8 Non-Abelian Gauge Theory: Perturbative Quantization

We're now ready to consider the quantum theory of Yang–Mills. In the first few sections, we'll treat the path integral formally as an integral over infinite dimensional spaces, without worrying about imposing a regularization. We'll turn to questions about using renormalization to make sense of these formal integrals in section 8.3.

To specify Yang–Mills theory, we had to pick a principal G bundle $P \to M$ together with a connection ∇ on P. So our first thought might be to try to define the Yang–Mills partition function as

$$Z_{\mathrm{YM}}[(M,g),\mathrm{g}_{\mathrm{YM}}] \stackrel{?}{=} \int_{\mathcal{A}} \mathcal{D}A\ \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar} \tag{8.1}$$

where

$$S_{\mathrm{YM}}[\nabla] = -\frac{1}{2\mathrm{g}_{\mathrm{YM}}^2}\int_M \mathrm{tr}(F_\nabla \wedge *F_\nabla) \tag{8.2}$$

as before, and $\mathcal{A}$ is the space of all connections on P.

To understand what this integral might mean, first note that, given any two connections ∇ and ∇', the 1-parameter family

$$\nabla^\tau = \tau\nabla + (1-\tau)\nabla' \tag{8.3}$$

is also a connection for all $\tau \in [0,1]$. For example, you can check that the *rhs* has the behaviour expected of a connection under any gauge transformation. Thus we can find a path in $\mathcal{A}$ between any two connections. Since $\nabla' - \nabla \in \Omega^1_M(\mathfrak{g})$, we conclude that $\mathcal{A}$ is an infinite dimensional affine space whose tangent space at any point is $\Omega^1_M(\mathfrak{g})$, the infinite dimensional space of all $\mathfrak{g}$–valued covectors on M. In fact, it's easy to write down a flat (L^2-)metric on $\mathcal{A}$ using the metric on M:

$$\mathrm{d}s^2_{\mathcal{A}} = -\int_M \mathrm{tr}(\delta A \wedge *\,\delta A) = \frac{1}{2}\int_M g^{\mu\nu}\,\delta A^{\mathrm{a}}_\mu\,\delta A^{\mathrm{a}}_\nu\ \sqrt{g}\,\mathrm{d}^d x\ . \tag{8.4}$$

In other words, given any two tangent vectors $(a_1,a_2) \in \Omega^1_M(\mathfrak{g})$ at the point $\nabla \in \mathcal{A}$,

$$\mathrm{d}s^2_{\mathcal{A}}(a_1,a_2) = -\int_M \mathrm{tr}(a_1 \wedge *a_2)\,, \tag{8.5}$$

independent of where in $\mathcal{A}$ we are. This is encouraging: $\mathcal{A}$ just looks like an infinite dimensional version of $\mathbb{R}^n$, with no preferred origin since there is no preferred connection on P.

We might now hope that the path integral (8.1) means formally that we should pick an arbitrary base–point $\nabla_0 \in \mathcal{A}$, then write any other connection $\nabla = \nabla_0 + A$, with the measure DA indicating that we integrate over all $A \in \Omega^1_M(\mathfrak{g})$ using the translationally invariant measure on $\mathcal{A}$ associated to the flat metric (8.4). (Such an infinite dimensional flat measure does not exist — we're delaying this worry for now.) For a connection $\nabla = \nabla_0 + A$, the action becomes

$$\begin{aligned} S_{\mathrm{YM}}[\nabla] &= -\frac{1}{2\mathrm{g}_{\mathrm{YM}}^2}\int_M \mathrm{tr}(F_\nabla \wedge *F_\nabla) \\ &= -\frac{1}{2\mathrm{g}_{\mathrm{YM}}^2}\int_M \mathrm{tr}(F_{\nabla_0} \wedge *F_{\nabla_0}) \ -\ \frac{1}{2\mathrm{g}_{\mathrm{YM}}^2}\int_M \mathrm{tr}(\nabla_0 A + A^2)\wedge *(\nabla_0 A + A^2)\,. \end{aligned} \tag{8.6}$$

For example, on a topologically trivial bundle a standard choice would be to pick the trivial connection $\nabla_0 = d$ as base–point. Then $F_{\nabla_0} = 0$ and the action takes the familiar form

$$S_{\mathrm{YM}}[d + A] = -\frac{1}{2\mathrm{g}_{\mathrm{YM}}^2} \int_M \mathrm{tr}(dA + A^2) \wedge *(dA + A^2) \tag{8.7}$$

as above. The path integral (8.1) would be interpreted as an integral over all gauge fields A. However, in some circumstances we'll meet later (even when P is topologically trivial), it will be useful to choose a different base–point ∇_0 for which $F_{\nabla_0} \neq 0$, known as a *background field*. In this case, the first term on the *rhs* of (8.6) is the action for the background field and comes out of the path integral as an overall factor, while the remaining action for A involves the covariant derivative with respect to the background field.

Of course, there's a problem. By construction, the Yang–Mills action was invariant under gauge transformations, so the integrand in (8.1) is degenerate along gauge orbits. Consequently, the integral will inevitably diverge because we're vastly overcounting. You met this problem already in the case of QED during the Michaelmas QFT course. There, as here, the right thing to do is to integrate just over physically inequivalent connections — those that are *not* related by a gauge transform. In other words, the correct path integral for Yang–Mills should be of the form

$$Z_{\mathrm{YM}}[(M, g), \mathrm{g}_{\mathrm{YM}}] = \int_{\mathcal{A}/\mathcal{G}} D\mu \; \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar} \tag{8.8}$$

where $\mathcal{G}$ is the space of all gauge transformations, so that $\mathcal{A}/\mathcal{G}$ denotes the space of all gauge equivalence classes of connections: we do not count as different two connections that are related by a gauge transformation. Note that this definition means *gauge 'symmetry' does not exist* in Nature! We've taken the *quotient* by gauge transformations in constructing the path integral, so the resulting object has no knowledge of any sort of gauge transformations. They were simply a *redundancy* in our construction. The same conclusion holds if we compute correlation functions of any gauge invariant quantities, whether they be local operators built from gauge invariant combinations of matter fields, or Wilson loops running around some curves in space.

However, we're not out of the woods. Whilst $\mathcal{A}$ itself was just an affine space, the space $\mathcal{A}/\mathcal{G}$ is much more complicated. For example, it has highly non–trivial topology investigated by Atiyah & Jones, and by Singer. Certainly $\mathcal{A}/\mathcal{G}$ is *not* affine, so we don't yet have any understanding of what the right measure $D\mu$ to use on this space is, even formally. In the case of electrodynamics, you were able to avoid this problem (at least in perturbation theory on $\mathbb{R}^4$) by picking a gauge, defining the photon propagator and just getting on with it. The non–linear structure of the non–Abelian theory means we'll have to consider this step in more detail.

8.1 A ghost story

The way to proceed was found by Feynman, de Witt, and by Faddeev & Popov. To understand what they did, let's warm up with a finite–dimensional example.

Suppose we have a function $S : \mathbb{R}^2 \to \mathbb{R}$ defined at any point on the (x, y)-plane, and suppose further that this function is invariant under rotations of the plane around the origin. We think of $S(x, y)$ as playing the role of our 'action' for 'fields' (x, y), while rotations represent 'gauge transformations' leaving this action invariant. Of course, rotational invariance implies that $S(x, y) = h(r)$ in this example, where $h(r)$ is some function of the radius. We easily compute

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mathrm{e}^{-S(x,y)/\hbar} = 2\pi \int_0^\infty \mathrm{d}r\, r\,\mathrm{e}^{-h(r)/\hbar} \tag{8.9}$$

which will make sense for sufficiently well–behaved $f(r)$. The factor of $2\pi = \mathrm{vol}(\mathrm{SO}(2))$ appears here because the original integral was rotationally symmetric: it represents the redundancy in the expression on the left of (8.9).

In the case of Yang–Mills, if we integrated over the space $\mathcal{A}$ of all connections rather than over $\mathcal{A}/\mathcal{G}$, the redundancy would be infinite: while the volume $\mathrm{vol}(G)$ computed using the Haar measure on G is finite if the structure group G is compact, the volume $\mathrm{vol}(\mathcal{G})$ of the space of all gauge transformations is infinite — heuristically, you can think of it as a copy of $\mathrm{vol}(G)$ at each point of M. What we'd like to do is understand how to keep the analogue of $\int_0^\infty \mathrm{d}r\, r\,\mathrm{e}^{-h(r)}$ in the gauge theory case, without the redundancy factor. However, neither the right set of gauge invariant variables (analogous to r) nor the right measure on $\mathcal{A}/\mathcal{G}$ (generalizing $r\,\mathrm{d}r$) are obvious in the infinite dimensional case.

