Symmetries in Classical Field Theory: A Brief Overview

Abstract

In this pedagogical review we introduce classical relativistic field theory, Noether's theorem, and continuous symmetries, illustrating the formalism with examples including real and complex scalar fields, Lorentz and internal $U(1)$ transformations, Maxwell and Proca theory, and simple scale-invariant models.

1 Introduction

The aim of this note is to serve as an introduction to classical field theory. We begin with an overview of classical fields themselves. We then derive Noether's theorem, and finally, we look at some applied examples.

A quick note on conventions within this text: we opt to use natural units such that $c = 1 = \hbar$, and we choose our spacetime manifold to be Minkowski space $\mathbb{R}^{1,3}$ with metric signature $(- + + +)$, i.e. $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$ (in opposition to the particle physicists' choice). We also denote a global inertial coordinate system via $x^\mu = (t, \mathbf{x})$. A consequence of this "mostly plus" convention, which we will track carefully throughout, is that the kinetic terms of bosonic Lagrangians carry an overall minus sign. For example, the free scalar reads $\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m^2\phi^2$, so that the energy density comes out positive.

These notes were inspired by the lecture course that the author took at Cambridge, as well as the excellent textbook by Matthew D. Schwartz.

2 What is a Field?

The prototypical examples of classical relativistic field theories are Maxwell's Electromagnetism and Gravitation (note that Newton's theory can be described in terms of fields, but it is not relativistic). Maxwell's theory of electromagnetism is actually the first example of a relativistic field theory, as it is naturally Lorentz covariant. In order to introduce the theory properly, we must define what we mean by a field. In modern physics, they are often best understood as a smooth section of a bundle.

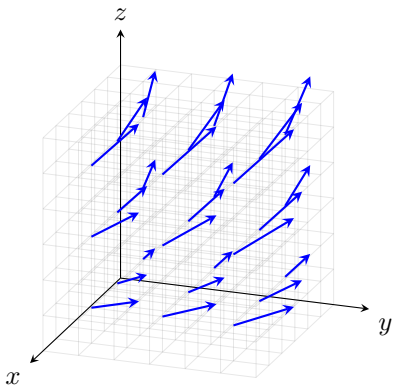


Figure 1: A sample vector field $F(x, y, z) = (-y, x, z)$ in $\mathbb{R}^3$.

Definition 2.1 (Fibre bundle)**.** *A fibre bundle with typical fibre F over a manifold M is a quadruple*

$$(E, M, \pi, F),$$

where E is also a smooth manifold, $\pi : E \to M$ is a smooth surjection, and for every $x \in M$ there exists an open neighbourhood $U \subseteq M$ of x and a diffeomorphism (a local trivialisation*)*

$$\varphi : \pi^{-1}(U) \longrightarrow U \times F$$

such that

$$\mathrm{pr}_1 \circ \varphi \;=\; \pi\big|_{\pi^{-1}(U)},$$

with

$$\mathrm{pr}_1 : U \times F \to U\,.$$

The fibre over x is $E_x := \pi^{-1}(x)$.

The fibre of a fibre bundle E_x can be locally trivialised to product space

$$E_x \cong \{x\} \times F\,.$$

With this, a smooth section is then the object formed from allowing x to vary smoothly. We can make this more formal via the following:

Definition 2.2. *Given a fibre bundle (E, M, π, F), and a local trivialising chart (U, φ) containing $x \in M$, a (local) smooth section is the map*

$$f : U \to E\,,$$

such that $\pi(f(x)) = x$ for all $x \in U$. Locally, we can trivialise with φ and construct the map

$$\varphi \circ f : U \to U \times F\,,$$

often written as

$$(\varphi \circ f)(x) = (x, \sigma(x)),$$

for some map σ. We can then identify the section f with the map $\sigma : U \to F$.

It is best to solidify these concepts with some examples. One common example is the temperature distribution in a room (at a fixed time). It can be represented by a scalar field, which is a map $T : \mathbb{R}^3 \to \mathbb{R}$. In that case, it is a smooth section of the (trivial) real line bundle. Similarly, the flow of a fluid field can be represented by a vector field, which would then be the smooth section of the tangent bundle: a map $X : \mathbb{R}^3 \to T\mathbb{R}^3$, where $T\mathbb{R}^3$ is the tangent bundle associated with $\mathbb{R}^3$, with $\pi : T\mathbb{R}^3 \to \mathbb{R}^3$ such that $\pi \circ X = \mathrm{id}_{\mathbb{R}^3}$. A visual example of such a flow can be seen in Figure 1. In this example, we take $M = \mathbb{R}^3$ and $T\mathbb{R}^3 \cong \mathbb{R}^3 \times \mathbb{R}^3$. The vector field is then identified as the map $F : \mathbb{R}^3 \to \mathbb{R}^3$. This rigorous definition of a field has many advantages: for instance, it allows us to generalise easily to different kinds of fields, and it also makes the later discussion of quantisation more systematic (although we will not pursue those details here).

For our purposes, however, this full definition is not often required. We will identify a field (in the scalar field theory case) as the map $\phi : M \to \mathbb{R}/\mathbb{C}$, which is a smooth section of the real or complex line bundle over M (analogous to the temperature distribution). However, when we turn to classical Maxwell theory, we will have to discuss what we mean by the field more carefully. We now review some aspects of Lagrangian mechanics.

A remark on style: the treatment of what a field is, as a smooth section of a fibre bundle (as seen above), is deliberately more formal than the remainder of these notes. As a student the author found the usual definition of a field unsatisfyingly vague, and so sets it on firmer ground here; once that picture is in place, later sections adopt the lighter, coordinate-based level of rigour customary for a first course.

2.1 Lagrangian mechanics

You are likely familiar with the Lagrangian for a point particle, usually introduced as

$$L = T - V,$$

where T is the kinetic energy and V is the potential energy. In field theory, however, the more fundamental object is the *Lagrangian density*.

Definition 2.3 (Lagrangian density)**.** *Given a field ϕ on a spacetime (M, g), a (classical) Lagrangian density is a function*

$$\mathcal{L} = \mathcal{L}\big(x, \phi(x), \partial\phi(x), \dots\big), \qquad x \in M,$$

which is a scalar under coordinate transformations.

The Lagrangian density allows one to define the *classical action* of a theory. It encodes the dynamics of the theory and can be defined as follows:

Definition 2.4 (Classical action)**.** *Given a field ϕ on a spacetime (M, g) the classical action is the functional*

$$S[\phi] \;=\; \int_M \mathrm{d}^n x \,\sqrt{|g(x)|}\,\mathcal{L}\big(x, \phi, \partial\phi, \dots\big)\,,$$

where $g = \det(g_{\mu\nu})$. The factor $\sqrt{|g|}$ provides the invariant volume element, ensuring that the action $S[\phi]$ is independent of the choice of coordinates on M.

As we take our spacetime manifold to be Minkowski space, it is trivial to notice that

$$|\eta| = 1\,.$$

Therefore, it is often omitted in QFT courses unless explicitly investigating curved spacetimes. As stated, the action encodes the dynamics of the theory; this is achieved via the *stationary action principle*. The stationary action principle states that the path a particle takes between two fixed points in configuration space is one that extremises the action. The same idea applies in the context of fields: instead of particle trajectories we now speak about field configurations, and the physical evolution between two configurations is determined by requiring that the action functional be stationary.

We shall now make this explanation less informal by first considering the trajectory of a point particle and then generalising to the case of a field.

