7 Non–Abelian Gauge Theory: Classical Aspects

Perhaps the single most important class of quantum field theory to understand is non–Abelian Gauge Theory. The starting-point for any non–Abelian gauge theory is the statement that our world is not just a manifold M, but rather a *principal G-bundle*. In this section we'll begin by looking at these from a geometric and topological perspective.

7.1 Principal bundles and vector bundles

$P \to M$. These words mean that P is a manifold that comes with a projection map $\pi : P \to M$, such that for any $x \in M$, $\pi^{-1}(x) \cong G$ for some Lie group G. The space $\pi^{-1}(x)$ is known as the *fibre* of P over x and is often denoted by P_x, while the space M is called the *base*. You should think of P as M with a copy of G attached at each point (see figure **??**). In physics, the Lie group G is known as the *gauge group*, while in maths it's often called the *structure group*. For example, electromagnetism is the case $G = U(1)$, while for reasons nobody really understands[67] the Standard Model has $G = SU(3) \times SU(2) \times U(1)$.

Principal bundles come with a natural (right) group action $G : P \to P$ that preserves the fibres. In other words, if $p \in P$ is a point in the fibre over $x \in M$ then acting with a group element $g \in G$ gives another point $pg \in P$, with the property that $\pi(pg) = \pi(p) = x$ so that pg and p both lie in the *same* copy of the fibre. Thus the group action allows you to move around within each copy of G, but does not move you around in M.

To get more of a handle on these abstract ideas, it's useful to describe the situation just in a small region. Given an open set $U \subset M$, a *local trivialization* is a choice of isomorphism

$$\Phi : \pi^{-1}(U) \to U \times G \tag{7.1}$$

and so gives a way to identify $P|_U$ with $U \times G$. Explicitly, if we're given a point $p \in \pi^{-1}(U) \subset P$, then we can always write $\Phi(p) = (\pi(p), \phi(p))$ where $\pi(p) \in U \subset M$ is just whatever point p projects to, and $\phi(p)$ is some group element. Exactly which group element we get will of course depend on exactly how we choose Φ, but we require that this is compatible with the action of G on the bundle itself. In other words, if $G : p \to pg$, then

$$\Phi(pg) = (\pi(pg), \phi(pg)) = (\pi(p), \phi(p)g)\,. \tag{7.2}$$

Notice that while it's true *locally* that any principal bundle looks like $U \times G$, this might not be true globally. The simplest example is to take $G = \mathbb{R}$ thought of as a one–dimensional Abelian group, and $M = S^1$. Then both the cylinder $S^1 \times \mathbb{R}$ and the Möbius strip are principal R-bundles — they both look locally like $U \times \mathbb{R}$ — but topologically they are different.

Now suppose that $\{U_\alpha\}$ are a collection of open sets in M. Given a local trivialization Φ_α of $\pi^{-1}(U_\alpha)$ on each region U_α, if the open sets overlap then we can ask how the

[67] As you'll learn if you're taking the Part III Standard Model course, the running of the coupling constants for each of the three semi-simple factors, together with the particular representations of $SU(3)\times SU(2)\times U(1)$ in which quarks and leptons lie, suggests that this group may just be a low-energy remnant of a larger 'grand unified gauge group'. This grand unified group is often thought to be $SU(5)$. Or $SU(5) \times U(1)$. Or perhaps $SO(10)$. Or maybe E_6. Like I said, no one really knows.

trivializations are related on their common domain of definition $\pi^{-1}(U_\alpha \cap U_\beta)$. First, let's just consider what happens at one point $p \in U_\alpha \cap U_\beta$. By definition, both trivializations involve projecting p to the same point in M, but $\phi_\alpha(p)$ may be a different group element than $\phi_\beta(p)$. Nonetheless, since both $\phi_\alpha(p)$ and $\phi_\beta(p)$ are certainly both in G, we must be able to find a group element $t_{\alpha\beta} \in G$ such that

$$\phi_\beta(p) = \phi_\alpha(p)\, t_{\alpha\beta}\,. \tag{7.3}$$

This group element allows us to relate our two local trivializations at some point p. If we wish to compare our trivializations throughout $\pi^{-1}(U_\alpha \cap U_\beta)$ then we must allow $t_{\alpha\beta}$ to vary. Thus, given a pair of open sets U_α and U_β, we define a *transition function* to be a map

$$T_{\alpha\beta} : U_\alpha \cap U_\beta \to G\,. \tag{7.4}$$

which we can think of as a G-valued function $t_{\alpha\beta}(x)$, defined at each $x \in U_\alpha \cap U_\beta$. So that we can compare Φ_α with Φ_β as well as compare Φ_β with Φ_α, we ask that the transition functions are invertible: $t_{\beta\alpha}(x) = t^{-1}_{\alpha\beta}(x)$. We also impose the compatibility relation $T_{\alpha\gamma} = T_{\alpha\beta} \circ T_{\beta\gamma}$ on triple overlaps $U_\alpha \cap U_\beta \cap U_\gamma$, which says that the result of comparing Φ_α with Φ_γ is the same as first comparing Φ_α with Φ_β and then comparing Φ_β with Φ_γ in any region where all three are defined. Finally, it's natural to ask that all these transition functions vary smoothly over $U_\alpha \cap U_\beta$, so we will.

In physics, the most common case of all this is when $U_\alpha = U_\beta$ are actually the *same* set U, and we're just comparing two different ways of identifying $\pi^{-1}(U)$ with $U \times G$. In this case, the local trivializations are thought of as choices of *gauge*, while the transition function is usually called a *gauge transformation*. For example, a familiar case might be to choose $M \cong \mathbb{R}^{3,1}$ and $G \cong U(1)$, whereupon for each x we could write $t(x) = \mathrm{e}^{\mathrm{i}\lambda(x)} \in U(1)$ with $\lambda(x)$ a gauge parameter in electrodynamics. Another example that should be familiar from General Relativity is to take M to be (curved) space-time and G to be $GL(d, \mathbb{R})$. In this case, a local trivialization is a choice of *coordinate system* above an open patch U, whereas the transition functions $T_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(d, \mathbb{R})$ are general coordinate transformations. Of course, in any open region of M there could be many valid coordinate systems, and a large part of the Principle of Relativity is the statement that the actual *physics* doesn't depend on which coordinates (= local trivialization) we use. You probably also know that spaces with non-trivial topology (*e.g.* just a circle S^1 or sphere S^d) cannot be described by just a single set of coordinates, which is why all our constructions are defined only locally.

