\phi|^2) + \frac{1}{2g^2}(\vec{E}^2 + \vec{B}^2) \right] \quad (5.58)$$

where

$$E_i = -\partial_0 A_i , \quad B_i = \frac{1}{2}\epsilon_{ijk}F_{jk} \quad (5.59)$$

are the electric and magnetic field.

5.5 $U(1)$ Wilson line and Wilson loop

Let us conclude this chapter with an appetizer of geometric aspects that we will hopefully return to later. A good reference for this section is section 15.1 of the book by Peskin and Schroeder [Peskin, 1995].

We start by recalling that if ϕ is a charged scalar (of charge 1 for definiteness), then its partial derivative is not gauge covariant, that is, it does not transform under a well-defined representation of the $U(1)$ gauge group. You have seen this explicitly in the first term, when you worked out how $\partial_\mu\phi$ transforms under a $U(1)$ gauge transformation (5.41). One can fix this problem by introducing the gauge covariant derivative $D_\mu\phi = (\partial_\mu - iA_\mu)\phi$, which transform covariantly as a field of charge 1 under the gauge transformation (5.41). Hopefully this is all clear by now at a technical level. But why is this, conceptually?

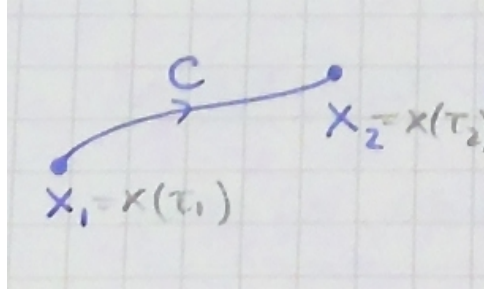
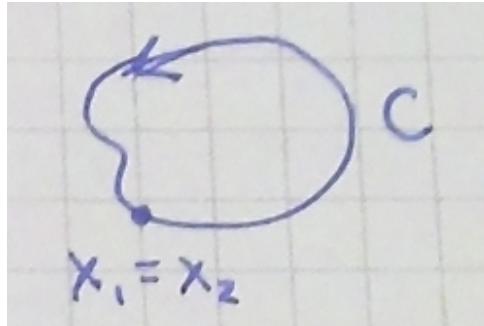
To analyze all the partial derivatives in one fell swoop, let us consider the **total differential** of $\phi(x)$,

$$d\phi(x) = \lim_{\epsilon \rightarrow 0} \frac{\phi(x + \epsilon dx) - \phi(x)}{\epsilon} = \partial_\mu\phi(x)dx^\mu , \quad (5.60)$$

where I have introduced an infinitesimal book-keeping parameter ϵ in front of the line increment dx^μ , so that I could write the total differential as a limit. The final expression, which writes the total differential of $\phi(x)$ as the 4-vector $\partial_\mu\phi(x)$ contracted with the differential increment dx^μ , follows from Taylor expanding the numerator inside the limit and by taking the limit (see Calculus and AMV).

The reason why the total differential (5.60) of ϕ (and hence its partial derivatives) does not transform covariantly under gauge transformations is that the two terms that we are sub-

¹⁹This is allowed because A_0 has no kinetic term which involves its time derivative in (5.39). A_0 is therefore non-dynamical: it can be determined at all times from A_i and the values of other charged fields in the theory.

Figure 5.4: An open curve from point x_1 to point x_2 .Figure 5.5: A closed curve (or ‘loop’) with base-point $x_1 = x_2$.

tracting inside the limit have different gauge transformation properties

$$\begin{aligned}\phi(x + \epsilon dx) &\mapsto e^{i\alpha(x + \epsilon dx)}\phi(x + \epsilon dx) \\ \phi(x) &\mapsto e^{i\alpha(x)}\phi(x),\end{aligned}$$

because $\alpha(x + \epsilon dx) \neq \alpha(x)$.

This problem can be fixed by introducing the ‘Wilson line’, or, as we will learn in later chapters, the mathematical notion of ‘parallel transport’.

Let C be an open curve (or a path) from point x_1 to point x_2 , see figure 5.4. Mathematically, this is a smooth map from an interval to space-time $\mathbb{R}^{1,3}$

$$\begin{aligned}C : \quad I = [\tau_1, \tau_2] &\mapsto \mathbb{R}^{1,3} \\ \tau &\mapsto x^\mu(\tau)\end{aligned}$$

with $x(\tau_1) = x_1$ and $x(\tau_2) = x_2$ at the endpoints.

The **Wilson line** (of charge 1) along the path C is defined to be

$$W_C(x_2, x_1) := \exp \left[i \int_{x_1, C}^{x_2} A_\mu(x) dx^\mu \right] \equiv \exp \left[i \int_{\tau_1}^{\tau_2} A_\mu(x(\tau)) \dot{x}^\mu(\tau) d\tau \right], \quad (5.61)$$

where the first integral is the line integral from x_1 to x_2 along C , and the second integral is its expression in the parametrization $x^\mu(\tau)$. If C is a closed path (or a ‘loop’), namely if $x_1 = x_2$ as in figure 5.5, then

$$W_C := \exp \left[i \oint_C A_\mu(x) dx^\mu \right] \quad (5.62)$$

is called the **Wilson loop** (of charge 1) along the curve C . By standard results from multivariate calculus, the line integral $\oint_C A_\mu(x) dx^\mu$ only depends on the curve C and not on the base-point $x_1 = x_2$.

Under a $U(1)$ gauge transformation (5.41), we claim that the Wilson line (5.61) transforms as²⁰

$$W_C(x_2, x_1) \mapsto e^{i\alpha(x_2)} W_C(x_2, x_1) e^{-i\alpha(x_1)}. \quad (5.63)$$

Proof.

$$\begin{aligned} W_C(x_2, x_1) &= e^{i \int_{x_1, C}^{x_2} A_\mu dx^\mu} \mapsto e^{i \int_{x_1, C}^{x_2} (A_\mu + \partial_\mu \alpha) dx^\mu} \\ &= e^{i \int_{x_1, C}^{x_2} A_\mu dx^\mu} e^{i \int_{x_1, C}^{x_2} \partial_\mu \alpha dx^\mu} \\ &= W_C(x_2, x_1) e^{i(\alpha(x_2) - \alpha(x_1))} \\ &= e^{i\alpha(x_2)} W_C(x_2, x_1) e^{-i\alpha(x_1)}. \end{aligned}$$

To go from the second to the third line, we have used the fact that $\partial_\mu \alpha dx^\mu = d\alpha(x)$ is an exact differential, so its integral along a curve C only receives contribution from the boundary terms.

A corollary of the gauge transformation (5.63) is that the $U(1)$ Wilson loop (5.62) is gauge invariant. To see that, simply set $x_1 = x_2$, or use the fact that the integral of an exact differential along a closed curve vanishes.

Now we can combine the gauge transformation of a charged scalar field and of a Wilson line of the same charge (here 1 for simplicity) to find that the gauge transformation of the product of the charged scalar $\phi(x_1)$ at $x = x_1$ and the Wilson line from x_1 to x_2 along C

$$\begin{aligned} W_C(x_2, x_1) \phi(x_1) &\mapsto e^{i\alpha(x_2)} W_C(x_2, x_1) e^{-i\alpha(x_1)} e^{i\alpha(x_1)} \phi(x_1) \\ &= e^{i\alpha(x_2)} W_C(x_2, x_1) \phi(x_1) \end{aligned} \quad (5.64)$$

is by the same phase $e^{i\alpha(x_2)}$ as for $\phi(x_2)$.

²⁰For the gauge group $G = U(1)$, which we are considering here, the Wilson line and the gauge transformations $e^{i\alpha(x_i)}$ commute, so we could have written the gauge transformation of the Wilson line simply as

$$W_C(x_2, x_1) \mapsto e^{i(\alpha(x_2) - \alpha(x_1))} W_C(x_2, x_1).$$

I wrote the result like (5.63) for comparison to the case of a non-abelian gauge group, which we will study later.

Therefore it makes sense to consider the **total covariant differential**

$$D\phi(x) = \partial_\mu\phi(x)dx^\mu := \lim_{\epsilon \rightarrow 0} \frac{\phi(x + \epsilon dx) - W_{dC}(x + \epsilon dx, x)\phi(x)}{\epsilon}, \quad (5.65)$$

where we have inserted the Wilson line along an infinitesimal line element dC connecting x to $x + \epsilon dx$ in front of $\phi(x)$. This ensures that the terms which are subtracted inside the limit have the same gauge transformation property.

Expanding to first order in ϵ ,

$$\phi(x + \epsilon dx) = \phi(x) + \epsilon\partial_\mu\phi(x)dx^\mu + O(\epsilon^2)$$

and

$$\begin{aligned} W_{dC}(x + \epsilon dx, x) &= \exp \left[i \int_{x, dC}^{x+\epsilon dx} A_\mu(x')dx'^\mu \right] \\ &= \exp \left[iA_\mu(x)\epsilon dx^\mu + O(\epsilon^2) \right] \\ &= 1 + i\epsilon A_\mu(x)dx^\mu + O(\epsilon^2), \end{aligned} \quad (5.66)$$

and substituting in (5.65) we find

$$\begin{aligned} D\phi(x) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left[\phi(x) + \epsilon\partial_\mu\phi(x)dx^\mu - \phi(x) - i\epsilon A_\mu(x)\phi(x)dx^\mu + O(\epsilon^2) \right], \\ &= (\partial_\mu\phi(x) - i\epsilon A_\mu(x)\phi(x)) dx^\mu \equiv D_\mu\phi(x)dx^\mu, \end{aligned} \quad (5.67)$$

which precisely reproduces the previous definition (5.40) of the covariant derivative of a scalar field $\phi(x)$ of charge 1!

REMARKS:

1. In QM, the Wilson line $W_C(x_2, x_1)$ is the phase picked up by the wave-function of a charged point particle slowly (more precisely, ‘adiabatically’) moving from x_1 to x_2 along a path C in the presence of a gauge field.
2. The Wilson loop (5.62) is gauge invariant and therefore physically observable. It is the phase picked up by the wave-function of a charged point particle slowly moving along a loop C . This phase controls the Aharonov-Bohm effect in QM, a subtle and unexpected form of quantum interference which arises because the wave-function couples directly to the gauge potential A_μ rather than to the physical electric and magnetic fields \vec{E} , \vec{B} .

If the loop C is the boundary of a surface Σ , then by a higher-dimensional version of Stokes’

theorem (see Differential Geometry III) one has

$$\begin{aligned} \oint_C A_\mu(x) dx^\mu &= \frac{1}{2} \int_\Sigma F_{\mu\nu}(x) dx^\mu \wedge dx^\nu \\ &\equiv \frac{1}{2} \int_{x^{-1}(\Sigma)} F_{\mu\nu}(x(\sigma)) \left(\frac{\partial x^\mu(\sigma)}{\partial \sigma^1} \frac{\partial x^\nu(\sigma)}{\partial \sigma^2} - \frac{\partial x^\nu(\sigma)}{\partial \sigma^1} \frac{\partial x^\mu(\sigma)}{\partial \sigma^2} \right) d\sigma^1 d\sigma^2 \end{aligned} \quad (5.68)$$

where $x^\mu(\sigma) \equiv x^\mu(\sigma^1, \sigma^2)$ is a parametrization of the surface Σ .²¹ The previous formula is a higher-dimensional analogue of Stokes' theorem

$$\oint_C \vec{A} \cdot d\vec{l} = \int_\Sigma (\nabla \times \vec{A}) \cdot \hat{n} d^2\sigma = \int_\Sigma \vec{B} \cdot \hat{n} d^2\sigma, \quad (5.69)$$

which is used in electromagnetism to relate the circulation of the vector potential \vec{A} along C to the magnetic flux through a surface with boundary C . The formula (5.68) tells us that the field strength $F_{\mu\nu}$ encodes the value of infinitesimal Wilson loops.

If the loop C is not contractible to a point, it may happen that $A_\mu \neq 0$ and therefore

$$\oint_C A_\mu dx^\mu \neq 0$$

even if the field strength $F_{\mu\nu} = 0$ vanishes everywhere in the region probed by a quantum-mechanical particle (or by a charged scalar field). Examples of spaces which allow these phenomenon are $\mathbb{R}^2 \setminus p$, for loops which encircle the removed point p , or the torus T^n , for loops that wind non-trivially around a circle direction in the torus.

Time permitting, we will return to the Aharonov-Bohm effect later. For an accessible summary, see section 10.5.3 of [Nakahara, 2003], up to equation (10.100).

²¹ $F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ is called a differential 2-form. It can be shown that the surface integral of a differential 2-form is independent under reparametrizations of the surface that preserve its orientation, much like line integrals of a differential 1-form $A = A_\mu dx^\mu$.

Chapter 6

Applications of abelian gauge theories

In this chapter we will study non-trivial field configurations of abelian gauge theories (with and without charged matter fields) whose existence and stability is ensured by topology. We will see that this topology is associated to maps from a circle to a circle, and we will investigate some of the underlying geometry and physics.

6.1 Global vortices

Two good references for this section of the notes are section 3.1 of [Weinberg, 2012] and sections 7.1, 7.2 of [Manton and Sutcliffe, 2004].

Vortices are topological solitons, finite energy localized solutions of the field equations whose shape is protected by topology, which look like particles in **2 space and 1 time dimensions**.¹

Three-dimensional Minkowski spacetime has coordinates $(x^0 = t, x^1, x^2)$. It is convenient to introduce a complex coordinate

$$z = x^1 + ix^2 = re^{i\theta} \tag{6.1}$$

for space \mathbb{R}^2 , viewed as the complex plane \mathbb{C} . We consider a complex scalar field ϕ of charge 1 under a $U(1)$ internal global symmetry. The Lagrangian density has the standard canonical kinetic term $-(\partial^\mu \bar{\phi})(\partial_\mu \phi)$ and a $U(1)$ -symmetric scalar potential

$$V(\bar{\phi}, \phi) = U(|\phi|^2) = \frac{\lambda}{2}(|\phi|^2 - v^2)^2, \tag{6.2}$$

¹We are always free to add extra spatial dimensions on which the solutions do not depend. If we add one extra dimension to get to 3+1 spacetime dimensions, vortices look like strings. If we add another extra dimensions to get to 4+1 spacetime dimensions, vortices look like membranes. And so on and so forth.

where $\lambda, v > 0$ are constants. See figure 5.1.

We will be interested in static field configurations, which have $\dot{\phi} \equiv \partial_0 \phi = 0$. Their energy is

$$E = \int d^2x [|\nabla\phi|^2 + U(|\phi|^2)] . \quad (6.3)$$

The vacua of the system (static minimum energy configurations) are given by

$$\phi = |\phi| e^{i \arg(\phi)} = v e^{i\alpha} \quad (6.4)$$

where $\alpha \sim \alpha + 2\pi$ is an arbitrary constant that labels the vacuum. These configurations parametrize the **vacuum manifold** (or ‘**moduli space of vacua**’)

$$\mathcal{V} \cong S^1 , \quad (6.5)$$

which is a circle of radius v in field space (*i.e.* ϕ space).

It is not the vacua that we are interested in, however. Rather, we will focus on less trivial static field configurations of finite energy, which look as follows if we draw the vector field $\vec{\phi} = (\text{Re } \phi, \text{Im } \phi) \equiv (\phi_1, \phi_2)^2$ on the spatial plane \mathbb{R}^2 :

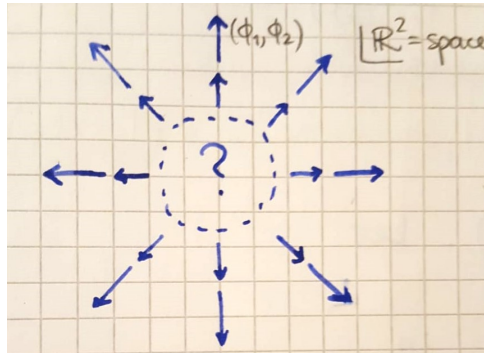


Figure 6.1: Sketch of the field $\vec{\phi} = (\phi_1, \phi_2)$ for a vortex solution.

We demand that the magnitude $|\vec{\phi}|$ of $\vec{\phi}$ (or equivalently $|\phi|$ in the complex representation for the scalar field) tends to v at spatial infinity, which is a necessary condition to have finite energy, and that the vector $\vec{\phi}$ rotates by 2π as we follow the circle at spatial infinity.

REMARKS:

²Note that in this representation $\vec{\phi}$ transforms as a doublet (or vector) under the $SO(2)$ internal global symmetry. We are using the obvious isomorphism between $U(1)$ and $SO(2)$.

1. Such a field configuration cannot be deformed continuously to the vacuum, if we keep the asymptotics at spatial infinity fixed. (In order to deform to a vacuum, where $\vec{\phi}$ is constant and points in a fixed direction, we would need to provide infinite energy.)
2. Field configurations which obey this asymptotics requirement can be deformed continuously until the energy (6.3) is minimized. The resulting minimum energy configuration solves the EoM [Ex 14] and is called a **global vortex**.³

There is some interesting **topology** underlying these field configurations. Given a smooth configuration on ϕ , we can define the **vorticity** / **vortex number** / **winding number** associated to a loop C

$$N[C] := \frac{1}{2\pi} \oint_C \nabla \arg(\phi) \cdot d\vec{l} \equiv \frac{1}{2\pi} \oint_C \partial_i \arg(\phi) dx^i, \quad (6.6)$$

which is the circulation of the vector field $\nabla \arg(\phi)$ along the loop C . (Notation: $d\vec{l} = (dx^1, dx^2)$ is the infinitesimal line element.)

REMARKS:

1. If $\phi \neq 0$ along C , then $\arg(\phi)$ is well defined and (6.6) counts the number of full rotations that $\arg(\phi)$ makes during one counterclockwise circuit along C , which is an **integer**.
2. If we deform C continuously, the vortex number $N[C]$ also changes continuously. But it's an integer, so it can only be continuous by being a **constant**. (One says that $N[C]$ is topologically conserved.)
3. There is an **exception** to the previous fact: if C passes through a point P where $\phi = 0$, then $\arg(\phi)$ is ill-defined, and so is the vortex number (6.6). (A similar argument applies to points where $\phi \rightarrow \infty$, if we allow them.) If C and C' are two loops which only differ by how they avoid P , as in figure 6.2, we have

$$N[C] - N[C'] = N[C_P] \quad (6.7)$$

where C_P is an infinitesimal loop that encircles the point P .

*** EXERCISE:**

[Ex 15]

³Global because it is invariant under a global internal $U(1)$ symmetry. Vortex because, as we will see later, the flow lines of the gradient of $\arg(\phi)$ resemble a vortex.

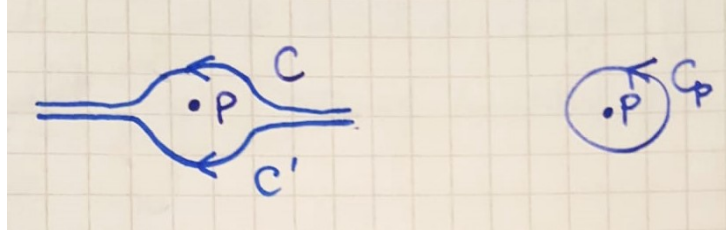


Figure 6.2: Infinitesimal deformation of a contour C to C' through a zero of ϕ , and $C_P = C - C'$ (in homology, really).

- (a) Use Stokes' theorem to write the vortex number as a surface integral, and show that it can only receive contributions from points where $\phi = 0$ or $1/\phi = 0$ (you can assume that ϕ is smooth otherwise).
- (b) Assume for simplicity that ϕ is a holomorphic function of z . Let z_0 be a zero of ϕ of order n if $n > 0$, and a pole of order $|n|$ if $n < 0$, that is $\phi(z) \approx c(z - z_0)^n$ near $z = z_0$, where $c \neq 0$. Show that $N[C_{z_0}] = n$ for any infinitesimal loop C_{z_0} which encircles z_0 counterclockwise.
- (c) Generalize the previous calculation to $\phi(z, \bar{z}) \approx c(z - z_0)^n \overline{(z - z_0)}^m$.