2.1.1 Trajectory of a point particle

Consider a point particle in one spatial dimension, with position $q(t)$ and Lagrangian

$$L = L\big(q(t), \dot{q}(t), t\big)\,,$$

where $\dot{q}$ denotes the derivative of the particle's position with respect to time. The action functional for this particle is

$$S[q] = \int_{t_1}^{t_2} \mathrm{d}t\, L\big(q(t), \dot{q}(t), t\big)\,.$$

Now we suppose that the particle's position is perturbed,

$$q(t) \to q(t) + \epsilon\eta(t)\,,$$

with $\epsilon \ll 1$ and $\eta(t)$ compactly supported on (t_1, t_2). With this, the Lagrangian is varied via

$$L \to L + \epsilon\eta(t)\frac{\partial L}{\partial q} + \epsilon\frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} + \mathcal{O}(\epsilon^2)\,.$$

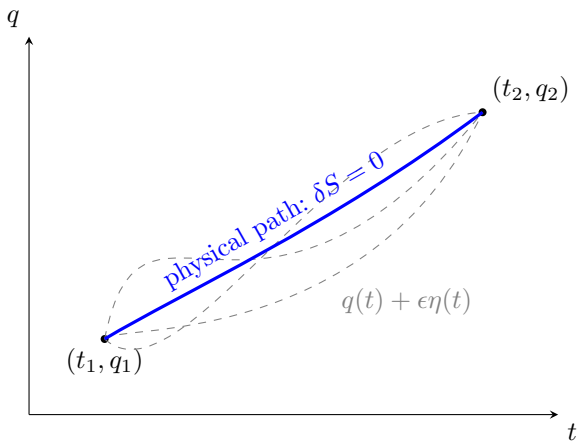


Figure 2: Many possible paths between fixed endpoints; the physical path is the one for which the action is stationary (minimal in simple cases).

Thus, the action becomes

$$\begin{aligned} S[q+\epsilon\eta] &= \int_{t_1}^{t_2} \mathrm{d}t \, \left(L + \epsilon\eta(t)\frac{\partial L}{\partial q} + \epsilon\frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} \right) \\ &= \int_{t_1}^{t_2} \mathrm{d}t \, L + \int_{t_1}^{t_2} \mathrm{d}t \, \left(\epsilon\eta(t)\frac{\partial L}{\partial q} + \epsilon\frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} \right) \\ &= S + \delta S \,, \end{aligned}$$

where

$$\delta S := \int_{t_1}^{t_2} \mathrm{d}t \, \left(\epsilon\eta(t)\frac{\partial L}{\partial q} + \epsilon\frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} \right) .$$

In order to extremise the action, we demand that $\delta S = 0$ (analogous to finding the extrema of a function);

$$\begin{aligned} \delta S &= \int_{t_1}^{t_2} \mathrm{d}t \, \left(\epsilon\eta(t)\frac{\partial L}{\partial q} + \epsilon\frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} \right) = 0 \\ &= \int_{t_1}^{t_2} \mathrm{d}t \, \epsilon\eta(t)\frac{\partial L}{\partial q} + \int_{t_1}^{t_2} \mathrm{d}t \, \epsilon\frac{\mathrm{d}}{\mathrm{d}t}\left(\eta(t)\frac{\partial L}{\partial \dot{q}} \right) - \int_{t_1}^{t_2} \mathrm{d}t \, \epsilon\eta(t)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}} \right) = 0 \\ &= \int_{t_1}^{t_2} \mathrm{d}t \, \epsilon\eta(t) \left(\frac{\partial L}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}} \right) \right) = 0 \,. \end{aligned}$$

To achieve the last line, we used the fundamental theorem of calculus and the fact η is compactly supported. The variation of S is required to be zero, and since η was entirely arbitrary, this must hold for all such η; thus, we find

$$\boxed{\frac{\partial L}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}} \right) = 0 \,.}$$

This is the *Euler-Lagrange* (E-L) equation ; it is boxed to emphasise its importance. This equation gives the equation of motion for the particle, i.e. its trajectory. A visual example of this can be seen in Figure 2.

Example: Assume we have a free particle of mass m and a Lagrangian

$$L = \frac{1}{2} m \dot{q}^2 \,,$$

then, the E-L equation yields,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \left(m \dot{q}(t) \right) = 0 \\ \implies &\; m \ddot{q}(t) = 0 \,. \end{aligned}$$

This is nothing but Newton's 2$^{\text{nd}}$ law, where the total force is 0. Solving this equation, one finds that the trajectory of a particle is a linear function

$$q(t) = C_1 t + C_2 \,,$$

where C_1, C_2 are typically fixed with either Neumann or Dirichlet boundary conditions.

2.1.2 The Euler-Lagrange equations for a field

The preceding argument extends naturally to fields. The configuration space is now infinite-dimensional: each point in this space corresponds to a complete field configuration $\phi(x)$ (for all x in spacetime). [1] Following essentially the same argument, one finds that a field $\phi(x)$ with Lagrangian density $\mathcal{L}(\phi, \partial\phi)$ satisfies the Euler–Lagrange equation

$$\boxed{\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) = 0 \,,}$$

where $\partial_\mu := \frac{\partial}{\partial x^\mu}$ and repeated indices μ are summed over. In global inertial coordinates $x^\mu = (t, \mathbf{x})$, one may write $\partial_\mu = (\partial_t, \nabla)$. If a field satisfies the E-L equations it is said to be *on-shell*.

Example: Consider the Lagrangian density of a *massive scalar field*

$$\mathcal{L} = -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 \,,$$

[1] This viewpoint is closely related to the idea of "theory space" in renormalisation group (RG) flows, where each point represents a different field theory, parametrised by its couplings. The RG flow can then be viewed as a vector field on this space, flowing towards fixed points.

where, in keeping with our mostly-plus signature, the kinetic term carries an overall minus sign so that the energy is bounded below. Using the E-L equations,

$$\begin{aligned} -m^2\phi + \partial_\mu(\partial^\mu\phi) &= 0 \\ \implies \left(\partial_\mu\partial^\mu - m^2\right)\phi &= 0\,. \end{aligned}$$

The final line is nothing but the *Klein-Gordon* equation: the relativistic wave equation. Writing $\Box \equiv \partial_\mu\partial^\mu = -\partial_t^2 + \nabla^2$ and inserting a plane-wave solution $\phi \sim e^{i(\mathbf{p}\cdot\mathbf{x} - Et)}$, we read off the dispersion relation

$$E^2 = \mathbf{p}^2 + m^2\,,$$

the relativistic energy–momentum relation for a particle of mass m. (Had we instead kept the "mostly minus" Lagrangian $+\frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m^2\phi^2$ together with the signature $(-+++)$, the kinetic term would have the wrong sign and the dispersion would come out tachyonic; the overall minus is precisely what makes the two conventions consistent.)

2.1.3 Lorentz invariance

Our field theories had better be Lorentz invariant. To encode this, we require the action functional to be invariant under a Lorentz transformation. To understand this requirement, consider, for example, the worldline of a particle in spacetime, as described by the Euler-Lagrange equations (derived from the action). It is a geometric curve that does not depend on which inertial observer is describing it. The coordinate representation of this curve, $q(t)$, will, of course, look different to different observers: the perceived position of the object may vary depending on the observer's location and velocity. However, the underlying curve in spacetime is the same; only the coordinates used to describe it change.

2.1.4 The Lorentz group

A Lorentz transformation is one that preserves the invariant line element (or quadratic form in Minkowski space). Under the transformation $x^\mu \to x'^\mu = \Lambda^\mu_\nu x^\nu$, we have

$$\begin{aligned} \mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x'^\mu\mathrm{d}x'^\nu &= \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \\ &= \Lambda^\mu_\rho\eta_{\mu\nu}\Lambda^\nu_\sigma\mathrm{d}x^\rho\mathrm{d}x^\sigma\,, \end{aligned}$$

for the equality to hold, we require

$$\Lambda^\intercal\eta\Lambda = \eta\,.$$

We then find

$$\begin{aligned} (\det(\Lambda))^2 &= 1 \\ \implies \det(\Lambda) &= \pm 1\,. \end{aligned}$$

The Lorentz group with $\det(\Lambda) = 1$ is technically the pseudo-orthogonal group $SO(1,3)$. Note that the full group manifold for the Lorentz group is non-compact, with several disconnected components (proper/improper, time-reversal etc). The defining relation forces $|\Lambda^0{}_0| \geq 1$, and we often additionally require that

$$\Lambda^0{}_0 \geq 1\,,$$

which corresponds to time "pointing forwards"; this is called the proper orthochronous Lorentz group, and denoted by $SO^+(1,3)$.