Returning to (8.9), suppose C is any curve traveling out from the origin that intersects every circle of constant radius exactly once. More specifically, let $f(\mathbf{x})$ be some function such that i) for any point $\mathbf{x} \in \mathbb{R}^2$ there exists a rotation $R \in \mathrm{SO}(2)$ so that $f(R\mathbf{x}) = 0$, and ii) f is non–degenerate on the orbits; *i.e.*, $f(R\mathbf{x}) = f(\mathbf{x})$ iff R is the identity[76] in $\mathrm{SO}(2)$. The curve

$$C = \{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) = 0\}$$

then intersects every orbit of the rotation group exactly once, and we can think of $C \subset \mathbb{R}^2$ as a way to embed the orbit space $(\mathbb{R}^2 - \{0\})/\mathrm{SO}(2) \cong \mathbb{R}_>$ in the plane (see figure 11). This non–degeneracy property means that $f(\mathbf{x})$ itself is certainly not rotationally invariant. In anticipation of the application to Yang–Mills theory, we call $f(\mathbf{x})$ the **gauge fixing function** and the curve C it defines the **gauge slice**.

Now consider the integral

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\delta(f(\mathbf{x}))\,\mathrm{e}^{-S(x,y)/\hbar} \tag{8.10}$$

over all of $\mathbb{R}^2$. Clearly, the δ-function restricts this integral to the gauge slice. However, the actual value we get depends on our choice of specific function $f(\mathbf{x})$; for example, even replacing $f \to cf$ for some constant c (an operation which preserves the curve C) reduces the integral by a factor of $1/|c|$. Thus we cannot regard (8.10) as an integral over the moduli space $(\mathbb{R}^2 - \{0\})/\mathrm{SO}(2)$ — it also depends on exactly how we embedded this moduli space inside $\mathbb{R}^2$.

[76] Technically, we should restrict to $\mathbb{R}^2 - \{0\}$ to ensure this condition holds. For smooth functions $S(x, y)$ this subtlety won't affect our results and I'll ignore it henceforth.

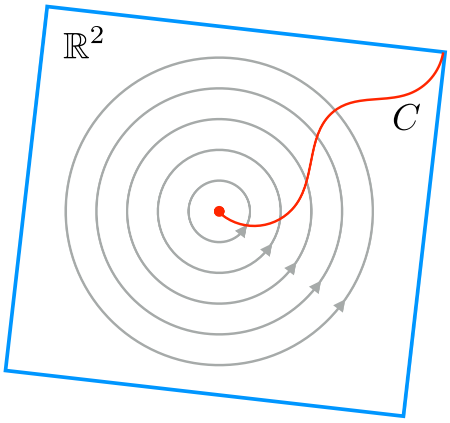


Figure 11: *The gauge slice C should be chosen to be transverse to the gauge orbits, which for SO(2) are the circles shown in grey.*

The problem arose because the δ-function changes as we change $f(\mathbf{x})$. To account for this, define

$$\Delta_f(\mathbf{x}) = \frac{\partial}{\partial\theta} f(R_\theta \mathbf{x})\bigg|_{\theta=0} \tag{8.11}$$

where the right hand side means we compute the rate of change of f with respect to a rotation R_θ through angle θ, evaluated at the identity $\theta = 0$. Notice that we only need to know how an *infinitesimal* rotation acts in order to compute this. It's clear that the new integral

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|\Delta_f(\mathbf{x})|\,\delta(f(\mathbf{x}))\,\mathrm{e}^{-S(x,y)/\hbar} \tag{8.12}$$

involving the modulus of Δ_f doesn't change if we rescale f by a constant factor as above. Nor does it change if we rescale f by a non–zero, r-dependent factor $c(r)$, which means that (8.12) is completely independent of the choice of function used to define the gauge slice C. In fact, I claim that (8.12) is actually *independent of the particular gauge slice itself*. To see this, let f_1 and f_2 be any two different gauge–fixing functions. Since the curves $C_{1,2}$ they define each intersect every orbit of SO(2) uniquely, we can always rotate C_1 into C_2, provided we allow ourselves to rotate by different amounts at different values of the radius r. Thus we must have

$$f_2(\mathbf{x}) \propto f_1(R_{12}(r)\mathbf{x}) \tag{8.13}$$

for some r-dependent rotation $R_{12}(r)$ and where the proportionality factor depends at most on the radius. By rescaling invariance,

$$|\Delta_{f_2}(\mathbf{x})|\,\delta(f_2(\mathbf{x})) = |\Delta_{f_1}(\mathbf{x}')|\,\delta(f_1(\mathbf{x}')) \tag{8.14}$$

where we've defined $\mathbf{x}' := R_{12}\mathbf{x}$ for any point $\mathbf{x} \in \mathbb{R}^2$, whether it lies on our curves or not. Now, the statement that the action $S(\mathbf{x})$ is rotationally invariant means that it takes the same value all around every circle of constant radius, so $S(\mathbf{x}) = S(\mathbf{x}')$. Similarly,

$$\mathrm{d}x'\,\mathrm{d}y' = \mathrm{d}x\,\mathrm{d}y \tag{8.15}$$

because again this measure is rotationally invariant at every value of r [77]. Putting all this together, the integral in (8.12) is independent of the choice of gauge slice $\mathbb{R}_> \hookrightarrow \mathbb{R}^2$, as we wished to show.

As a concrete example, suppose we choose C to be the x-axis, defined by the zero–set of $f(\mathbf{x}) = y$. With this choice, $f(R_\theta \mathbf{x}) = y\cos\theta - x\sin\theta$ where R represents anti-clockwise rotation through θ. Thus

$$\Delta_f(\mathbf{x}) = \frac{\partial}{\partial\theta}(y\cos\theta - x\sin\theta)\bigg|_{\theta=0} = -x \tag{8.16}$$

and therefore our integral (8.12) becomes

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|\Delta_c(\mathbf{x})|\,\delta(f(\mathbf{x}))\,\mathrm{e}^{-S(x,y)/\hbar} = \int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|x|\,\delta(y)\,\mathrm{e}^{-S(x,y)/\hbar} = \int_{\mathbb{R}} \mathrm{d}x\,|x|\,\mathrm{e}^{-h(|x|)\hbar} \tag{8.17}$$

where in the last step we've used the fact that since $S(x,y)$ was rotationally invariant, along the line $y=0$ it can only depend on $|x|$. Since $|x|$ is an even function of x, we have

$$\int_{\mathbb{R}} \mathrm{d}x\,|x|\,\mathrm{e}^{-h(|x|)/\hbar} = 2\int_0^\infty \mathrm{d}r\,r\,\mathrm{e}^{-h(r)/\hbar}\,, \tag{8.18}$$

which disagrees with the radial part of our original integral by a factor of 2. What's gone wrong is that circles of constant r intersect the x-axis twice — when $x>0$ and when $x<0$ — and our gauge fixing condition $y=0$ failed to account for this; in other words, it slightly failed the non–degeneracy property. We'll see below that this glitch is actually a model of something that also happens in the case of Yang–Mills theory.

To recap, what we've achieved with all this is that, for any non–degenerate gauge–fixing function f, we can write the desired integral over the space of orbits $(\mathbb{R}^2 - \{0\})/\mathrm{SO}(2) \cong \mathbb{R}_>$ as

$$\int_{\mathbb{R}_>} \mathrm{d}r\,r\,\mathrm{e}^{-h(r)/\hbar} = \frac{1}{|Z_{\mathrm{SO}(2)}|}\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|\Delta_f(\mathbf{x})|\,\delta(f(\mathbf{x}))\,\mathrm{e}^{-S(x,y)/\hbar}\,. \tag{8.19}$$

where $|Z_{\mathrm{SO}(2)}| = 2$ is the number of elements of the centre of SO(2). The point is that the expression on the *rhs* refers only to functions and coordinates on the affine space $\mathbb{R}^2$, and uses only the standard measure $\mathrm{d}x\,\mathrm{d}y$ on $\mathbb{R}^2$. When the gauge orbits have dimension >1 we must impose several gauge fixing conditions f^{a}, one for each transformation parameter θ^{a}. Then we take the integral to include a factor

$$|\Delta_f(\mathbf{x})|\prod_{\mathrm{a}}\delta(f^{\mathrm{a}}(\mathbf{x})) \tag{8.20}$$

[77] Again, it's a good idea to check you're comfortable with this assertion by writing

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha(r) & \sin\alpha(r) \\ -\sin\alpha(r) & \cos\alpha(r) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

and explicitly working out the transformation of the measure, allowing for the fact that the angle $\alpha(r) = \alpha(\sqrt{x^2+y^2})$ depends on the radius. You'll find the measure is nonetheless invariant.

where now Δ_f is the **Faddeev–Popov determinant**

$$\Delta_f(\mathbf{x}) = \det\left(\frac{\partial f^{\mathrm{a}}(R_\theta \mathbf{x})}{\partial \theta^{\mathrm{b}}}\right) \tag{8.21}$$

for a generic set of variables $\mathbf{x} \in \mathbb{R}^n$ where the action is invariant under some transformation $\mathbf{x} \to R_\theta \mathbf{x}$ (not necessarily a rotation) depending on parameters θ^{a}. Again, this will allow us to write an integral over the space of orbits of these transformations in terms of an integral over the affine space $\mathbb{R}^n$. These are things we have access to in the gauge theory case[78] where the affine space in question is the space $\mathcal{A}$ of all gauge fields, and the transformation group is the space $\mathcal{G}$ of all gauge transformations. Armed with these ideas, we now turn to the case of gauge theory.