We say that there are:

$$\begin{aligned} &N[C_P] \text{ vortices sitting at } P \text{ if } N[C_P] > 0 \\ &-N[C_P] \text{ antivortices sitting at } P \text{ if } N[C_P] < 0. \end{aligned}$$

4. From the previous remarks we conclude that

$$N[C] = \#(\text{vortices}) - \#(\text{antivortices}) \text{ enclosed by } C. \tag{6.8}$$

Letting $C = S^1_\infty$, the circle at spatial infinity (parametrized by θ as $r \rightarrow \infty$), we find the **total vortex number**

$$N \equiv N[S^1_\infty] = \text{total } \#(\text{vortices}) - \text{total } \#(\text{antivortices}). \tag{6.9}$$

5. **Topology** is the branch of mathematics that studies shapes up to continuous deformations, and maps between them. Topologically, $C \simeq S^1$ and

$$\arg(\phi)|_C : C \simeq S^1 \longrightarrow S^1$$

counts how many times the spatial loop C winds around the circle parametrized by $\arg(\phi)$ in field space (or 'target space'). Continuous Maps from S^1 to S^1 are labelled by an integer, the winding number of the map. This is what $N[C]$ represents mathematically.

Now let's look for a **static solution of the field equations**

$$\nabla^2 \phi - \lambda(|\phi|^2 - v^2) = 0 \quad (6.10)$$

with total vortex number $N = 1$. Solving this equation in general is hard, so let's simplify our lives by making the **ansatz**

$$\phi(\vec{x}) = f(r)e^{i\theta}, \quad (6.11)$$

which clearly has unit vortex number. Note that (6.11) is invariant under a combined rotation in the spatial plane \mathbb{R}^2 and a phase rotation of ϕ .

If we further demand that ϕ be invariant under a reflection about the x^1 axis combined with complex conjugation of ϕ , that is

$$\phi(r, \theta) = \overline{\phi(r, -\theta)}, \quad (6.12)$$

it follows that

$$f(r) \in \mathbb{R}. \quad (6.13)$$

Substituting (6.11) and (6.13) in the equation of motion (6.10), one finds the ODE

$$f'' + \frac{1}{r}f' - \frac{1}{r^2}f + \lambda(v^2 - f^2)f = 0 \quad (6.14)$$

for the radial profile $f(r)$.

Proof. [Ex 16].

We impose the **boundary conditions**

$$f(0) = 0, f(\infty) = v \quad (6.15)$$

which are necessary for ϕ to be non-singular at the origin and for the integral of the energy density to converge at spatial infinity. With these boundary conditions, the ODE (6.14) can be solved numerically. However we run into a problem: even though $U(|\phi|^2) \rightarrow 0$ at spatial infinity, the **static energy** (6.3) is **infinite** because $\nabla\phi \not\rightarrow 0$ at spatial infinity, due to the θ -dependence in (6.11).

*** EXERCISE:**

[Ex 17]

1. Let $\phi(\vec{x}) = \rho(\vec{x}) \exp(i\alpha(\vec{x}))$, where $\rho(\vec{x}), \alpha(\vec{x})$ are real functions. Show that

$$E = \int d^2x \left[(\nabla\rho)^2 + \rho^2(\nabla\alpha)^2 + \frac{\lambda}{2}(\rho^2 - v^2)^2 \right].$$

2. Let $\phi = f(r) \exp(i\theta)$ where (r, θ) are polar coordinates on the spatial plane, and $f(r)$ is a real function which obeys the boundary conditions $f(0) = 0$ and $f(\infty) = v$. Show that

$$\rho^2 (\nabla \alpha)^2 = \frac{f^2}{r^2}$$

and use the boundary conditions to show that this causes a logarithmic divergence of the energy E as $r \rightarrow \infty$. That is, let

$$E_R \equiv \int_{r \leq R} d^2x \mathcal{E}$$

and show that $E_R \sim \log R$ as $R \rightarrow \infty$.

6.2 Derrick's theorem

The previous negative result is not an accident, but rather a particular case of a general theorem due to Derrick [Derrick, 1964]. See section 3.2 of [Weinberg, 2012] for this section.

Consider a **scalar field theory in D space dimensions**, with Lagrangian density

$$\mathcal{L} = -\frac{1}{2} G_{ab}(\phi) \partial_\mu \phi^a \partial_\mu \phi^b - V(\phi) \quad (6.16)$$

where the matrix $G_{ab}(\phi)$ is positive definite for all values of the scalar fields ϕ^a and $V(\phi)$ vanishes at its global minima. (The latter can always be arranged by a shift of the potential.) Then any **finite energy static solution** of the field equations is a **stationary point** of the **static energy functional**

$$\begin{aligned} E[\phi] &= E_K[\phi] + E_V[\phi], \\ E_K[\phi] &= \int d^Dx \frac{1}{2} G_{ab}(\phi) \partial_i \phi^a \partial_i \phi^b \\ E_V[\phi] &= \int d^Dx V(\phi). \end{aligned} \quad (6.17)$$

Spatial indices $i = 1, 2, 3$ are summed over in the kinetic static energy E_k , and we do not distinguish upper and lower indices. Internal indices a, b are also summed over. Both the kinetic static energy E_k and the potential static energy E_V are non-negative, by our assumption.

* EXERCISE:

[Ex 18.1] Prove the claim made above (6.17). [**Hint:** treat $E[\phi]$ as an action for static field configurations, and derive its Euler-Lagrange equations.]

Take a static solution of the equations of motion, $\phi(x) = \phi_1(x)$, and consider the one-parameter family of field configurations

$$\phi(x) = \phi_\lambda(x) := \phi_1(\lambda x) \quad (6.18)$$

labelled by the constant $\lambda > 0$. Then

$$\begin{aligned} E[\phi_\lambda] &= E_K[\phi_\lambda] + E_V[\phi_\lambda] \\ &= \lambda^{2-D} E_K[\phi_\lambda] + \lambda^{-D} E_V[\phi_\lambda]. \end{aligned} \quad (6.19)$$

Proof. [Ex 18.2].

We know that the static solution $\phi_1(x)$ of the equations of motion is a stationary point of the functional $E[\phi]$, that is $\delta E[\phi] = 0$ at $\phi = \phi_1$ for any infinitesimal variation $\delta\phi$. If we restrict to the one-parameter family $\phi_\lambda(x)$, this means that $E[\phi_\lambda]$, which is a function of λ , is stationary at $\lambda = 1$. In equations,

$$\begin{aligned} 0 &= \frac{d}{d\lambda} E[\phi_\lambda] \Big|_{\lambda=1} \\ &= (2-D)\lambda^{1-D} E_K[\phi_1] - D\lambda^{-1-D} E_V[\phi_1] \Big|_{\lambda=1} \\ &= (2-D)E_K[\phi_1] - DE_V[\phi_1]. \end{aligned} \quad (6.20)$$

What do we learn from this?

- **D=1:** We find

$$E_K[\phi_1] = E_V[\phi_1], \quad (6.21)$$

which allows finite energy solutions of the field equations which depend non-trivially on the space coordinates, as long as their kinetic energy is equal to their potential energy.

- **D=2:** We find

$$E_V[\phi_1] = 0, \quad (6.22)$$

which implies that $\phi = \phi_1$ minimizes the scalar potential for all x . This rules out a global vortex solution, which does not minimize the scalar potential everywhere (in particular, $\phi = 0$ at the core of the vortex).

- **D>2:** The two terms in the last line of (6.20) cannot have opposite signs. Hence the only solution to (6.20) is

$$E_K[\phi_1] = E_V[\phi_1] = 0, \quad (6.23)$$

which means that ϕ_1 is a vacuum solution (constant and a global minimum of the scalar potential).

The conclusion is **Derrick's theorem**: there are no non-trivial (that is, non-constant) finite energy static solutions to scalar field theories of the form (6.16) in $D > 1$ space dimensions.

If we want to find non-trivial static solutions in more than one space dimensions, we need to modify our assumptions and allow fields of different type. One can add spinors, but a similar argument applies. The **way out** is to introduce **gauge fields**.

6.3 Gauged vortices

For this part see section 3.3 of [Weinberg, 2012], sections 7.1 and 7.3 of [Manton and Sutcliffe, 2004] and subsection 2.5.2 of [Tong, 2018].

Let's return to our model for vortices in $2 + 1$ spacetime dimensions, but now let's **gauge the $U(1)$ symmetry**. The theory that we obtain is called the **Abelian Higgs model** and has Lagrangian density

$$\mathcal{L} = -\frac{1}{4g^2} F^{\mu\nu} F_{\mu\nu} - \overline{D^\mu \phi} D_\mu \phi - \frac{\lambda}{2} (|\phi|^2 - v^2)^2, \quad (6.24)$$

where λ and v are positive constants and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu \\ D_\mu \phi &= \partial_\mu \phi - iA_\mu \phi. \end{aligned} \quad (6.25)$$

In a gauge where $A_0 = 0$ (known as 'temporal gauge', or 'Weyl gauge'), the **energy of static field configurations** (which have $\partial_0 \phi = 0, \partial_0 A_i = 0$), is

$$E = \int d^2x \left[\sum_{j=1}^2 |D_j \phi|^2 + \frac{\lambda}{2} (|\phi|^2 - v^2)^2 + \frac{1}{2g^2} B^2 \right] \quad (6.26)$$

where

$$B = F_{12} = \partial_1 A_2 - \partial_2 A_1 \quad (6.27)$$

is the magnetic field, which is a scalar under spatial rotations in two dimensions.⁴

Proof. This follows from [Ex 11] in the problem sheet, specialized to two space dimensions and static fields.

REMARKS:

⁴If we add a third spatial dimension, B becomes the magnetic field transverse to the plane.

1. The vacua of the system are zero energy field configurations and obey

$$\begin{aligned} (1) \quad D_j \phi &\equiv \partial_j \phi - i A_j \phi = 0 \\ (2) \quad |\phi| &= v \\ (3) \quad B &= 0 . \end{aligned} \tag{6.28}$$

Equation (2) sets $\phi = v e^{i\alpha(\vec{x})}$, and substituting in equation (1) we can solve for

$$A_j = -i \partial_j \log \phi = \partial_j \alpha , \tag{6.29}$$

which is a ‘**pure gauge**’ configuration, namely a gauge field configuration which is gauge equivalent to 0. This implies that the field strength vanishes:

$$B = F_{12} = 0 , \tag{6.30}$$

consistently with equation (3).⁵

2. A **necessary condition** for the **static energy** E to be **finite** is that the **fields** ϕ, A_μ **approach a vacuum at spatial infinity**. It would be nice if we could argue that ϕ has a limiting form on the circle at spatial infinity S_∞^1 , but this is not possible because of gauge invariance. However, from the expression for the static energy in polar coordinates (r, θ)

$$E = \int_0^\infty r dr \int_0^{2\pi} d\theta \left[|D_r \phi|^2 + \frac{1}{r^2} |D_\theta \phi|^2 + \frac{\lambda}{2} (|\phi|^2 - v^2)^2 + \frac{1}{2g^2} \frac{1}{r^2} F_{r\theta}^2 \right] \tag{6.31}$$

we see that we must have $D_r \phi = \partial_r \phi - i A_r \phi \rightarrow 0$ as $r \rightarrow \infty$, which implies that $A_r - \partial_r \arg(\phi) \rightarrow 0$ and $\partial_r |\phi| \rightarrow 0$. Let’s now switch to the **radial gauge** in which $A_r = 0$, which we can do by applying a gauge transformation with group element $e^{i\beta}$ given by

$$\beta(r, \theta) = \int_r^\infty dr' A_r(r', \theta) , \tag{6.32}$$

which is smooth and obeys $\beta = 0$ at spatial infinity. This gauge transformation also changes $\phi \mapsto e^{i\beta} \phi$ and $A_\theta \mapsto A_\theta + \partial_\theta \beta$, but they both remain smooth if they were smooth in the first place, which we assume. In the radial gauge, the boundary condition that $D_r \phi \rightarrow 0$ as $r \rightarrow \infty$ simplifies to $\partial_r \phi \rightarrow 0$, so ϕ has a limiting value along each radial line, and we can set

$$\begin{aligned} \lim_{r \rightarrow \infty} \phi(r, \theta) &= v e^{i\alpha_\infty(\theta)} \\ \lim_{r \rightarrow \infty} A_\theta(r, \theta) &= \frac{d}{d\theta} \alpha_\infty(\theta) , \end{aligned} \tag{6.33}$$

⁵A side remark: Any choice of a vacuum value for the complex scalar field breaks the abelian $U(1)$ gauge symmetry. This is called spontaneous breaking of the $U(1)$ gauge symmetry, since the symmetry is preserved by the action but is broken by the state of the system. This is in contrast to explicit symmetry breaking, where a term that breaks the symmetry is added to the action. In quantum field theory, the spontaneous breaking of a continuous internal global symmetry implies the existence of massless particles, called Nambu-Goldstone bosons. The spontaneous breaking of a gauge symmetry instead to the famous Higgs mechanism in quantum field theory, in which the gauge boson (the quantum of the gauge field) gains a mass in a way that is consistent with gauge invariance, by ‘eating’ the degree of freedom of the would-be Goldstone boson. This is the reason for the name ‘abelian Higgs model’.

where we denoted the argument of ϕ on the circle at spatial infinity by $\alpha_\infty(\theta)$. In the second line of (6.33) we used the boundary condition $\lim_{r \rightarrow \infty} D_\theta \phi = 0$ to solve for the limit of A_θ in terms of the limit of ϕ .

3. The **total vortex number / winding number**

$$N = N[S_\infty^1] := \frac{1}{2\pi} \oint_{S_\infty^1} \nabla \alpha_\infty \cdot d\vec{l} = \frac{1}{2\pi} \int_0^{2\pi} d\theta \frac{d\alpha_\infty(\theta)}{d\theta} = \frac{1}{2\pi} (\alpha_\infty(2\pi) - \alpha_\infty(0)) \quad (6.34)$$

is **gauge invariant** and proportional to the **total magnetic flux** $\Phi(B)$ through the spatial plane:

$$N = \frac{1}{2\pi} \oint_{S_\infty^1} \vec{A} \cdot d\vec{l} = \frac{1}{2\pi} \int_{\mathbb{R}^2} d^2x B = \frac{1}{2\pi} \Phi(B). \quad (6.35)$$

We used (6.33) in the first equality and Stokes' theorem in the second equality. This means in particular that the magnetic flux through the spatial plane is quantized in units of 2π .⁶

4. The asymptotic profile $\alpha_\infty(\theta)$ of $\arg \phi$ on S_∞^1 defines a continuous map

$$\alpha_\infty : S_\infty^1 \rightarrow S^1. \quad (6.36)$$

Continuous maps from a circle to a circle fall into disjoint classes labelled by the integer N , the winding number, which counts the number of times $\alpha_\infty(S_\infty^1)$ winds around S^1 . The precise mathematical statement is that the **1st homotopy group** (or the **fundamental group**) of S^1 is the integers:

$$\pi_1(S^1) = \mathbb{Z}. \quad (6.37)$$

Now that we have understood the global topological nature of the field configurations, the next question to ask is whether we can find **static solutions** of the equations of motion with nontrivial winding number, for example $N = 1$. In the temporal and radial gauge ($A_0 = A_r = 0$), we can use the **ansatz**

$$\begin{aligned} \phi &= v e^{i\theta} f(vr) \\ A_j &= \epsilon_{jk} \hat{x}^k \frac{a(vr)}{r} \end{aligned} \quad (6.38)$$

where f and a are real functions of a single variable vr . We rescale the radial coordinate by v because that's the only (inverse length) scale in the problem.

REMARKS:

⁶For those who are studying QM: there is a hidden reduced Planck constant \hbar which I set to 1.

1. The ansatz (6.38) is rotationally symmetric: the effect of a spatial rotation can be undone by a gauge transformation.
2. The ansatz (6.38) is invariant under a reflection about the x^1 -axis combined with complex conjugation of ϕ .
3. We require that

$$f, a \xrightarrow[r \rightarrow 0]{} 0, \quad f, a \xrightarrow[r \rightarrow \infty]{} 1, \quad (6.39)$$

to ensure **regularity** at the origin (the center of the vortex) and the **boundary conditions** at infinity.

4. Using the ansatz (6.38), the equations of motion reduce to a system of two ODEs for f and a [Ex 19]. The solutions are not known analytically, but they are easy to find **numerically** and look like this:

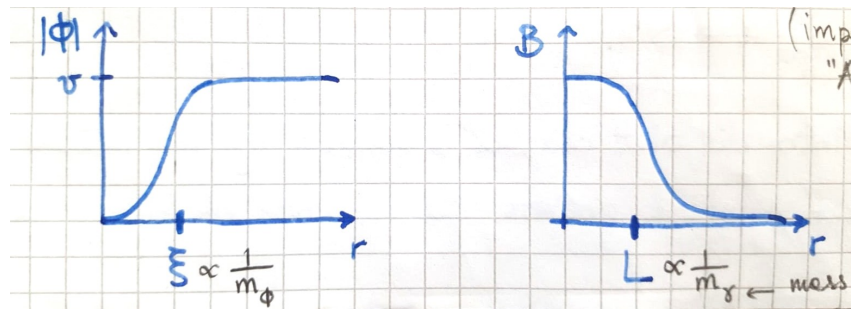


Figure 6.3: Qualitative behaviours of the magnitude of the scalar field $|\phi|$ and the magnetic field B for the Nielsen-Olesen vortex.

They are called **Nielsen-Olesen vortices** and play an important role in several areas of physics, including the theory of superconductors (where they are known as **Abrikosov vortices**).

5. The physics of multiple vortices turns out to depend on the ratio between the length scales $\xi \propto \frac{1}{m_\phi}$ and $L \propto \frac{1}{m_\gamma}$ associated to (quanta of) the ϕ field and of the electromagnetic field, due to a competition between an attractive force felt in regions where $\phi \approx 0$, and a repulsive force felt in regions where $B \neq 0$.
 - Type I ($\xi > L$): attraction wins. Separate vortices combine into one big multi-vortex, inside which we are just in a normal vacuum.
 - Type II ($\xi < L$): repulsion wins. If there are many vortices, the energy is minimized

when the individual vortices form a triangular lattice, known as the **Abrikosov lattice**.

- $\xi = L$: attraction and repulsion cancel, leading to a zero net force. This happens when $\lambda = g^2$ in the Lagrangian density (6.24). We study this case next.

6.3.1 Bogomol'nyi vortices

If $\lambda = g^2$, we can massage the static energy (6.26), using a trick originally due to Bogomol'nyi [Bogomol'nyi, 1976], to find a lower bound for the energy for every fixed value of the vortex number N . The argument involves **completing the square**, which means writing $a^2 + b^2 = (a \pm b)^2 \mp 2ab$, twice, and goes as follows:

$$\begin{aligned} E &= \int d^2x \left[\sum_{j=1}^2 |D_j \phi|^2 + \frac{g^2}{2} (|\phi|^2 - v^2)^2 + \frac{1}{2g^2} B^2 \right] \\ &= \int d^2x \left[|D_1 \phi - i\epsilon_1 D_2 \phi|^2 + i\epsilon_1 (\overline{D_1 \phi} D_2 \phi - \overline{D_2 \phi} D_1 \phi) \right. \\ &\quad \left. + \frac{1}{2g^2} (B - \epsilon_2 g^2 (|\phi|^2 - v^2))^2 + \epsilon_2 B (|\phi|^2 - v^2) \right] \end{aligned} \quad (6.40)$$

where ϵ_1, ϵ_2 are signs (i.e. $\epsilon_1^2 = \epsilon_2^2 = 1$). Integrating by parts and using the boundary conditions to drop boundary terms, we can write

$$\begin{aligned} \int d^2x \left(\overline{(D_1 \phi)} (D_2 \phi) - \overline{(D_2 \phi)} (D_1 \phi) \right) &= - \int d^2x \bar{\phi} [D_1, D_2] \phi \\ i \int d^2x \bar{\phi} F_{12} \phi &= i \int d^2x B |\phi|^2. \end{aligned} \quad (6.41)$$

Hence

$$\begin{aligned} E &= \int d^2x \left[|(D_1 - i\epsilon_1 D_2) \phi|^2 + \frac{1}{2g^2} (B - \epsilon_2 g^2 (|\phi|^2 - v^2))^2 \right. \\ &\quad \left. - \epsilon_1 B |\phi|^2 + \epsilon_2 B (|\phi|^2 - v^2) \right], \end{aligned} \quad (6.42)$$

where we notice that the first line is non-negative, being a sum of squares. Picking the two signs to be equal, $\epsilon_1 = \epsilon_2 \equiv \epsilon$, the $B|\phi|^2$ terms cancel out in the second line and we find the lower bound

$$E \geq -\epsilon v^2 \int d^2x B = -2\pi \epsilon v^2 N. \quad (6.43)$$

Since this holds for both values of ϵ , we can take $\epsilon = -\text{sign}(N)$ to obtain the **Bogomol'nyi bound**

$$E = 2\pi v^2 |N|, \quad (6.44)$$

which is a lower bound for the energy in terms of the absolute value of the vortex number (a topologically conserved charge).