2.1.5 Viable Lagrangians

The fact that our action must be Lorentz invariant is very powerful. This is because it constrains the viable Lagrangian densities; overall, each term must transform as a Lorentz scalar. For example, consider the (schematic) Lagrangian density

$$\mathcal{L} = \partial_\mu \phi + m^2 \phi^2\,.$$

This is, of course, drawn from a completely general local dependence on x with arbitrary non-linear terms, i.e.,

$$\mathcal{L}(x, \phi, \partial\phi, \ldots, \partial^k \phi) = F\Big(x, \phi, \big(\partial^\alpha \phi(x)\big)_{|\alpha| \leq k}\Big)\,,$$

with

$$\partial^\alpha \phi := \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} \ldots \partial^{\alpha_n}} \phi\,,$$

$k \in \mathbb{N}$, $\alpha = (\alpha_1, \cdots, \alpha_n) \in \mathbb{N}^n$, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and F some arbitrary smooth function. The above extends to arbitrary (but finite) k. The first term above is not even a Lorentz scalar, it carries a free index μ, but it serves to illustrate the point. Under a Lorentz transformation

$$x^\mu \to x'^\mu = \Lambda^\mu_\nu x^\nu\,,$$

the field transforms as

$$\phi \to \phi(\Lambda^{-1} x), \quad \partial_\mu \phi \to \Lambda_\mu{}^\nu \partial_\nu \phi\,,$$

thus, our Lagrangian transforms via

$$\mathcal{L} \to \mathcal{L}' = \Lambda_\mu{}^\nu \partial_\nu \phi + m^2 \phi^2\,.$$

The first term transforms as a vector component, while $m^2\phi^2$ is a scalar[2]; their sum does not transform as a scalar, so $\mathcal{L}$ is not Lorentz invariant. However, look at the similar Lagrangian density

$$\mathcal{L} = -\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2\,,$$

[2]This is described in terms of representations of the Lorentz group. For example, the derivative term transforms under the "vector representation of the group," while the field transforms under the "trivial representation". This subject is slightly beyond the scope of this text; however, we strongly suggest that the reader read further into the subject.

which is (up to the overall normalisation) the typical Lagrangian of a massive scalar field. It transforms as

$$\begin{aligned}\mathcal{L} \to \mathcal{L}' &= -\Lambda_\mu{}^\nu \Lambda^\mu{}_\rho \eta^{\tau\rho} \partial_\nu \phi \partial_\tau \phi - m^2 \phi^2 \\ &= -\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2 \\ &= \mathcal{L}\,,\end{aligned}$$

where we used the defining relation for a Lorentz transformation. Thus, this does indeed transform as a scalar and is a viable Lagrangian. In general, viable Lagrangians are ones with indices fully contracted with invariant tensors (such as $\eta_{\mu\nu}$ or $\epsilon_{\mu\nu\rho\sigma}$), so that each term in $\mathcal{L}$ is a scalar built from the fields and their derivatives. One often also demands spacetime translation invariance, which forbids any explicit dependence of $\mathcal{L}$ on the coordinates x^μ.

Let us now show that a general (viable) Lagrangian leads to a Lorentz invariant action. To do this, notice that since $\mathcal{L}$ is viable, it must transform as a scalar. Under a Lorentz transformation, the corresponding Jacobian is

$$J^\mu_\nu := \frac{\partial x'^\mu}{\partial x^\nu} = \Lambda^\mu_\nu\,,$$

with

$$|\det(J)| = |\det(\Lambda)| = 1\,,$$

thus, we have

$$S \to S' = \int \mathrm{d}^4 x' \,|\det(\Lambda)|\, \mathcal{L} = \int \mathrm{d}^4 x\, \mathcal{L} = S\,.$$

3 Noether's Theorem

Noether's theorem is arguably one of the most important theorems in theoretical physics. Before we write it down we must first define what we mean by a continuous symmetry:

Definition 3.1. *A continuous symmetry is a smooth, parameter-dependent transformation of the fields (and possibly of the spacetime coordinates) that leaves the action invariant, up to boundary terms.*

An example of this would be the Lorentz transformations discussed in the previous section. Infinitesimally, we may write

$$\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu + \mathcal{O}(\omega^2)\,,$$

where ω^μ_ν is small and satisfies

$$\omega_{\mu\nu} = -\omega_{\nu\mu}\,.$$

This matrix, as well as being anti-symmetric, has 6 independent components corresponding to 3 spatial rotations as well as 3 boosts. We shall return to this

example soon. It is useful to keep two independent distinctions in mind. First, a symmetry is *global* if its parameters are constant over spacetime, and *local* (or *gauged*) if they are allowed to depend on the spacetime point x. Second, a transformation may act only on the fields an *internal* symmetry (such as the $U(1)$ phase rotation we meet below) or also on the spacetime coordinates, a *spacetime* symmetry, such as the Lorentz transformations above. In the form stated here, Noether's (first) theorem applies to continuous global symmetries of either kind.

Noether's theorem (informally) states

Theorem 3.1. *Every continuous symmetry gives rise to a conserved current* j^μ *such that*

$$\partial_\mu j^\mu = 0\,.$$

Given suitable boundary conditions, each current gives rise to a conserved charge

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

Let us derive the conserved current from a general transformation. Begin with the infinitesimal transformation

$$\phi \to \phi' = \phi + \delta\phi\,,$$

under this change, we require

$$\delta S = S[\phi'] - S[\phi] = 0\,.$$

This requirement is equivalent to demanding

$$\begin{aligned}\delta\mathcal{L} &= \mathcal{L}' - \mathcal{L} \\ &= \mathcal{L} + \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu(\delta\phi) - \mathcal{L} \\ &= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu(\delta\phi) = \partial_\mu F^\mu\,,\end{aligned}$$

where F^μ is some smooth function of the fields and their derivatives. The last line is the requirement that the variation in the Lagrangian be at most a *surface term* (total derivative). This ensures that it vanishes in the action, given sufficient boundary conditions. We may rearrange the above and use the E-L (assuming the field is on-shell) equations to find

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi - F^\mu\right) = 0\,,$$

the term inside the brackets is defined to be the conserved current

$$j^\mu := \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi - F^\mu\,,$$

which clearly satisfies

$$\partial_\mu j^\mu = 0\,.$$

This allows us to define the conserved charge as stated in Noether's theorem. Let us show that it is indeed conserved; take the time derivative of the charge and notice

$$\begin{aligned}\frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial x^0} &= \int_V \mathrm{d}^3x\,\partial_0 j^0 \\ &= -\int_V \mathrm{d}^3x\,\partial_i j^i \\ &= -\int_{\partial V} \mathrm{d}S\,n_i j^i \\ &= 0\,,\end{aligned}$$

with $i = 1, 2, 3$, and n_i a unit normal vector. The last line is true, provided that the fields vanish at infinity or are compactly supported on the volume V.

An important case of conserved current is the *energy momentum* (E-M) tensor. It is the conserved current generated by infinitesimal spacetime translations of the form

$$x^\mu \to x^\mu + \epsilon^\mu\,,$$

and is defined as follows:

$$(j^\mu)_\nu = T^\mu_\nu := \partial_\nu\phi\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} - \delta^\mu_\nu\mathcal{L}\,,$$

such that

$$\partial_\mu T^\mu_\nu = 0\,.$$

This formalises the conservation of momentum and energy that we expect for matter. For common examples, it is symmetric, i.e.,

$$T^{\mu\nu} = T^{\nu\mu}\,.$$

There are cases where it is not symmetric however; in these cases we may define the *Belinfante–Rosenfeld tensor*

$$\Theta^{\mu\nu} = T^{\mu\nu} - \partial_\rho T^{\rho\mu\nu}\,,$$

such that $T^{\rho\mu\nu} = -T^{\mu\rho\nu}$ (i.e., an antisymmetric tensor), and

$$\partial_\mu\Theta^{\mu\nu} = 0\,.$$

We can always define such a tensor, as the E-M tensor, as defined in General relativity, is always symmetric. **Example:** Consider the infinitesimal Lorentz transformation,

$$x^\mu \to x^\mu + \omega^\mu_\nu x^\nu\,,$$

with the field transforming as

$$\phi(x) \to \phi(x) - \alpha\omega^\mu_\nu x^\nu \partial_\mu \phi \,,$$

using Taylor's theorem. This transformation yields the following transformation of the Lagrangian