Let me also point out that the mathematical picture of principal bundles is also the inspiration for efforts to recover the Standard Model from higher dimensional theories, initiated in the 1930s by Kaluza and later by Klein. For example, the Lie group $SU(2)$ is isomorphic to S^2, so a principal $SU(2)$-bundle over 'our' space–time M can be thought of as a six dimensional space–time where every point of M comes with a copy of S^2. If the sphere is very small, then present technology won't allow us to concentrate enough energy in a small region so as to excite spherical harmonics on the S^2, since these oscillations will involve some very high frequency set by the inverse radius. Thus, at low energies,

this theory should involve just the constant modes on S^2 and so look like a theory on M. We might hope to find some remnant of the spherical extra dimensions by examining how low–energy fields transform under higher dimensional coordinate transformations. You can find out more about these ideas in the Extra Dimensions course in Part III.

7.1.1 Vector bundles from representations

As you learned in the Symmetries, Particles & Fields course, whenever we're given a Lie group it's often a good idea to look at the *representations* of this group. Recall that a representation ρ is a choice of vector space (usually either $\mathbb{R}^r$ or $\mathbb{C}^r$) together with map

$$\rho : G \to \mathrm{Mat}(r;\mathbb{R}) \qquad \text{or} \qquad \rho : G \to \mathrm{Mat}(r;\mathbb{C}) \tag{7.5}$$

to the space of $r \times r$ matrices (with real or complex values) that tells us how elements of G act on the vector space. This map should be compatible with the group structure in the sense that

$$\rho(gh) = \rho(g) \circ \rho(h)\,, \tag{7.6}$$

where on the left gh denotes multiplication in G, whereas the *rhs* denotes matrix multiplication.

For example, the fundamental representation of the rotation group $SO(3)$ represents elements $g \in SO(3)$ by 3×3 matrices that are orthogonal and have unit determinant; these matrices tell us how the components of a standard vector $\mathbf{v} \in \mathbb{R}^3$ change as we rotate. The same group also has a spinor representation where an element $g \in SO(3)$ is represented by a 2×2 unitary matrix U_g with unit determinant, often written in terms of the Pauli sigma matrices σ as $\mathrm{e}^{\mathrm{i}\alpha_g\cdot\sigma}$, where the parameters α_g depend on which group element we're considering. This representation tells us how the two complex components of a spinor (such as an electron wavefunction) change under rotations.

Now, if we have a principal bundle then we have not just a single copy of a Lie group G, but a whole family of copies, one at each point of M. If we pick a representation ρ, we thus get a whole family of vector spaces. This structure is known as a *vector bundle* $E \to M$. In a vector bundle the fibre $\pi^{-1}(x)$ at each point $x \in M$ is now a vector space — the one we got when we chose a G-representation. Thus vector space is abstract – *a priori*, it comes with no preferred basis. Locally, we choose a basis by choosing a local trivialization $\Phi : \pi^{-1}(U) \cong U \times \mathbb{C}^r$ which is a way of identifying the abstract vector space $E|_x$ with $\mathbb{C}^r$ for all $x \in U$. Given a pair of overlapping open sets U_α and U_β, with trivializations $\Phi_{\alpha,\beta}$ on each, the transition functions $T_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{Mat}(r;\mathbb{C})$ are matrices which tell us how the basis of $E|_x$ given by the trivialization Φ_α is related to the basis given by Φ_β.

If the principal bundle we started with is a subgroup of $GL(r;\mathbb{C})$ then these transition functions will preserve some structure on the vector space. For example, if $G = U(r) \subset GL(r;\mathbb{C})$ and we choose the fundamental r-dimensional (complex) representation, then the transition functions will be *unitary* matrices preserving the inner product $\sum_{a=1}^{1} |z^a|^2$ on each fibre, while if $G = SU(r)$ then the transition functions will additionally have *unit*

determinant, and so will also preserve the top holomorphic form[68]

$$\epsilon_{a_1\ldots a_r}\mathrm{d}z^{a_1}\wedge\mathrm{d}z^{a_2}\wedge\cdots\wedge\mathrm{d}z^{a_r}$$

on each fibre.

Vector bundles are of relevance to physics because a charged matter field is a *section* of E. This is a map

$$s : M \to E \tag{7.7}$$

that obeys $\pi\circ s = \mathrm{id}$. Given a choice of trivialization, we can think of this section as the assignment of a vector $s^a(x)\in\mathbb{C}^r$ (for a complex vector bundle) to each point x. If we change our local trivialization using a gauge transform, then the particular vector components we get will change according to

$$s_\beta(x) = s_\alpha(x)\,t_{\alpha\beta}(x) \tag{7.8}$$

where $t_{\alpha\beta}(x)$ are the transition functions. We'll sometimes write $\Omega^0_M(E)$ to denote the space of all smooth sections of $E\to M$. Note that the statement that a section is a map $s : M\to E$ implies that, once we've picked a gauge, the resulting field *depends only on* $x\in M$. This is where the notion of a gauge theory differs from Kaluza–Klein theory, where the fields are allowed to vary over the whole higher–dimensional space.

As a simple example, consider a complex scalar field on M. Usually, we think of this as just a function $\phi : M\to\mathbb{C}$ where $\phi : x\mapsto\phi(x)$ is the value of the field at $x\in M$. However, if this scalar carries electric charge q, then it doesn't really have any preferred 'value' because, as you learnt in electrodynamics, we can change $\phi(x)\to \mathrm{e}^{\mathrm{i}q\lambda(x)}\phi(x)$ by a gauge transform. The correct interpretation of our charged scalar is that ϕ is really section of a vector bundle $E\to M$ associated to the principal $U(1)$ bundle of electromagnetism. One we pick a local trivialization — *i.e.* pick a gauge — then we can think of $\phi(x)$ as a (one component) vector, at least for some open region $U\subset M$. However, there's no preferred way to choose this gauge, and making different choices (changing gauge) will cause ϕ to transform as is familiar.

The electromagnetic example is special because the only irreducible representations of $U(1)$ are one–dimensional; they're just labelled by the *charge* q of the field, with $q=0$ being the trivial representation. As you saw in Symmetries, Particles & Fields, non-Abelian groups typically have (infinitely) many different irreps, labelled by Young tableaux. In practice, most of the cases we meet in physics, the representations are fairly straightforward; we rarely encounter much beyond the fundamental representation (*e.g.* quarks transform in the **3** of the SU(3) associated to 'colour'), the anti-fundamental (*e.g.* antiquarks in the $\overline{\mathbf{3}}$ of colour SU(3)) and the adjoint representation (*e.g.* gluons).