In Yang–Mills theory, we can fix the gauge redundancy by picking a particular connection in each gauge equivalence class — in other words, by picking an embedding of $\mathcal{A}/\mathcal{G} \hookrightarrow \mathcal{A}$ specified by some gauge–fixing functional $f[A]$. The most common choices of gauge fixing functional are *local*, in the sense that $f[A]$ depends on the value of the gauge field just at a single point $x \in M$. Heuristically, we then restrict to $f[A(x)] = 0$ at every point $x \in M$ by inserting "$\delta[f] = \prod_{x\in M} \delta(f[A(x)])$" in the path integral. We'll consider how to interpret this infinite–dimensional δ-function below. The Faddeev–Popov determinant is then

$$\Delta_f = \det \frac{\delta f^{\mathrm{a}}[A^\lambda(x)]}{\delta \lambda^{\mathrm{b}}(y)}\,, \tag{8.22}$$

where $A^\lambda = A + \nabla\lambda$ denotes an infinitesimal gauge transformation of A with parameters $\lambda^a(x)$ valued in the adjoint. Like the δ-functional $\delta[f]$, this determinant is now of an infinite dimensional matrix; we'll consider what this determinant means momentarily. With these ingredients, our Yang–Mills path integral can be written as

$$\int_{\mathcal{A}/\mathcal{G}} D\mu\ \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar} = \int_{\mathcal{A}} DA\ |\Delta_f|\,\delta[f]\,\mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar}\,, \tag{8.23}$$

where the factor of $|\Delta_f|\,\delta[f]$ restricts us to an arbitrary gauge slice, but leaves no dependence on any particular choice of slice, as above. Again, the advantage of the rhs is that it refers only to the naïve integral measure over all connections.

For some purposes, especially in perturbation theory, it's useful to rewrite the $\delta[f]$ and Δ_f factors in a way that makes them amenable to treatment via Feynman diagrams. Taking our lead from Fourier analysis, we introduce a new bosonic scalar field h (sometimes called a Nakanishi–Lautrup field) and write

$$\delta[f] = \int Dh\,\mathrm{e}^{-S_{\mathrm{gf}}[h,\nabla]/\hbar}\,, \tag{8.24}$$

where

$$S_{\mathrm{gf}}[h,\nabla] = -\mathrm{i}\int_M \mathrm{tr}(h * f[A]) = \frac{\mathrm{i}}{2}\int_M h^{\mathrm{a}} f^{\mathrm{a}}[A]\,\sqrt{g}\,\mathrm{d}^d x \tag{8.25}$$

[78] Modulo, as always, the problem that there is no Lebesgue measure on $\mathcal{A}$: this is what we'll treat with renormalization.

is the **gauge–fixing action**. The idea is that h is a Lagrange multiplier — performing the path integral over h imposes $f[A(x)] = 0$ throughout M. Notice that since we needed one gauge–fixing condition for every gauge parameter, we take h to lie in the adjoint representation, $h \in \Omega^0_M(\mathfrak{g})$. This does *not* imply that (8.24) is gauge invariant: indeed it *cannot* be if it is to fix a gauge! For the Faddeev–Popov determinant Δ_f, recall that if M is an $n \times n$ matrix and $(c^i, \bar{c}_j)$ are n-component Grassmann variables, then $\det(M) = \int \mathrm{d}^n c\, \mathrm{d}^n \bar{c}\, \exp(\bar{c}_j M^j{}_i c^i)$. Applying the same idea here, we have[79]

$$\det \frac{\delta f^{\mathrm{a}}[A^\lambda(x)]}{\delta \lambda^{\mathrm{b}}(y)} = \int Dc\, D\bar{c}\; \mathrm{e}^{-S_{\mathrm{gh}}[\bar{c},c,\nabla]/\hbar} \tag{8.26}$$

where

$$S_{\mathrm{gh}}[\bar{c}, c, \nabla] = \int_{M\times M} \bar{c}^{\mathrm{a}}(x) \frac{\delta f^{\mathrm{a}}[A^\lambda(x)]}{\delta \lambda^{\mathrm{b}}(y)} c^{\mathrm{b}}(y) \tag{8.27}$$

and the fields $(c^{\mathrm{a}}, \bar{c}^{\mathrm{a}})$ are fermionic scalars, again valued in the adjoint representation of G. They're known as **ghosts** (c) and **antighosts** ($\bar{c}$). Putting everything together, our Yang–Mills path integral can finally be written as

$$\int_{\mathcal{A}/\mathcal{G}} \mathcal{D}\mu\; \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]/\hbar} = \int \mathcal{D}A\, \mathcal{D}c\, \mathcal{D}\bar{c}\, \mathcal{D}h\; \exp\left(-\frac{1}{\hbar}(S_{\mathrm{YM}}[\nabla] + S_{\mathrm{gh}}[\bar{c}, c, \nabla] + S_{\mathrm{gf}}[h, \nabla])\right) \tag{8.28}$$

where the integral on the *rhs* is formally to be taken over the space of all fields $(\nabla, \bar{c}, c, h)$. Everything on the right is now written in terms of an integral over the naıve, affine space of all connections, together with the space of ghost, antighost and Nakanishi–Lautrup fields, weighted by some action $S[A, c, \bar{c}, h]$. Thus we can hope to compute it perturbatively using Feynman rules.

Let's now make all this more concrete by seeing how it works in an example. An important, frequently occurring choice of gauge is *Lorenz*[80] *gauge*: we pick the trivial connection ∂ as a base-point by writing $\nabla = \partial + A = \partial - \mathrm{i}A^{\mathrm{a}} t_{\mathrm{a}}$, and impose that the A^{a} obey

$$f^{\mathrm{a}}[A] = \partial^\mu A^{\mathrm{a}}_\mu(x) = 0 \tag{8.29}$$

for all $x \in M$ and for all $\mathrm{a} = 1, \ldots, \dim(G)$. An obvious reason to want to work in Lorenz gauge is that in the important case $(M, g) = (\mathbb{R}^d, \delta)$, it respects the $\mathrm{SO}(d)$ invariance of the flat Euclidean metric. Under an infinitesimal gauge transformation with parameters λ^{a}, the gauge field transforms as

$$A^{\mathrm{a}}_\mu \mapsto A^{\mathrm{a}}_\mu + \partial\lambda^{\mathrm{a}} + f^{\mathrm{a}}_{\mathrm{cd}} A^{\mathrm{c}}_\mu \lambda^{\mathrm{d}} \tag{8.30}$$

so the gauge variation of the Lorenz gauge–fixing condition is

$$\begin{aligned} f^{\mathrm{a}}[A^\lambda] &= \partial^\mu (A^{\mathrm{a}}_\mu + \partial\lambda^{\mathrm{a}} + f^{\mathrm{a}}_{\mathrm{cd}} A^{\mathrm{c}}_\mu \lambda^{\mathrm{d}}) \\ &= \partial^\mu A^{\mathrm{a}}_\mu + \partial^\mu(\partial_\mu \lambda^{\mathrm{a}} + f^{\mathrm{a}}_{\mathrm{cd}} A^{\mathrm{c}}_\mu \lambda^{\mathrm{d}})\,. \end{aligned} \tag{8.31}$$

[79]One can show that the determinant is positive–definite, at least in a neighbourhood of the trivial connection. Thus, for the purposes of perturbation theory around the trivial background, we can drop the modulus sign. Non–perturbatively we must be more careful.

[80]Poor Ludvig Lorenz. Eternally outshone by Hendrik Lorentz to the point of having his work misattributed.

Consequently, the matrix appearing in the Faddeev–Popov determinant is

$$\begin{aligned}\frac{\delta f^{\rm a}[A^\lambda(x)]}{\delta\lambda^b(y)} &= (\partial^\mu\partial_\mu\delta^{\rm a}_{\rm b} + f^{\rm a}_{\rm cb}A^{\rm c}_\mu)\,\delta^d(x-y)\\ &= (\partial^\mu\nabla_\mu)^{\rm a}{}_{\rm b}\,\delta^d(x-y)\end{aligned}\tag{8.32}$$

where the differential operator $\partial^\mu\nabla_\mu$ acts formally on the variable x appearing in the δ-functions[81]. These δ-functions arose because our gauge–fixing condition was local: the object $\partial^\mu(\nabla_\mu\lambda)$ lives at one point x, so we get nothing if we vary it wrt to changes in λ at some other point. Using this result in the ghost action yields

$$\begin{aligned}S_{\rm gh}[\bar c, c,\nabla] &= -\int_{M\times M}\bar c_{\rm a}(x)\,(\partial^\mu\nabla_\mu)^{\rm a}{}_{\rm b}\,\delta^d(x-y)\,c^{\rm b}(y)\,{\rm d}^dy\,{\rm d}^dx\\ &= -\int_M \bar c_{\rm a}(\partial^\mu\nabla_\mu c)^{\rm a}\,{\rm d}^dx = \int_M(\partial^\mu\bar c_{\rm a})(\nabla_\mu c)^{\rm a}\,{\rm d}^dx\\ &= -2\int_M {\rm tr}\,(d\bar c\wedge *\nabla c)\end{aligned}\tag{8.33}$$

where in the first step we integrated out y using the δ-function, recalling that the differential operators only care about x. Altogether, in Lorenz gauge we have the action

$$S[\nabla,c,\bar c,h] = -\frac{1}{2{\rm g}^2_{\rm YM}}\int_M {\rm tr}\,(F\wedge *F) - 2\int_M {\rm tr}\,(d\bar c\wedge *\nabla c) - {\rm i}\int_M {\rm tr}\,(h\ d{*}A)\tag{8.34}$$

Except for the strange spin/statistics of the ghost fields and the mixture of covariant and normal derivatives, this is now a perfectly respectable, local action for scalar fields coupled to the gauge field. Notice that in the Abelian case where the adjoint representation is trivial, the Lorenz gauge ghost action would be $\int_M d\bar c\wedge *dc$ and in particular would be independent of the gauge field A. Thus the ghosts would have completely decoupled, which is why you didn't meet them last term.