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L}(\phi, \partial\phi) - \alpha\omega^\mu_\nu x^\nu \partial_\mu \phi \frac{\partial \mathcal{L}}{\partial \phi} - \alpha \partial_\rho (x^\mu \omega^\nu_\mu \partial_\nu \phi) \frac{\partial \mathcal{L}}{\partial(\partial_\rho \phi)} \,.$$

Then, the variation is

$$\begin{aligned} \delta\mathcal{L} &= -\alpha\omega^\mu_\nu x^\nu \left(\partial_\mu \phi \frac{\partial \mathcal{L}}{\partial \phi} + \partial_\rho \partial_\mu \phi \frac{\partial \mathcal{L}}{\partial(\partial_\rho \phi)} \right) \\ &= -\alpha\omega^\mu_\nu x^\nu \partial_\mu \mathcal{L} = -\alpha \partial_\mu (x^\nu \omega^\mu_\nu \mathcal{L}) \end{aligned}$$

using the anti-symmetry of ω and the chain rule on $\mathcal{L}$. This transformation gives rise to the current

$$\begin{aligned} j^\mu &= -\omega^\rho_\nu x^\nu \partial_\rho \phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} + \omega^\mu_\nu x^\nu \mathcal{L} \\ &= -\omega^\rho_\nu x^\nu \left(\underbrace{\partial_\rho \phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} - \delta^\mu_\rho \mathcal{L}}_{\text{E-M Tensor}} \right) . \end{aligned}$$

As stated, we can identify the expression within the brackets as the energy-momentum tensor. In our example, we find

$$j^\mu = -\omega^\rho_\nu x^\nu T^\mu_\rho \,.$$

Let us identify the conserved charge generated by the 3 spatial rotations. For rotations, we restrict ω to spatial indices $i, j = 1, 2, 3$, and notice that a general spatial rotation takes the form

$$\omega_{ij} = \epsilon_{ijk}\theta n_k \,, \;\; \omega_{0i} = -\omega_{i0} = 0 \,,$$

with θ small, and a unit vector n_k identifying the axis of rotation. As a quick aside, look at the case for a rotation about x^3. Take the unit vector to be $n_k = (0, 0, 1)$, yielding

$$\omega_{ij} = \theta\epsilon_{ij3} = \theta \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} ,$$

which, when we substitute into the expression for Λ, yields the correct infinitesimal rotation matrix around x_3.

With this expression for ω, the conserved charge is

$$\begin{aligned}
Q &= -\int \mathrm{d}^3x \, \left(\omega^0_\nu x^\nu T^0_0 + \omega^i_0 x^0 T^0_i + \omega^i_j x^j T^0_i\right) \\
&= -\int \mathrm{d}^3x \, \omega_{ij} x^j T^{0i} \\
&= -\frac{1}{2}\int \mathrm{d}^3x \, \left(\omega_{ij} x^i T^{0j} - \omega_{ji} x^j T^{0i}\right) \\
&= \frac{1}{2}\int d^3x \left(-\theta\epsilon_{ijk} x^i T^{0j} + \theta\epsilon_{jik} x^j T^{0i}\right) n_k \\
&= \frac{\theta}{2}\epsilon_{ijk}\int \mathrm{d}^3x \, \left(x^j T^{0k} - x^k T^{0j}\right) n_i \\
&= \theta Q_i n_i \,,
\end{aligned}$$

where

$$Q_i := \frac{1}{2}\epsilon_{ijk}\int \mathrm{d}^3x \, \left(x^j T^{0k} - x^k T^{0j}\right) \,.$$

Physically, this charge corresponds to the angular momentum charge, generated by a rotation through θ.

These charges are generated by rotations; we are ignoring 3 extra viable Lorentz transformations: boosts. An infinitesimal Lorentz boost takes the form

$$\omega_{i0} = \xi n_i, \quad \omega_{0i} = -\xi n_i, \quad \omega_{ij} = 0 \,, \quad \omega_{00} = 0$$

where ξn_i is an infinitesimal spatial vector representing the *rapidity*. Using the same calculations as before, we find

$$\begin{aligned}
Q &= -\int \mathrm{d}^3x \; \omega^0_i x^i T^0_0 \\
&= -\int \mathrm{d}^3x \; \xi n_i x^i T^{00} \\
\implies \xi Q^i n_i &= -\xi \int \mathrm{d}^3x \; x^i T^{00} \; n_i \\
\implies Q^i &:= -\int \mathrm{d}^3x \; x^i T^{00} \,.
\end{aligned}$$

Let us compute the time derivative of these charges,

$$\begin{aligned}\frac{\mathrm{d}Q^i}{\mathrm{d}t} &= -\int \mathrm{d}^3x \ \left(\frac{\partial x^i}{\partial x^0}T^{00} + x^i\frac{\partial T^{00}}{\partial x^0}\right)\\ &= \int \mathrm{d}^3x \ x^i\frac{\partial T^{j0}}{\partial x^j}\\ &= \underbrace{\int \mathrm{d}^3x \ \partial_j(x^iT^{j0})}_{=0} - \int \mathrm{d}^3x \ T^{i0}\\ &= -\int \mathrm{d}^3x \ T^{i0}\,,\end{aligned}$$

where we assumed x^iT^{j0} decays sufficiently fast at infinity. There are two important charges that arise due to the E-M tensor; these are the conserved energy and momentum

$$E := \int \mathrm{d}^3x \ T^{00}, \quad P^i := \int \mathrm{d}^3x \ T^{i0}\,.$$

With these, we then clearly identify

$$\frac{\mathrm{d}Q^i}{\mathrm{d}t} = -P^i\,.$$

Momentum conservation implies

$$\frac{\mathrm{d}P^i}{\mathrm{d}t} = 0\,,$$

hence P^i is constant and we may integrate to obtain

$$Q^i(t) = -P^it + Q^i(0)\,.$$

Physically, this implies that the boost "charge" is not conserved in the naive sense; rather, it evolves linearly in time, and represents the shifting centre of energy of the system that we expect between different inertial frames. (The genuinely conserved quantity is the combination $K^i = tP^i + \int \mathrm{d}^3x\, x^iT^{00}$, whose constancy is precisely the statement that the centre of energy moves uniformly.)

4 Complex Scalar Fields

The description of complex scalar fields is essentially the same as that of real scalar fields; however, our fields are now the maps $\psi : M \to \mathbb{C}$. Complex scalar fields also come with their own symmetries, which are not possessed by real scalar fields.

Take, for example, the Lagrangian density

$$\mathcal{L} = -\partial_\mu\psi^*\partial^\mu\psi - m^2\psi^*\psi\,,$$

this is essentially the same as the Lagrangian we looked at earlier, however, now we have two different fields

$$\psi, \quad \psi^* .$$

Inspired by complex numbers, if we take the pointwise "norm" of the complex scalar field to be

$$\begin{aligned} \|\psi\|^2 &:= (\psi, \psi) \\ &= \psi^*(x)\psi(x) . \end{aligned}$$

We could then write the Lagrangian density as

$$\mathcal{L} = -\|\partial\psi\|^2 - m^2\|\psi\|^2 ,$$

which looks almost identical to the real case.[3] It also shows that this Lagrangian is a Lorentz scalar (as required).

As stated, we treat the two fields as separate objects. This means that we actually obtain a set of *two* Euler-Lagrange equations for this field: we find (check yourself),

$$\begin{aligned} \partial_\mu \partial^\mu \psi^* - m^2 \psi^* &= 0 , \\ \partial_\mu \partial^\mu \psi - m^2 \psi &= 0 . \end{aligned}$$

One immediately notices that these are the complex analogues of the Klein-Gordon equation derived earlier (with the same relative sign $(\partial_\mu \partial^\mu - m^2)$); this is good, as the separate fields obey the same equation (as they are both scalar fields). When the Lagrangian is written in terms of the norm, it gives a hint as to what extra symmetries complex scalar fields possess. Recall that given a complex number z, we can define a new complex number via

$$z' = e^{i\theta} z , \quad \theta \in \mathbb{R} .$$

This corresponds to a rotation; to see this, write z in its polar form. Notice that the norm of the new vector is

$$\begin{aligned} \|z'\|^2 &= (z', z') \\ &= z^* e^{-i\theta} e^{i\theta} z \\ &= z^* z \\ &= \|z\|^2 . \end{aligned}$$

Thus, z and z' have the same norm; z' is simply the rotated version of z. The set of all rotations that preserve the norm in $\mathbb{C}$ actually forms a group[4]: *the*

[3] Here $\|\partial\psi\|^2$ is a slight abuse of notation for the spacetime-contracted bilinear $\partial_\mu \psi^* \partial^\mu \psi$; unlike the internal norm $\|\psi\|^2 = \psi^* \psi$ this is not positive-definite, since the Minkowski metric is indefinite.