Above, we've constructed vector bundles from principal bundles by making a choice of representation. Finally, let me mention that we can also go the other way and construct a principal bundle starting from a vector bundle, at least in the case of a matrix Lie group

[68] Don't worry if you don't know what this means.

(those that are subgroups of $GL(r)$). For given any rank r vector bundle $E \to M$, we define the *frame bundle* to be the principal $GL(r)$-bundle whose fibre $\pi^{-1}(x)$ is the collection of all basis vectors (= frames) in the fibre E_x. There is a natural action of $GL(r)$ on this frame bundle, relating any pair of frames. If we wish to construct a principal G-bundle for a subgroup of $GL(r)$ then as above we just require that our basis vectors are compatible with some extra structure. For example, in the real case we obtain a principal $O(r)$-bundle by asking that our basis vectors are orthogonal, and a principal $SO(r)$ bundle by asking that they also define a fixed volume element.

The most common Lie groups that arise in physics are indeed matrix Lie groups, so the two viewpoints are equivalent. However, in some exotic theories (especially string theory and some grand unified theories) exceptional Lie groups such as E_6 play an important role, so the fundamental picture is really that of principal bundles. With this caveat, I'll mostly use vector bundles from now on.

7.1.2 Connections and curvature

So far, we have described a vector bundle E as just a collection of vector spaces parametrized by a base space M. If we wish to write down the kinetic terms of any matter field, we will need derivatives and we'd usually write

$$v^\mu \partial_\mu \phi(x) \stackrel{?}{=} \lim_{\epsilon \to 0} \frac{\phi(x+\epsilon v) - \phi(x)}{\epsilon}\,, \tag{7.9}$$

where v^μ is a vector at x. However, for a charged field this expression is meaningless as it stands, because the two terms on the right live in completely different spaces: $\phi(x+\epsilon v)$ lives in the fibre $E_{x+\epsilon v}$ while $\phi(x)$ lives in a different copy E_x of the fibre.

To make sense of this, we need a way to compare vectors in different fibres, which is what a *connection*, or *covariant derivative* provides. This is a linear map

$$\nabla : \Omega^0_M(E) \to \Omega^1_M(E) \tag{7.10}$$

from the space of sections to the space of 1-forms (covectors) on M with values in E. The connection is defined by the properties of *linearity*:

$$\nabla(\alpha_1 s_1 + \alpha_2 s_2) = \alpha_1 \nabla(s_1) + \alpha_2 \nabla(s_2) \tag{7.11}$$

for any two sections s_1, s_2 and constants $\alpha_1 \alpha_2$, and the *Leibniz rule*:

$$\nabla(fs) = df\, s + f\nabla(s)\,, \tag{7.12}$$

where $f \in C^\infty(M)$ is a smooth function. More specifically, for every tangent vector v on M, the connection defines a derivative $v \cdot \nabla s = v^\mu \nabla_\mu(s)$, thought of as the derivative of our section $s(x)$ in the direction of v, and then the Leibniz rule says $v \cdot \nabla(fs) = f v^\mu \nabla_\mu s + (v^\mu \partial_\mu f) s$, where ∂_μ is the standard partial derivative of the function f. Notice that if ∇ and ∇' are any two connections, then the difference obeys

$$(\nabla - \nabla')(fs) = f(\nabla - \nabla')s\,. \tag{7.13}$$

Thus $(\nabla - \nabla')$ maps $\Omega^0_M(E) \to \Omega^1_M(E)$ in a way that is linear over *functions* $f \in C^\infty(M)$. Hence the difference between any two connections is an element of $\mathrm{Hom}(E, E \otimes T^*M) \cong \mathrm{End}(E) \otimes T^*M$.

To understand what this means, let's again look in a small region. Suppose we have a trivialization $\Phi : E|_U \to U \times \mathbb{C}^r$. Then in this region, any section $s : U \to E$ can be thought of as a vector–valued function on U, *i.e.* given $s : M \to E$, we can write $\Phi \circ s : U \to U \times \mathbb{C}^r$ with $\Phi(s(x)) = (x, s_U(x))$ for some vector $s_U(x)$. Then, within U we have

$$\Phi(\nabla s)(x) = (x, \nabla s_U) \qquad \text{where} \qquad (\nabla s)_U = ds_U + A_U s_U \tag{7.14}$$

where A_U is a section of $\mathrm{End}(\mathbb{C}^r) \otimes T^*M|_U$ that is independent of the particular section s, but depends on our choice of trivialization Φ. The object A_U is thus a matrix–valued 1-form (or covector) known to mathematicians as a *connection 1-form*. In physics, A_U is the gauge field.

To make this more familiar, suppose U_α and U_β are each open sets in M with overlap $U_\alpha \cap U_\beta \neq \emptyset$. Suppose we're given trivializations $\Phi_\alpha : E|_{U_\alpha} \to U_\alpha \times \mathbb{C}^r$ and $\Phi_\beta : E|_{U_\beta} \to U_\beta \times \mathbb{C}^r$, and let $g_{\beta\alpha} : U_\beta \cap U_\alpha \to \mathrm{End}(\mathbb{C}^r)$ denote the transition function. In physics, we'd say we're working in one gauge (given by Φ_α) in U_α, and in another gauge in U_β, and that $g_{\beta\alpha}$ is the gauge transformation that takes us between the two gauges. We have $s_\beta = g_{\beta\alpha} s_\alpha$ and similarly $(\nabla s)_\beta = g_{\beta\alpha}(\nabla s_\alpha)$, since these are sections of E and $E \otimes T^*M$, respectively. It now follows that the gauge fields A_α and A_β on the two open patches must be related by

$$A_\beta = -g_{\beta\alpha} dg^{-1}_{\beta\alpha} + g_{\beta\alpha} A_\alpha g^{-1}_{\beta\alpha} \,. \tag{7.15}$$

As a special case, U_β and U_α might both be the *same* region, in which case (7.15) tells us how A changes under a change of trivialization on a given region. Notice also that if $\mathrm{rk}(E) = r = 1$, then $g(x) \in \mathrm{End}(\mathbb{C}^r)$ is just a single function at each point of $x \in U_\beta \cap U_\alpha$ which we can write as $\mathrm{e}^{\mathrm{i}\lambda}$ whereupon (7.15) reduces to $A_\beta = A_\alpha - \mathrm{i} d\lambda$, which is the familiar behaviour under a gauge transformation of the vector potential in electromagnetism.