8.1.1 Can we fix a gauge?

Recall that in the finite dimensional case, we needed our gauge fixing function to obey two conditions: that we can indeed always find a gauge transformation such that $f[A]=0$ holds, and that once we've found it, this gauge is unique so that in particular starting on the gauge slice and performing any gauge transformation takes us off the slice.

Let's start by considering whether we *can* always solve (8.29). In other words, let $\nabla\in\mathcal{A}$ be some arbitrary connection and let $\mathcal{G}(\nabla)$ denote the orbit of ∇ under $\mathcal{G}$. Then we must show that there is always *some* $\nabla'\in\mathcal{G}(\nabla)$ whose connection 1-form obeys (8.29). You've seen how to do this in the Abelian case of electrodynamics: you noted that under a gauge transform, $A\to A' = A - d\lambda$ for some λ. The condition that A be in Lorenz gauge says

$$0 = \partial^\mu A'_\mu = \partial^\mu A_\mu - \Delta\lambda\tag{8.35}$$

where Δ is the Laplacian on M. Regarding this as a condition on λ, we must solve $\Delta\lambda = u$ where $u(x) = \partial^\mu A_\mu(x)$ is essentially arbitrary. This can always be done provided

[81]Recall that, when treated properly as distributions, δ-functions are infinitely differentiable!

u is orthogonal to the kernel of the adjoint of the Laplacian in the L^2 norm on (M, g). However, the Laplacian is self–adjoint ($\int_M \phi * \Delta\psi = \int_M (\Delta\phi) * \psi$) and $\ker \Delta$ consists of constant functions, since if $u \in \ker \Delta$ then

$$0 = \int_M u * \Delta u = -\int_M du \wedge *du = -\|du\|^2 \tag{8.36}$$

whenever u has compact support. Hence $du = 0$ so u is constant. Thus, for any generic electromagnetic potential A_μ, we can find a gauge transform that puts it in Lorenz gauge. In the non–Abelian case, things are more complicated because the gauge transform of a connection is non–linear: $A \to A + \nabla_A \lambda$, whose value depends on which A we start with. It turns out that we can always solve (8.29), but the proof is considerably more complicated — one was found by Karen Uhlenbeck in 1982 (at least for some common choices of M), and an alternative proof was later found by Simon Donaldson.

We must still check that (8.29) singles out a *unique* representative, so that we count each gauge equivalence class only once. Encouragement comes from the fact that connections obeying (8.29) are orthogonal to connections that are pure gauge, with respect to the L^2-metric (8.4) on $\mathcal{A}$. For if a_1 is a tangent vector at the point $\nabla \in \mathcal{A}$ that obeys $\nabla * a_1 = 0$, while $a_2 = \nabla\lambda$ is also a tangent vector at ∇ that points in the direction of an infinitesimal gauge transform, then

$$ds^2_{\mathcal{A}}(a_1, a_2) = \int_M \mathrm{tr}(a_1 \wedge *a_2) = \int_M \mathrm{tr}(a_1 \wedge *\nabla\lambda) = -\int_M \mathrm{tr}((\nabla * a_1)\,\lambda) = 0 \tag{8.37}$$

using the Lorenz condition. Thus changing our connection in a way that obeys Lorenz gauge takes us in a direction that is *orthogonal* to the orbits of the gauge group. This certainly shows that starting from any base-point and integrating over all gauge fields along the slice incorporates only gauge inequivalent connections while we're *near* our base-point.

However, as in the finite dimensional example where the line $y = 0$ intersected each circle on constant radius twice, it doesn't guarantee that some other connection, far away along the gauge slice, isn't secretly gauge equivalent to one we've already accounted for. This troubling possibility is known as the **Gribov ambiguity**, after Vladimir Gribov who first pointed it out and showed it actually occurs in the case of Coulomb gauge $\partial^i A_i = 0$ (the indices just run over $\mathbb{R}^3 \subset \mathbb{R}^{3,1}$). Somewhat later, Iz Singer showed that the Gribov ambiguity is in fact *inevitable*: no matter which gauge condition you pick, the gauge orbit *always* intersects the gauge slice more than once (at least for most reasonable M). To show this, Singer noted that $\mathcal{A}$ is itself an infinite dimensional principle bundle over the space $\mathcal{B} := \mathcal{A}/\mathcal{G}$ where the group $\mathcal{G}$ of all gauge transformations plays the role of the *structure* group. A gauge slice amounts to a global section of this bundle — *i.e.*, the choice of a unique point in $\mathcal{A}$ for each point in $\mathcal{B}$. A result I won't prove states that a principal bundle only admits a global section when it's topologically trivial, so the existence of a global gauge slice would imply

$$\mathcal{A} \overset{?}{\cong} \mathcal{B} \times \mathcal{G}\,. \tag{8.38}$$

Since $\mathcal{A}$ is an affine space, clearly $\pi_k(\mathcal{A}) = 0$ for all $k > 0$ (*i.e.* $\mathcal{A}$ itself is topologically trivial and has no non–contractible cycles). However, Singer computed that $\pi_k(\mathcal{G}) \neq 0$

for at least some $k > 0$ which says that there *are* some non–contractible cycles in the space on the *rhs* of (8.38). Thus $\mathcal{A} \neq \mathcal{B} \times \mathcal{G}$, so $\mathcal{A}$ is non–trivial as a principal bundle over $\mathcal{B}$, and no global gauge choice exists. In practice, we'll work perturbatively, meaning we never venture far enough from our chosen base-point connection to meet any Gribov copies. Non–perturbatively, we'd have to cover $\mathcal{A}/\mathcal{G}$ with different coordinate patches, pick different gauges in each one and then piece them together at the end. I'm not aware of anyone actually trying to do this.

8.2 BRST transformations

The ghosts and gauge–fixing terms were introduced in order to allow us to integrate over the naïve affine space $\mathcal{A}$ of all connections, whilst still removing the gauge redundancy inherent in our description of Yang–Mills theory. As in electromagnetism, only the $d-2$ transverse polarization states of the gluon should be physical, but our action describes these gluons in terms of A_μ, which has d components. We've seen how the ghosts and Nakanishi–Lautrup fields are responsible for imposing a gauge in the path integral, thus removing the unphysical states. However, from at least two points of view, we appear to have made matters *worse*.

Firstly, we expect that renormalization group flow will generate an infinite series of interactions — every possible term that is not forbidden by symmetries of the original action and regularized measure — all but finitely many of which are strongly suppressed at low energies. The gauge–fixing and ghost terms in the previous section are *not* invariant under gauge transformations, so how can we guarantee that all manner of gauge non–invariant terms will not be generated in the quantum effective action? In particular, what's to stop *e.g.* a mass term[82] $\sim \operatorname{tr} A_\mu A^\mu$ from being generated automatically in the quantum theory?

Secondly, our path integral is not just over the space of connections, but also over the space of all ghost, antighost and Nakanishi–Lautrup fields. On a manifold with boundary, or a non-compact manifold such as $\mathbb{R}^d$, we must specify boundary or asymptotic conditions not just on the gauge field A_μ, but also on $\{c, \bar{c}, h\}$. Thus, in the extended theory our states $\Psi[A, \bar{c}, c, h]$ will depend on the boundary values of all these fields, rather than just the states $\Psi[A]$ we expect in pure gauge theory. Worse still, this extended space of states $\mathcal{H}_0$ cannot be a Hilbert space, because the fermionic scalars $\{\bar{c}, c\}$ violate the spin–statistics theorem, leading to a non–unitary theory. The unphysical, longitudinally polarized gauge bosons likewise have negative norm.