[4] Once we introduce a binary group operation $\cdot : G \times G \to G$, given by $e^{i\theta_1} \cdot e^{i\theta_2} = e^{i(\theta_1 + \theta_2)}$.

unitary group $U(1)$. In general, the group $U(n)$ is the group of $n \times n$ matrices $U \in U(n)$ such that

$$U^\dagger U = I_n \,,$$

where I_n is the $n \times n$ identity matrix. For our case, $n = 1$, which corresponds to a complex scalar as seen above.

We can apply the same reasoning to fields: suppose our field picks up a *global phase*

$$\psi \to e^{i\alpha}\psi, \quad \alpha \ \in \mathbb{R} \,,$$

where due to the fact that $e^{i2\pi} = 1$, we have $\alpha \sim \alpha + 2\pi k, k \in \mathbb{Z}$. Note that this transformation corresponds to the 1-parameter subgroup $U(\alpha) : \mathbb{R} \to U(1)$, which can be thought of as a curve (more precisely, an arc) on our group manifold $U(1)$ (which is the circle S^1). The Lagrangian transforms as

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} \,.$$

Such a transformation is clearly a symmetry of this Lagrangian and the corresponding theory; this theory is said to have the symmetry group $U(1)$. From Noether's theorem, taking α to be infinitesimal, we find the corresponding conserved current

$$j^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)}\delta\psi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi^*)}\delta\psi^* \,,$$

where:

$$\delta\psi = i\alpha\psi$$
$$\delta\psi^* = -i\alpha\psi^* \,.$$

Using $\partial\mathcal{L}/\partial(\partial_\mu\psi) = -\partial^\mu\psi^*$ and $\partial\mathcal{L}/\partial(\partial_\mu\psi^*) = -\partial^\mu\psi$, and dropping the constant α (which does not affect conservation), the current can then be explicitly written as

$$j^\mu = i\left(\psi^*\,\partial^\mu\psi - \psi\,\partial^\mu\psi^*\right) \,.$$

(Its overall sign and normalisation are conventional, and would be fixed by the electric charge q once the field is coupled to electromagnetism.) This current is conserved on-shell; this can be checked directly using the field equations satisfied by ψ and ψ^*.

What happens if we take α to instead be local? This implies that the transformation takes the form

$$\psi \to e^{i\alpha(x)}\psi \,, \quad \alpha : \mathbb{R}^{1,3} \to \mathbb{R} \,.$$

The Lagrangian density now transforms as

$$\begin{aligned}
\mathcal{L} \to \mathcal{L}' &= -\partial_\mu\left(\psi^* e^{-i\alpha(x)}\right)\partial^\mu\left(e^{i\alpha(x)}\psi\right) - m^2\psi^*\psi \\
&= -\left(\partial_\mu\psi^* - i\,\partial_\mu\alpha(x)\,\psi^*\right)e^{-i\alpha(x)}\,\left(\partial^\mu\psi + i\,\partial^\mu\alpha(x)\,\psi\right)e^{i\alpha(x)} - m^2\psi^*\psi \\
&= -\left(\partial_\mu\psi^* - i\partial_\mu\alpha(x)\psi^*\right)\left(\partial^\mu\psi + i\partial^\mu\alpha(x)\psi\right) - m^2\psi^*\psi \,.
\end{aligned}$$

What we have found above implies that the Lagrangian is not invariant under such a transformation; however, we know that the theory itself admits $U(1)$ as a global symmetry group. There is a way to remedy this; we must introduce the *covariant derivative*.

4.1 Covariant Derivative

The covariant derivative, as the name suggests, transforms covariantly under group transformations; for example, under $\psi \to U(\alpha)\psi$ we would like our covariant derivative, denoted D_μ, to transform via

$$D_\mu \psi \to U(\alpha) D_\mu \psi \,.$$

Essentially, it transforms in the same way as the field itself. This is actually what defines the covariant derivative:

Definition 4.1. *Given a local $U(1)$ transformation, the covariant derivative is*

$$D_\mu := \partial_\mu + iA_\mu \,, \quad \textit{with } A_\mu \to A_\mu - \partial_\mu \alpha(x) \,,$$

such that

$$\begin{aligned} D_\mu \psi \to D'_\mu(U(\alpha)\psi) &= (\partial_\mu + iA_\mu - i\partial_\mu \alpha(x))(U(\alpha)\psi) \\ &= U(\alpha) i\partial_\mu \alpha(x)\psi + U(\alpha)\partial_\mu \psi + U(\alpha) iA_\mu \psi - U(\alpha) i\partial_\mu \alpha(x)\psi \\ &= U(\alpha)(\partial_\mu \psi + iA_\mu \psi) = U(\alpha) D_\mu \psi \,. \end{aligned}$$

The field A_μ is often referred to as the gauge field.

We are not being overly cautious with the above definition; we often also introduce a charge q, however, we have elected to set $q = 1$. The important point is that requiring our derivative transforms like the field itself forces us to introduce a new field (which is also dynamical with its own equations of motion) A_μ. This isn't just a neat mathematical trick, but an actual physically observable effect; one important example of this is the Aharonov–Bohm effect.

If you have done any GR, this construction and the definition of the covariant derivative will likely look familiar; there is a reason for this, which we shall cover briefly below[5]. Regardless, with this definition of the covariant derivative, we may write our Lagrangian in terms of the covariant derivative instead,

$$\mathcal{L} = -(D_\mu \psi)^* D^\mu \psi - m^2 \psi^* \psi \,.$$

Under a local $U(1)$ transformation, following the computations above,

$$\begin{aligned} \mathcal{L} \to \mathcal{L}' &= -(D_\mu \psi)^* e^{-i\alpha(x)} e^{i\alpha(x)} (D^\mu \psi) - m^2 \psi^* \psi \\ &= \mathcal{L} \,. \end{aligned}$$

[5]There is a beautiful geometric view of Gauge theories in terms of Principal G-bundles, which describes this structure and also relates closely to the mathematical formulation of general relativity.

Thus, the Lagrangian, expressed in terms of the covariant derivative, is invariant. This implies that our naïve guess of the form of the Lagrangian in terms of regular derivatives was the incorrect one, and that the true theory must be described in terms of this more geometric description. This observation leads to the study of Gauge theories. We elect not to cover gauge theories here, they are best suited to a second course in QFT.

5 Maxwell and Proca Theory

Maxwell theory is likely well known to you by this point; for completeness, however, we shall state the Maxwell equations in their familiar differential form (for our purposes the integral formulation will be less insightful):

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad \text{(Gauss)} \qquad \nabla \cdot \mathbf{B} = 0 \quad \text{(Monopoles)}$$
$$-\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{E} \quad \text{(Faraday)} \qquad \nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \quad \text{(Ampère)} \,.$$

These equations have a very nice geometric interpretation in terms of sources and sinks, where a source can be thought of as a positive charge with field lines flowing from it, and a sink can be viewed as a negative charge, into which the field lines flow. It is well known that Maxwell theory is a gauge theory; the theory can be naturally described in terms of a vector gauge potential $\mathbf{A}$ and scalar gauge potential ϕ, such that