Given a connection $\nabla : \Omega^0_M(E) \to \Omega^1_M(E)$, we can extend its definition to sections of $E \otimes \wedge^p T^*M$. Such sections are to be thought of as p-forms[69] with values in sections of E. The space of such is written $\Omega^p_M(E)$. This extension is also called ∇, and is again defined by

$$\begin{aligned} &\text{linearity} && \nabla(\alpha_1 \mathfrak{s}_1 + \alpha_2 \mathfrak{s}_2) = \alpha_1 \nabla \mathfrak{s}_1 + \alpha_2 \nabla \mathfrak{s}_2 \\ &\text{Leibniz} && \nabla(s\omega) = \nabla(s) \wedge \omega + s\, d\omega \end{aligned} \tag{7.16}$$

where $\mathfrak{s}_{1,2} \in \Omega^p_M(E)$ are p-form sections, while $s \in \Omega^0_M(E)$ is a regular section as above, $\omega \in \Omega^p_M$ is a p-form on M and $\alpha_{1,2}$ are constants.

The construction above shows that the connection ∇ behaves as a sort of exterior derivative, generalized the usual de Rham exterior derivative d to the case of sections of

[69] A p-form on M can be thought of as a tensor with p contravariant indices, antisymmetrized on these indices. We write Ω^p_M for the space of such p-forms. Thus, if $\omega \in \Omega^p_M$ then in local coordinates on M we have $\omega = \omega_{\mu_1\mu_2\cdots\mu_p}(x)\, dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_p}$.

vector bundles. However, there is a crucial difference. While $d^2 = 0$ automatically, it is not in general true that $\nabla^2 = 0$. To see this, note that

$$\begin{aligned}\nabla(\nabla(s\omega)) &= \nabla((\nabla s)\wedge\omega + s\,d\omega)\\ &= (\nabla^2 s)\wedge\omega - (\nabla s)\wedge d\omega + (\nabla s)\wedge d\omega + s d^2\omega\\ &= (\nabla^2 s)\wedge\omega\end{aligned} \tag{7.17}$$

where the second and third terms have cancelled and the last term is identically zero by the nilpotency of the de Rham exterior derivative. This calculation shows that $\nabla^2 : \Omega^p_M(E) \to \Omega^{p+2}_M(E)$ is linear over multiplication of the section by an arbitrary form,

$$\nabla^2(\mathfrak{s}\wedge\omega) = (\nabla^2\mathfrak{s})\wedge\omega\,.$$

It therefore must correspond to multiplication by some section $F_\nabla \in \Omega^2_M(\mathrm{End}(E))$. In particular, if $s \in \Omega^0_M(E)$, we have

$$\nabla^2(s) = F_\nabla s\,. \tag{7.18}$$

The $\mathrm{End}(E)$–valued 2-form F_∇ is called the *curvature* of the connection.

To understand this more explicitly, let's again choose a local trivialization $\Phi : E|_U \to U\times\mathbb{C}^r$, with $(\nabla s)_U = ds_U + A_U s_U$. Then we have

$$\begin{aligned}(\nabla^2 s)_U &= \nabla(ds_U + A_U s_U)\\ &= d^2 s_U + d(A_U s_U) + A_U\wedge(ds_U + A_U)s_U\\ &= (dA_U + A_U\wedge A_U)s_U\end{aligned} \tag{7.19}$$

and indeed all the derivatives of s_U itself cancel out. Thus on $E|_U$ with the trivialization given by Φ, we can identify the curvature as

$$\begin{aligned}(F_\nabla)_U &= dA_U + A_U\wedge A_U = (\partial_\mu A_\nu + A_\mu A_\nu)\,dx^\mu\wedge dx^\nu\\ &= \frac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])\,dx^\mu\wedge dx^\nu = \frac{1}{2}F_{\mu\nu}dx^\mu\wedge dx^\nu\end{aligned} \tag{7.20}$$

where $F_{\mu\nu}$ are the curvature components. Recall that A_U is a matrix–valued 1-form, so that the commutator term $[A_\mu, A_\nu]$ does not in general vanish.

Given that $\nabla^2(\mathfrak{s}) = F_\nabla\mathfrak{s}$ for any section $\mathfrak{s}\in\Omega^p_M(E)$, so that ∇^2 acts as a purely multiplicative operator, it's interesting to compute what happens when we act with ∇ for a third time. On the one hand, we have

$$\nabla^3(\mathfrak{s}) = \nabla(F_\nabla\mathfrak{s}) = \nabla(F_\nabla)\mathfrak{s} + F_\nabla\nabla(\mathfrak{s}) \tag{7.21}$$

while on the other hand,

$$\nabla^3(\mathfrak{s}) = \nabla^2(\nabla\mathfrak{s}) = F_\nabla\nabla\mathfrak{s}\,. \tag{7.22}$$

The two equations (7.21) & (7.46) are compatible iff

$$\nabla(F_\nabla) = 0\,, \tag{7.23}$$

which is known as the *Bianchi identity* for the curvature F_∇. In a local trivialization where $\nabla = d + A_U$ and $F_\nabla = dA_U + A_U \wedge A_U$, the Bianchi identity can be seen explicitly by calculating

$$\begin{aligned}\nabla(F_\nabla)|_U &= dF_\nabla + A_U \wedge F_\nabla - F_\nabla \wedge A_U \\ &= d(dA_U + A_U \wedge A_U) + A_U \wedge (dA_U + A_U \wedge A_U) - (dA_U + A_U \wedge A_U) \wedge A_U \\ &= dA_U \wedge A_U - A_U \wedge dA_U + A_U \wedge dA_U + A_U^3 - dA_U \wedge A_U - A_U^3 \\ &= 0\,,\end{aligned} \tag{7.24}$$

where we've used the fact that, in a local trivialization, the covariant derivative acts as $\nabla\phi = d\phi + A \wedge \phi - (-)^p \phi \wedge A$ on any section $\phi \in \Omega^p_M(\mathrm{End}(E))$. This agrees with our global argument above.

A section $\mathfrak{s}$ is said to be **flat** if $\nabla\mathfrak{s} = 0$. Generically, vector bundles do not have any non-trivial[70] flat sections, since $\nabla\mathfrak{s} = 0$ implies

$$0 = \nabla^2\mathfrak{s} = F\mathfrak{s}\,, \tag{7.25}$$

so the existence of a non–vanishing flat section in some region U implies that the curvature vanishes throughout U. This has an important application in the theory of superconductivity. A superconductor is just a material in which it is energetically favoured for there to exist a non-vanishing charged field. For example, we may have an electrically charged complex scalar field s whose energy is given by

$$E[s] = \int_U \mathrm{d}^3x \left[\frac{1}{2}|\nabla s|^2 + \frac{\lambda}{2}(|s|^2 - a^2)^2\right] \tag{7.26}$$

where $U \subset \mathbb{R}^3$ is the region of space occupied by the superconductor, and $a, \lambda > 0$ are constants. The minimum of this energy is obtained when $\nabla s = 0$ and $|s| = a$. Consequently, in the ground state, the curvature F must vanish inside the superconductor. The space–space components of F is the magnetic field B, and the expulsion of magnetic fields from the interior of superconductors is known as the **Meissner effect**. It's responsible for levitation effects much favoured in Physics Open Day demonstrations and now put to use in the fastest train in the world, running from Shanghai to the Pudong International airport.