To resolve these puzzles, we'll need to look at a remarkable set of transformations that turn out to be symmetries of the full, gauge–fixed action (8.34). These are known as **BRST transformations**, after their discoverers Becchi, Rouet, Stora and (independently) Tyurin. They act on the fields as

$$\begin{aligned} \delta A_\mu &= \epsilon(\nabla_\mu c) \qquad\qquad & \delta\bar{c} &= \mathrm{i}\epsilon h \\ \delta c &= -\frac{\epsilon}{2}[c, c] \qquad\qquad & \delta h &= 0 \end{aligned} \tag{8.39}$$

[82]Here I mean just a naïve perturbative mass term for the gauge boson (or gluon), not the more sophisticated appearance of a mass gap in the non–perturbative theory, or phase transition to a massive theory by means of the Higgs mechanism.

where ϵ is a constant, Grassmann (*i.e.*, fermionic) parameter. Note that $[c,c] = -\mathrm{i}f^{\mathrm{a}}_{\mathrm{bc}}c^{\mathrm{b}}c^{\mathrm{c}}t_{\mathrm{a}}$ is in fact *symmetric* in the two ghost fields because exchanging them produces *two* minus signs: one from the Lie bracket and one from the Grassmann statistics. Letting Ψ^i denote any of the fields $\{A_\mu, c, \bar{c}, h\}$, we'll often write these transformations as $\delta\Psi = \epsilon\,\mathcal{Q}\Psi$ so that $\mathcal{Q}\Psi$ represents the *rhs* of (8.39) with the anticommuting parameter ϵ stripped away. $\mathcal{Q}\Psi^i$ thus has opposite statistics to Ψ^i itself.

The expression for δA shows that, as far as the gauge field itself is concerned, these BRST transformations act just like a gauge transformation $A_\mu \to A_\mu + \nabla_\mu\lambda$ with infinitesimal gauge parameter $\lambda(x) = \epsilon c$ given in terms of the ghost field. It follows that any gauge–invariant function of the connection alone, such as the original Yang–Mills action $S_{\mathrm{YM}}[\nabla]$, is invariant under the transformations (8.39). To see that the rest of the action is also invariant under (8.39), we'll first show that $[\delta_1, \delta_2] = 0$, where $\delta_{1,2}$ are transformations with parameters $\epsilon_{1,2}$, so that the BRST transformations form an Abelian group. In fact, we can write this as

$$[\delta_1,\delta_2]\Psi^i = \delta_1(\epsilon_2\,\mathcal{Q}\Psi^i) - \delta_2(\epsilon_1\,\mathcal{Q}\Psi^i) = -(\epsilon_1\epsilon_2 - \epsilon_2\epsilon_1)\mathcal{Q}^2\Psi^i = -2\epsilon_1\epsilon_2\,\mathcal{Q}^2\Psi^i \tag{8.40}$$

using the fact that the parameters are fermionic and so anticommute with $\mathcal{Q}$. Therefore the statement $[\delta_1, \delta_2] = 0$ amounts to the statement that the transformation $\Psi \mapsto \mathcal{Q}\Psi$ is nilpotent. For the Nakanishi–Lautrup field h, this assertion is trivial. Similarly, for the antighost $\bar{c}$ we have

$$\mathcal{Q}^2\bar{c} = \mathrm{i}(\mathcal{Q}h) = 0 \tag{8.41}$$

since h itself is invariant. For the gauge field,

$$\mathcal{Q}^2 A_\mu = \mathcal{Q}(\nabla_\mu c) = [\nabla_\mu c, c] - \frac{1}{2}\nabla_\mu([c,c]) \tag{8.42}$$

where the first term is the variation of the connection acting on c in the adjoint (*i.e.* the variation of A_μ in the $[A_\mu, c]$ term in $\nabla_\mu c$), and the second is the variation of the ghost we produced the first time $\mathcal{Q}$ acted. This vanishes since $\nabla_\mu[c,c] = [\nabla_\mu c, c] + [c, \nabla_\mu c] = 2[\nabla_\mu c, c]$ using the symmetry of this bracket, as explained above. Thus

$$\mathcal{Q}^2 A_\mu = 0\,. \tag{8.43}$$

Finally, for the ghost itself we have

$$\mathcal{Q}^2 c = -\frac{1}{2}\mathcal{Q}\left([c,c]\right) = -[\mathcal{Q}c, c] = \frac{1}{2}[[c,c],c] \tag{8.44}$$

again using the fact that $[c,c]$ is symmetric. In detail, this is $[[c,c],c] = \mathrm{i}f^{\mathrm{a}}_{\mathrm{bc}}f^{\mathrm{c}}_{\mathrm{de}}c^{\mathrm{d}}c^{\mathrm{e}}c^{\mathrm{b}}\,t_{\mathrm{a}}$, so the structure constants are totally antisymmetrized over the indices $\mathrm{b}, \mathrm{d}, \mathrm{e}$. Thus it vanishes by the Jacobi identity

$$f^{\mathrm{a}}_{\mathrm{bc}}f^{\mathrm{c}}_{\mathrm{de}} + f^{\mathrm{a}}_{\mathrm{dc}}f^{\mathrm{c}}_{\mathrm{eb}} + f^{\mathrm{a}}_{\mathrm{ec}}f^{\mathrm{e}}_{\mathrm{bd}} = 0\,. \tag{8.45}$$

Thus $\mathcal{Q}^2\Psi^i = 0$ for any single field $\Psi^i \in \{A, c, \bar{c}, h\}$.

Now let's show that the BRST transformation is nilpotent even when acting on an arbitrary functional $\mathcal{O}(A, c, \bar{c}, h)$ of the fields. We compute

$$\mathcal{Q}^2\mathcal{O} = \mathcal{Q}\left((\mathcal{Q}\Psi^i)\,\frac{\delta\mathcal{O}}{\delta\Psi^i}\right) = (\mathcal{Q}^2\Psi^i)\,\frac{\delta\mathcal{O}}{\delta\Psi^i} - \mathcal{Q}\Psi^i\,\mathcal{Q}\Psi^j\frac{\delta^2\mathcal{O}}{\delta\Psi^j\,\delta\Psi^i}\,. \tag{8.46}$$

The first term vanishes by our calculations above. To see that the second term also vanishes, split the sums over all fields (labelled by i, j) into separate sums over bosonic fields $\Psi^i \in \{A_\mu, h\}$ and fermionic fields $\Psi^i \in \{c, \bar{c}\}$. In the case that i and j both refer either to bosonic or fermionic fields, the term cancels because $\mathcal{Q}\Psi^i$ has opposite statistics to Ψ^i itself, so that pre–factor is symmetric if the second derivatives are antisymmetric, and *vice–versa*. The mixed terms cancel among themselves.

We're now in position to see why the full, gauge–fixed action is BRST invariant. Firstly, as mentioned above the original Yang–Mills action is BRST invariant because it is a gauge invariant function of A_μ alone. For the remaining terms, we note

$$\int \mathcal{Q}\,\mathrm{tr}(\bar{c}f[A])\,\mathrm{d}^dx = \int\left[\mathrm{i}\,\mathrm{tr}(hf[A]) - \mathrm{tr}\left(\bar{c}\frac{\delta f[A^\lambda]}{\delta\lambda}c\right)\right]\,\mathrm{d}^dx = S_{\mathrm{gf}}[h, A] + S_{\mathrm{gh}}[A, c, \bar{c}]\,, \tag{8.47}$$

so the gauge–fixing and ghost terms in the action are the BRST transformation of $\mathrm{tr}(\bar{c}f[A])$. Since BRST transformations are nilpotent, it follows that these terms are BRST invariant for any gauge–fixing functional $f[A]$. Combined with the gauge invariance of the original Yang–Mills action, this shows that BRST transformations preserve the full Yang–Mills gauge–fixed action. Provided we regularize the path integral measure in a way that preserves this (as will be true perturbatively in dimensional regularization), BRST symmetry will be a symmetry of the quantum theory, and all new terms that are generated by RG flow will also be constrained by BRST invariance[83]. In particular, terms that depend only on the original gauge field will be constrained to be gauge invariant, preventing the appearance of a mass term $\sim A^2$ even at the quantum level.

8.2.1 Ward–Takahashi identities for BRST transformations

It's important to realise that BRST invariance is a *global* symmetry of the action, valid for constant fermionic parameters ϵ in (8.39). Just as in section 6.3, such global symmetries impose selection rules on correlators. Consider the correlation function of operators $\mathcal{O}_i$ that are not necessarily BRST invariant. As for any (global) Ward–Takahashi identity, we have[84]

$$0 = \sum_{i=1}^{n}\left\langle(\mathcal{Q}\mathcal{O}_i)\prod_{j\neq i}\mathcal{O}_j\right\rangle\,, \tag{8.48}$$

[83]Because the BRST transformations do not act linearly on the fields, the form of these BRST transformations in the quantum effective action can be different to the classical form (8.39). See section **??**, or Weinberg, *The Quantum Theory of Fields*, vol. 2, chapter 17 for further details.