The static solutions which saturate the bound, that is which have $E = 2\pi v^2|N|$, obey the **1st order Bogomol'nyi equations**

$$\begin{aligned} (D_1 - i\epsilon D_2)\phi &= 0 \\ B &= \epsilon g^2(|\phi|^2 - v^2) \end{aligned} \quad (6.45)$$

where $\epsilon = -\text{sign}(N)$. They are called **Bogomol'nyi vortices / antivortices**, according to the sign of the total vortex number N . It can be proven that a solution that contains both vortices and antivortices does not solve the Bogomol'nyi equations. The general solution to the Bogomol'nyi equations can be found numerically and has $2|N|$ real parameters, the positions of the centres of the N vortices (or $-N$ antivortices) in the plane.

6.4 The Dirac monopole (à la Wu and Yang)

For this topic, see sections 1.9, 9.4.1 and 10.5.2 of [Nakahara, 2003].

If we extend the Abelian Higgs model to 3+1 spacetime dimensions, static gauged vortices look like static strings extending along the coordinate x^3 that parametrizes the extra dimension: the magnetic field is localized in the (x^1, x^2) plane, near the core of the vortex strings, much like as it were in the interior of a solenoid extending along x^3 . So we can localize the magnetic field along a line. This leads to the following natural **question: can we have a magnetic field localized near a point in space \mathbb{R}^3 ?** The resulting putative configuration is called a **magnetic monopole**, to contrast it with the magnetic dipoles which are physically realized and observed in real world magnets and have two poles.

We can already ask the question of the mathematical existence of magnetic monopoles in **pure electromagnetism**. The immediate answer that comes to mind is that no, magnetic monopoles are forbidden by Maxwell's equations

$$\begin{aligned} \partial_\nu F^{\mu\nu} &= J^\mu \\ \partial_\nu \tilde{F}^{\mu\nu} &= 0 \end{aligned} \quad (6.46)$$

where $\tilde{F}^{\mu\nu} := \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$ is the dual field strength which is obtained from the original field strength by the replacement $(\mathbf{E}, \mathbf{B}) \rightarrow (\mathbf{B}, -\mathbf{E})$. The vacuum Maxwell equations which are obtained by setting to zero the sources for the electric and magnetic fields in the right-hand side, are invariant under the electric-magnetic duality that sends $(\mathbf{E}, \mathbf{B}) \mapsto (\mathbf{B}, -\mathbf{E})$ or equivalently $F_{\mu\nu} \mapsto \tilde{F}_{\mu\nu}$. But the sources break this symmetry: in the first equation of (6.46) we have the electric current 4-vector J^μ , but there is no analogous magnetic current 4-vector \tilde{J}^μ in the second equation. It is precisely the absence of a magnetic current 4-vector in the Maxwell equations that allows us to write the field strength in terms of a gauge field. For static field configurations, we have

$$\mathbf{B} = \nabla \times \mathbf{A} \quad \implies \quad \nabla \cdot \mathbf{B} = 0, \quad (6.47)$$

with no magnetic charge density $\tilde{\rho}$ in the right-hand side to source the magnetic field \mathbf{B} .

The previous argument seems to suggest that if we accept Maxwell's equations as the correct mathematical description of the phenomena of electromagnetism, then pointlike electric charges are allowed, but pointlike magnetic charges are not. But Dirac [Dirac, 1931] found a loophole in this reasoning and was able to describe a magnetic monopole, which is dubbed the **Dirac monopole** since. Or almost... Dirac's argument involves a so-called Dirac string, which has a localized magnetic flux inside it, much like an infinitesimally thin solenoid. The Dirac string ends at a point, from which a radial magnetic field emanates, analogously to the electric field that emanates from an electrically charged point particle. That's the Dirac monopole. The location of the Dirac string turns out to be unphysical, as it can be moved around by performing a gauge transformation, but the endpoint of the string, which is the center of the monopole, is physical. Then by a quantum-mechanical consideration (requiring that the wave-function of a charged particle is single-valued when the particle loops around the Dirac string, which is equivalent to requiring that the Wilson line around the Dirac string is equal to 1) it follows that the magnetic charge is quantized. Note that in Dirac's point of view there is no pointlike magnetic charge really, just the endpoint of a movable Dirac string coming in from infinity. The magnetic flux through a 2-sphere that surrounds the endpoint of the Dirac string is zero, because the magnetic flux that enters the sphere from the Dirac string is equal and opposite to the flux that exits the sphere having emanated from the endpoint of the Dirac string (or the Dirac monopole).

The explanation of the Dirac monopole with the Dirac string can be confusing. Luckily, one can improve on Dirac's intuition, reinterpreting it in more geometric terms, to actually describe a genuine pointlike magnetic charge. This was achieved by Wu and Yang [Wu and Yang, 1975, Wu and Yang, 1976b, Wu and Yang, 1976a], and it's their modern description of the Dirac monopole that we will present here. The key point that will allow us to introduce a magnetic monopole is to remove from space \mathbb{R}^3 a point, the position of the monopole, which we will set to be the origin O in what follows. Then, while $\nabla \cdot \mathbf{B} = 0$ everywhere in $\mathbb{R}^3 \setminus O$, we can still have a non-vanishing magnetic flux through any 2-sphere surrounding the location of the monopole, which is measured by the **magnetic charge**

$$m = \frac{1}{2\pi} \int_{S^2} \mathbf{B} \cdot d^2\boldsymbol{\sigma}, \quad (6.48)$$

where $d^2\boldsymbol{\sigma}$ is the infinitesimal area element of the sphere, see figure 6.4.

REMARK:

We could equivalently work on \mathbb{R}^3 and use Gauss' theorem to rewrite $\nabla \cdot \mathbf{B} = 0$ on $\mathbb{R}^3 \setminus O$ together with (6.48) as **[Ex 21]**

$$\nabla \cdot \mathbf{B} = 2\pi m \delta^{(3)}(\mathbf{x}) \quad \text{in } \mathbb{R}^3, \quad (6.49)$$

but it is preferable to work in $\mathbb{R}^3 \setminus O$, which allows us to use gauge fields.

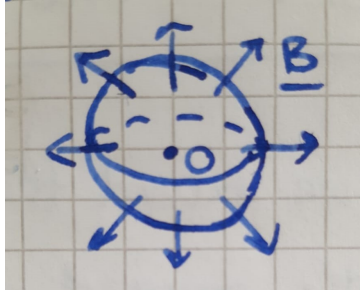


Figure 6.4: Magnetic flux produced by a magnetic monopole at the origin.

Using **polar coordinates** in \mathbb{R}^3 , we have the identities

$$\nabla \frac{1}{r} = -\frac{\mathbf{x}}{r^3}, \quad \Delta \frac{1}{r} = -4\pi \delta^{(3)}(\mathbf{x}), \quad (6.50)$$

where $r = |\mathbf{x}|$ and $\Delta \equiv \nabla^2$ is the Laplacian. Then we can solve (6.49) by

$$\mathbf{B} = \frac{m}{2} \frac{\mathbf{x}}{r^3} = \frac{m}{2} \frac{1}{r^2} \hat{\mathbf{x}}, \quad (6.51)$$

similarly to how we describe pointlike electric charges.

What about the vector potential or gauge field \mathbf{A} ? We cannot write a smooth \mathbf{A} which is defined everywhere in \mathbb{R}^3 , such that $\mathbf{B} := \nabla \times \mathbf{A}$ obeys (6.49), because then we would have $\nabla \cdot (\nabla \times \mathbf{A}) = 0$. Next, we can try to write a smooth \mathbf{A} which is defined everywhere in $\mathbb{R}^3 \setminus O$, such that $\mathbf{B} := \nabla \times \mathbf{A}$ obeys $\nabla \cdot \mathbf{B} = 0$. But this fails too. Indeed, consider for instance the vector potential \mathbf{A}^+ given by⁷

$$A_x^+ = -\frac{m}{2} \frac{y}{r(r+z)}, \quad A_y^+ = +\frac{m}{2} \frac{x}{r(r+z)}, \quad A_z^+ = 0. \quad (6.52)$$

The corresponding magnetic field is **[Ex 21]**

$$\nabla \times \mathbf{A}^+ = \frac{m}{2} \frac{\mathbf{x}}{r^3} \quad (6.53)$$

as we hoped, but unfortunately this only holds where (6.52) is defined, namely on \mathbb{R}^3 minus the origin and the negative z axis. We can try harder, but we will only be able to move the semi-infinite open path where the gauge field is ill-defined (different choices are related by singular gauge transformations).

The reason why it is not possible to find a globally defined gauge field on $\mathbb{R}^3 \setminus O \cong \mathbb{R}_{>0} \times S^2$ is that in this space there is a two-sphere surrounding the origin, and the two-sphere is a differentiable manifold which requires at least two charts (or patches) with the topology of an

⁷The subscript + is simply a label, the reason for which will become clear later.

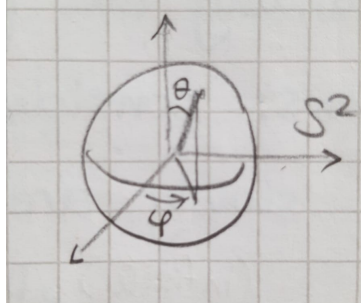


Figure 6.5: Spherical coordinates.

open disc. Working in polar coordinates (r, θ, φ) , see figure 6.5, we can take the two patches on S^2 to be⁸

$$\begin{aligned} U_+ &= \{(\theta, \varphi) \mid 0 \leq \theta < \frac{\pi}{2} + \epsilon\} \\ U_- &= \{(\theta, \varphi) \mid \frac{\pi}{2} - \epsilon < \theta \leq \pi\} \end{aligned} \quad (6.54)$$

for a constant $\epsilon \in (0, \pi)$. The two patches overlap in a region

$$U_+ \cap U_- = \{(\theta, \varphi) \mid \frac{\pi}{2} - \epsilon < \theta < \frac{\pi}{2} + \epsilon\} \quad (6.55)$$

near the equator, which has the topology of an open interval (parametrized by θ) times as circle (parametrized by φ). Then we can view \mathbf{A}^+ , defined in (6.52) in terms of Cartesian coordinates, as a gauge field defined in the northern patch U^+ . We now need to define a gauge field in the southern patch U^- , and to figure out how \mathbf{A}^+ and \mathbf{A}^- are related on the overlap $U_+ \cap U_-$. The key idea is that on the overlap $U_+ \cap U_-$ the two gauge fields are allowed to differ by a gauge transformation, since field configurations which are related by a gauge transformation are physically equivalent. On the southern patch U_- we can take the gauge field to be \mathbf{A}^- , defined by

$$A_x^- = +\frac{m}{2} \frac{y}{r(r-z)}, \quad A_y^- = -\frac{m}{2} \frac{x}{r(r-z)}, \quad A_z^- = 0, \quad (6.56)$$

which also has magnetic field

$$\nabla \times \mathbf{A}^- = \frac{m}{2} \frac{\mathbf{x}}{r^3} \quad (6.57)$$

where it is defined.

Since the gauge fields \mathbf{A}^+ and \mathbf{A}^- lead to the same gauge invariant magnetic field $\mathbf{B}^+ := \nabla \times \mathbf{A}^+ = \nabla \times \mathbf{A}^- =: \mathbf{B}^-$ in the overlap region $U_+ \cap U_-$ where they are both defined, we

⁸These spherical coordinates are ill-defined near the poles, but this won't be important for what follows. One can find a set of well-defined coordinates in the two patches, for example the stereographic coordinates that you encountered in the first term. What matters is that there is no single set of coordinates which cover the whole S^2 .

might expect them to be gauge equivalent. To see this explicitly, it is easier to switch to polar coordinates. Using differential form notation we find⁹ [Ex 22]

$$\begin{aligned} A^\pm &= A_x^\pm dx + A_y^\pm dy + A_z^\pm dz = A_r^\pm dr + A_\theta^\pm d\theta + A_\varphi^\pm d\varphi \\ &= \frac{m}{2}(\pm 1 - \cos\theta) d\varphi. \end{aligned} \quad (6.58)$$

Then we find that on the overlap of the two patches $U_+ \cap U_-$ the two gauge fields differ by

$$A^+ - A^- = m d\varphi = d(m\varphi) \equiv d\alpha_{+-} \equiv -ig_{+-}^{-1} dg_{+-} \quad (6.59)$$

where the **transition function**, namely the parameter of the $U(1)$ gauge transformation that relates the gauge fields in the two patches, is

$$g_{+-}(\varphi) = e^{i\alpha_{+-}(\varphi)} = e^{im\varphi} \in U(1). \quad (6.60)$$

Since $\varphi \sim \varphi + 2\pi$, $g_{+-}(\varphi)$ is **single-valued** (or periodic) if we do one lap around the φ circle (e.g., the equator) if and only if **the magnetic charge m is an integer**:

$$g_{+-}(\varphi + 2\pi) = g_{+-}(\varphi) \iff m \in \mathbb{Z}. \quad (6.61)$$

We learn that the quantization of the magnetic charge follows from carefully considering gauge fields defined locally on the two patches of S^2 , and gluing them consistently by $U(1)$ gauge transformations in the overlap of the two patches. The $U(1)$ -valued transition function g_{+-} on the overlap tells us how to relate gauge transformation parameters g_\pm on U_+ and U_- along the overlap: $g_+ = g_{+-}g_-$, or equivalently $\alpha_+ = \alpha_- + \alpha_{+-}$. Mathematically, the $U(1)$ gauge transformation parameters define sections of a so called principal $U(1)$ bundle over S^2 ; the gauge fields A^\pm are (local) connections for this principal $U(1)$ bundles. If you want to learn about the definition of these bundles, their sections and connections, and how they provide a mathematical definition of gauge groups and gauge fields, see the bonus chapter 9.

REMARK:

In this formulation we can calculate the magnetic flux through the 2-sphere surrounding the origin (the position of the magnetic monopole) as follows. Call U_N and U_S the northern and southern hemisphere respectively, which are the limits as $\epsilon \rightarrow 0$ of U^\pm , so that the overlap reduces to the equator S_{eq}^1 . Then the contributions of the two hemispheres to the magnetic

⁹The vanishing of A^+ at the north pole and of A^- at the south pole is what ensures that they are well defined there, even if the polar coordinates are ill defined. This can be checked explicitly by switching to stereographic coordinates or to Cartesian coordinates.



Figure 6.6: Oriented hemispheres and their oriented boundaries.

flux add up:

$$\begin{aligned}
 \frac{1}{2\pi}\Phi_{S^2}(\mathbf{B}) &= \frac{1}{2\pi} \int_{S^2} \mathbf{B} \cdot d^2\boldsymbol{\sigma} = \frac{1}{2\pi} \int_{U_N} \mathbf{B}^+ \cdot d^2\boldsymbol{\sigma} + \frac{1}{2\pi} \int_{U_S} \mathbf{B}^- \cdot d^2\boldsymbol{\sigma} \\
 &= \frac{1}{2\pi} \int_{U_N} (\nabla \times \mathbf{A}^+) \cdot d^2\boldsymbol{\sigma} + \frac{1}{2\pi} \int_{U_S} (\nabla \times \mathbf{A}^-) \cdot d^2\boldsymbol{\sigma} \\
 &= \frac{1}{2\pi} \oint_{S_{\text{eq}}^1} \mathbf{A}^+ \cdot d\mathbf{l} - \frac{1}{2\pi} \oint_{S_{\text{eq}}^1} \mathbf{A}^- \cdot d\mathbf{l} \\
 &= \frac{1}{2\pi} \oint_{S_{\text{eq}}^1} (\mathbf{A}^+ - \mathbf{A}^-) \cdot d\mathbf{l} = \frac{1}{2\pi} \oint_{S_{\text{eq}}^1} (A^+ - A^-) \\
 &= \frac{1}{2\pi} \oint_{S_{\text{eq}}^1} d\alpha_{+-} = \frac{m}{2\pi} \int_0^{2\pi} d\varphi = m.
 \end{aligned} \tag{6.62}$$

To go from the second to the third line we used Stokes' theorem. The relative minus sign between the two terms is there because the two hemispheres have opposite orientations, so that $\partial U_N = S_{\text{eq}}^1$ but $\partial U_S = -S_{\text{eq}}^1$ (the equatorial circle with the opposite orientation), see figure (6.6). This reproduces the desired result (6.48).

This is very nice! We can describe a **static solution of Maxwell's equations** which is a pointlike magnetic charge (**magnetic monopole**) by excising the location of the monopole from space and exploiting the geometry and topology of gauge fields over $\mathbb{R}^3 \setminus O$ (or equivalently of S^2). But unfortunately it is not hard to see that a **Dirac monopole has infinite energy [Ex 22]**. As we will see, this problem can be fixed if we embed the $U(1)$ gauge group into a bigger nonabelian gauge group, such as $SU(2)$.

Chapter 7

Non-abelian gauge theories

In this chapter we will learn how to formulate gauge theories with a non-abelian (that is, non-commutative) gauge group. Non-abelian gauge theories are named **Yang-Mills theories**, after Chen-Ning Yang and Robert Mills, who developed the formalism in 1954 [Yang and Mills, 1954].

The formalism of Yang and Mills became prominent in the late 1960s, and has remained central in modern physics ever since. Non-abelian gauge theories are the **language of the Standard Model** of Particle Physics, and have also established very fruitful **interactions between Physics and Maths**, which have led to numerous developments in both subjects and quite a few Nobel prizes and Fields medals.

We will spend the rest of the term studying the geometry (and some topology) underlying non-abelian gauge field configurations. But let's start by introducing the main characters.

7.1 Compact Lie algebras

This section is mostly a review of material from the previous term, but I will introduce new conventions following the Physics literature. I will also introduce some new terminology and definitions along the way. There will be a number of exercise that I recommend attempting to make sure that you understand the concepts. An excellent reference is section 1.8.1 of Argyres' supersymmetry lecture notes [Argyres, 2001], of which this section is a shameless rip-off.

We start by recalling that a **Lie algebra** \mathfrak{g} is a vector space¹ endowed with an additional

¹We will only consider finite-dimensional vector spaces. We refer to the dimension of this vector space as the dimension of the Lie algebra, which we denote as $\dim \mathfrak{g}$.

structure, the **Lie bracket**

$$[,] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g} ,$$

which is antisymmetric and bilinear. The vector space has a **basis** $\{t_a\}_{a=1}^{\dim \mathfrak{g}}$ of so called **generators** t_a . In this basis the Lie bracket reads

$$[t_a, t_b] = i f_{ab}^c t_c \quad (a, b, c = 1, \dots, \dim \mathfrak{g}) \quad (7.1)$$

where f_{ab}^c are real **structure constants**, which express the component of the Lie bracket $[t_a, t_b]$ along the generator t_c . Repeated indices are summed over. Note the i in the right-hand side in my conventions. I'll return to why that might be useful shortly. In the rest of the course we will content ourselves with matrix Lie algebras, in which case the Lie bracket is simply the commutator of two matrices, that you are familiar with from Linear Algebra. You can always keep that in mind whenever I use the term Lie bracket. But the abstract definition of Lie algebras and the Lie bracket (7.1) is more general.²

The associativity of the Lie bracket is expressed by the **Jacobi identity**

$$[[t_a, t_b], t_c] + [[t_b, t_c], t_a] + [[t_c, t_a], t_b] = 0 \quad (7.2)$$

which expressed in the basis of generators is the identity

$$f_{ab}^d f_{dc}^e + f_{bc}^d f_{da}^e + f_{ca}^d f_{db}^e = 0 \quad (7.3)$$

for the structure constants.