7.1.3 Holonomy

Suppose $I = [0, 1]$, and $\gamma : I \to M$ is a smooth map, with image curve $C \subset M$. Given a vector bundle $E \to M$, we define the pullback bundle γ^*E to be the bundle over I whose fibre at each point $t \in [0, 1]$ is $E_{\gamma(t)}$. This pullback bundle is necessarily flat since there is no room for any curvature (a 2-form) on I.

A section $s : M \to E$ is said to be **parallel transported** along C if it obeys

$$\gamma^*(\nabla s) = 0\,, \tag{7.27}$$

[70] The zero section $\mathfrak{s} \equiv 0$ is trivially flat.

or, more explicitly,

$$\frac{d\gamma^\mu(t)}{dt}\left(\partial_\mu s + A_\mu(\gamma(t))s\right) = 0 \tag{7.28}$$

where the way the connection acts is determined by the representation (*i.e.* the particular vector space) in which s takes its values. Suppose that $x = \gamma(0)$ and $y = \gamma(1)$ are the endpoints of C, and that we are given some value s_0 in E_x which we wish to parallel transport along C.

Consider first the Abelian case, with s in the fundamental representation (charge +1). We solve (7.27) explicitly to find

$$s_1 = U_C(x,y)s_0\,, \tag{7.29}$$

where the **parallel transporter** $U_C(x,y) : E_x \to E_y$ is defined by

$$U_C(x,y) := \exp\left(-\int_I \gamma^* A\right) = \exp\left(-\int_0^1 A_\mu(\gamma(t))\frac{d\gamma^\mu(t)}{dt}\,\mathrm{d}t\right)\,. \tag{7.30}$$

In the physics literature, this is known as a **Wilson line**. Notice that under the gauge transform $A \to A + d\lambda$ we have

$$\int_0^1 \gamma^* A \to \int_0^1 \gamma^* A + \int_0^1 \gamma^*(d\lambda) \quad = \int_0^1 \gamma^* A + \lambda(\gamma(1)) - \lambda(\gamma(0)) \tag{7.31}$$

so that the parallel transporter behaves as

$$U_C(x,y) \to \mathrm{e}^{-\lambda(y)}\, U_C(x,y)\, \mathrm{e}^{\lambda(x)} \tag{7.32}$$

showing that s_1 in (7.29) transforms as it should.

In the non-Abelian case, because the connection 1-forms A do not commute, we can only solve (7.27) iteratively, finding

$$\begin{aligned} U_C(x,y) &= 1 - \left(\frac{d}{dt}\right)^{-1}\left(\frac{d\gamma^\mu}{dt}A_\mu(t)\,U\right) \\ &= 1 - \int_0^1 \mathrm{d}t \frac{d\gamma^\mu}{dt}A_\mu(t) + \int_0^1 \mathrm{d}t \int_0^t \mathrm{d}t'\, \frac{d\gamma^\mu}{dt}A_\mu(t)\,\frac{d\gamma^\nu}{dt'}A_\nu(t') - \cdots \end{aligned} \tag{7.33}$$

We frequently write

$$U_C(x,y) = P\exp\left(-\int \gamma^* A\right) \tag{7.34}$$

as shorthand for this infinite series. It follows from the definition (7.27) that $\gamma^\mu \nabla_\mu U_C(x,y) = 0$ and that under the gauge transform $A \mapsto A^g = gAg^{-1} + \mathrm{d}g\, g^{-1}$ the non–Abelian Wilson line behaves as

$$U_C(x,y) \mapsto g(x)U_C(x,y)g^{-1}(y)\,. \tag{7.35}$$

For a closed loop, where $x = y$, we define the **Wilson loop** or **holonomy** by

$$W(C) = \mathrm{tr}P\exp\left(-\oint \gamma^* A\right) \tag{7.36}$$

where the trace is taken in the representation corresponding to the vector bundle $E \to M$ on which we were performing our parallel transport. Wilson loops are the basic observables in any gauge theory as they exist irrespective of any matter content.

7.2 Classical Yang–Mills theory

The first and most important example of a non–Abelian gauge theory was introduced to physics 1954 by Chen Ning Yang and Robert Mills, and then almost completely ignored for nearly a decade. From a phenomenological point of view, the importance of Yang–Mills theory arises because (as you're surely aware) the Standard Model — the most fundamental description of Nature we currently possess — is at it's heart a non–Abelian gauge theory based on $SU(3) \times SU(2) \times U(1)$, though it took much hard work and many further new ideas before this became apparent. From the perspective of a theoretical physicst, Yang–Mills is theory important also because it's the *only* QFT in $d = 4$ that might a continuum limit, as realized by Coleman & Gross in 1973. In mathematics, Yang–Mills theory is at the heart of Simon Donaldson's exploration of the wild world of four–manifolds. More recently, it's even been related to the (geometric) Langlands Program. In a deep sense, Yang–Mills theory is the right four–dimensional analogue of geodesics in $d = 1$ and harmonic maps in $d = 2$.

7.2.1 The Yang–Mills action

To describe Yang–Mills theory, we pick a d–dimensional (pseudo-)Riemannian manifold (M, g) complete with a choice of metric g and associated Hodge star $* : \Omega^p(M) \to \Omega^{d-p}(M)$. The Yang–Mills action is then defined to be

$$S_{\mathrm{YM}}[\nabla] = -\frac{1}{2\mathrm{g}_{\mathrm{YM}}^2} \int_M \mathrm{tr}(F_\nabla \wedge *F_\nabla) \tag{7.37}$$

where tr denotes the Killing form on the Lie algebra $\mathfrak{g}$ of the structure group and where g_{YM} is a coupling constant. The Yang–Mills action is thus just the (square of) the L^2-norm of F_∇ with respect to the standard volume element on M supplied by the metric g. Picking a Hermitian basis $\{t_\mathrm{a}\}$ of $\mathfrak{g}$, obeying $[t_\mathrm{a}, t_\mathrm{b}] = \mathrm{i} f^\mathrm{c}_{\mathrm{ab}} t_\mathrm{c}$ in terms of the structure constants $f^\mathrm{c}_{\mathrm{ab}}$ and normalized so that $\mathrm{tr}(t_\mathrm{a} t_\mathrm{b}) = \delta_{\mathrm{ab}}/2$, we write