[84]Here we're assuming that all the operators are bosonic (though they may be built up using fermionic fields), otherwise bringing the operator that is to be varied to the left may introduce some signs.

where $(\mathcal{Q}\mathcal{O}_i)$ is the variation of the i^{th} operator under the BRST transformations (8.39). In particular, if the operators $\mathcal{O}_2, \ldots, \mathcal{O}_n$ do happen to be BRST invariant, then

$$0 = \left\langle (\mathcal{Q}\mathcal{O}_1) \prod_{j=2}^{n} \mathcal{O}_j \right\rangle \tag{8.49}$$

which says that the correlation function of a BRST exact operator $\mathcal{Q}\mathcal{O}$ with any number of BRST invariant operators vanishes.

Furthermore, we can now see that correlation functions of BRST invariant operators will be independent of our choice of gauge–fixing condition $f[A]$, just as for the partition function. This is because $f[A]$ appears only in the BRST exact term $\mathcal{Q}\,\mathrm{tr}(hf[A])$ in the action. In particular, the difference between the gauge–fixed actions with two different choices of gauge–fixing condition, $f_1[A]$ and $f_2[A]$, is

$$S_1 - S_2 = \int \mathcal{Q}\,\mathrm{tr}\,\bar{c}(f_1[A] - f_2[A])\,\mathrm{d}^d x = \mathcal{Q}V_{12}\,, \tag{8.50}$$

so that if the $\mathcal{O}_i$ are BRST invariant,

$$\begin{aligned}
\left\langle \prod_i \mathcal{O}_i \right\rangle_1 &= \int \mathrm{e}^{-S_1[A,c,\bar{c},h]/\hbar} \prod_i \mathcal{O}_i = \int \mathrm{e}^{-(S_2[A,c,\bar{c},h]+\mathcal{Q}V_{12})/\hbar} \prod_i \mathcal{O}_i \\
&= \int \mathrm{e}^{-S_2[A,c,\bar{c},h]/\hbar} \prod_i \mathcal{O}_i \; + \int \mathrm{e}^{-S_2[A,c,\bar{c},h]/\hbar} (\mathcal{Q}R_{12}) \prod_i \mathcal{O}_i \\
&= \left\langle \prod_i \mathcal{O}_i \right\rangle_2 .
\end{aligned} \tag{8.51}$$

where $\langle \cdots \rangle_{1,2}$ refers to the correlation function computed using the actions $S_{1,2}$, respectively. In deriving this equation, we used the fact that

$$\mathrm{e}^{-\mathcal{Q}V_{12}/\hbar} = 1 + \mathcal{Q}R_{12} \qquad \text{where} \qquad R_{12} = -\frac{1}{\hbar}V_{12} + \frac{1}{2\hbar^2}V_{12}(\mathcal{Q}V_{12}) - \cdots \tag{8.52}$$

and the Ward identity (8.39) with $\mathcal{O}_1 = V_{12}$.

The BRST transformations (8.39) correspond to a global symmetry of the gauge–fixed action, unlike the gauge redundancy of the unfixed action. Hence, by Noether's theorem, we can derive an expression for the conserved **BRST charge**. A short calculation shows that this is given by[85]

$$Q_{\mathrm{BRST}} = -\int_N n^\mu\,\mathrm{tr}\left(\frac{1}{\mathrm{g}_{\mathrm{YM}}^2} F_{\mu\nu}\nabla^\nu c + h\nabla_\mu c + \frac{1}{2}\partial_\mu \bar{c}\,[c,c]\right)\sqrt{g}\,\mathrm{d}^{d-1}x \tag{8.53}$$

in the case of the Lorenz gauge action (8.34), where n^μ is the unit normal to the hypersurface N (*e.g.*, N could be a constant time slice). There is also a ghost number charge

$$Q_{\mathrm{gh}} = \int_N n^\mu\,\mathrm{tr}\,(\bar{c}\,\nabla_\mu c - \partial_\mu \bar{c}\,c)\;\mathrm{d}^{d-1}x \tag{8.54}$$

[85] In a more sophisticated treatment, I'd point out that under canonical quantization this charge is really the Chevalley–Eilenberg differential of the infinite–dimensional Lie group $\mathcal{G}$ of all gauge transformations, acting on the space of fields.

corresponding to the U(1) symmetry transformations

$$c \mapsto \mathrm{e}^{\mathrm{i}\theta} c \qquad\qquad \bar{c} \mapsto \mathrm{e}^{-\mathrm{i}\theta} \bar{c} \tag{8.55}$$

with all other fields invariant, where θ is a constant (bosonic) parameter. The gauge–fixed action is real if the fields obey

$$(A^{\mathrm{a}}_\mu)^\dagger = A^{\mathrm{a}}_\mu\,, \qquad (c^{\mathrm{a}})^\dagger = c^{\mathrm{a}}\,, \qquad (\bar{c}^{\mathrm{a}})^\dagger = -\bar{c}^{\mathrm{a}} \qquad \text{and} \qquad (h^{\mathrm{a}})^\dagger = h^{\mathrm{a}}\,,$$

whereupon $Q^\dagger_{\mathrm{BRST}} = Q_{\mathrm{BRST}}$ and $Q^\dagger_{\mathrm{gh}} = Q_{\mathrm{gh}}$. As with any Noether charge, in the quantum theory these charges act on states. This charge generates BRST transformations, in the sense that

$$\mathcal{Q}\mathcal{O} = [Q_{\mathrm{BSRT}}, \hat{\mathcal{O}}] \tag{8.56}$$

in the canonical picture. In particular, $Q^2_{\mathrm{BRST}} = 0$ and $Q_{\mathrm{BRST}}|\Omega\rangle = 0$ saying that the vacuum $|\Omega\rangle$ is BRST invariant.

8.2.2 BRST cohomology and the physical Hilbert space

We're now ready to understand the rôle BRST transformations play in the canonical framework. Let's start by considering the case of pure gauge theory. Just as in section 3.1.1, if our space M has non-empty boundary then by varying $S_{\mathrm{YM}}[\nabla]$ we obtain a (pre-)symplectic potential

$$\Theta = -\frac{1}{\mathrm{g}^2_{\mathrm{YM}}} \int_{\partial M} n^\mu \,\mathrm{tr}(F_{\mu\nu}\,\delta A^\nu)\,\sqrt{g}\,\mathrm{d}^{d-1}x \tag{8.57}$$

on the space of solutions to the classical Yang–Mills equations, generalizing the usual expression

$$\theta = m\dot{\mathbf{x}}\cdot\,\delta\mathbf{x}|_{\partial I} = \mathbf{p}\cdot\,\delta\mathbf{x}|_{\partial I} \tag{8.58}$$

on the boundary of the worldline I of a point particle. In particular, if N is a constant time slice then $n^\mu F_{\mu\nu} = (0, E_i)$ is the (chromo-)electric field which plays the rôle of a momentum conjugate to the spatial part of the connection A_i. The 1-form (8.57) is degenerate along gauge orbits, but it descends to give a symplectic potential on the space $T^*(\mathcal{A}_N/\mathcal{G}_N)$; the cotangent bundle to the space of gauge equivalence classes of connections on N. Upon quantization, we obtain the equal–time commutation relations

$$\Big[A^{\mathrm{a}}_i(\mathbf{x}), E^{\mathrm{b}}_j(\mathbf{y})\Big] = \mathrm{g}^2_{\mathrm{YM}}\,\delta_{ij}\,\delta^{\mathrm{ab}}\,\delta^{(d-1)}(\mathbf{x}-\mathbf{y}) \tag{8.59}$$

and we (formally) take the Hilbert space to be $\mathcal{H} = L^2(\mathcal{A}_N/\mathcal{G}_N)$. Note that there is no momentum conjugate to A_0; this is the problem of gauge redundancy in the Hamiltonian framework. The Hamiltonian of the gauge theory is represented by

$$H = \frac{1}{2\mathrm{g}^2_{\mathrm{YM}}}\int_N \mathrm{tr}(\mathbf{E}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{B})\,\mathrm{d}^{d-1}x = \frac{1}{2}\int_N \mathrm{tr}\left(-\mathrm{g}^2_{\mathrm{YM}}\frac{\delta^2}{\delta A_i^2} + \frac{1}{\mathrm{g}^2_{\mathrm{YM}}}\mathbf{B}\cdot\mathbf{B}\right)\mathrm{d}^{d-1}x \tag{8.60}$$

acting on gauge invariant functions of the connection on N.[86]

[86]Note that this Hamiltonian is not so very different from an infinite–dimensional generalization of a harmonic oscillator – all the complexities of Yang–Mills theory lie in the fact that we must consider only gauge equivalence classes.