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

This freedom in our choices of $\mathbf{A}$ and ϕ means that there are many choices of gauge, with the Lorenz and Coulomb gauges being most common. The quantities $\mathbf{A}$ and ϕ can be naturally combined into a contravariant 4-vector (gauge) potential

$$A^\mu = (\phi, \mathbf{A}) \,,$$

whose covariant counterpart, in our mostly-plus signature, is $A_\mu = \eta_{\mu\nu} A^\nu = (-\phi, \mathbf{A})$. We now return to our original formal definition of a field in terms of bundles. As we saw, a scalar field is described in terms of a smooth section of the trivial (real or complex) line bundle over our spacetime manifold M. The quantity we have just defined (i.e. the gauge field 4-vector) is best described as a *Lie-algebra-valued one-form*. This sounds horrifying, yes, but it is a very natural and geometric viewpoint once de-mystified. We will only introduce this conceptually in order to not get lost in the mathematical weeds. In terms of the bundle description, our 4-vector potential is described by a connection on a (principal) G-bundle, such that the fibre looks like

$$E_x \cong \{x\} \times G \,,$$

which assigns to each point $x \in M$ a copy of the Lie group[6] G. The point of this is to encode gauge redundancy mathematically. Recall that in Maxwell theory, we have a redundancy in our 4-potential description; we can always add a smooth function φ such that

$$A_\mu \to A_\mu + \partial_\mu \varphi \,,$$

with $\partial_\mu := (\partial_t, \nabla)$, that leaves the observables, the $\mathbf{E}$ and $\mathbf{B}$ fields, unchanged. This redundancy can be formalised by our G-bundle. This is because the G-bundle, by definition, does not have a preferred identity. When we then choose a local section $f : U \to E$, which is essentially choosing a gauge, we may always take fg with $g \in G$ (the gauge group). This geometric picture describes exactly our gauge redundancy within gauge theories.

The Lie algebra of G, $\mathfrak{g}$, is the linearised version of G, imagine taking the derivative at the identity. Our object A naturally takes values in this linear space, meaning it outputs Lie algebra elements. In the geometric picture, A is what is called a *connection*. This is a quantity that allows us to compare derivatives at different points on a manifold. The problem is simple: if we have a vector at one point (typically a tangent vector to some curve on the manifold), we cannot directly compare it to a vector at a different point, since they live in different vector spaces. The connection tells us how to relate these vectors at different points via *parallel transport*. Parallel transport works as follows: we take a tangent vector and transport it along a curve on the manifold. If we transport the vector around a closed loop, we can then compare it to the original vector. In general, the transported vector will differ from the original; this difference is what the connection encodes. Consider, for example, a tangent vector on a sphere. If we parallel transport this vector around a closed curve, as illustrated in Figure 3, the vector is rotated relative to its initial direction. This rotation, which depends on the path taken, is precisely what the connection describes. Technically the last example is the holonomy determined by the connection, which is the way of describing this difference along closed loops specifically. For the sphere example, the holonomy group is the set of all linear maps that represent parallel transport around closed curves based at the point x, denoted

$$\mathrm{Hol}_x = \{P_\gamma : T_x S^2 \to T_x S^2 \mid \gamma : I \subset \mathbb{R} \to S^2 \text{ is a closed curve based at } x\} \,.$$

This group (one can see it is a group with multiplication given by composition) is isomorphic to $SO(2)$, since parallel transport preserves the metric and orientation, so the transformations are rotations (determinant $+1$) rather than reflections; hence why it is $SO(2)$, and not $O(2)$. We shall not go too much deeper into this topic as it easily deserves its own set of rigorous notes. We give these facts in order to provide a sense of the geometric formalism. Finally, A being a one-form (a differential form which acts on vectors as a linear map) is

[6] this is a group that is also a smooth manifold; group operations, as well as their inverses, are smooth maps.

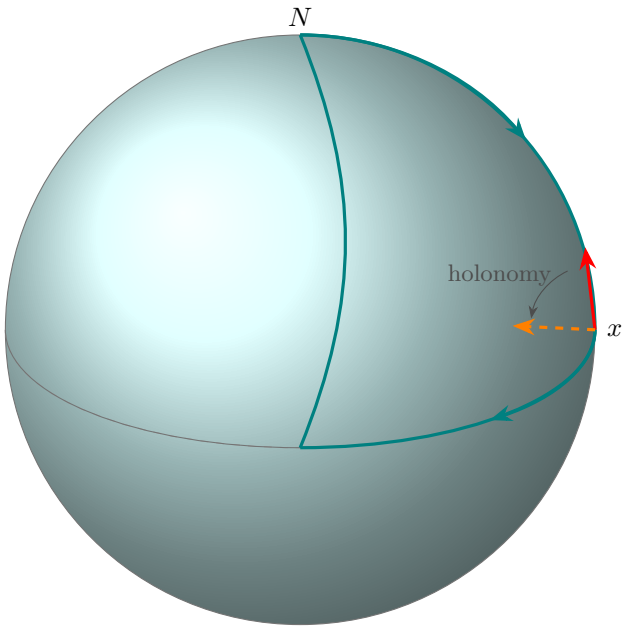


Figure 3: Parallel transport of a tangent vector (red) around a geodesic triangle on the sphere S^2. Carrying the vector along $x \to N \to B \to x$ returns it rotated (orange, dashed) relative to its starting orientation; the rotation angle is the *holonomy* of the Levi-Civita connection and equals the enclosed solid angle. This path dependence is exactly the information stored in a connection.

intimately tied to the geometry of the manifold. For our case, simply keep in mind that it is a map

$$A : T_x M \to \mathfrak{g}\,,$$

with $x \in M$.

5.1 Maxwell Lagrangian

With that brief overview of how one actually defines vector fields in Maxwell theory, we now turn to investigating the Maxwell Lagrangian along with its symmetries. The Maxwell Lagrangian is defined in terms of the field strength tensor[7]

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

If you continue to look at gauge theory there will be a slightly different version of this equation. This is the non-abelian case, which includes a bracket term, the Lie bracket of the Lie algebra of the gauge group G. Since, for Maxwell theory, $G = U(1)$, which is abelian, we ignore this term as it vanishes. The field strength is an anti-symmetric quantity (check) with $F_{\mu\nu} = -F_{\nu\mu}$, with components given by the E and B field components. In our mostly-plus conventions, with $A_\mu = (-\phi, \mathbf{A})$, one finds $F_{i0} = E_i$ and $F_{ij} = \epsilon_{ijk} B_k$, i.e.

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

(Readers using the opposite, "mostly minus" metric will find the overall sign of this matrix flipped; the physical fields $\mathbf{E}, \mathbf{B}$ are of course convention independent.) This field strength naturally encodes the *homogeneous* half of Maxwell's equations via the *Bianchi identity*, which states

$$\partial_\gamma F_{\mu\nu} + \partial_\mu F_{\nu\gamma} + \partial_\nu F_{\gamma\mu} = 0\,;$$

we believe it would be very insightful if you checked that this single identity reproduces Faraday's law and the absence of magnetic monopoles, $\nabla \cdot \mathbf{B} = 0$. It is rather enlightening. (The remaining, *inhomogeneous* pair are the Gauss and Ampère laws. These are not identities but genuine equations of motion, derived below.) Regardless, with this tensor defined we may write the free Maxwell Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}\,,$$

notice that, via our previous discussion, this is a viable Lagrangian as it is Lorentz invariant. (A short computation gives $F_{\mu\nu}F^{\mu\nu} = 2(\mathbf{B}^2 - \mathbf{E}^2)$, so that $\mathcal{L} = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2)$, with the correct positive electric kinetic energy.) The factor of $1/4$ is a convention. Let us inspect how our gauge transformation above affects

[7] Which is actually the curvature associated with the connection form discussed above.

this Lagrangian (it should hopefully be completely gauge invariant).

Example: Take the general gauge transformation as above

$$A_\mu \to A_\mu + \partial_\mu \xi \,,$$

with ξ some smooth function. With this, it is clear that

$$F_{\mu\nu} \to F'_{\mu\nu} = F_{\mu\nu} + \partial_\mu \partial_\nu \xi - \partial_\nu \partial_\mu \xi = F_{\mu\nu} \,,$$

which implies that the Lagrangian is invariant.