$$F = -\mathrm{i} F^\mathrm{a} t_\mathrm{a} = -\mathrm{i} F^\mathrm{a}_{\mu\nu}(x)\, dx^\mu \wedge dx^\nu\, t_\mathrm{a}$$

to expand the curvature in this basis. The Yang–Mills action then becomes

$$S_{\mathrm{YM}}[\nabla] = \frac{1}{4\mathrm{g}_{\mathrm{YM}}^2} \int_M F^{\mathrm{a}\mu\nu} F^\mathrm{a}_{\mu\nu}\, \sqrt{g}\, \mathrm{d}^d x \tag{7.38}$$

which we see is the natural generalization of the Maxwell action

$$S_{\mathrm{Max}}[\nabla] = \frac{1}{4e^2} \int_M F \wedge *F = \frac{1}{4e^2} \int_M F^{\mu\nu} F_{\mu\nu}\, \sqrt{g}\, \mathrm{d}^d x \tag{7.39}$$

of electromagnetism, to which it reduces when $G = U(1)$ and $(M, g) = (\mathbb{R}^{3,1}, \delta)$. I've written the action as $S_{\mathrm{YM}}[\nabla]$ to emphasize that we should treat the action as a function of the connection, not of the curvature. Again, this is familiar to you from deriving Maxwell's (vacuum) equations as the Euler–Lagrange equations of (7.39).

From the point of view of physics, the most important difference between the Yang–Mills action for a non–Abelian group G and the Abelian (Maxwell) case is that, for non–Abelian G, *the Yang–Mills field interacts with itself.* We can see this by restricting to a local trivialization on $U \subset M$ where $\nabla = d + A$ and $F = dA + A^2$. On this patch the Yang–Mills action becomes[71]

$$\begin{aligned} S_{\rm YM}[A] &= -\frac{1}{2{\rm g}_{\rm YM}^2}\int_U {\rm tr}((dA+A^2)\wedge *(dA+A^2)) \\ &= -\frac{1}{2{\rm g}_{\rm YM}^2}\int_U {\rm tr}(dA\wedge *dA + 2AA\wedge *dA + AA\wedge *AA)\,. \end{aligned} \tag{7.40}$$

Thus we see that the action contains a cubic vertex of the schematic form $\sim AA\,dA$, and a quartic vertex $\sim A^4$. Thus, even in the absence of any charged matter, Yang–Mills theory is a non–trivial interacting theory.

At the classical level, these self–interactions make themselves felt via the Euler–Lagrange equations that follow from (7.37). Varying the connection $\nabla \to \nabla + \delta a$ where $\delta a \in \Omega^1_M({\rm End}(E))$, to first order in δa we have that

$$F_{\nabla+\delta a} - F_\nabla = \nabla(\delta a), \tag{7.41}$$

or in other words, $\delta F_{\mu\nu} = \nabla_{[\mu}\delta a_{\nu]}$. Therefore, varying the action gives[72]

$$\delta S_{\rm YM} = \frac{1}{{\rm g}_{\rm YM}^2}\int_M {\rm tr}(\nabla\delta a\wedge *F_\nabla) = -\frac{1}{{\rm g}_{\rm YM}^2}\int_M {\rm tr}(\delta a\wedge \nabla * F_\nabla) + \text{boundary terms} \tag{7.42}$$

so that the Euler–Lagrange equations are

$$\nabla * F_\nabla = 0\,, \tag{7.43}$$

or equivalently

$$0 = \nabla^\mu F^{\rm a}_{\mu\nu} = \partial^\mu F^{\rm a}_{\mu\nu} + f^{\rm a}_{\rm bc}A^{{\rm b}\mu}F^{\rm c}_{\mu\nu} \tag{7.44}$$

in terms of a local trivialization of P, where we write $A_\mu = -{\rm i}A^{\rm a}_\mu t_{\rm a}$ using the basis $\{t_{\rm a}\}$. However you wish to write them, these are known as the **Yang–Mills equations**. We recall from (7.23) that the curvature of any connection automatically obeys the Bianchi identity $\nabla F_\nabla = 0$, or

$$\nabla_\mu F_{\nu\lambda} + \nabla_\nu F_{\lambda\mu} + \nabla_\lambda F_{\mu\nu} = 0 \tag{7.45}$$

or

$$\partial_\mu F^{\rm a}_{\nu\lambda} + \partial_\nu F^{\rm a}_{\lambda\mu} + \partial_\lambda F^{\rm a}_{\mu\nu} + f^{\rm a}_{\rm bc}\left(A^{\rm b}_\mu F^{\rm c}_{\nu\lambda} + A^{\rm b}_\nu F^{\rm c}_{\mu\lambda} + A^{\rm b}_\lambda F^{\rm c}_{\mu\nu}\right) = 0 \tag{7.46}$$

to be completely explicit. Mathematically, connections whose curvature obeys the Yang–Mills equation (7.43) are *critical points* of the function $S_{\rm YM}[\nabla]$ defined on the space of *all* connections on $P \to M$.

[71] Henceforth, I'll drop the subscript on A_U, taking the local patch for granted. We'll look at this action in components below.

[72] As an exercise, you should go through this derivation for yourself, using the second line of (7.40) and checking you're happy with where all the indices go. Once you've done this and checked you are happy, I hope you'll begin to appreciate the usefulness of the form notation.

Unlike the vacuum Maxwell equations, the Yang–Mills equations and the Bianchi identity (7.46) are *non–linear* p.d.e.s. The non–linearities arise because of the presence of A in the covariant derivative ∇ and the non–linear terms $\sim A^2$ in the curvature itself. The reason this happened is that the curvature F transforms in the adjoint representation, with $F \to gFg^{-1}$ under a gauge transformation $g(x)$. So any sort of differential equation obeyed by F will have to involve covariant derivatives, as these are the only derivatives that make geometric sense. Electrodynamics avoids this complication because the adjoint representation of an Abelian group is actually trivial, as we can see by the manipulation $gFg^{-1} = Fgg^{-1} = F$, which is allowed since all group elements commute. As a consequence of the non–linearity, unlike beams of light or radio waves in the Abelian case, we do not expect superposition of solutions, and propagating waves of Yang–Mills fields do not pass through one another freely.