An r -dimensional **representation** (rep) of \mathfrak{g} is a realization of the generators $\{t_a\}$ as a set of $r \times r$ **matrices** satisfying (7.1), where now $[,]$ is interpreted as the **commutator** of matrices: $[A, B] = AB - BA$. We will often denote an r -dimensional representation as \mathbf{r} and its generators as $t_a^{(\mathbf{r})}$. If there are multiple representations with the same dimension we will distinguish them by primes or other notation. We might omit the subscript $t_a^{(\mathbf{r})}$ when it is clear from the context which representation we are discussing.

A **compact Lie algebra** is one which can be represented by finite-dimensional **hermitian** matrices:³

$$t_a^\dagger = t_a . \quad (7.4)$$

REMARK:

The imaginary unit $i = \sqrt{-1}$ in the right-hand side of the Lie bracket (7.1), which is common in

²In Killing and Cartan's classification of simple Lie algebras, written in its definitive form by Dynkin, the matrix Lie algebras form four infinite series, the so called classical Lie algebras $A_n = su(n+1)$, $B_n = so(2n+1)$, $C_n = usp(2n)$, $D_n = so(2n)$. But there are a few more exceptional Lie algebras which are not of matrix type: E_6, E_7, E_8, F_4, G_2 . See [Cahn, 2014] for a down-to-earth introduction to the subject, and [Harris et al., 1991] for a more advanced perspective.

³This is the result of a theorem, if one starts from standard definition of compactness, but we will take it to be our definition of compact Lie algebra.

the physics literature, is there to ensure the hermiticity of generators of compact Lie algebras (7.4). This convention is convenient to manifest the reality (and positive definiteness) of energy functionals or other physical quantities. The maths literature (and Andreas) typically uses $\tilde{t}_a = it_a$ as generators, which are anti-hermitian for compact Lie algebras: $\tilde{t}_a^\dagger = -\tilde{t}_a$.

It is a theorem that any compact Lie algebra can be decomposed into the direct sum of $u(1)$ Lie algebras and of simple Lie algebras:

$$\mathfrak{g} = (\oplus_{i=1}^h u(1)) \oplus (\oplus_{i=1}^l \mathfrak{g}_i) = u(1) \oplus \cdots \oplus u(1) \oplus \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_l . \quad (7.5)$$

Let's recall $u(1)$ and simple Lie algebras in turn:

1. The $u(1)$ Lie algebra is the compact Lie algebra with a single generator t . By the anti-symmetry of the Lie bracket, we have

$$[t, t] = 0 , \quad (7.6)$$

so the algebra is **abelian**. Its **irreducible representations** (irreps) are 1-dimensional

$$t = q\mathbb{1} \quad (q \in \mathbb{Z}) \quad (7.7)$$

where the integer q is called the **charge** of the representation and $\mathbb{1}$ is the identity operator.

2. a **simple Lie algebra**⁴ is characterised by structure constants such that

$$f_{ab}^c \neq 0 \quad \forall a \quad (7.8)$$

(or equivalently for all b , or for all c , it turns out).

One can show that there is a basis of generators, which we will adopt from now on, such that

$$\mathrm{tr}_{\mathbf{r}}(t_a^{(\mathbf{r})} t_b^{(\mathbf{r})}) = C(\mathbf{r}) \delta_{ab} , \quad (7.9)$$

The real number $C(\mathbf{r})$, which is positive for representations of compact Lie algebras, is called the **quadratic invariant** of the representation \mathbf{r} .

REMARKS:

⁴Here are some definitions. A **Lie subalgebra** \mathfrak{h} of a Lie algebra \mathfrak{g} is a subspace $\mathfrak{h} \subseteq \mathfrak{g}$ that is closed under the Lie bracket: $[\mathfrak{h}, \mathfrak{h}] \subseteq \mathfrak{h}$, which is a shorthand for $[X, Y] \in \mathfrak{h}$ for all $X, Y \in \mathfrak{h}$. An **ideal** $\mathfrak{i} \subseteq \mathfrak{g}$ is a subalgebra that obeys the stronger condition $[\mathfrak{i}, \mathfrak{g}] \subseteq \mathfrak{i}$, namely $[X, Y] \in \mathfrak{i}$ for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{i}$. Examples of ideals of \mathfrak{g} which always exist are the subalgebra consisting of the zero element, which is called the zero ideal, and the Lie algebra \mathfrak{g} itself. An ideal is called proper if it is not the full algebra \mathfrak{g} . A Lie algebra is called **simple** if it is non-abelian and contains no nonzero proper ideals. A Lie algebra is called **semisimple** if it is isomorphic to a direct sum of simple Lie algebras.

1. There is a **normalization** ambiguity: rescaling

$$(t_a, f_{ab}{}^c, C(\mathbf{r})) \longrightarrow (\lambda t_a, \lambda f_{ab}{}^c, \lambda^2 C(\mathbf{r})) \quad (7.10)$$

by a constant $\lambda \neq 0$ leaves all the previous equations invariant.

2. For the **adjoint representation** $\mathbf{r} = \text{adj}$, which we will review later, equation (7.9) defines the **Killing form**⁵

$$\begin{aligned} K: \quad \mathfrak{g} \times \mathfrak{g} &\rightarrow \mathbb{R} \\ (v, w) &\mapsto K(v, w) := \text{tr}(\text{ad}_v \circ \text{ad}_w) \end{aligned} \quad (7.11)$$

where $v = v^a t_a$, $w = w^a t_a$, and ad_x denotes the adjoint action of x on the Lie algebra. The Killing form K is bilinear and symmetric.

3. We can use δ_{ab} / its inverse δ^{cd} to **lower/raise** Lie algebra **indices**.⁶

*** EXERCISE:**

[Ex 23.1] Show that f_{abc} is totally antisymmetric in its indices:

$$f_{abc} = -f_{bac} = -f_{cba} . \quad (7.12)$$

[Ex 23.2] Show that for any representation \mathbf{r} of a simple Lie algebra

$$\text{tr}_{\mathbf{r}}(t_a^{(\mathbf{r})}) = 0 , \quad (7.13)$$

where $\text{tr}_{\mathbf{r}}$ is the trace in the representation \mathbf{r} , which is nothing but the trace of $r \times r$ matrices.

We can obtain a **Lie group** G from a **Lie algebra** \mathfrak{g} by applying the **exponential map**⁷

$$\begin{aligned} \text{exp:} \quad \mathfrak{g} &\rightarrow G \\ \alpha = \alpha^a t_a &\mapsto g = e^{i\alpha^a t_a} \end{aligned} \quad (7.14)$$

⁵Named after Wilhelm Killing. No humans or animals were harmed in the production of this lecture course.

⁶More precisely, one should use the Killing form and its inverse to lower and raise indices. Since we can pick a basis in which the Killing form is proportional to the identity and since the normalization is arbitrary, we do not lose much by using δ_{ab} instead of the Killing form. It may sound silly to distinguish upper and lower indices if we are raising and lowering them using the identity matrix. I am mentioning this distinction because in a different basis of the Lie algebra the Killing form might be a less trivial non-degenerate symmetric matrix K_{ab} . In such a basis you would use this matrix K_{ab} to lower Lie algebra indices, and its inverse K^{ab} to raise indices. (This is completely analogous to raising/lowering spacetime indices with the Minkowski metric $\eta_{\mu\nu}$ and its inverse $\eta^{\mu\nu}$.)

⁷More precisely, the exponential map of the Lie algebra produces a subgroup whose elements are continuously connected to the identity (*aka* the connected component of the identity).

where $\alpha^a \in \mathbb{R}$.

Substituting the abstract generators t_a of the Lie algebra by their realizations $t_a^{(\mathbf{r})}$ in a representation \mathbf{r} , we obtain **representations of the Lie group** G . In an r -dimensional rep \mathbf{r} , the group element g is realized as an $r \times r$ **unitary** matrix:

$$g : \phi \mapsto (g\phi) := r(g) \cdot \phi = e^{i\alpha^a t_a^{(\mathbf{r})}} \cdot \phi, \quad (7.15)$$

where ϕ and its transformed $(g\phi)$ are r -vectors, while $r(g) = \exp \left[i\alpha^a t_a^{(\mathbf{r})} \right]$ is an $r \times r$ matrix. In components,

$$g : \phi^j \mapsto (g\phi)^j = r(g)^j_k \phi^k = (e^{i\alpha^a t_a^{(\mathbf{r})}})^j_k \phi^k, \quad (7.16)$$

We will refer to the action (7.15)-(7.16) of a group element on ϕ as the **finite transformation** of ϕ .

Conversely, we recover the action of the Lie algebra from the action of the Lie group if the group parameters α^a are infinitesimal. Then to linear order $\phi \mapsto \phi + \delta_\alpha \phi$, with

$$\delta_\alpha \phi^j = i\alpha^a (t_a^{(\mathbf{r})})^j_k \phi^k, \quad (7.17)$$

which we will refer to as the **infinitesimal transformation** of ϕ .

Next we define the **complex conjugate** representation $\bar{\mathbf{r}}$ of a representation \mathbf{r} as the representation with complex conjugate representation matrix:

$$\bar{r}(g) := r(g)^* \equiv \overline{r(g)}, \quad (7.18)$$

where we use star or bar interchangeably to denote complex conjugation.

[Ex 24] * EXERCISE:

1. Show explicitly that if the r -vector ϕ transforms in irrep \mathbf{r} , then its complex conjugate $\phi^* \equiv \bar{\phi}$, which is also an r -vector, transforms in irrep $\bar{\mathbf{r}}$.

2. Show that, as $r \times r$ matrices,

$$t_a^{(\bar{\mathbf{r}})} = -(t_a^{(\mathbf{r})})^T, \quad (7.19)$$

where the subscript T denotes the transposition of a matrix.

3. Denote $\bar{\phi}_j := (\phi^j)^*$ and construct the **row** r -vector $\phi^\dagger = \bar{\phi}^T = (\bar{\phi}_1, \dots, \bar{\phi}_r)$. Show that the inner product $\phi^\dagger \phi$ is invariant under the action of G .

Next we introduce the **adjoint representation** adj , which is the $(\dim \mathfrak{g})$ - dimensional irrep defined by

$$(t_a^{(\text{adj})})^b_c = i f_{ac}^b \quad (b, c = 1, \dots, \dim \mathfrak{g}) . \quad (7.20)$$

[Ex 25] * EXERCISE:

1. Check that (7.20) defines a representation of \mathfrak{g} .
2. Recall that the **adjoint action** of the Lie algebra \mathfrak{g} on itself is given by

$$\begin{aligned} \text{ad: } \mathfrak{g} &\rightarrow \mathfrak{g} \\ y &\mapsto \text{ad}_x(y) := [x, y] \end{aligned} \quad (7.21)$$

for all Lie algebra elements $x \in \mathfrak{g}$. Show that

$$\text{ad}_{t_a}(y^b t_b) = (t_a^{(\text{adj})})^b_c y^c t_b . \quad (7.22)$$

[Remark: this has a nice interpretation: the adjoint representation (7.20) is nothing but the adjoint action of the Lie algebra on itself, expressed in a basis.

3. Show that the quadratic invariant of the adjoint representation is

$$C(\text{adj}) = \frac{f_{abc} f^{abc}}{\dim \mathfrak{g}} ,$$

where Lie algebra indices are raised (/lowered) using δ^{ab} (/ δ_{ab}).

Finally we introduce the notion of fundamental representation. Consider a **matrix group** G , that is a group whose elements are square matrices and where the group composition law is matrix multiplication. Let n be the size of the matrices. The **fundamental representation** (or defining representation) of a matrix group G is the representation in which G acts by matrix multiplication:

$$r(g)^i_j = g^i_j \quad (7.23)$$

where $g \in G$ is a matrix. We denote the fundamental representation by fund or by \mathfrak{n} according to its dimension.

[Ex 26] * EXERCISE:

Let $\text{Mat}_n(F)$ denote $n \times n$ matrices whose entries are in the field F , and $\mathbb{1}_n$ the $n \times n$ identity matrix. The **'classical' compact simple Lie groups** are

$$\begin{aligned} SU(N) &= \{g \in \text{Mat}_N(\mathbb{C}) \mid g^\dagger g = \mathbb{1}_N, \det g = 1\} \\ SO(N) &= \{g \in \text{Mat}_N(\mathbb{R}) \mid g^T g = \mathbb{1}_N, \det g = 1\} \\ USp(2N) &= \{g \in \text{Mat}_{2N}(\mathbb{C}) \mid g^\dagger g = \mathbb{1}_{2N}, g^T J g = J\} \end{aligned}$$

where the $(2N) \times (2N)$ antisymmetric matrix

$$J = \begin{pmatrix} 0_N & \mathbb{1}_N \\ -\mathbb{1}_N & 0_N \end{pmatrix}$$

is called the symplectic form.

1. Characterize the Lie algebras $su(N)$, $so(N)$, and $usp(2N)$ as vector spaces of matrices subject to certain linear conditions, which you should find.

[**Hint:** You can assume that a group element takes the exponential form $g = \exp(i\alpha^a t_a)$ and Taylor expand for infinitesimal α .]

2. Find the generators of the fundamental representation fund and its complex conjugate rep $\overline{\text{fund}}$ (the so called antifundamental representation) for $G = SU(N)$, $SO(N)$, $USp(2N)$.
3. For $G = SO(N)$, $USp(2N)$, show that fund and $\overline{\text{fund}}$ are **equivalent** representations, namely

$$t_a^{(\overline{\text{fund}})} = V t_a^{(\text{fund})} V^{-1} \quad \forall a$$

for some invertible matrix V .⁸

7.2 Non-abelian gauge theories: fields

This section introduces the cast of characters which we will use in the next section to formulate actions which are invariant under non-abelian gauge transformations. The cast of characters will consist of:

- **Charged fields** (scalars or spinors), collectively denoted as ϕ , transforming in a representation \mathbf{r} ⁹ of the gauge group G ;¹⁰
- Their **covariant derivatives** $D_\mu \phi$;
- The **gauge field** A_μ , which is hidden inside the covariant derivative;
- The **field strength** $F_{\mu\nu}$ of the gauge field,

⁸For compact Lie groups and algebras, V is unitary: $V^\dagger = V^{-1}$.

⁹If the representation \mathbf{r} is irreducible we think of ϕ as a single field; if representation \mathbf{r} is reducible, namely it is the direct sum of multiple irreps of G , then we think of ϕ as describing multiple charged fields.

¹⁰**Note:** from now on I will ignore the distinction between the Lie group G and the gauge group \mathcal{G} , which consists of coordinate-dependent elements of G . I will simply use G for the gauge group.

and their gauge transformations.

References for this section are section 1.8.1 of [Argyres, 2001] and section 2.1 of [Tong, 2018].

We will be more general later, but let us start slowly and assume that the gauge group G is a **classical** group (e.g. $SU(N)$), whose elements are matrices, and that the charged field ϕ transforms in the fundamental representation fund (that is \mathbf{N} for $SU(N)$). This means that the gauge transformation of the charged field ϕ is

$$\phi \mapsto g\phi = e^{i\alpha^a t_a} \phi \quad (7.24)$$

where ϕ is a column vector (N -dimensional for $G = SU(N)$, that is $\phi = (\phi^j)_{j=1}^N \in \mathbb{C}^N$), the Lie algebra generators t_a are matrices ($N \times N$ hermitian traceless for $G = SU(N)$), and the group element g is also a matrix ($N \times N$ unitary and with unit determinant for $G = SU(N)$), which acts on ϕ by matrix multiplication. Recall that both the field $\phi = \phi(x)$ and the group element $g = g(x)$, and therefore the gauge parameter $\alpha = \alpha(x)$, depend on the space-time point x .

Given the charged field ϕ , we define its (gauge) **covariant derivative**

$$D_\mu \phi := \partial_\mu \phi - iA_\mu \phi \quad (7.25)$$

where the **gauge field** A_μ is now a **matrix**, which will turn out to be an element of the Lie algebra to ensure the consistency of its gauge transformation:

$$A_\mu = A_\mu^a t_a . \quad (7.26)$$

We require that under the non-abelian gauge transformation (7.24) the covariant derivative transforms in the same way as ϕ :

$$D_\mu \phi \mapsto g D_\mu \phi . \quad (7.27)$$

Viewing the covariant derivative¹¹

$$D_\mu := \mathbb{1} \partial_\mu - iA_\mu \quad (7.28)$$

as a **matrix-valued** differential operator, which in components reads

$$(D_\mu)^j_k = \delta^j_k \partial_\mu - i(A_\mu)^j_k ,$$

¹¹Here $\mathbb{1}$ is the identity matrix of the same size as A_μ , e.g. the $N \times N$ identity matrix for $G = SU(N)$. It is customary to omit the identity matrix from the notation and simply write, and I'll follow that convention and only restore $\mathbb{1}$ when it helps to understand what is going on. If you are formally minded and want to be very precise, you might write the covariant derivative as

$$D_\mu = \partial_\mu \otimes \mathbb{1} - iA_\mu^a(x) \otimes t_a ,$$

which acts on the tensor product $C^\infty(U) \otimes V$ of the vector space $C^\infty(U)$ of smooth functions defined on a patch U of space-time and of the finite-dimensional vector space V associated to the fundamental representation. We won't need to worry about such level of abstraction and formality.

we require the gauge transformation

$$D_\mu \mapsto gD_\mu g^{-1}. \quad (7.29)$$

In terms of the gauge field, the gauge transformation of the covariant derivative is

$$\begin{aligned} \partial_\mu - iA_\mu &\mapsto \partial_\mu - iA'_\mu = g(\partial_\mu - iA_\mu)g^{-1} \\ &= g(\partial_\mu g^{-1}) + gg^{-1}\partial_\mu - igA_\mu g^{-1}. \end{aligned} \quad (7.30)$$

Note that the gauge group element g and the gauge field A_μ are matrices now, so they do not commute: their **order matters!**

Comparing the initial and final result, we obtain the following gauge transformation for the gauge field A_μ :

$$\begin{aligned} A_\mu &\mapsto A'_\mu = gA_\mu g^{-1} + ig(\partial_\mu g^{-1}) \\ &= gA_\mu g^{-1} - i(\partial_\mu g)g^{-1}, \end{aligned} \quad (7.31)$$

where I have used parenthesis to make it clear that all objects are (matrix-valued) functions,¹² not differential operators. I have used the identity

$$0 = (\partial_\mu \mathbb{1}) = (\partial_\mu (gg^{-1})) = (\partial_\mu g)g^{-1} + g(\partial_\mu g^{-1}) \quad (7.32)$$

to go from the first line to the second line.

REMARKS:

1. The first term in the gauge transformation (7.31) of the gauge field A_μ is the **adjoint action** of the Lie group G on a Lie algebra element. This clarifies why A_μ belongs to the Lie algebra $\mathfrak{g} = \text{Lie}(G)$.
2. The second term in (7.31) is a correction term to the adjoint action, which involves a derivative. This is also an element of the Lie algebra \mathfrak{g} , since $(\partial_\mu g)g^{-1} = \partial_\mu g|_{g=\mathbb{1}}$, which is the very definition of an element of the Lie algebra of G as a tangent vector to the identity element of the group.

Finally, in analogy with the $G = U(1)$ case, we define the **field strength**

$$F_{\mu\nu} := i[D_\mu, D_\nu]. \quad (7.33)$$

As in the $U(1)$ case, in the above definition we view both sides as differential operators, except that now they are matrix-valued. As we will see shortly, despite appearance $F_{\mu\nu}$ turns out to

¹²The derivative of a (matrix-valued) function is a (matrix-valued) function.

be a **multiplicative** operator, which means that it is a (matrix-valued) function that simply acts by (matrix) multiplication, no differentiations are involved.