5.1.1 The Maxwell field equations

So far we have only built the action; the dynamics follow from the Euler–Lagrange equations for A_μ. Coupling the field to an external current J^μ through the standard minimal term, the Lagrangian is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu \,.$$

A short computation using $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha$ gives the canonical conjugate momentum

$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = -\tfrac{1}{2} F^{\alpha\beta} \left(\delta^\mu_\alpha \delta^\nu_\beta - \delta^\mu_\beta \delta^\nu_\alpha \right) = -F^{\mu\nu} \,,$$

while $\partial \mathcal{L} / \partial A_\nu = -J^\nu$. The Euler–Lagrange equation

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

therefore yields the inhomogeneous Maxwell equations in covariant form,

$$\boxed{\partial_\mu F^{\mu\nu} = J^\nu \,.}$$

Writing out the components reproduces Gauss's law (for $\nu = 0$) and the Ampère–Maxwell law (for $\nu = i$), exactly the two equations that contain the sources, while the homogeneous pair is supplied automatically by the Bianchi identity above. The action principle thus delivers precisely the half of Maxwell's equations that the geometry did not.

Two structural features deserve emphasis. First, the matter coupling is gauge invariant only up to a boundary term: under $A_\mu \to A_\mu + \partial_\mu \alpha$,

$$J^\mu A_\mu \;\longrightarrow\; J^\mu A_\mu + J^\mu \partial_\mu \alpha = J^\mu A_\mu + \partial_\mu (J^\mu \alpha) - (\partial_\mu J^\mu)\, \alpha \,,$$

so the action is invariant for *arbitrary* $\alpha(x)$ if and only if the current is conserved, $\partial_\mu J^\mu = 0$. Gauge symmetry and charge conservation are thus two faces of the same structure; equivalently, conservation is the consistency condition of the field equations, since $\partial_\nu \partial_\mu F^{\mu\nu} \equiv 0$. Second, the time component A_0 is

non-dynamical: no $\dot{A}_0$ appears in $\mathcal{L}$ (because $F_{00} = 0$), so its Euler–Lagrange equation is not an evolution equation but the constraint $\partial_i F^{i0} = J^0$, i.e. Gauss's law. Together with the one-parameter gauge freedom, this removes two of the four components of A_μ, leaving the $\boxed{2}$ transverse polarisations of the photon. We will see that switching on a mass changes this counting.

5.1.2 The energy–momentum tensor

Let us now investigate the energy–momentum tensor of this theory. Under an infinitesimal translation $x^\mu \mapsto x^\mu + \epsilon^\mu$ the Noether (canonical) tensor is

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\rho)}\,\partial_\nu A_\rho - \delta^\mu{}_\nu\,\mathcal{L}.$$

Recalling $\partial\mathcal{L}/\partial(\partial_\mu A_\rho) = -F^{\mu\rho}$ from above, we obtain

$$T^{\mu\nu} = -F^{\mu\rho}\partial^\nu A_\rho + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}, \tag{1}$$

where we raised the index using the Minkowski metric η. The important point to notice about this tensor is that it is *not* symmetric. To see this, notice that the antisymmetric part is

$$T^{\mu\nu} - T^{\nu\mu} = -F^{\mu\rho}\partial^\nu A_\rho + F^{\nu\rho}\partial^\mu A_\rho, \tag{2}$$

which does not vanish in general, so $T^{\mu\nu} \neq T^{\nu\mu}$. As well as this it is *not* gauge invariant. For example, under the gauge transformation $A_\mu \mapsto A_\mu + \partial_\mu\chi$, one has $\delta F_{\mu\nu} = 0$ and

$$\delta T^{\mu\nu} = -F^{\mu\rho}\partial^\nu\partial_\rho\chi, \tag{3}$$

which is generically nonzero. Thus the canonical tensor is gauge dependent. This causes an issue, as the energy-momentum tensor contains observables, which implies it must be gauge invariant. As we mentioned previously in our discussion about energy-momentum tensors, we can solve this issue by introducing an "improved" tensor which is symmetric (and gauge invariant). Define

$$\Theta^{\mu\nu} = T^{\mu\nu} - F^{\rho\mu}\partial_\rho A^\nu. \tag{4}$$

Using the derived expression for the E-M tensor and the fact $F^{\rho\mu} = -F^{\mu\rho}$,

$$\begin{aligned}
\Theta^{\mu\nu} &= -F^{\mu\rho}\partial^\nu A_\rho - F^{\rho\mu}\partial_\rho A^\nu + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} \\
&= -F^{\mu\rho}\partial^\nu A_\rho + F^{\mu\rho}\partial_\rho A^\nu + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} \\
&= F^{\mu\rho}\big(\partial_\rho A^\nu - \partial^\nu A_\rho\big) + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} \\
&= F^{\mu\rho}F_\rho{}^\nu + \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}.
\end{aligned} \tag{5}$$

This is the standard Maxwell stress–energy tensor (up to sign conventions for the metric). We should now check whether this tensor is indeed symmetric and gauge invariant as we require. Let $M^{\mu\nu} = F^{\mu\rho}F_{\rho}{}^{\nu}$, then

$$M^{\nu\mu} = F^{\nu\rho}F_{\rho}{}^{\mu} = F^{\nu}{}_{\gamma}\eta^{\gamma\rho}F_{\rho}{}^{\mu} = F^{\mu\gamma}F^{\nu}{}_{\gamma} = M^{\mu\nu}\,,$$

where the last step used the symmetry of η and relabelling. Thus, the improved tensor is fully symmetric. Good. Now, we should also check gauge invariance. To do so, notice that $\Theta^{\mu\nu}$ depends only on $F_{\mu\nu}$ and $\eta_{\mu\nu}$, and $F_{\mu\nu}$ is gauge invariant, hence

$$\delta\Theta^{\mu\nu} = 0. \tag{6}$$

This new tensor is all well and good, however, does it actually produce the same conserved currents that arise due to Noether's theorem? At first glance it seems like it would not, as the expression is rather different from our initial, defective, E-M tensor. Let us now show that they do indeed lead to the same charges and conserved currents. For the free theory, the equations of motion are $\partial_\mu F^{\mu\nu} = 0$. Both tensors are conserved on-shell:

$$\partial_\mu T^{\mu\nu} = 0, \qquad \partial_\mu \Theta^{\mu\nu} = 0. \tag{7}$$

Moreover, their difference is a total divergence on-shell. Indeed, from (4),

$$\Theta^{\mu\nu} - T^{\mu\nu} = -F^{\rho\mu}\partial_\rho A^\nu. \tag{8}$$

Using $F^{\rho\mu} = -F^{\mu\rho}$,

$$-F^{\rho\mu}\partial_\rho A^\nu = \partial_\rho\big(F^{\mu\rho}A^\nu\big) - (\partial_\rho F^{\mu\rho})A^\nu, \tag{9}$$

so on-shell ($\partial_\rho F^{\mu\rho} = 0$) one has

$$\Theta^{\mu\nu} = T^{\mu\nu} + \partial_\rho\big(F^{\mu\rho}A^\nu\big). \tag{10}$$

We may define the conserved four-momentum currents $(j^\mu)_\nu := T^\mu{}_\nu$ and $(\mathcal{J}^\mu)_\nu := \Theta^\mu{}_\nu$. Integrating the $\mu = 0$ component over a constant-time slice Σ_t and assuming the fields fall off fast enough at spatial infinity, the spatial integral of the total divergence term in (10) becomes a boundary term and vanishes:

$$P^\nu := \int_{\Sigma_t} d^3x\, j^{0\nu} = \int_{\Sigma_t} d^3x\, \mathcal{J}^{0\nu}. \tag{11}$$

Thus, $T^{\mu\nu}$ and $\Theta^{\mu\nu}$ define the same four conserved charges (energy and three-momentum), even though $\Theta^{\mu\nu}$ is the physically nicer object: symmetric, gauge invariant, and traceless (check yourself). We will return to that tracelessness, which is the fingerprint of scale invariance, at the very end.

5.2 Proca Theory

The Maxwell field is massless, and this is no accident: it is tied directly to gauge invariance, because the only Lorentz-invariant quadratic term one could add as a mass, $A_\mu A^\mu$, is *not* gauge invariant and is therefore forbidden in a gauge theory. It is natural to ask what happens if we add it anyway. The resulting theory describes a massive vector (spin-1) field and is known as *Proca theory*. It is the prototype for every massive vector boson in physics: the $W^\pm$ and Z^0 mediators of the weak interaction are the most important realisations, and Proca's equation was historically proposed as a phenomenological model of a massive photon (now constrained to $m_\gamma \lesssim 10^{-18}\,\mathrm{eV}$).