Famously, General Relativity is also a geometric theory in which the field (Einstein) equations are a system of non–linear pdes. Indeed, there are many parallels between the two theories, most of which can be made apparent by treating General Relativity as a theory of connections on the tangent bundle $TM \to M$. Now I'm sure you can all write down several non–trivial solutions of the Einstein equations, probably including the Schwarzschild metric, various homogeneous cosmological models, the Kerr metric and perhaps a few others. Likewise, I certainly expect you've solved Maxwell's equations in the presence of any number of weird charge configurations, including electrical circuits, solenoids and all manner of other things. However, I doubt that many of you know any non–trivial solutions of the Yang–Mills equations at all.

Why not? The answer turns out to be revealing[73]. Turning the question around, the reason you *do* know lots of solutions of the Maxwell or Einstein equations is simply that the role these equations play in Nature has been understood for over a century. Their weak field, Newtonian approximations have been known considerably longer, and the phenomena they describe are apparent in everyday life. By contrast, our technology has not yet reached the point where we can perform any experiment in which the classical Yang–Mills equations are relevant.

The reason this is so is an effect known as the *mass gap*. Skipping ahead of our story, the path integral for Yang–Mills will roughly take the form $\int DA\, \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar}$. Because the coupling constant appears only as an overall factor in the Yang–Mills action (7.37), it plays the same role as $\hbar$; the path integral depends on g_{YM} and $\hbar$ only through the combination $\mathrm{g}^2_{\mathrm{YM}}\hbar$ (at least in pure Yang–Mills). Thus we should expect that the quantum theory is well–approximated by the classical limit if $\mathrm{g}_{\mathrm{YM}} \to 0$ so that the theory is 'weakly coupled'. However, we've seen already that coupling constants can run depending on the scale at which we examine the physics. Below, we'll find that in a non-Abelian theory (with not too much matter), the gauge coupling actually *increases* as we view physics at lower and lower energy scales. In this region, Yang–Mills theory is an inherently *quantum* theory and can yield results that are very far from the classical story. For example, the energy

[73] No, it's not that the Yang–Mills equations are 'harder' to solve than the Einstein equations — quite the contrary.

scale at which QCD becomes strongly coupled is ~ 200 MeV. This is accessible to particle accelerators, but at everyday energy scales the role of Yang–Mills theory in Nature cannot be described without talking about the quantum theory.

In fact, when C.N. Yang first presented his work with Mills, he was strongly criticized by Pauli. Pauli had spotted that there is no term $\sim A^2$ in the Yang–Mills action (it would not be compatible with gauge invariance), so quanta of the Yang–Mills field will be massless. He concluded that, as in both electromagnetism and gravity, these particles will be responsible for some long range force[74]. Needless to say, in Nature no other such forces exist: this is why the Yang–Mills equations were only contemplated in the middle of the 20^{th} century. (Yes, strong and weak nuclear forces were known about earlier, but they're certainly not long range.) Pauli's conclusion was correct in the weak coupling approximation, but the classical Yang–Mills equations are a very poor guide to the low–energy physics.

7.2.2 Minimal coupling

We can also write down actions describing the coupling of Yang–Mills theory to charged matter. The simplest of these are the *minimally coupled* actions, which essentially says that you just take your favourite action for uncharged matter, and then replace all derivatives by gauge covariant derivatives.

For example, suppose we have a scalar field ϕ that lives in the adjoint representation of the gauge group so that $\phi \to h\phi h^{-1}$ under a gauge transformation $h(x)$. With minimal coupling, the kinetic terms for this scalar are

$$S_{\text{kin}}[\phi, \nabla] = -\int_M \text{tr}(\nabla\phi \wedge *\nabla\phi) \tag{7.47}$$

and locally on $U \subset M$ this is

$$S_{\text{kin}}[\phi, A] = \frac{1}{2}\int_U g^{\mu\nu}\left(\partial_\mu\phi^{\text{a}} + f^{\text{a}}_{\text{bc}}\, A^{\text{b}}_\mu\, \phi^{\text{c}}\right)\left(\partial_\nu\phi^{\text{a}} + f^{\text{a}}_{\text{de}}\, A^{\text{d}}_\nu\, \phi^{\text{e}}\right)\, \sqrt{g}\, \text{d}^d x \tag{7.48}$$

in terms of a local trivialization, with components $A_\mu = -\text{i}A^{\text{a}}_\mu t_{\text{a}}$ and $\phi = -\text{i}\phi^{\text{a}} t_{\text{a}}$ in the basis $\{t_{\text{a}}\}$. We can also construct potential terms for ϕ of the form

$$S_{\text{pot}}[\phi] = \int_U V(\phi)\, \sqrt{g}\, \text{d}^d x \tag{7.49}$$

where $V(\phi)$ is any gauge invariant polynomial in ϕ. For example, since ϕ is in the adjoint representation we can think of it as a matrix and then a simple choice would be

$$V(\phi) = \frac{m^2}{2}\text{tr}(\phi\phi) + \frac{\lambda}{4!}\text{tr}(\phi\phi\phi\phi) \tag{7.50}$$

where the traces ensure that $V(h\phi h^{-1}) = V(\phi)$.

[74] Recall that in four dimensions, a particle of mass m gives rise to a potential $V(r) \sim \text{e}^{-mr}/r$. For any finite m the resulting force is negligible at distances $\gg 1/m$ from the centre of the potential, but when $m = 0$ the force can be felt right across the Universe.

Similarly, the minimally coupled action for a massive Dirac spinor ψ transforming the in fundamental representation of G and a conjugate spinor transforming in the antifundamental

$$A \mapsto gAg^{-1} - g^{-1}\mathrm{d}g\,, \qquad \psi \mapsto g\psi\,, \qquad \bar\psi \mapsto \bar\psi g^{-1} \tag{7.51}$$

is[75]

$$S_{\mathrm{Dir}}[\psi, \nabla] = \int_{\mathbb{R}^d} \bar\psi(\not\nabla + m)\psi \ \mathrm{d}^d x \tag{7.52}$$

where ∇ acts in the fundamental representation. Explicitly, if $G = SU(N)$ and $i, j = 1, \ldots, N$ label a basis of $\mathbb{C}^N$, then in the coupling to ψ, we write the gauge field as $A_\mu = -\mathrm{i}A^{\mathrm{a}}_\mu t_{\mathrm{a}}$, where the $\{t_a\}$ are a Hermitian basis of $\mathfrak{g}$ in the fundamental representation. We can think of these as $N \times N$ Hermitian matrices, whereupon the action becomes

$$S_{\mathrm{Dir}}[\psi, \nabla] = \int_{\mathbb{R}^d} \bar\psi_i\, \delta^i{}_j(\not\partial + m)\psi^j - \mathrm{i}\bar\psi_i\gamma^\mu(A^{\mathrm{a}}_\mu t_{\mathrm{a}})^i{}_j\psi^i \ \mathrm{d}^d x \tag{7.53}$$

where γ^μ are the Dirac γ-matrices. Note that S_{Dir} reduces to the usual electron action (5.49) in the Abelian case $G = U(1)$ of QED, where the only generator t_{a} is the 1×1 identity matrix.