By construction, under a gauge transformation (7.24) the field strength transforms as

$$F_{\mu\nu} \mapsto gF_{\mu\nu}g^{-1}. \quad (7.34)$$

Proof. We simply need to use the gauge transformation property (7.29) and basic properties of the commutator:

$$\begin{aligned} F_{\mu\nu} = i[D_\mu, D_\nu] &\mapsto F'_{\mu\nu} = i[gD_\mu g^{-1}, gD_\nu g^{-1}] \\ &= g[D_\mu, D_\nu]g^{-1} = gF_{\mu\nu}g^{-1}. \end{aligned}$$

Calculating the commutator in (7.33), we find the following expression for the field strength:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]. \quad (7.35)$$

Proof. Restoring the identity matrix $\mathbb{1}$ for clarity (feel free to omit it if you are comfortable without it),

$$\begin{aligned} -iF_{\mu\nu} &= [D_\mu, D_\nu] = [\mathbb{1}\partial_\mu - iA_\mu, \mathbb{1}\partial_\nu - iA_\nu] \\ &= [\mathbb{1}\partial_\mu, \mathbb{1}\partial_\nu] - i[\mathbb{1}\partial_\mu, A_\nu] - i[A_\mu, \mathbb{1}\partial_\nu] - [A_\mu, A_\nu] \\ &= 0 - i(\partial_\mu A_\nu) + i(\partial_\nu A_\mu) - [A_\mu, A_\nu] \\ &= -i(\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]). \end{aligned}$$

REMARK:

The finite gauge transformations (7.29) of the covariant derivative D_μ and (7.34) of the field strength $F_{\mu\nu}$ is by the adjoint action of the Lie group on the Lie algebra. This means that D_μ and $F_{\mu\nu}$ transform in the **adjoint representation** adj of G .

[Ex 28] * EXERCISE:

By considering infinitesimal gauge transformations ($|\alpha^a| \ll 1$)

$$g = e^{i\alpha^a t_a} \equiv e^{i\alpha} = 1 + i\alpha + O(\alpha^2) \quad (7.36)$$

and Taylor expanding finite gauge transformations to leading order in $\alpha \in \mathfrak{g} = \text{Lie}(G)$, show that the **infinitesimal gauge variations** of the fields are

$$\begin{aligned} \delta_\alpha \phi &= i\alpha\phi \\ \delta_\alpha A_\mu &= i[\alpha, A_\mu] + \partial_\mu \alpha \\ \delta_\alpha F_{\mu\nu} &= i[\alpha, F_{\mu\nu}], \end{aligned} \quad (7.37)$$

where $\phi \mapsto \phi + \delta_\alpha \phi + O(\alpha^2)$ and so on.

REMARKS:

1. The field strength $F_{\mu\nu}$ transforms in the adj rep of \mathfrak{g} under infinitesimal gauge transformations.
2. The gauge field A_μ doesn't quite transform in adj, as the first term in its variation suggests, because of the additional derivative term, which we have already encountered when we studied $\mathfrak{g} = u(1)$. People often say (and I might also say in the future) that A_μ transforms in the adjoint representation adj, but that's an abuse of terminology.
3. On the other hand the covariant derivative D_μ does transform in the adj representation.

Everything that we have seen so far generalizes to an arbitrary Lie group G and a charged field ϕ transforming in an r -dimensional representation \mathfrak{r} . Now ϕ is a column vector with r components, and we simply need to replace the group element g in previous formulae by the appropriate $r \times r$ representation matrix

$$r(g) = \exp[i\alpha^a t_a^{(\mathfrak{r})}]. \quad (7.38)$$

For instance

$$D_\mu \phi = \partial_\mu \phi - iA_\mu \phi := (\mathbb{1}_r \partial_\mu - iA_\mu^a t_a^{(\mathfrak{r})}) \phi, \quad (7.39)$$

and

$$\begin{aligned} F_{\mu\nu} \phi &= i[D_\mu, D_\nu] \\ &= (\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]) \phi \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f_{bc}^a A_\mu^b A_\nu^c) t_a^{(\mathfrak{r})} \phi, \end{aligned} \quad (7.40)$$

where it is understood that if ϕ transforms in the representation \mathfrak{r} , then

$$\begin{aligned} A_\mu \phi &:= A_\mu^a t_a^{(\mathfrak{r})} \phi \\ F_{\mu\nu} \phi &:= F_{\mu\nu}^a t_a^{(\mathfrak{r})} \phi \end{aligned} \quad (7.41)$$

etc. Similarly, I should warn you that it is customary to simply write $g\phi$, to mean the abstract action of g on ϕ in the appropriate representation, rather than the explicit multiplication $r(g)\phi$ by the representation matrix $r(g)$. Of course one needs to specify the representation \mathfrak{r} beforehand, or it wouldn't be clear what $g\phi$ means.

In components,

$$(A_\mu \phi)^i = A_\mu^a (t_a^{(\mathfrak{r})})^i_j \phi^j \quad (i, j = 1, \dots, r) \quad (7.42)$$

etc.

[Ex 29] * EXERCISE:

Show that, if $G = U(1)$, all the equations written so far in this section reduce to those introduced in chapter 5, both for the charge 1 representation, which is analogous to the fundamental representation, and for the more general charge q representation.

[Ex 30] * EXERCISE:

Consider a field ϕ in the adj representation, with components ϕ^a , where $a = 1, \dots, \dim \mathfrak{g}$.

1. Show that

$$(A_\mu \phi)^a = i f_{bc}^a A_\mu^b \phi^c \quad (7.43)$$

and similarly for $(F_{\mu\nu} \phi)^a$.

2. Let $\Phi := \phi^a t_a$, and $A_\mu = A_\mu^a t_a$, $F_{\mu\nu} = F_{\mu\nu}^a t_a$ as usual. Show that

$$(A_\mu \phi)^a t_a = [A_\mu, \Phi] \quad (7.44)$$

and similarly for $F_{\mu\nu} \phi$. Show that therefore

$$\begin{aligned} D_\mu \Phi &= \partial_\mu \Phi - i[A_\mu, \Phi] \\ [D_\mu, D_\nu] \Phi &= -i[F_{\mu\nu}, \Phi]. \end{aligned} \quad (7.45)$$

The lesson here is that the action of the adjoint representation on itself is by commutators (or Lie brackets). This is simply a consequence of what we have seen in Ex 25.

7.3 Non-abelian gauge theories: action and EoM

Let us start by constructing a **gauge invariant action** for the (Lie algebra valued) **non-abelian gauge field** $A_\mu = A_\mu^a t_a$. This is easy: since the field strength $F_{\mu\nu} = F_{\mu\nu}^a t_a$ transforms as

$$F_{\mu\nu} \mapsto g F_{\mu\nu} g^{-1} \quad (7.46)$$

under a gauge transformation, it follows immediately that $\text{tr}(F_{\mu\nu} F^{\mu\nu})$ is gauge invariant and can therefore be used as a term in the Lagrangian density.

Proof. Under a gauge transformation,

$$\begin{aligned} \text{tr}(F_{\mu\nu} F^{\mu\nu}) &= \text{tr}(g F_{\mu\nu} g^{-1} g F^{\mu\nu} g^{-1}) \\ &= \text{tr}(g^{-1} g F_{\mu\nu} g^{-1} g F^{\mu\nu}) = \text{tr}(F_{\mu\nu} F^{\mu\nu}). \end{aligned}$$

where we have used the cyclic property of the trace.

REMARKS:

1. In the previous equation the trace is over the vector space on which the group element g acts naturally by matrix multiplication. Therefore $\text{tr} \equiv \text{tr}_{\text{fund}}$, the trace over the fundamental representation. This is what we will mean by the trace tr unless we specify a representation.

2. We could equally well use any other irreducible representation \mathbf{r} , in which case $F_{\mu\nu}^{(\mathbf{r})} = F_{\mu\nu}^a t_a^{(\mathbf{r})}$, where we have specified the representation explicitly in the superscript. Under a gauge transformation

$$F_{\mu\nu}^{(\mathbf{r})} \mapsto r(g) F_{\mu\nu}^{(\mathbf{r})} r(g)^{-1}, \quad (7.47)$$

therefore $\text{tr}_{\mathbf{r}}(F_{\mu\nu} F^{\mu\nu}) := \text{tr}_{\mathbf{r}}(F_{\mu\nu}^{(\mathbf{r})} F^{(\mathbf{r})\mu\nu})$ (we will use both notations interchangeably) is also gauge invariant, by the same logic as above.

3. Any two such choices are proportional to one another:

$$\begin{aligned} \text{tr}_{\mathbf{r}}(F_{\mu\nu}^{(\mathbf{r})} F^{(\mathbf{r})\mu\nu}) &= F_{\mu\nu}^a F^{b\ \mu\nu} \text{tr}_{\mathbf{r}}(t_a^{(\mathbf{r})} t_b^{(\mathbf{r})}) \\ &= F_{\mu\nu}^a F^{b\ \mu\nu} C(\mathbf{r}) \delta_{ab} \\ &= C(\mathbf{r}) F_{\mu\nu}^a F^{a\ \mu\nu}, \end{aligned} \quad (7.48)$$

where we have used (7.9). Then

$$\frac{1}{C(\mathbf{r})} \text{tr}_{\mathbf{r}}(F_{\mu\nu}^{(\mathbf{r})} F^{(\mathbf{r})\mu\nu}) = F_{\mu\nu}^a F^{a\ \mu\nu} \quad (7.49)$$

is independent of the choice of representation \mathbf{r} .

We are now ready to define the **Yang-Mills action**

$$\begin{aligned} S_{YM}[A] &= \int d^4x \mathcal{L}_{YM}, \\ \mathcal{L}_{YM} &= -\frac{1}{2g_{YM}^2} \text{tr}(F_{\mu\nu} F^{\mu\nu}), \end{aligned} \quad (7.50)$$

where again $\text{tr} \equiv \text{tr}_{\text{fund}}$ and we work in a **normalization** where

$$C(\text{fund}) = \frac{1}{2}, \quad (7.51)$$

so, being explicit with the representation used,

$$\begin{aligned} \mathcal{L}_{YM} &= -\frac{1}{4g_{YM}^2} \frac{1}{C(\mathbf{r})} \text{tr}_{\text{fund}}(F_{\mu\nu}^{(\text{fund})} F^{(\text{fund})\mu\nu}) \\ &= -\frac{1}{4g_{YM}^2} F_{\mu\nu}^a F^{a\ \mu\nu}. \end{aligned} \quad (7.52)$$

g_{YM} is called the **Yang-Mills coupling constant**¹³ and controls the strength of the interactions. (To see that, it helps to rescale $A_\mu \rightarrow g_{YM} A_\mu$.)

[Ex 32] * EXERCISE:

Show that, for any irreducible representation \mathbf{r} ,

$$\mathcal{L}_{YM} = -\frac{1}{2g_{YM}^2} \frac{1}{T(\mathbf{r})} \text{tr}_{\mathbf{r}}(F_{\mu\nu}^{(\mathbf{r})} F^{(\mathbf{r})\mu\nu}), \quad (7.53)$$

where the **Dynkin index**

$$T(\mathbf{r}) := \frac{C(\mathbf{r})}{C(\text{fund})} \quad (7.54)$$

of the irreducible representation \mathbf{r} is invariant under changes of normalization (7.10) of the Lie algebra.

It turns out that there is a second gauge invariant term that one can add to the action. It is the **theta term**

$$S_\theta[A] = \int d^4x \mathcal{L}_\theta, \quad (7.55)$$

$$\mathcal{L}_\theta = \frac{\theta}{16\pi^2} \text{tr} \left(F_{\mu\nu} \tilde{F}^{\mu\nu} \right),$$

where θ is called the **theta angle** (more about why it is an angle in a later chapter), and

$$\tilde{F}^{\mu\nu} := \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} \quad (7.56)$$

is the **dual field strength**. In (7.56), $\epsilon^{\mu\nu\rho\sigma}$ is the completely antisymmetric tensor in four indices, with $\epsilon^{0123} = 1$.

To summarize, the most general gauge invariant action (with two derivatives) which contains a kinetic term for the non-abelian gauge field A_μ , as well as interaction terms, is

$$S_{\text{gauge}}[A] = S_{YM}[A] + S_\theta[A],$$

$$\mathcal{L}_{\text{gauge}} = \mathcal{L}_{YM} + \mathcal{L}_\theta = -\frac{1}{2g_{YM}^2} \text{tr}(F_{\mu\nu} F^{\mu\nu}) + \frac{\theta}{16\pi^2} \text{tr} \left(F_{\mu\nu} \tilde{F}^{\mu\nu} \right). \quad (7.57)$$

[Ex 33] * EXERCISE:

¹³It's constant in the sense that it does not depend on space-time. In quantum field theory, g_{YM} develops a dependence on the energy scale at which we are probing the system, so 'constant' is a misnomer. With that in mind, even though it's not relevant for this course, I'll typically call g_{YM} simply the 'Yang-Mills coupling'.

1. Express the Lagrangian density $\mathcal{L}_{\text{gauge}}$ in terms of A_μ^a and the structure constants f_{ab}^c , and identify quadratic terms involving derivatives of the gauge field, and cubic and quartic terms in A_μ , which represent interactions.
2. Show that the theta term (7.55) can be written as a surface (or ‘boundary’) term:

$$S_\theta = \frac{\theta}{8\pi^2} \int d^4x \partial_\mu K^\mu, \quad (7.58)$$

$$K^\mu = \epsilon^{\mu\nu\rho\sigma} \text{tr} \left(A_\nu \partial_\rho A_\sigma - \frac{2i}{3} A_\nu A_\rho A_\sigma \right).$$

3. Show that the **equations of motion** (EoM) obtained from the action S_{gauge} are

$$D_\mu F^{\mu\nu} \equiv \partial_\mu F^{\mu\nu} - i[A_\mu, F^{\mu\nu}] = 0. \quad (7.59)$$

4. Show, without using the EoM, that the **Bianchi identity**

$$D_\mu \tilde{F}^{\mu\nu} = 0. \quad (7.60)$$

holds.

If in addition to the gauge field A_μ there are also **charged fields** ϕ transforming in a representation \mathbf{r} (reducible or irreducible), then we can write a gauge invariant action for them using covariant derivatives. For instance for $G = SU(N)$, we have

$$S_{\text{matter}}[\phi, \phi^\dagger, A] = \int d^4x \mathcal{L}_{\text{matter}}, \quad (7.61)$$

$$\mathcal{L}_{\text{matter}} = -(D_\mu \phi)^\dagger D^\mu \phi - V(\phi, \phi^\dagger),$$

where we require the scalar potential V to be gauge invariant, that is, $V \mapsto V$ under non-abelian gauge transformations. This generalizes to other classical groups G by using the appropriate inner product in the kinetic term.

[Ex 34] * EXERCISE:

Consider the action

$$S[\phi, \bar{\phi}, A] = S_{YM}[A] + S_\theta[A] + S_{\text{matter}}[\phi, \bar{\phi}, A].$$

1. Show that the **EoM** are

$$D_\mu D^\mu \phi = \frac{\partial V}{\partial \phi^\dagger} \quad (7.62)$$

$$D_\nu F^{\mu\nu} = g_{YM}^2 J^\mu$$

for a current $J_\mu = J_\mu^a t_a$ that you should find.

2. Show that under a gauge transformation the current J^μ transforms as

$$J^\mu \mapsto g J^\mu g^{-1}, \quad (7.63)$$

and that J^μ is **covariantly conserved**, namely

$$D_\mu J^\mu = 0. \quad (7.64)$$

7.4 * Non-abelian Wilson lines and Wilson loops

This is a bonus section on advanced non-examinable material (hence the * in the section title), which I am including for completeness.

I will write this section up later, when I have time. In the meantime, if you are interested in this topic please see my handwritten lecture notes.

Chapter 8

Applications of non-abelian gauge theories

In this chapter we will study interesting smooth localised field configurations in nonabelian gauge theories, whose existence is ensured by topology, and study some of their geometric and physical properties. We will encounter 't Hooft-Polyakov monopoles, which are magnetically charged particle-like objects, and instantons, which are localized in (Euclidean) space-time. Good references for this chapter are [Manton and Sutcliffe, 2004] and [Weinberg, 2012].

8.1 The 't Hooft-Polyakov monopole

In 1974 Gerard 't Hooft and Aleksandr M. Polyakov discovered that nonabelian gauge theories with scalar fields transforming in the adjoint representation admit smooth magnetic monopoles as static finite energy solutions of their equations of motion [Hooft, 1974, Polyakov, 1974].

The field theory of interest is the so-called **Georgi-Glashow model** (or $SU(2)$ adjoint Higgs model) [Georgi and Glashow, 1972] a field theory in three space and one time dimension, with $G = SU(2)$ gauge group, a scalar field Φ transforming in the (3-dimensional) adjoint representation, which we represent as a 2×2 traceless hermitian matrix. The Lagrangian density is

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2g_{YM}^2} \text{tr}(F_{\mu\nu}F^{\mu\nu}) - \text{tr}((D_\mu\Phi)(D^\mu\Phi)) - V(\Phi) , \\ V(\Phi) &= \lambda \left(\frac{1}{2} \text{tr}(\Phi^2) - v^2 \right)^2 , \end{aligned} \tag{8.1}$$

where $\lambda, v > 0$ are constants and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \\ D_\mu \Phi &= \partial_\mu \Phi - i[A_\mu, \Phi] . \end{aligned} \quad (8.2)$$

We can calculate the Hamiltonian (or energy) density \mathcal{H} as the Legendre transform of the Lagrangian density \mathcal{L} , and from it the total energy $E = \int d^3x \mathcal{H}$ of the system, which is by construction gauge invariant (as should be the case for all physically observable quantities). We will be interested in static field configurations, so we can drop all time derivatives ∂_0 . It is then convenient to work in the temporal gauge $A_0 = 0$, which we can always achieve by a suitable gauge transformation, so that we can drop all time covariant derivatives D_0 . In the **temporal gauge**, the **energy of static field configurations** is

$$E = \int d^3x \left[\frac{1}{g_{YM}^2} \text{tr}(B_i B_i) + \text{tr}((D_i \Phi)(D_i \Phi)) + V(\Phi) \right] , \quad (8.3)$$

where $B_i = \frac{1}{2} \epsilon_{ijk} F_{jk}$ are the components of the nonabelian magnetic field \mathbf{B} . $i = 1, 2, 3$ runs over spatial Euclidean indices (which we write up or down since the spatial metric is δ_{ij}), and as usual repeated indices are summed over.

The energy is the integral of a sum of squares, and is minimized by setting

$$\mathbf{B} = 0 , \quad \mathbf{D}\Phi = 0 , \quad \text{tr}(\Phi^2) = 2v^2 . \quad (8.4)$$

The first vector equation tells us that $F_{ij} = 0$, so the vector potential $\mathbf{A} = (A_1, A_2, A_3)$ is ‘pure gauge’: $A_j = ih(\partial_j h^{-1}) = -i(\partial_j h)h^{-1}$ for a function $h(\mathbf{x})$ which takes values in $SU(2)$. The second vector equation tells us that the adjoint scalar field Φ is covariantly constant. The final scalar equation tells us that Φ minimizes the scalar potential. By a gauge transformation we can set $\mathbf{A} = 0$, then the second equation sets Φ to be constant. Letting $\Phi = \phi^a \sigma_a$, where (σ_a) are the Pauli matrices, we find that

$$\text{tr}(\Phi^2) = 2v^2 \quad \Leftrightarrow \quad (\phi^1)^2 + (\phi^2)^2 + (\phi^3)^2 = v^2 , \quad (8.5)$$

so the **vacuum manifold** is a 2-sphere of radius v :

$$\begin{aligned} \mathcal{V} &= \{ \Phi = \phi^a \sigma_a \in su(2) \mid \text{tr}(\Phi^2) = 2v^2 \} \\ &= \{ \boldsymbol{\phi} = (\phi^1, \phi^2, \phi^3) \in \mathbb{R}^3 \mid \boldsymbol{\phi}^2 = v^2 \} \cong S^2 . \end{aligned} \quad (8.6)$$

By a constant gauge transformation, we can take

$$\Phi = \begin{pmatrix} v & 0 \\ 0 & -v \end{pmatrix} = v\sigma_3 \quad \boldsymbol{\phi} = (0, 0, v) . \quad (8.7)$$

Any choice of vacuum breaks the gauge group $G = SU(2)$ down to a subgroup $H = U(1)$ which leaves the vacuum invariant. This type of symmetry breaking, where the action is

invariant under a symmetry (here elements of G which are not elements of H), but the vacuum (or ground state) is not, is called **spontaneous symmetry breaking** in physics.