The Proca Lagrangian simply adds the forbidden mass term to Maxwell's,

$$\boxed{\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}m^2 A_\mu A^\mu}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ exactly as before. The kinetic term is untouched, but the mass term explicitly breaks gauge invariance: under $A_\mu \to A_\mu + \partial_\mu \alpha$,

$$A_\mu A^\mu \;\longrightarrow\; A_\mu A^\mu + 2A^\mu \partial_\mu \alpha + \partial_\mu \alpha\, \partial^\mu \alpha \neq A_\mu A^\mu \,.$$

Consequently A_μ is now a genuinely physical field rather than a redundant potential; there is no longer any gauge freedom with which to discard components. We shall see that it carries exactly one more polarisation than the photon.

5.2.1 Equations of motion and the automatic Lorenz constraint

Varying the action as we did for Maxwell, the only new contribution is $\partial\mathcal{L}/\partial A_\nu = -m^2 A^\nu$, so the Euler–Lagrange equations give the *Proca equation*

$$\boxed{\partial_\mu F^{\mu\nu} - m^2 A^\nu = 0\,.}$$

Unlike Maxwell theory, this single tensor equation already carries a built-in constraint. Taking its four-divergence and using the antisymmetry of $F^{\mu\nu}$, which makes $\partial_\nu\partial_\mu F^{\mu\nu} \equiv 0$,

$$\partial_\nu\big(\partial_\mu F^{\mu\nu} - m^2 A^\nu\big) = -m^2\,\partial_\nu A^\nu = 0 \quad\Longrightarrow\quad \partial_\nu A^\nu = 0 \qquad (m \neq 0)\,.$$

The Lorenz condition $\partial_\nu A^\nu = 0$, which in Maxwell theory is an *optional* gauge choice we are free to impose, here it arises *dynamically*: it is forced on us by the equations of motion and cannot be removed, precisely because no gauge freedom remains. Substituting it back, and using $\partial_\mu F^{\mu\nu} = \Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = \Box A^\nu$, the Proca equation collapses to four Klein–Gordon equations subject to one constraint,

$$(\Box - m^2)A^\nu = 0\,, \qquad \partial_\nu A^\nu = 0\,.$$

Each component therefore propagates with the dispersion $E^2 = \mathbf{p}^2 + m^2$, exactly as a free particle of mass m should.

5.2.2 Counting the degrees of freedom

The constraint is the key to counting physical polarisations. For a plane wave $A^\mu(x) = \epsilon^\mu(p)\, e^{ip\cdot x}$ with $p^2 = -m^2$ on shell, the constraint $\partial_\nu A^\nu = 0$ becomes the purely algebraic condition

$$p_\mu \epsilon^\mu = 0\,,$$

which removes one of the four would-be polarisation states and leaves 3 independent physical polarisations; the three spin projections ($s_z = -1, 0, +1$) of a massive spin-1 particle. This is to be contrasted with the massless photon, where gauge invariance removes a *second* state (in addition to the timelike one) and only the 2 transverse polarisations survive. The polarisation sum over the three Proca modes is the projector onto the subspace orthogonal to p_μ,

$$\sum_{\lambda=1}^{3} \epsilon_\mu^{(\lambda)}\, \epsilon_\nu^{(\lambda)*} = \eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2}\,,$$

to be compared with the photon's $\sum_{\lambda=1}^{2} \epsilon_\mu^{(\lambda)} \epsilon_\nu^{(\lambda)*} = \eta_{\mu\nu}$+(gauge-dependent terms). The explicit $1/m^2$ already warns us that the massless limit needs care.

5.2.3 The propagator and the massless limit

Anticipating quantisation, we look ahead and present the momentum-space inverse of the Proca kinetic operator $(\Box - m^2)$ (the propagator) is

$$D_{\mu\nu}(p) = \frac{-i\left(\eta_{\mu\nu} + p_\mu p_\nu/m^2\right)}{p^2 + m^2 - i\varepsilon}\,.$$

Two features stand out. The pole at $p^2 = -m^2$ confirms a propagating state of mass m, and the numerator carries the same singular $p_\mu p_\nu/m^2$ term as the polarisation sum. Naively this suggests a violent discontinuity between Proca theory and Maxwell theory as $m \to 0$. However, when the field couples to a *conserved* current J^μ with $p_\mu J^\mu = 0$, the dangerous term drops out of every physical amplitude $J^\mu D_{\mu\nu} J'^\nu$, and the limit is in fact smooth: a massive photon coupled to conserved charges reproduces ordinary electromagnetism as $m \to 0$, the extra longitudinal mode simply decoupling. It is worth stressing that the analogous statement *fails* for a massive spin-2 field (massive gravity): there the extra modes retain a finite coupling to the trace of the stress tensor, producing the celebrated van Dam–Veltman–Zakharov (vDVZ) discontinuity. The vector case is the benign cousin of that famous story.

5.2.4 The Stückelberg trick

The abrupt change in the degree-of-freedom count (3 for $m \neq 0$ versus 2 for $m = 0$) can be made smooth by *restoring* gauge invariance with an auxiliary scalar. Introducing the *Stückelberg field* φ through the replacement $A_\mu \to A_\mu + \frac{1}{m}\partial_\mu \varphi$ in the mass term gives

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \frac{1}{2} m^2 A_\mu A^\mu - m\, A^\mu \partial_\mu \varphi - \frac{1}{2}\partial_\mu\varphi\, \partial^\mu \varphi\,,$$

which is invariant under the *simultaneous* transformations

$$A_\mu \to A_\mu + \partial_\mu \lambda \,, \qquad \varphi \to \varphi - m\lambda \,.$$

Choosing the *unitary gauge* $\varphi = 0$ recovers the Proca Lagrangian exactly, confirming that no physics has been added, φ is pure bookkeeping. Its advantage is that the massless limit is now manifest: as $m \to 0$ the mixing term $-m\, A^\mu \partial_\mu \varphi$ disappears and the Lagrangian factorises into a free Maxwell field (2 polarisations) plus a decoupled massless scalar (1 polarisation), making the counting $3 = 2 + 1$ completely transparent. This $A_\mu A^\mu \to (A_\mu + m^{-1}\partial_\mu \varphi)^2$ construction generalises to the non-abelian case and is the classical skeleton of the Higgs mechanism, in which the scalar φ is "eaten" to give the gauge boson its mass.

5.2.5 Scale invariance and its breaking

Finally, the Maxwell/Proca pair furnishes a clean illustration of *scale invariance* and its violation. Under a rigid rescaling $x^\mu \to \lambda x^\mu$, a field of scaling dimension Δ transforms as $A_\mu(x) \to \lambda^{-\Delta} A_\mu(\lambda^{-1}x)$; for a vector potential in four dimensions $\Delta = 1$. The massless Maxwell action $-\frac{1}{4}\int \mathrm{d}^4x\, F_{\mu\nu}F^{\mu\nu}$ is then invariant, because the four powers of length supplied by d^4x are exactly cancelled by the four inverse powers carried by $F_{\mu\nu}F^{\mu\nu}$. This classical scale invariance (which is actually in fact full *conformal* invariance in $d = 4$) is encoded in the tracelessness of the improved stress tensor established above,

$$\Theta^\mu{}_\mu = 0 \,.$$

However, when the Proca mass is added, it spoils it. The parameter m introduces a fixed length scale, which is the Compton wavelength $1/m$, and a short computation of the Proca stress tensor gives a non-vanishing trace,

$$\Theta^\mu{}_\mu \;\propto\; m^2 A_\mu A^\mu \neq 0 \,.$$

The non-zero trace is the unmistakable signature of broken scale invariance, with the mass playing the role of the symmetry-breaking parameter. This same pattern, namely a dimensionful coupling generating a non-zero trace of the stress tensor, recurs throughout field theory: from the mass terms of the scalar examples earlier in these notes, through to the running couplings and trace (conformal) anomaly of the full quantum theory, where scale invariance is broken even when no mass is present in the classical Lagrangian. That, however, is a story for a course in quantum field theory.