We must now understand how to recover this space from the gauge–fixed Yang–Mills action including ghosts. In fact, this is simplest to see using **axial gauge** where we set one component of the gauge field to zero. We achieve this using the gauge–fixing functional $f[A] = n^\mu A_\mu$, so the ghost and gauge–fixing parts of the action become

$$-\mathcal{Q}\int_M \mathrm{tr}(\bar{c}\, n^\mu A_\mu)\,\sqrt{g}\,\mathrm{d}^d x = -\mathrm{i}\int_M \mathrm{tr}(h\, n^\mu A_\mu)\,\sqrt{g}\,\mathrm{d}^d x + \int_M \mathrm{tr}(\bar{c}\, n^\mu \nabla_\mu c)\,\sqrt{g}\,\mathrm{d}^d x\,. \quad (8.61)$$

Upon integrating out the Nakanishi–Lautrup field and imposing the gauge condition $n^\mu A_\mu = 0$, we obtain the ghost action

$$S_{\mathrm{gh}}[c,\bar{c}] = \int_M \mathrm{tr}(\bar{c}\, n^\mu \partial_\mu c)\,\sqrt{g}\,\mathrm{d}^d x \quad (8.62)$$

The fact that the ghosts and gauge fields have decoupled is a special feature of axial gauges. Varying the action $S_{\mathrm{YM}}[\nabla] + S_{\mathrm{gh}}[c,\bar{c}]$, with this gauge condition we obtain the boundary term

$$\Theta = \int_{\partial M} \frac{1}{\mathrm{g}_{\mathrm{YM}}^2}\, n^\mu\, \mathrm{tr}(F_{\mu\nu}\,\delta A^\nu) + \mathrm{tr}(\bar{c}\,\delta c)\,\sqrt{g}\,\mathrm{d}^{d-1}x \quad (8.63)$$

leading upon quantization to the relations

$$\begin{aligned} \left[A_i^{\mathrm{a}}(\mathbf{x}), E_j^{\mathrm{b}}(\mathbf{y})\right] &= \mathrm{i}\,\mathrm{g}_{\mathrm{YM}}^2\,\delta_{ij}\,\delta^{\mathrm{ab}}\,\delta^{(d-1)}(\mathbf{x}-\mathbf{y}) \\ \left\{c^{\mathrm{a}}(\mathbf{x}), \bar{c}^{\mathrm{b}}(\mathbf{y})\right\} &= \mathrm{i}\,\delta^{\mathrm{ab}}\,\delta^{(d-1)}(\mathbf{x}-\mathbf{y}) \end{aligned} \quad (8.64)$$

among the fields. The gauge field commutation relations are exactly the same as before in (8.59), while the anticommutation relations appropriate for the fermionic ghost fields show that the antighost plays the rôle of momentum conjugate to the ghost. Since the ghost action is purely its kinetic term, the ghost Hamiltonian vanishes and we have

$$H = \frac{1}{2}\int_N \mathrm{tr}\left(-\mathrm{g}_{\mathrm{YM}}^2\frac{\delta^2}{\delta A_i^2} + \frac{1}{\mathrm{g}_{\mathrm{YM}}^2}\mathbf{B}\cdot\mathbf{B}\right)\mathrm{d}^{d-1}x \quad (8.65)$$

as before. However, in the 'position representation' we should build our QFT wavefunctions from the boundary values of the remaining components A_i of the gauge field, together with the boundary values of the ghost c. The remaining degrees of freedom $E_i = n^\mu \partial_\mu A_i$ and $\bar{c}$ are treated as differential operators on the space of fields[87]. Thus the space of states may be identified as $\widetilde{\mathcal{H}} \cong \mathcal{O}(\mathcal{A}_N)\otimes \wedge\tilde{\mathfrak{g}}^*$, where $\mathcal{A}_N$ is the space of connections on N, $\mathcal{O}(\mathcal{A}_N)$ is the space of functions on $\mathcal{A}_N$, $\tilde{\mathfrak{g}}$ is the (infinite dimensional) Lie algebra of gauge transformations on N and $\wedge\tilde{\mathfrak{g}}^*$ is the space of functions of the ghosts. This is because a generic $\Psi[A_i, c] \in \widetilde{\mathcal{H}}$ may be expanded in powers of the fermionic ghosts as

$$\Psi[A_i, c] = \sum_{n=0}^{\infty}\int_{\otimes^n N} c^{a_1}(x_1)\,c^{a_2}(x_2)\cdots c^{a_n}(x_n)\,\Psi^{(n)}_{a_1 a_2\ldots a_n}[A_i; x_1,\cdots,x_n]\prod_i \mathrm{d}^{d-1}x_i\,,$$

[87]Here, we are taking n^μ to be *both* the unit normal to the quantization hypersurface N, and the vector that picks out our choice of axial gauge. In the case of Lorentzian signature, where N is a constant time hypersurface in Minkowksi space, this is often called **temporal gauge**.

with coefficients $\Psi^{(n)}$ that are valued in $\mathcal{O}(\mathcal{A}_N)$-valued elements of $\wedge^n \tilde{\mathfrak{g}}^*$, the dual space to n powers of the fermionic ghosts.) We say the space $\tilde{\mathcal{H}}$ is **graded** by ghost number and write $\tilde{\mathcal{H}} = \bigoplus_n \tilde{\mathcal{H}}^{(n)}$, where $\mathcal{H}^{(n)}$ is the space of terms in the above expansion that are n^{th} order monomials in the ghosts.

The full space $\widetilde{\mathcal{H}}$ is not a Hilbert space, because the anticommutation relations imposed on the fermionic scalar ghosts violates the spin–statistics theorem. (Recall from the Michaelmas QFT course that fermionic fields had to have odd half-integer spin in any unitary theory.) However, elements of the ghost number zero part $\mathcal{H}^{(0)}$ are just function of the gauge field $A|_N$, so BRST transformations act on them as gauge transformations on N. In particular, BRST invariant functions Ψ that are independent of ghosts correspond to elements of the 'standard' Hilbert space $L^2(\mathcal{A}_N/\mathcal{G}_N)$. As a corrollary, suppose we wish to compute the correlation function of a set of BRST-invariant local operator insertions on a manifold M with boundary ∂M. If $\partial M =$ then we have shown above that these correlators are independent of the choice of gauge–fixing condition. This remains true in the presence of a non-trivial boundary only if the states associated with each boundary are themselves BRST invariant.

The above discussion was somewhat abstract, so it may be useful to see the significance of BRST cohomology in a down-to-earth example. Let us consider Maxwell theory in 3+1 dimensional Minkowski space, with gauge condition $f[A] = \partial^\mu A_\mu$. In this Abelian case the adjoint representation is trivial, and BRST transformations become

$$\mathcal{Q}A_\mu = \partial_\mu c\,, \qquad \mathcal{Q}c = 0\,, \qquad \mathcal{Q}\bar{c} = \partial^\mu A_\mu/\xi\,, \tag{8.66}$$

where we have

We expand our fields in terms of momentum space as

$$\begin{aligned} A_\mu(x) &= \frac{1}{(2\pi)^{3/2}}\int \frac{\mathrm{d}^3p}{(2p_0)^{1/2}}\left[a_\mu(\mathbf{p})\,\mathrm{e}^{\mathrm{i}p\cdot x} + a^\dagger_\mu(\mathbf{p})\,\mathrm{e}^{-\mathrm{i}p\cdot x}\right] \\ c(x) &= \frac{1}{(2\pi)^{3/2}}\int \frac{\mathrm{d}^3p}{(2p_0)^{1/2}}\left[c(\mathbf{p})\,\mathrm{e}^{\mathrm{i}p\cdot x} + c^\dagger(\mathbf{p})\,\mathrm{e}^{-\mathrm{i}p\cdot x}\right] \\ \bar{c}(x) &= \frac{1}{(2\pi)^{3/2}}\int \frac{\mathrm{d}^3p}{(2p_0)^{1/2}}\left[\bar{c}(\mathbf{p})\,\mathrm{e}^{\mathrm{i}p\cdot x} + \bar{c}^\dagger(\mathbf{p})\,\mathrm{e}^{-\mathrm{i}p\cdot x}\right]\,. \end{aligned} \tag{8.67}$$

Using the operator Q_{BRST} to apply the transformations (8.66) on the field modes, and matching coefficients of $\mathrm{e}^{\mathrm{i}p\cdot x}$ on both sides, we obtain

$$\begin{aligned} &[Q_{\mathrm{BRST}}, \bar{c}(\mathbf{p})] = -p_\mu c(\mathbf{p})\,, && [Q_{\mathrm{BRST}}, a^\dagger_\mu(\mathbf{p})] = p_\mu c^\dagger(\mathbf{p})\,, \\ &[Q_{\mathrm{BRST}}, c(\mathbf{p})] = 0\,, && [Q_{\mathrm{BRST}}, c^\dagger(\mathbf{p})] = 0\,, \\ &[Q_{\mathrm{BRST}}, \bar{c}(\mathbf{p})] = p^\mu a_\mu(\mathbf{p})/\xi && [Q_{\mathrm{BRST}}, \bar{c}^\dagger(\mathbf{p})] = p^\mu a^\dagger_\mu(\mathbf{p})/\xi\,. \end{aligned} \tag{8.68}$$