REMARK:

If a continuous internal global symmetry is spontaneously broken, then there is a massless scalar field (called Nambu-Goldstone bosons) for each spontaneously broken symmetry generator [Nambu, 1960, Goldstone, 1961] If the symmetry is gauged, as it is here, the would-be Nambu-Goldstone bosons are not physical as they can be absorbed by a gauge transformation, but the gauge fields associated to the spontaneously broken gauge symmetry gain a mass, which is otherwise forbidden by gauge invariance. This is called the (Anderson-Brout-Englert-Guralnik-Hagen-) **Higgs** (-Kibble) **mechanism** [Anderson, 1963, Englert and Brout, 1964, Higgs, 1964, Guralnik et al., 1964], which is a key phenomenon in the Standard Model of Particle Physics.¹ If you want to understand these statements in the Georgi-Glashow model, try to solve the advanced parts of [Ex 36], which are indicated by a star. This is usually taught in Quantum Field Theory, but as you will see if you solve the exercise, it is already there in classical field theory.

In order for the energy (8.3) to be finite, we demand the boundary conditions

$$\mathbf{B} \rightarrow 0, \quad \mathbf{D}\Phi \rightarrow 0, \quad \text{tr}(\Phi^2) \rightarrow 2v^2 \quad \text{as } |\mathbf{x}| \rightarrow \infty, \quad (8.8)$$

so the fields must tend to a vacuum at spatial infinity. Note: this can be a different vacuum for each direction. As in the abelian Higgs model, we can use the gauge redundancy to work in a radial gauge, where $A_r = 0$. Then the limits of the fields as $r \rightarrow \infty$ with (θ, φ) fixed exist. In particular, the limit of the adjoint scalar field at spatial infinity defines a map

$$\begin{aligned} \Phi_\infty: \quad S_\infty^2 &\rightarrow \mathcal{V} \cong S^2 \\ (\theta, \varphi) &\mapsto \Phi_\infty(\theta, \varphi) := \lim_{r \rightarrow \infty} \Phi(r, \theta, \varphi), \end{aligned} \quad (8.9)$$

which is characterized by an **integer**, the **topological degree of the map**, which is a generalization of the winding number for maps from S^1 to S^1 .²

$$\nu = \frac{1}{8\pi v^3} \int_{S_\infty^2} \epsilon_{ijk} \phi_\infty \cdot (\partial_j \phi_\infty \times \partial_k \phi_\infty) d^2 \sigma_i, \quad (8.10)$$

where $\phi_\infty = ((\phi_\infty)^1, (\phi_\infty)^2, (\phi_\infty)^3)$. Note: the prefactor of v^{-3} is there because the target (image) of ϕ_∞ is a 2-sphere of radius v .

¹In the standard model, the gauge Lie algebra is $\mathfrak{g} = su(3) \oplus su(2) \oplus u(1)$. The fluctuations of the corresponding gauge fields are the gluons which mediate the strong force (for $su(3)$), the weak bosons which mediate the weak force (for $su(2)$), and an abelian gauge field which mediates the ‘hyper-charge’ force (for $u(1)$). A scalar field called the Higgs field is charged under $su(2) \oplus u(1)$ and is non-vanishing in the ground state, spontaneously breaking $su(2) \oplus u(1)$ down to a $u(1)$ subalgebra, whose associated gauge field is the electromagnetic gauge field. The fluctuations of the gauge fields associated to the broken symmetry are the massive electroweak bosons which have been observed at the LEP experiment at CERN: two charged electroweak bosons W_μ^\pm and a neutral electroweak boson Z_μ . So all in all $\mathfrak{g} = su(3) \oplus su(2) \oplus u(1)$, and $\mathfrak{h} = su(3) \oplus u(1)$.

²Mathematically, this is because $\Pi_2(S^2) = \mathbb{Z}$.

[Ex 37.4] * EXERCISE:

Define

$$F_{\mu\nu}^{U(1)} := \frac{1}{2v} \text{tr}(\Phi_\infty F_{\mu\nu}) \quad (8.11)$$

to be the field strength of the unbroken $H = U(1)$ subgroup of the gauge group $G = SU(2)$. Show that the magnetic charge

$$m^{U(1)} := \frac{1}{2\pi} \int_{S_\infty^2} \mathbf{B}^{U(1)} \cdot d^2\vec{\sigma} \quad (8.12)$$

of this unbroken $U(1)$ is proportional to the topological degree ν of Φ_∞ , and find the proportionality factor.

As an example, the map

$$\Phi_\infty = v \hat{x} \cdot \boldsymbol{\sigma}, \quad (8.13)$$

where $\hat{x} = \mathbf{x}/|\mathbf{x}| = \mathbf{x}/r$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, has degree $\nu = 1$. This is the identity map from S^2 to S^2 , up to an overall constant factor that takes care of the radius of the target sphere. We note incidentally that we can write (8.13) as

$$\Phi_\infty = v e^{-i\alpha} \sigma_3 e^{i\alpha} \quad (8.14)$$

with

$$\alpha = \frac{\theta}{2} (-\sin \varphi \sigma_1 + \cos \varphi \sigma_2) = \frac{\theta}{2} e^{-i\varphi\sigma_3/2} \sigma_2 e^{i\varphi\sigma_3/2}. \quad (8.15)$$

So Φ_∞ reduces to the constant vacuum with $\Phi = v\sigma_3$ in (8.7), if we perform a gauge transformation with parameter $g = e^{i\alpha}$. Note however that this gauge transformation is singular at $\theta = \pi$, the south pole of the 2-sphere, where φ is ill-defined. (The gauge transformation is regular at the north pole $\theta = 0$, thanks to the θ prefactor in α . This statement can be checked by switching to local coordinates which are well-defined at either pole.)

We are now ready to introduce the '**t Hooft-Polyakov 'hedgehog' ansatz**, so called because the vector field $\boldsymbol{\phi}$ points in the radial direction and looks a bit like a hedgehog. We assume that the adjoint scalar and the gauge field (written as a matrix-valued differential $A = A_\mu dx^\mu$) take the form

$$\begin{aligned} \Phi &= \frac{\mathbf{x} \cdot \boldsymbol{\sigma}}{r^2} H(vr) \\ A &= \sigma_a \epsilon_{aij} \frac{x_i dx_j}{r^2} [1 - K(vr)]. \end{aligned} \quad (8.16)$$

Note that the dependence on the angular polar coordinates in space \mathbb{R}^3 is correlated with the behaviour in the internal space in which the fields take values. We also assume the asymptotics

$$\underline{\xi \equiv vr \rightarrow \infty} : \quad H(\xi) - \xi \rightarrow 0, \quad K(\xi) \rightarrow 0 \quad (8.17)$$

at spatial infinity, to satisfy the boundary conditions (8.8) which are needed for the energy to be finite,³ and

$$\xi \equiv vr \rightarrow 0 : \quad H(\xi) = O(\xi), \quad K(\xi) - 1 = O(\xi) \quad (8.18)$$

to ensure regularity (smoothness) at the centre of the monopole, and finiteness of the energy at short distances from the centre.

Note that the adjoint scalar field approaches (8.13) at spatial infinity, which has topological degree 1. The magnetic field also approaches an abelian magnetic monopole for the unbroken gauge group $H = U(1)$ at spatial infinity. Indeed, if one applies the above singular gauge transformation, the gauge field $A_\mu^{U(1)}$ looks precisely like a Dirac monopole in the northern patch (or for $\theta \neq \pi$). One can find an analogous singular gauge transformation to obtain the Dirac monopole in the southern patch (or for $\theta \neq 0$).

One can substitute the 't Hooft-Polyakov ansatz (8.16) in the equations of motion, to find a system of two coupled ODE's for the functions $H(\xi)$ and $K(\xi)$. Together with the boundary conditions (8.17)-(8.18), this defines a well-posed boundary value problem which can be solved numerically. This shows the existence of a finite energy static solution which describes a magnetically charged object of finite size.

As in the case of gauged vortices in the abelian Higgs model, we can use a Bogomol'nyi-type argument to find a lower bound for the energy in each topological sector, namely for field configurations with given topological degree for the adjoint scalar, or equivalently magnetic charge for the unbroken $U(1)$ gauge field. This is called the Bogomol'nyi-Prasad-Sommerfield (or BPS) bound [Bogomol'nyi, 1976, Prasad and Sommerfield, 1975]. It is in the context of magnetic monopoles that the Bogomol'nyi trick was first developed. The argument follows the same logic of completing squares and reducing to a surface term that we have already encountered when studying gauged vortices in the abelian Higgs model. We write

$$\begin{aligned} E &= \int d^3x \left[\frac{1}{g_{YM}^2} \text{tr}(\mathbf{B}^2) + \text{tr}((\mathbf{D}\Phi)^2) + V(\Phi) \right] \\ &\geq \int d^3x \text{tr} \left(\left(\frac{1}{g_{YM}} \mathbf{B} \mp \mathbf{D}\Phi \right)^2 \pm \frac{2}{g_{YM}} \mathbf{B} \cdot \mathbf{D}\Phi \right) \\ &\geq \pm \frac{2}{g_{YM}} \int d^3x \text{tr}(\mathbf{B} \cdot \mathbf{D}\Phi) = \pm \frac{2}{g_{YM}} \int d^3x \text{tr}(\mathbf{D} \cdot (\Phi \mathbf{B})) \quad (8.19) \\ &= \pm \frac{2}{g_{YM}} \int d^3x \nabla \cdot \text{tr}(\Phi \mathbf{B}) = \pm \frac{2}{g_{YM}} \int_{S_\infty^2} \text{tr}(\Phi_\infty \mathbf{B}) \cdot d^2\sigma \\ &= \pm \frac{4v}{g_{YM}} \int_{S_\infty^2} \mathbf{B}^{U(1)} \cdot d^2\sigma = \pm \frac{8\pi v}{g_{YM}} m^{U(1)}. \end{aligned}$$

³It can be shown that the solution approaches the limiting values exponentially fast, much faster than is needed for the integral (8.3) to converge.

Going from the first to the second line, we dropped the contribution of the (non-negative) potential energy and completed a square. We then dropped the square to get to the third line, and then used the Bianchi identity $\mathbf{D} \cdot \mathbf{B} = 0$. Going to the fourth line we took the gauge covariant divergence outside the trace, and replaced it by a standard divergence since the trace is gauge invariant. Then we used Gauss' theorem (*aka* divergence theorem) to rewrite the lower bound as a surface integral, which in the last line we related to the magnetic charge of the unbroken $H = U(1)$ subgroup of the gauge group, defined in (8.12). We have deduced the **BPS bound**

$$E \geq \frac{8\pi v}{g_{YM}} |m^{U(1)}|, \quad (8.20)$$

which is a lower bound for the energy in terms of the magnetic charge.

The **bound is saturated**, that is $E = \frac{8\pi v}{g_{YM}} |m^{U(1)}|$, if and only if

$$\lambda \rightarrow 0 \quad \text{keeping } v \text{ fixed}, \quad (8.21)$$

which is called the **BPS limit**, and the fields satisfy the 1st order **Bogomol'nyi equation**

$$\mathbf{B} = \text{sign}(m^{U(1)}) g_{YM} \mathbf{D}\Phi. \quad (8.22)$$

Solutions to the Bogomol'nyi equations for monopoles come in infinite families, parametrized by continuous parameters also known as moduli. For $G = SU(2)$, the moduli space of n BPS monopoles (solutions of the Bogomol'nyi equations with total magnetic charge $m^{U(1)} = n > 0$) has $4n$ real dimensions.

8.2 Instantons

This is for another year.

Chapter 9

Bundles, connections, curvature and sections*

This is a bonus chapter that sketches some of the differential geometry that underlies gauge theories. We won't have time in the lectures for this advanced material, which is best learned in a different module. I include it here for completeness for students who would like to learn more. This material will not be examined.

So far we have learned how to formulate gauge theories in terms of gauge invariant actions for the gauge field and (potentially) charged fields. Our goal in this chapter will be understand how to describe gauge transformations, gauge fields, their field strengths, and charged fields geometrically. We will learn about fibre bundles, which are a consistent way of adding extra structure on top of a differentiable manifold.

I should warn you that the general formal definition is quite abstract, but I will try to build towards it slowly by successive generalizations. At the beginning I will give you a flavour of the abstract “intrinsic” approach, which defines concepts without making reference to a coordinate system. This can be hard to grasp, and this is not a course on differential geometry, so we will spend most of our time working in the “extrinsic” approach, which uses local coordinates. The extrinsic approach has the disadvantage that one needs to make sure that no definitions depend on the choice of coordinates used, but the advantage of being more explicit and accessible to beginners. This will be more than sufficient for our purposes.

This chapter is largely based on lectures 2 and 5 in Ooguri's lecture course on Mathematics for Theoretical Physicists [Ooguri, 2010]. Other references which cover the same material in more detail are [Eguchi et al., 1980, Nakahara, 2003, Naber and Naber, 1997].

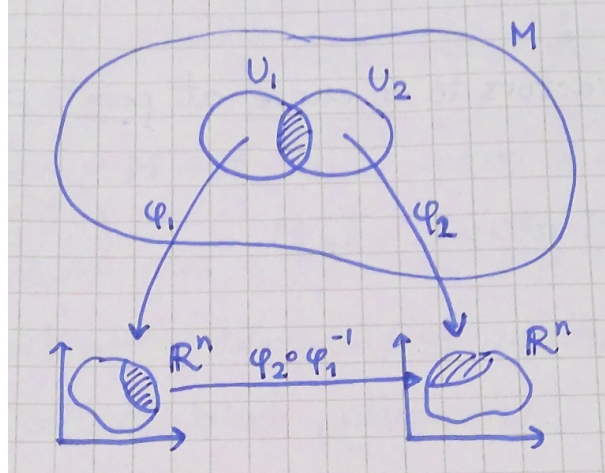


Figure 9.1: The basic data of a differentiable manifold.

9.1 The tangent bundle

Recall the definition of a differentiable manifold M (of dimension n) from the first term, see figure 9.1. It consists of a countable atlas $\{(U_i, \varphi_i)_{i \in I}\}$ of coordinate charts (or patches) (U_i, φ_i) , where U_i is an open subset of M , $\varphi_i : U_i \rightarrow \mathbb{R}^n$ is an invertible map from U_i to an open subset of \mathbb{R}^n , and $M = \bigcup_{i \in I} U_i$. Given a point $p \in M$, its image under $\varphi_i(p) = (x_{(i)}^1, \dots, x_{(i)}^n)$ under φ_i gives the coordinates of point p in the patch U_i . We refer to these as local coordinates. If two patches U_i and U_j overlap on $U_i \cap U_j \neq \emptyset$, then we can use two sets of coordinates. For any pair of overlapping patches, We require the transition functions

$$\varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \rightarrow \varphi_j(U_i \cap U_j),$$

which are invertible, to be smooth. This makes M a differentiable manifold.

Next we give the intrinsic definition of a differentiable (real) function. A function

$$\begin{aligned} \hat{f}: M &\rightarrow \mathbb{R} \\ p &\mapsto \hat{f}(p) \end{aligned} \tag{9.1}$$

is **differentiable (/smooth)** if for all charts (U_i, φ_i) , its extrinsic expression in local coordinates

$$\begin{aligned} f_{(i)} := \hat{f} \circ \varphi_i^{-1}: \quad \varphi_i(U_i) &\rightarrow \mathbb{R} \\ x_{(i)} = (x_{(i)}^1, \dots, x_{(i)}^n) &\mapsto f_{(i)}(x_{(i)}) \end{aligned} \tag{9.2}$$

is a differentiable/smooth function of n real variables. The requirement that the transition functions $\varphi_j \circ \varphi_i^{-1}$ of a differentiable manifold are smooth ensures that if f is smooth in one set of local coordinates, it is smooth in all sets of local coordinates. We denote the set of smooth function on M by $C^\infty(M)$.

In the following, to avoid cluttering the notation, we will drop the subscripts which label the different patches, unless they are strictly necessary. Note that we have used hats to distinguish the intrinsically defined value $\hat{f}(p)$ of the function at a point in the manifold from its extrinsic description $f(x) = (\hat{f} \circ \varphi^{-1})(x)$ in terms of local coordinates $x = \varphi(p)$ in a coordinate chart (U, φ) .

Last term you defined tangent vectors to a curve C at a point p in the manifold M . You saw that the set of tangent vectors to all curves passing through the point p is an n -dimensional real vector space, which is the **tangent space** $T_p M$ of the manifold M at point p . Next, we would like to extend this construction from a single point p to the whole manifold M . Informally, we would like to define

$$TM = \bigcup_{p \in M} T_p M, \quad (9.3)$$

a “bundle” of the tangent spaces at all the points in the manifold. This is called the **tangent bundle** TM of M . The question is: how do we define this object properly? To gain intuition, it is useful to take an equivalent but complementary view of tangent vectors. (We will see how this is related to the definitions that you saw last term below.)

We define a **tangent vector field** v on M as a map

$$\begin{aligned} \hat{v}: C^\infty(M) &\rightarrow C^\infty(M) \\ \hat{f} &\mapsto \hat{v}(\hat{f}) \end{aligned} \quad (9.4)$$

which obeys the following two properties:

1. **linearity:** $\forall a_1, a_2 \in \mathbb{R}, \forall \hat{f}_1, \hat{f}_2 \in C^\infty(M)$,

$$\hat{v}(a_1 \hat{f}_1 + a_2 \hat{f}_2) = a_1 \hat{v}(\hat{f}_1) + a_2 \hat{v}(\hat{f}_2) \quad (9.5)$$

2. **Leibniz rule:** $\forall \hat{f}, \hat{g} \in C^\infty(M)$,

$$\hat{v}(\hat{f}\hat{g}) = \hat{v}(\hat{f})\hat{g} + \hat{f}\hat{v}(\hat{g}). \quad (9.6)$$

Tangent vector fields form a vector space, more about this later.

*** EXERCISE:**

Let \hat{v}, \hat{w} be tangent vector fields.

1. Show that $\hat{w} \circ \hat{v}$ is not a tangent vector field.

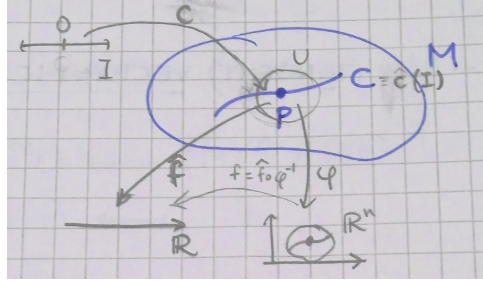


Figure 9.2: The data needed to define a tangent vector to a curve, applied to a function.

2. Show that $[\hat{w}, \hat{v}] = \hat{w} \circ \hat{v} - \hat{w} \circ \hat{v}$ is a tangent vector field.

Given a tangent vector field \hat{v} on M and a point $p \in M$, we can (re-)define a **tangent vector** $\hat{v}_p \in T_p M$ **at a point** p by evaluating everything at point p :¹

$$\begin{aligned} \hat{v}_p: C^\infty(M) &\rightarrow \mathbb{R} \\ \hat{f} &\mapsto \hat{v}_p(\hat{f}) := \left(\hat{v}(\hat{f}) \right) (p) \end{aligned} \tag{9.7}$$

See figure 9.2 for a depiction of the relevant data.