For simple operators such as $(\partial\phi)^2$ the replacement $(\partial\phi)^2 \to (\nabla\phi)^2$ is unambiguous; once we've declared in which representation ϕ transforms, there is a unique notion of the covariant derivative acting on that representation. However, you may feel uneasy about treating a more complicated operator such as $\partial^\mu\partial^\nu\phi\,\partial_\mu\partial_\nu\phi$. Clearly our replacement prescription should involve $\mathrm{tr}(\nabla^\mu\nabla^\nu\phi\,\nabla_\mu\nabla_\nu\phi)$, but since $[\partial_\mu, \partial_\nu] = 0$ whereas $[\nabla_\mu, \nabla_\nu] \neq 0$, how can we tell whether or not the operator we end up with should include the antisymmetric part of μ and ν? There is no unambiguous way to decide, but fortunately the issue is not very important: whether or not we include such terms in our initial action, if they are allowed by global symmetries then they will in any case be generated by quantum corrections to the effective action.

In the presence of charged matter, the Yang–Mills equations (7.43) are replaced by

$$\nabla^\mu F_{\mu\nu} = -\mathrm{g}^2_{\mathrm{YM}}\, J_\nu \tag{7.54}$$

where $J^\nu(x) = \delta S_{\mathrm{matter}}/\delta a_\nu(x)$ is the matter current. (The Bianchi identity still holds.) Notice that, since the matter action as a whole was invariant and a lives in the adjoint representation, the current J_μ also transforms in the adjoint. For example, in the case of our scalar above we have

$$J^{\mathrm{a}}_\nu = f^{\mathrm{a}}_{\mathrm{bc}}\phi^{\mathrm{b}}(\nabla_\nu\phi)^{\mathrm{c}} \tag{7.55}$$

whereas for the Dirac spinor

$$(J_\nu)^j_i = \mathrm{i}\,\bar\psi_i\gamma_\nu\psi^j \tag{7.56}$$

generalising the electric current $J_\nu = \mathrm{i}\bar\psi\gamma_\nu\psi$ in QED.

[75] I'll write this just for $(M, g) = (\mathbb{R}^d, \delta)$, which will be sufficient for our purposes. To do more we'd first need to discuss how to define spinors on a curved manifold.

In electromagnetism, the adjoint representation was trivial, so the electromagnetic current $J^{\rm em}$ satsified a standard conservation law $\partial^\mu J^{\rm em}_\mu = 0$. In the non–Abelian case, it only makes sense to differentiate J covariantly using ∇ acting in the adjoint representation. Using the equation of motion (7.54) we have

$$-\,\mathrm{g}_{\rm YM}^2\,(\nabla^\nu J_\nu)^{\rm a} = (\nabla^\nu(\nabla^\mu F_{\mu\nu}))^{\rm a} = \frac{1}{2}([\nabla^\nu,\nabla^\mu]F_{\mu\nu})^{\rm a} = \frac{\mathrm{i}}{4} f^{\rm a}_{\rm bc} F^{\rm b\,\mu\nu} F^{\rm c}_{\mu\nu} \tag{7.57}$$

where the second equality follows since $F_{\mu\nu}$ is antisymmetric in (μ,ν), the third equality uses the facts that the commutator of covariant derivatives is F and that these derivatives act on the original F in the adjoint representation. Finally, this expression vanishes by antisymmetry of the structure constants $f^{\rm a}_{\rm bc} = -f^{\rm a}_{\rm cb}$. Thus we have a *covariant conservation law*

$$(\nabla^\mu J_\mu)^{\rm a} = 0 \tag{7.58}$$

for our current in a non–Abelian theory.

7.3 Gauge transformations are not symmetries

The difference between the covariant conservation law (7.58) and the naïve conservation law $\partial^\mu J^{\rm Noether}_\mu = 0$ we found for Noether currents in section 6.1 has a profound consequence. Let's consider a special case of a gauge transformation, where the gauge parameter g is constant. Then matter fields ϕ, ψ and $\bar\psi$ in the adjoint, fundamental and anti-fundamental transform respectively as

$$\phi \mapsto g^{-1}\phi g\,, \qquad \psi \mapsto g\psi\,, \qquad \bar\psi \mapsto \bar\psi g^{-1}\,, \tag{7.59}$$

just as for any gauge transformation. However, the transformation of the

Let's consider a non–Abelian gauge theory with gauge group a compact Lie group G minimally coupled to a massive Dirac fermion ψ transforming in the fundamental representation of G. For the flat space theory, I hope you recall (*e.g.* from the Symmetries, Particles & Fields course) that this means we have a theory with action

$$S[\nabla,\psi,\bar\psi] = \frac{1}{2\mathrm{g}_{\rm YM}^2}\int_{\mathbb{R}^d} \mathrm{tr}(F^{\mu\nu}F_{\mu\nu})\,\mathrm{d}^d x + \int_{\mathbb{R}^d} \mathrm{i}\bar\psi(\not{D}+m)\psi\,\mathrm{d}^d x \tag{7.60}$$

where $\mathrm{g}_{\rm YM}^2$ is a coupling constant and

$$\begin{aligned} \not{D}\psi &= \gamma^\mu(\partial_\mu\psi + \mathrm{i}A_\mu\psi) \\ F_{\mu\nu} &= \Big(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^a_{bc}A^b_\mu A^c_\nu\Big)\,t_a \end{aligned} \tag{7.61}$$

is the Yang–Mills fieldstrength, with $\{t_a\}$ a Hermitian basis of the Lie algebra $\mathfrak{g}$ of G, normalised so that $\mathrm{tr}(t_a t_b) = \delta_{ab}/2$. Under a gauge transform with parameter $g(x) \in G$ we have

$$\begin{aligned} A \mapsto A^g &= gAg - \mathrm{d}g\,g^{-1} & \qquad F \mapsto F^g &= gFg^{-1} \\ \psi \mapsto \psi^g &= g\psi & \qquad \bar\psi \mapsto \bar\psi^g &= \bar\psi g^{-1}\,. \end{aligned} \tag{7.62}$$