You may ask: how is this definition of tangent vectors related to the definition in terms of tangents to a curve, that you encountered in the first term? Given a smooth curve through p , which is defined by a map from an interval I to the manifold M ,

$$\begin{aligned} c: I \subseteq \mathbb{R} &\rightarrow M \\ \tau &\mapsto c(\tau) \end{aligned} \tag{9.8}$$

with $c(0) = p$, we can define a tangent vector \hat{v}_p to the curve $C = c(I)$ by

$$\hat{v}_p(\hat{f}) = \left. \frac{d}{d\tau} \hat{f}(c(\tau)) \right|_{\tau=0}, \tag{9.9}$$

which is defined intrinsically for all smooth functions $\hat{f} \in C^\infty(M)$. See figure 9.2 To understand what is going on, let's express this in local coordinates x^μ in a chart (U, φ) , where the curve is parametrized by

$$(\varphi \circ c)(\tau) \equiv x(\tau) = (x^1(\tau), \dots, x^n(\tau)), \tag{9.10}$$

¹A tangent vector \hat{v}_p at a point p is also linear and obeys a form of the Leibniz rule:

1. $\hat{v}_p(a_1 \hat{f}_1 + a_2 \hat{f}_2) = a_1 \hat{v}_p(\hat{f}_1) + a_2 \hat{v}_p(\hat{f}_2)$
2. $\hat{v}_p(\hat{f} \hat{g}) = \hat{v}_p(\hat{f}) \hat{g}(p) + \hat{f}(p) \hat{v}_p(\hat{g})$,

as an immediate consequence of (9.5) and (9.6) for tangent vector fields. One can also define tangent vectors at a point more abstractly using the axioms in this footnote without making reference to tangent vector fields, and then introduce tangent vector fields from this.

and the function $\hat{f}(p)$ is represented as $f(x) = (\hat{f} \circ \varphi^{-1})(x)$:

$$\begin{aligned} \hat{v}_p(\hat{f}) &= \left. \frac{d}{d\tau} (\hat{f} \circ c)(\tau) \right|_{\tau=0} = \left. \frac{d}{d\tau} (\hat{f} \circ \varphi^{-1} \circ \varphi \circ c)(\tau) \right|_{\tau=0} \\ &= \left. \frac{d}{d\tau} f(x(\tau)) \right|_{\tau=0} = \dot{x}^\mu(\tau) \left. \frac{\partial f(x)}{\partial x^\mu} \right|_{x=x(0)=\varphi(p)}, \end{aligned} \quad (9.11)$$

where we used basic properties of the composition of functions, as well as the chain rule in the last equality (dots denote derivatives with respect to τ). We recognize the result as the **directional derivative** of the function f along the tangent to the curve at the point p , which has coordinates $x = x(0)$.

REMARKS:

1. When you described the tangent vector to a curve at a point p using local coordinates in the first term, $\dot{x}^\mu(0)$ were the **components of the tangent vector**.
2. To construct a **basis of the tangent space** $T_p M$, you used curves C_a which fixed all coordinates $x^\mu \neq a$ and varied only $x^a(\tau) = x^a(0) + \tau$. The components of the tangent vector e_a to such a curve are then $\dot{x}^\mu(0) = \delta_a^\mu$, and we have

$$e_a(\hat{f}) = \left. \frac{\partial}{\partial x^a} f(x) \right|_x = \varphi(p) = \left. \frac{\partial}{\partial x^a} \hat{f}(\varphi^{-1}(x)) \right|_x = \varphi(p), \quad (9.12)$$

or for short

$$e_a = (\partial_a)_p, \quad (9.13)$$

where $(\partial_a)_p$ is $\frac{\partial}{\partial x^a}$ when we work in local coordinates $x = \varphi(p)$.

In summary, we can write any tangent vector $\hat{v}_p \in T_p M$ intrinsically as

$$\hat{v}_p = \hat{v}^a (\partial_a)_p, \quad (9.14)$$

or extrinsically (in local coordinates) as

$$v = v^a \frac{\partial}{\partial x^a}, \quad (9.15)$$

where the components $\hat{v}^a = v^a$ are n real numbers.

Now let's consider a collection of tangent spaces over every point on M : the **tangent bundle**

$$TM = \bigcup_{p \in M} T_p M. \quad (9.16)$$

Using the isomorphism $T_p M \cong \mathbb{R}^n$ for all $p \in M$, we can view the tangent bundle locally as $U_i \times \mathbb{R}^n$. we can see that TM is naturally a **manifold** of dimension $2n$. For each coordinate

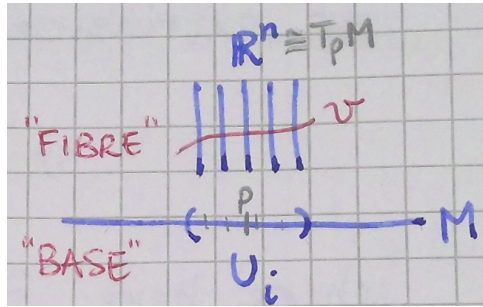


Figure 9.3: The tangent bundle and a tangent vector field.

chart (U_i, φ_i) on M , we define coordinates (x^μ, v^ν) on $\bigcup_{p \in U_i} T_p M$, where (x^μ) are coordinates on U_i , and we parametrize a tangent vector as

$$v = v^\nu \frac{\partial}{\partial x^\nu} . \tag{9.17}$$

We call M the **base** of the tangent bundle, and $\mathbb{R}^n \cong T_p M$ the **fibre** of the tangent bundle.²

A (smooth) **tangent vector field** is then (in local coordinates)

$$v = v^\mu(x) \frac{\partial}{\partial x^\mu} , \tag{9.18}$$

with components $v^\mu(x)$ which vary smoothly as p varies over M .

*** EXERCISE:**

Check that the local description (9.18) of a tangent vector field maps smooth functions to smooth functions, is linear, and obeys the Leibniz rule.

We say that a (smooth) tangent vector field (9.18) is a (smooth) **section** of the tangent bundle TM , and write $v \in \Gamma(TM)$. The reason for this terminology is as follows (see figure 9.3:

- **Locally**, TM is a **product space** $U_i \times \mathbb{R}^n$, where the fibre is $\mathbb{R}^n \cong T_p M$ for every p .
- The vector field v draws a graph $(x^\mu, v^\nu(x))$ in $\varphi_i(U_i) \times \mathbb{R}^n$, with **cuts** the fibres of the tangent bundle TM along the direction of the base M . Hence the term “section”.

²For the attentive reader: in order to equip TM with the structure of a *differentiable* manifold, we need to specify smooth transition functions for all its $2n$ coordinates, not just for the base coordinates. We will do that shortly, in equation (9.20).

What we have seen so far is a local description of the tangent bundle TM in a coordinate patch. When we change patch from U to \tilde{U} (on their overlap $U \cap \tilde{U}$) in the base M , the coordinates on M change as³

$$x^\mu \mapsto \tilde{x}^\mu = \tilde{x}^\mu(x). \quad (9.19)$$

In addition, we need to specify how the fibre coordinates change. We require the tangent space coordinates to change like

$$v^\mu \mapsto \tilde{v}^\mu = \frac{\partial \tilde{x}^\mu}{\partial x^\nu} v^\nu, \quad (9.20)$$

so that

$$v = v^\mu \frac{\partial}{\partial x^\mu} = \tilde{v}^\mu \frac{\partial}{\partial \tilde{x}^\mu} \quad (9.21)$$

is independent of the choice of coordinates.

Proof. Using the chain rule,

$$\frac{\partial}{\partial x^\mu} = \frac{\partial \tilde{x}^\nu}{\partial x^\mu} \frac{\partial}{\partial \tilde{x}^\nu} \quad \longrightarrow \quad v^\mu \frac{\partial}{\partial x^\mu} = v^\mu \frac{\partial \tilde{x}^\nu}{\partial x^\mu} \frac{\partial}{\partial \tilde{x}^\nu} = \tilde{v}^\nu \frac{\partial}{\partial \tilde{x}^\nu}. \quad (9.22)$$

Now recall that every vector space V has a dual vector space V^* , which is the space of linear functionals on V . Given a basis e_a of V , we can choose a basis e^{*a} of the dual space V^* by requiring that $e^{*a}(e_b) = \delta_b^a$. Then given $v = v^a e_a \in V$ and $w = w_a e^{*a}$, we have $w(v) = w_a v^a$. We can apply these ideas to the tangent space $T_p M$, and define its dual vector space, the **cotangent space** $T_p^* M$. An element ω of the cotangent space is a linear functional on the tangent space,

$$\begin{aligned} \omega: T_p M &\rightarrow \mathbb{R} \\ v &\mapsto \omega(v) \end{aligned} \quad (9.23)$$

such that for all coefficients $a_1, a_2 \in \mathbb{R}$ and for all tangent vectors $v_1, v_2 \in T_p M$,

$$\omega(a_1 v_1 + a_2 v_2) = a_1 \omega(v_1) + a_2 \omega(v_2). \quad (9.24)$$

The dual basis to the basis of partial derivatives $\left\{ \frac{\partial}{\partial x^\mu} \right\}$ for the tangent space $T_p M$ is the basis of **differentials** $\{dx^\mu\}$ for the cotangent space $T_p^* M$, where we require

$$dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta_\nu^\mu. \quad (9.25)$$

³(In terms of the invertible maps $\varphi: U \rightarrow \mathbb{R}^n$ and $\tilde{\varphi}: \tilde{U} \rightarrow \mathbb{R}^n$, the change of coordinates is given by the transition function $\tilde{\varphi} \circ \varphi^{-1}$:

$$x = \varphi(p) \mapsto \tilde{x} = \tilde{\varphi}(p) = (\tilde{\varphi} \circ \varphi^{-1})(x).$$

So we can write any **cotangent vector** $\omega \in T_p^*M$ as

$$\omega = \omega_\mu dx^\mu . \quad (9.26)$$

Under a change of coordinates (9.19) on M , we will require that the cotangent space coordinates transform as

$$\omega_\mu \mapsto \tilde{\omega}_\mu = \frac{\partial x^\nu}{\partial \tilde{x}^\mu} \omega_\nu , \quad (9.27)$$

so that

$$\omega = \omega_\mu dx^\mu = \tilde{\omega}_\mu d\tilde{x}^\mu \quad (9.28)$$

is independent of the choice of coordinates.

*** EXERCISE:**

1. Use the definition $df(x) = \frac{\partial f(x)}{\partial x^\mu} dx^\mu$ of the **differential** of a function to show that under a coordinate change (9.19)

$$dx^\mu \mapsto d\tilde{x}^\mu = \frac{\partial \tilde{x}^\mu}{\partial x^\nu} dx^\nu \quad (9.29)$$

and therefore

$$\omega = \omega_\mu dx^\mu \mapsto \tilde{\omega} = \tilde{\omega}_\mu d\tilde{x}^\mu = \omega_\nu dx^\nu = \omega . \quad (9.30)$$

2. Let $v = v^\mu \frac{\partial}{\partial x^\mu} \in T_pM$ and $\omega = \omega_\mu dx^\mu \in T_p^*M$. Show that

$$\omega(v) = \omega_\mu v^\mu \quad (9.31)$$

and that it is independent of the choice of coordinates:

$$\omega_\mu v^\mu = \tilde{\omega}_\mu \tilde{v}^\mu . \quad (9.32)$$

With all these data we can construct the **cotangent bundle**

$$T^*M = \bigcup_{p \in M} T_p^*M \quad (9.33)$$

as a collection of cotangent spaces over every point on M . For each coordinate chart (U_i, φ_i) on M , we require the cotangent bundle to locally look like $T^*U_i = \bigcup_{p \in U_i} T_p^*M \cong U_i \times \mathbb{R}^n$, with coordinates (x^μ, ω_ν) for the base and the fibre respectively. Under a change of coordinates (9.19) in the base M , the fibre coordinates change as in (9.27), so that $\omega = \omega_\mu dx^\mu$ is coordinate independent.

A (smooth) **cotangent vector field** is, in local coordinates,

$$\omega = \omega_\mu(x) dx^\mu , \quad (9.34)$$

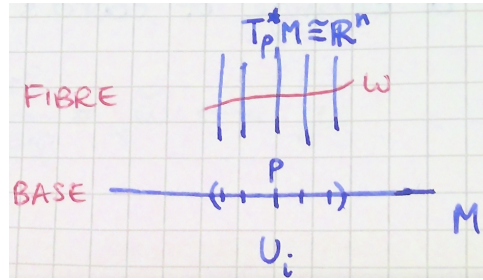


Figure 9.4: The cotangent bundle and a cotangent vector field.

where $\omega_\mu(x)$ are smooth functions. It is a (smooth) section of the cotangent bundle T^*M , and we write $\omega \in \Gamma(T^*M)$. See figure 9.4

REMARKS:

1. In **Lagrangian mechanics**, the generalised coordinates q^i and the generalised velocities v^j are coordinates on the **tangent bundle** TM of the configuration space M . The generalised coordinates q^i are coordinates on the base M , and the generalised velocities v^j are coordinates on the fibre T_pM . Under time evolution, the trajectory of the generalised coordinates traces a curve $(q^i(t))$ in the configuration space M , while the generalised velocities $(v^j(t)) = (\dot{q}^j(t))$ are the components of the tangent vector $v = v^j(t) \frac{\partial}{\partial q^j}$ to the trajectory.
2. In **Hamiltonian mechanics**, the generalised coordinates q^i and the generalised momenta p_j are coordinates on the **cotangent bundle** T^*M of the configuration space M , where we identify $p_j = \frac{\partial L}{\partial v^j} = \frac{\partial L}{\partial \dot{q}^j}$. Now $\theta = p_j(t) dq^j$ is a cotangent vector. The relation between Lagrangian and Hamiltonian can be written as

$$H = L - \dot{q}^i \frac{\partial L}{\partial \dot{q}^i} = L - \theta(v).$$

9.2 Fibre bundles

We can generalise the previous construction by replacing the tangent space T_pM or cotangent space T_p^*M by a more general fibre.

The simplest generalization is the notion of **vector bundle** E , which consists of a base $M = \bigcup_i U_i$ (of dimension $\dim M = n$) and of a fibre F which is a fixed vector space V (of dimension $\dim V = m$) over every point in M . Locally, the vector bundle E looks like $U_i \times V$, with coordinates (x, v) .

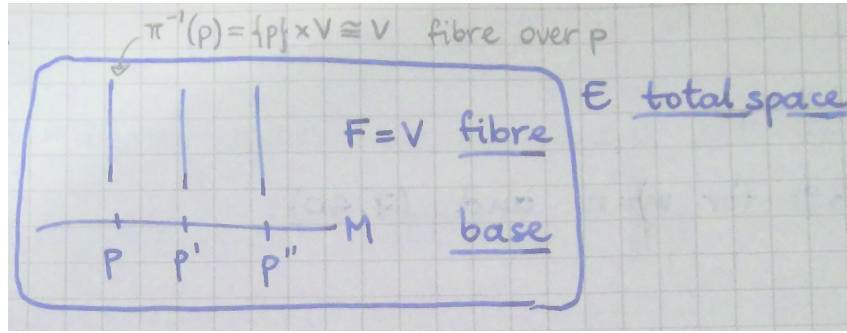


Figure 9.5: Schematic depiction of a vector bundle.

Mathematically, a differentiable manifold E is called a (smooth) **vector bundle** if:

1. There exists a **projection map**

$$\pi : E \rightarrow M \quad (9.35)$$

such that

$$\forall p \in M \quad \pi^{-1}(p) \cong V, \quad (9.36)$$

where V is a vector space.

2. There exist atlases of E and of M such that for all charts of M there exists a smooth map

$$\varphi : \pi^{-1}(U) \rightarrow U \times V, \quad (9.37)$$

which is called a **local trivialisation** of the vector bundle E over M .

Part 1 is a way of saying that the base M is part of the **total space** E , and that for each point in M we have a vector space V . Part 2 means that we can use local coordinates (x, v) for E , where x is a local coordinate for a point p in the base M , and v is a local coordinate of the fibre, the vector space $\pi^{-1}(p)$ associated to the point p . The structure of a vector bundle is summarized in Figure 9.5.

(To be precise, the vector bundle is the collection (E, M, π, V) of the total space E , the base M which is obtained by the projection map π , and the fibre V , which is the preimage of a point in the base under the projection map.)

To fully specify the vector bundle when we work in local coordinates, we need to state what happens to the fibre coordinates when we change coordinates in the base, from a neighbourhood U with coordinates x to a neighbourhood \tilde{U} with coordinates \tilde{x} . The change of coordinates in the base and the fibre is

$$\begin{aligned} x^\mu &\mapsto \tilde{x}^\mu = \tilde{x}^\mu(x) \\ v &\mapsto \tilde{v} = t(x)v, \end{aligned} \quad (9.38)$$

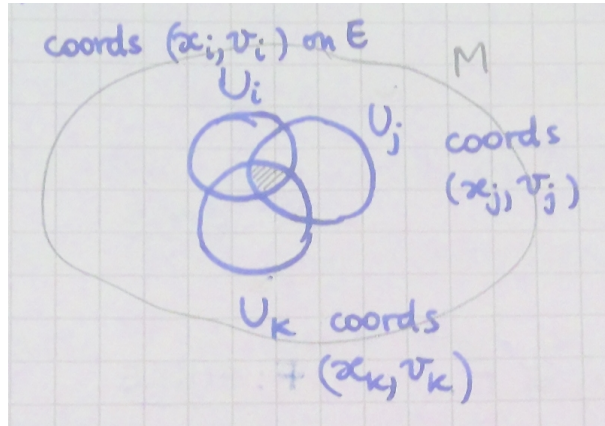


Figure 9.6: Triple overlap and coordinates in local trivialisations of a vector bundle.

where the **transition function** for the fibre is an x -dependent invertible linear transformation:⁴

$$t(x) \in GL(V) \equiv GL(m, \mathbb{R}) \tag{9.39}$$

There is a **consistency condition** associated to triple overlaps $U_i \cap U_j \cap U_k$, which ensures the uniqueness of the vector bundle. See Figure 9.6. Let (x_i, v_i) be local coordinates in $U_i \times V$, and likewise for j and k , and $t_{j \leftarrow i}(x_i)$ be the transition function for the fibre when we switch to the i -th trivialization to the j -th trivialization, and similarly for other transition functions.⁵ Then there are two ways of going from the i -th trivialization to the k -th trivialization: we can either go from i to k directly, or go from i to j and then from j to k . The results of the two processes are

$$v_k = t_{k \leftarrow i}(x_i)v_i = t_{k \leftarrow j}(x_j)v_j = t_{k \leftarrow j}(x_j)t_{j \leftarrow i}(x_i)v_i . \tag{9.40}$$

Demanding the compatibility of the two expressions for every vector v_i leads to the **cocycle condition**

$$t_{k \leftarrow i}(x_i) = t_{k \leftarrow j}(x_j(x_i))t_{j \leftarrow i}(x_i) . \tag{9.41}$$

It can be proven that there are no further compatibility conditions associated to quadruple or higher overlaps.

*** EXERCISE:**

Show that the transition functions for the tangent bundle TM and the cotangent bundle T^*M obey (9.39) and the cocycle condition (9.41).

⁴This is for a real vector bundle, in which the fibre V is a real vector space. If V is a vector space over a field \mathbb{F} , replace \mathbb{R} by \mathbb{F} .

⁵Here i, j, k are labels, not vector indices. In the notation used at the beginning of this chapter, I would have written $(x_{(i)}, v_{(i)})$ etc. I am omitting brackets here to avoid cluttering the notation.

REMARKS:

1. Unlike for TM and T^*M , the transition functions for the fibre of a general vector bundle are independent of the transition functions for the base.
2. We could take $\tilde{x}^\mu = x^\mu$, namely not change coordinates in the base, but still change coordinates in the fibre. Equations (9.39) and (9.41) must still hold.

Vocabulary: A (usually complex) vector bundle with one-dimensional fibre is called a **line bundle**.

We can generalize the previous structure further if we allow the fibre F to be a more general object than a vector space. We will restrict ourselves to considering fibres F which are differentiable manifolds themselves, even though this assumption can be relaxed further. Vector bundles are included as a special case, since a vector space is a differentiable manifold.

A differentiable manifold E is called a (smooth) **fibre bundle** if:

1. There exists a **projection map**

$$\pi : E \rightarrow M \quad (9.42)$$

such that

$$\forall p \in M \quad \pi^{-1}(p) \cong F. \quad (9.43)$$

2. There exist atlases of E and of M such that for all charts of M there exists a smooth map

$$\varphi : \pi^{-1}(U) \rightarrow U \times F, \quad (9.44)$$

which is called a **local trivialisation** of the fibre bundle E over M .

The interpretation is the same as for vector bundles, with the exception that the fibre need not be a vector space. In a local trivialisation, we can choose local coordinates (x, y) , where x is a local coordinate on the base M and y is a local coordinate on the fibre F . When we change coordinates in the base, the fibre coordinates must change appropriately, and the transition functions for the fibre must obey a cocycle condition.

The transition functions for the fibre are elements of a group, which is called the **structure group** of the fibre bundle E .⁶

EXAMPLE: Principal G -bundle P

⁶The transition functions for the base M are also elements of a group, the diffeomorphism group of M .

A principal bundle⁷ is a fibre bundle where the fibre is a Lie group, $F = G$, for example $G = U(1)$, $G = SU(2)$ or $G = SO(3)$. Let (x, h) be coordinates in (the image of) a local trivialization $U \times G$, and (\tilde{x}, \tilde{h}) be coordinates in $U \times G$, where h, \tilde{h} are elements of the group G . We require the transition function $t(x)$ for the fibre to be a group element itself, $t(x) = g(x) \in G$ for all x , which acts by group multiplication on the fibre coordinate:

$$(x^\mu, h) \mapsto (\tilde{x}^\mu(x), \tilde{h} = g(x)h) . \tag{9.45}$$

So for a principal G -bundle the fibre is the Lie group G , and the structure group is also G .

REMARKS:

1. This is called a ‘principal’ bundle because of its importance: it controls the structure of infinitely many vector bundles. Indeed, for each **representation** \mathbf{r} of G we have a vector space $V^{(\mathbf{r})}$ of dimension r and an action of the Lie group G on $V^{(\mathbf{r})}$ by a representation matrix $r(g)$. We can then define an **associated vector bundle** E with

$$\begin{array}{ll} \text{fibre} & F = V^{(\mathbf{r})} \\ \text{transitions functions} & t(x) = r(g(x)) \end{array} \tag{9.46}$$

so that under a change of coordinates

$$(x, v) \mapsto (\tilde{x}(x), \tilde{v} = r(g(x))v) . \tag{9.47}$$

The ‘associated vector bundle’ is associated to the principal bundle P and the representation \mathbf{r} .

2. We can start to observe a correspondence between Maths and Physics emerge:

Maths	Physics
Principal G -bundle (Section of) Associated vector bundle	Gauge symmetry G Charged field

We will complete this correspondence in the next section.

*** EXERCISE:**

$U(1)$ bundles over S^2 , or monopole bundles [\[SC: To be added\]](#)

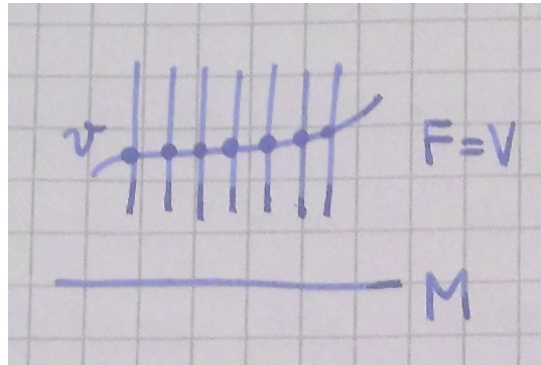


Figure 9.7: Schematic depiction of a vector field.

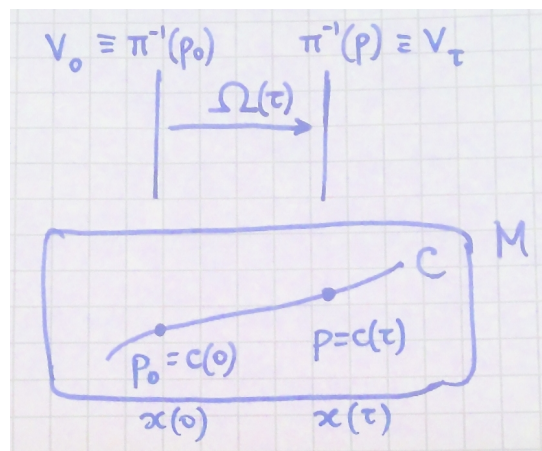


Figure 9.8: Schematic depiction of the notion of parallel transport.

9.3 Connection, holonomy and curvature

Let $v(x)$ be a smooth section of a vector bundle over M , written in local coordinates. See figure 9.7. Can we define partial derivatives of v , or directional derivatives of v along a curve C in M , which in local coordinates is parametrised by $x^\mu = x^\mu(\tau)$?

We immediately run into a problem: we cannot subtract vectors defined at infinitesimally close points, as we would do to define a derivative, because these two vectors belong to two different vector spaces. In order to define a notion of directional derivative, we need a way of comparing vectors defined at different points along the curve. Let $p_0 = c(0)$ and $p = c(\tau)$ be two points along the curve C , with coordinates $x(0)$ and $x(\tau)$ respectively. Associated to those two points we have two distinct (though isomorphic) vector spaces, $V_0 \equiv \pi^{-1}(p_0) = \pi^{-1}(c(0))$ and $V_\tau \equiv \pi^{-1}(p) = \pi^{-1}(c(\tau))$. We can compare elements of V_0 and elements of V_τ by introducing

⁷This is often misspelt as principle bundle. We shouldn't change our principles as we change coordinates in the base, therefore principle bundles are not a good idea.

a notion of **parallel transport** of vectors along the curve C , which is realised by an invertible linear map

$$\begin{aligned} \Omega(\tau): V_0 &\rightarrow V_\tau \\ v_0 &\mapsto \Omega(\tau)v_0 \end{aligned} \quad (9.48)$$

which obeys $\Omega(0) = \mathbb{1}$. See figure 9.8. Picking a basis of the vector space V , $\Omega(\tau)$ is a matrix in $GL(V)$.

More generally, we can compare vectors in the fibres above any two points $c(\tau)$ and $c(\tau')$ along the curve C by using the map

$$\Omega(\tau')\Omega^{-1}(\tau) : V_\tau \rightarrow V_{\tau'} . \quad (9.49)$$

By comparing the values of the vector field at infinitesimally close points, with coordinates $x^\mu = x^\mu(\tau)$ and $x^\mu(\tau + \epsilon d\tau)$, we can define the **covariant derivative** $\nabla_\mu v$ by

$$\nabla v = \nabla_\mu v dx^\mu := \lim_{\epsilon \rightarrow 0} \frac{v(x(\tau + \epsilon d\tau)) - \Omega(\tau + \epsilon d\tau)\Omega^{-1}(\tau)v(x(\tau))}{\epsilon} \quad (9.50)$$

where $dx^\mu = \dot{x}^\mu(\tau)d\tau$ in the parametrization of the curve. The parameter ϵ is a book-keeping device which I have introduced to keep track of infinitesimals and to define the limit.

REMARK:

The definition of the covariant derivative (9.50) of the vector field v depends on the local form of parallel transport Ω in an infinitesimal neighbourhood of τ . Letting

$$\Omega(\tau + \epsilon d\tau) = \Omega(\tau) - \epsilon \mathcal{A}(x(\tau))\Omega(\tau) + \mathcal{O}(\epsilon^2) , \quad (9.51)$$

the equation (9.50) becomes

$$\nabla v(x) = dv(x) + \mathcal{A}(x)v(x) , \quad (9.52)$$

where $\mathcal{A}(x)$, which is called the **connection** of the vector bundle, is a matrix-valued cotangent vector field (or equivalently, a matrix-valued differential form):

$$\mathcal{A}(x) = \mathcal{A}_\mu(x)dx^\mu , \quad (9.53)$$

with \mathcal{A}_m a matrix in $gl(V)$ for each μ and x .⁸

In components, the covariant derivative reads

$$\nabla_\mu v^\alpha(x) = \partial_\mu v^\alpha(x) + \mathcal{A}_\mu(x)^\alpha_\beta v^\beta(x) . \quad (9.54)$$

⁸Formally, \mathcal{A}_μ takes value in the Lie algebra of the structure group of the vector bundle.

The **connection** \mathcal{A} encodes the **infinitesimal** version of **parallel transport**.

Now consider a change of coordinates in the fibre only:

$$(x, v) \mapsto (x, \tilde{v} = t(x)v) . \quad (9.55)$$

Being a map from V_0 to V_τ , the parallel transport map $\Omega(\tau)$ transforms like

$$\Omega(\tau) \mapsto t(x(\tau))\Omega(\tau)t(x(0))^{-1}$$

under changes of coordinates in the fibres. Using the definition (9.50), it follows that $\nabla_\mu v$ transforms like v :

$$\nabla_\mu v(x) \mapsto t(x)\nabla_\mu v(x) , \quad (9.56)$$

or in terms of differential operators

$$\nabla_\mu \mapsto t(x)\nabla_\mu t(x)^{-1} . \quad (9.57)$$

This requires the connection to transform as follows:

$$\mathcal{A}_\mu \mapsto \tilde{\mathcal{A}}_\mu = t\partial_\mu t^{-1} + t\mathcal{A}_\mu t^{-1} . \quad (9.58)$$

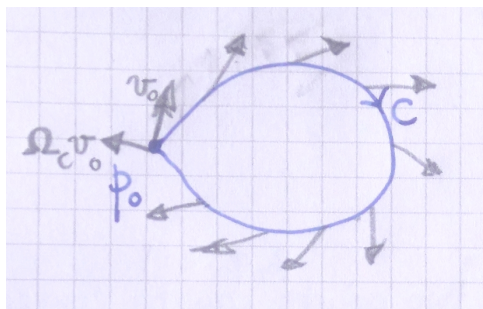
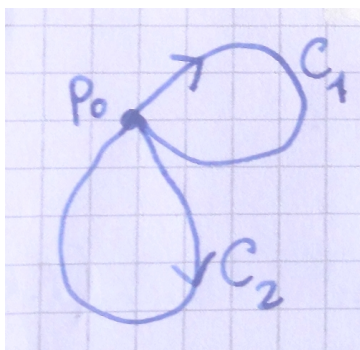
REMARKS:

1. This construction works for any vector bundle E . In fact, it works for any fibre bundle, with minor adjustments which I leave as an exercise for the interested reader.
2. When E is a vector bundle associated to a principal G -bundle G and a representation \mathbf{r} , the connection is $\mathcal{A}_\mu = -iA_\mu^{(\mathbf{r})}$, with $A_\mu^{(\mathbf{r})}$ the gauge field, acting in the representation \mathbf{r} . For a principal G -bundle, $\mathcal{A}_\mu = -iA_\mu$, where A_μ is the Lie algebra valued gauge field which transforms into $gA_\mu g^{-1} + ig\partial_\mu g^{-1}$. (The conversion factors of i are conventional: they are there because physicists have good reasons to like unitary and hermitian operators.)
3. When E is the tangent bundle, \mathcal{A}_μ is the affine connection which appears in differential geometry and general relativity, also known as Levi-Civita connection.

Now let's return to the finite version of parallel transport. Consider a closed curve (or **loop**) C in the base manifold M , starting and ending at the same point p_0 , which is called the **base point** of the loop. See figure 9.9. We can parallel transport a vector $v_0 \in \pi^{-1}(p_0)$ along the loop C . When we reach the end of the loop we obtain a new vector $\Omega_C v_0 \in \pi^{-1}(p_0)$, which is 'rotated' by a transformation⁹

$$\Omega_C \in GL(V) \quad (9.59)$$

⁹Recall that $\pi^{-1}(p_0) = V$.

Figure 9.9: Parallel transport and The holonomy Ω_C along the loop C .Figure 9.10: Concatenation of two paths C_1 and C_2 .

compared to the original vector v_0 . This is called the **holonomy** (of the connection \mathcal{A}_μ) along the loop C .

Holonomies along loops starting and ending at the same base point p_0 form a group, called the **holonomy group**, which is a subgroup of $GL(V)$. This is a consequence of the definition of parallel transport and of the fact that closed paths themselves form a group, where the composition law is the concatenation of paths. A bit more explicitly:

- If we concatenate two loops C_1 and C_2 to form a new loop $C_2 \circ C_1$ obtained by going along C_1 first and then along C_2 (see figure 9.10), we get

$$\Omega_{C_2 \circ C_1} = \Omega_{C_2} \Omega_{C_1} , \quad (9.60)$$

which is the **composition law** (or multiplication) in the holonomy group.

- The homotopy along the **trivial loop**, which doesn't move from the base point p_0 , is the **identity element** in the holonomy group.
- Given a loop C , we can define the loop $-C$ which traces the same curve with opposite

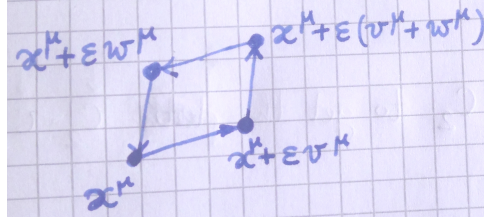


Figure 9.11: Curvature, from the holonomy along the perimeter of an infinitesimal parallelogram.

orientation. Then

$$\Omega_{-C} = \Omega_C^{-1} \tag{9.61}$$

is the inverse element of Ω_C in the holonomy group.

REMARK:

The holonomy group is generically non-abelian:

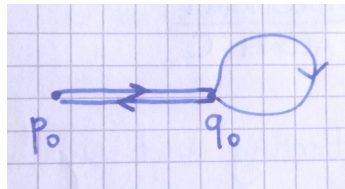
$$\Omega_{C_1}\Omega_{C_2} \neq \Omega_{C_2}\Omega_{C_1}. \tag{9.62}$$

If we parallel transport first along C_1 and then along C_2 , we'll usually get a different result than if we parallel transported first along C_2 and then along C_1 .¹⁰

*** EXERCISE:**

Let M be connected, that is, any two points $p_0, q_0 \in M$ can be connected by a curve in M . Show that the holonomy groups based at p_0 and at q_0 are isomorphic.

[**Hint:** think about the following picture:



]

The **curvature** $\mathcal{F}_{\mu\nu}$ is the holonomy along an infinitesimal loop. More precisely, consider an infinitesimal loop dC which is the perimeter of a parallelogram with vertices

$$x^\mu, \quad x^\mu + \epsilon v^\mu, \quad x^\mu + \epsilon(v^\mu + w^\mu), \quad x^\mu + \epsilon w^\mu, \tag{9.63}$$

¹⁰There are exceptions, for instance if the structure group of the fibre bundle is abelian, or if the connection vanishes. Hence the qualifier 'generically'.

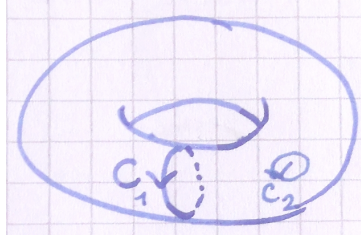


Figure 9.12: Contractible and non-contractible loops on a 2-torus.

as in figure 9.11. Then

$$\Omega_{dC} = \mathbb{1} + \epsilon^2 \mathcal{F}_{\mu\nu}(x) v^\mu w^\nu + \mathcal{O}(\epsilon^3), \quad (9.64)$$

where

$$\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu + [\mathcal{A}_\mu, \mathcal{A}_\nu]. \quad (9.65)$$

Proof. Exercise.

Under a change of coordinates in the fibre (9.55), the curvature transforms as follows:

$$\mathcal{F}_{\mu\nu} \mapsto t \mathcal{F}_{\mu\nu} t^{-1}. \quad (9.66)$$

REMARKS:

1. For a principal G -bundle, $\mathcal{F}_{\mu\nu} = -iF_{\mu\nu}$, where $F_{\mu\nu}$ is the field strength of A_μ . (Similarly, $\mathcal{F}_{\mu\nu} = -iF_{\mu\nu}^{(r)}$ for an associated vector bundle.)
2. Let us assume that the curvature vanishes. This does not mean that the connection vanishes. This has the surprising consequence that the holonomy can be non-trivial (that is, $\Omega_C \neq \mathbb{1}$) if the loop C is not contractible to a point. For instance, on a 2-torus T^2 (the surface of a doughnut), see figure 9.12, the holonomy along the loop C_1 , which is not contractible, can be non-trivial, whereas the holonomy along the loop C_2 , which is continuously contractible to a point, can be shown to be trivial.

Vocabulary:

if the **curvature vanishes**, $\mathcal{F}_{\mu\nu} = 0$, we say that \mathcal{A}_μ is a **flat connection**, or equivalently that the bundle E is flat. The holonomy of a flat connection is called **monodromy**.

There is a lot more that can be said, but this will be left to future courses. I'll conclude this chapter by summarizing the correspondence between the geometry of fibre bundles and the formulation of gauge theories in physics:

Geometry	Physics
Principal G -bundle P	Gauge symmetry G
Connection \mathcal{A}_μ of P	Gauge field (or vector potential) A_μ
Curvature $\mathcal{F}_{\mu\nu}$ of P	Field strength $F_{\mu\nu}$
(Section of) Associated vector bundle	Charged field
Covariant derivative ∇_μ	Gauge covariant derivative D_μ
Parallel transport	Wilson line
Trace of the holonomy	Wilson loop

Bibliography

- Anderson, P. W. (1963). Plasmons, gauge invariance, and mass. *Phys. Rev.*, 130:439–442.
- Argyres, P. C. (2001). An introduction to global supersymmetry. <https://homepages.uc.edu/~argyrepc/cu661-gr-SUSY/susy2001.pdf>. [Online; accessed 8-January-2022].
- Bogomol’nyi, E. B. (1976). Stability of classical solutions. *Yad. Fiz.*
- Cahn, R. (2014). *Semi-Simple Lie Algebras and Their Representations*. Dover Publications.
- Derrick, G. H. (1964). Comments on Nonlinear Wave Equations as Models for Elementary Particles. *Journal of Mathematical Physics*, 5(9):1252–1254.
- Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field,. *Proc. Roy. Soc. Lond. A*, 133(821):60–72.
- Eguchi, T., Gilkey, P. B., and Hanson, A. J. (1980). Gravitation, Gauge Theories and Differential Geometry. *Phys. Rept.*, 66:213.
- Englert, F. and Brout, R. (1964). Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323.
- Georgi, H. and Glashow, S. L. (1972). Unified weak and electromagnetic interactions without neutral currents. *Phys. Rev. Lett.*, 28:1494.
- Goldstone, J. (1961). Field Theories with Superconductor Solutions. *Nuovo Cim.*, 19:154–164.
- Griffiths, D. (2014). *Introduction to Electrodynamics*. Pearson Education.

- Guralnik, G. S., Hagen, C. R., and Kibble, T. W. B. (1964). Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587.
- Harris, W., Fulton, W., and Harris, J. (1991). *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York.
- Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509.
- Hooft, G. (1974). Magnetic monopoles in unified gauge theories. *Nuclear Physics B*, 79(2):276–284.
- Jackson, J. (2021). *Classical Electrodynamics*. John Wiley & Sons, Limited.
- Landau, L. and Lifshitz, E. (2013). *The Classical Theory of Fields*. Course of Theoretical Physics. Elsevier Science.
- Manton, N. and Sutcliffe, P. (2004). *Topological Solitons*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge.
- Naber, G. L. and Naber, G. L. (1997). *Topology, geometry, and gauge fields*. Springer.
- Nakahara, M. (2003). *Geometry, topology and physics*. Taylor and Francis.
- Nambu, Y. (1960). Quasi-Particles and Gauge Invariance in the Theory of Superconductivity. *Physical Review*, 117(3):648–663.
- Ooguri, H. (2010). Mathematics for theoretical physicists. https://ocw.u-tokyo.ac.jp/course_11330/?interface_language=en. [Online; accessed 20-January-2022].
- Peskin, M. (1995). *An introduction to quantum field theory*. Perseus Books.
- Polyakov, A. M. (1974). Particle Spectrum in Quantum Field Theory. *JETP Lett.*, 20:194–195.
- Prasad, M. K. and Sommerfield, C. M. (1975). Exact classical solution for the 't hooft monopole and the julia-zee dyon. *Phys. Rev. Lett.*, 35:760–762.

- Tong, D. (2006). Quantum field theory. <https://www.damtp.cam.ac.uk/user/tong/qft/qft.pdf>. [Online; accessed 4-January-2022].
- Tong, D. (2018). Gauge theory. <https://www.damtp.cam.ac.uk/user/tong/gaugetheory/gt.pdf>. [Online; accessed 4-January-2022].
- Weinberg, E. (2012). *Classical Solutions in Quantum Field Theory: Solitons and Instantons in High Energy Physics*. Cambridge Monographs on Mathematical Physics. Cambridge University Press.
- Wu, T. T. and Yang, C.-N. (1975). Some Remarks About Unquantized Nonabelian Gauge Fields. *Phys. Rev. D*, 12:3843–3844.
- Wu, T. T. and Yang, C. N. (1976a). Dirac Monopole Without Strings: Monopole Harmonics. *Nucl. Phys. B*, 107:365.
- Wu, T. T. and Yang, C. N. (1976b). Dirac's Monopole Without Strings: Classical Lagrangian Theory. *Phys. Rev. D*, 14:437–445.
- Yang, C.-N. and Mills, R. L. (1954). Conservation of isotopic spin and isotopic gauge invariance. *Physical review*, 96(1):191.