Suppose we have a state $|\psi\rangle$ that is BRST invariant, so $Q_{\mathrm{BRST}}|\psi\rangle = 0$. Then the state $|\epsilon,\psi\rangle = \epsilon^\mu a^\dagger_\mu(\mathbf{p})|\psi\rangle$ containing one extra photon will also be BRST invariant provided $\epsilon\cdot p = 0$, so that the photon is transverse. On the other hand, if $\epsilon^\mu \propto p^\mu$ then we have $\epsilon^\mu a^\dagger_\mu(\mathbf{p})|\psi\rangle \propto Q_{\mathrm{BRST}}\left(\bar{c}^\dagger(\mathbf{p})|\psi\rangle\right)$ and hence is BRST exact. Thus the state

$$|\epsilon + \alpha p, \psi\rangle = |\epsilon,\psi\rangle + \alpha p^\mu a^\dagger_\mu(\mathbf{p})|\psi\rangle = |\epsilon,\psi\rangle + Q_{\mathrm{BRST}}\left(\bar{c}^\dagger(\mathbf{p})|\psi\rangle\right) \tag{8.69}$$

is physically equivalent to the state $|\epsilon, \psi\rangle$. This is the usual notion of gauge equivalence classes for photons. Similarly, the state $|c, \psi\rangle = c^\dagger(\mathbf{p})|\psi\rangle$ with one extra ghost obeys

$$|c, \psi\rangle = Q_{\mathrm{BRST}} \left(\frac{\epsilon^\mu a_\mu^\dagger(\mathbf{p})|\psi\rangle}{\epsilon \cdot p} \right) \tag{8.70}$$

and so is inevitably BRST exact. So the physical Hilbert space of Maxwell theory is gauge invariant and free from ghosts. You'll explore BRST cohomology further in the final problem set.

8.3 Perturbative renormalization of Yang–Mills theory

Even though perturbation theory is of limited use at low energies, we'll begin our study of quantum Yang–Mills theory by trying to see what we can learn from it. As an incentive, we should expect that the perturbative description *will* be useful at high energies (where the renormalized coupling turns out to be small), so understanding perturbation theory will allow us to probe the UV behaviour of Yang–Mills theory.

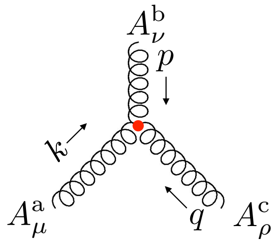
8.3.1 Feynman rules in R_ξ gauges

Let's start by taking a closer look at the propagators and interactions that result from our gauge–fixed action. For simplicity, we'll consider just pure Yang–Mills theory on $(M, g) = (\mathbb{R}^d, \delta)$, where the principal bundle P is trivial[88]. We can then write $F^{\mathrm{a}}_{\mu\nu} = \partial_\mu A^{\mathrm{a}}_\nu - \partial_\nu A^{\mathrm{a}}_\mu + f^{\mathrm{a}}_{\mathrm{bc}} A^{\mathrm{b}}_\mu A^{\mathrm{c}}_\nu$ globally, and rescaling $A \mapsto \mathrm{g}_{\mathrm{YM}} A$ so as to obtain canonically normalized gauge kinetic terms, in gory detail the Yang–Mills action becomes

$$\begin{aligned} S_{\mathrm{YM}}[A] = \int \frac{1}{4}(\partial^\mu A^{\nu\mathrm{a}} - \partial^\nu A^{\mu\mathrm{a}})(\partial_\mu A^{\mathrm{a}}_\nu - \partial_\nu A^{\mathrm{a}}_\mu)\, \mathrm{d}^d x \\ + \int \mathrm{g}_{\mathrm{YM}} f^{\mathrm{a}}_{\mathrm{bc}} A^{\mu\mathrm{b}} A^{\nu\mathrm{c}} \partial_\mu A^{\mathrm{a}}_\nu + \frac{\mathrm{g}^2_{\mathrm{YM}}}{4} f^{\mathrm{a}}_{\mathrm{bc}} f^{\mathrm{a}}_{\mathrm{de}} A^{\mu\mathrm{b}} A^{\nu\mathrm{c}} A^{\mathrm{d}}_\mu A^{\mathrm{e}}_\nu \, \mathrm{d}^d x \end{aligned} \tag{8.71}$$

We see that the self-interactions among the gauge fields are described by a three–gluon vertex

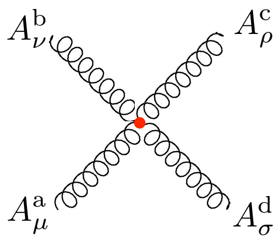
[88]More correctly, to ensure the action remains finite we should impose boundary conditions that the curvature vanishes as $|x| \to \infty$, so that the connection is pure gauge on a large S^{d-1}. It is then possible to have non–trivial bundles, even though the space $\mathbb{R}^d$ itself is contractible, classified by the homotopy group $\pi_{d-1}(G)$. For example, all semi–simple Lie groups G have non–trivial $\pi_3(G)$, the case that is relevant to instantons on $\mathbb{R}^4$.



that involves the derivative of the gauge field. This vertex is cyclically symmetric in the three participating gluons and is given by

$$\Gamma^{\mathrm{abc}}_{\mu\nu\rho}(k,p,q) = -\frac{\mathrm{g}_{\mathrm{YM}}}{\hbar}\, f^{\mathrm{abc}}\left[(k-p)_\rho\delta_{\mu\nu} + (p-q)_\mu\delta_{\nu\rho} + (q-k)_\nu\delta_{\rho\mu}\right] \tag{8.72}$$

in momentum space. The four–gluon vertex



is independent of the momentum, and is given by

$$\begin{aligned}\Gamma^{\mathrm{abcd}}_{\mu\nu\rho\sigma} = &-\frac{\mathrm{g}^2_{\mathrm{YM}}}{\hbar}\, f^{\mathrm{abe}} f^{\mathrm{cde}}\left(\delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\mu\sigma}\delta_{\nu\rho}\right) \;-\; \frac{\mathrm{g}^2_{\mathrm{YM}}}{\hbar}\, f^{\mathrm{ace}} f^{\mathrm{bde}}\left(\delta_{\mu\nu}\delta_{\sigma\rho} - \delta_{\mu\sigma}\delta_{\nu\rho}\right) \\ &-\; \frac{\mathrm{g}^2_{\mathrm{YM}}}{\hbar} f^{\mathrm{ade}}\, f^{\mathrm{bce}}\left(\delta_{\mu\nu}\delta_{\sigma\rho} - \delta_{\mu\rho}\delta_{\nu\sigma}\right).\end{aligned} \tag{8.73}$$

Note that this interaction is symmetric under the simultaneous exchange $A^{\mathrm{a}}_\mu \leftrightarrow A^{\mathrm{b}}_\nu$ and $A^{\mu\,\mathrm{c}} \leftrightarrow A^{\nu\,\mathrm{d}}$, which cancels the symmetry factor of 1/4.

To understand the gluon propagator we must include the gauge–fixing term. As above, we choose to work in Lorenz gauge so that the gauge–fixing functional is $f^{\mathrm{a}}[A] = \partial^\mu A^{\mathrm{a}}_\mu$. As it stands, the gauge–fixing term $\mathrm{i}\int \mathrm{tr}(h\,\partial^\mu A_\mu)$ is a little awkward to work with; the field h is non–dynamical, but integrating it out introduces a δ–function into the path integral that we don't know how to handle. For this reason, it will be convenient to add the BRST-exact term

$$-\,\mathrm{i}\frac{\xi}{4}\int \mathcal{Q}\,\mathrm{tr}(\bar{c}h)\,\mathrm{d}^d x = \frac{\xi}{4}\int \mathrm{tr}(hh)\,\mathrm{d}^d x \tag{8.74}$$

to the action, where ξ is an arbitrary constant. Since this term is BRST exact, its presence doesn't affect the value of any correlation function of BRST invariant operators. Integrating out the auxiliary field h, by completing the square we obtain

$$\int \mathcal{D}h\,\exp\left(-\frac{1}{\hbar}\int \mathrm{i}\,\mathrm{tr}(h\,\partial^\mu A_\mu) + \frac{\xi}{4}\mathrm{tr}(h^2)\,\mathrm{d}^d x\right) = \exp\left(-\frac{1}{\hbar\xi}\int \mathrm{tr}(\partial^\mu A_\mu\,\partial^\nu A_\nu)\,\mathrm{d}^d x\right), \tag{8.75}$$

together with the constraint that $h = \partial^\mu A_\mu$. (We used this constraint in analysing the content of BRST transformations in Maxwell theory at the end of the last section.) Thus the effect of the path integral over the Nakanishi–Lautrup field is simply to modify the kinetic term of the gauge field. Combining this with the kinetic part of the original Yang–Mills action, the gluon kinetic terms are

$$
\begin{aligned}
&\frac{1}{4}\int (\partial_\mu A^{\mathrm{a}}_\nu - \partial_\nu A^{\mathrm{a}}_\mu)(\partial^\mu A^{\nu\,\mathrm{a}} - \partial^\nu A^{\mu\,\mathrm{a}})\,\mathrm{d}^d x + \frac{1}{2\xi}\int \partial^\mu A^{\mathrm{a}}_\mu\,\partial^\nu A^{\mathrm{a}}_\nu\,\mathrm{d}^d x \\
&= \frac{1}{2}\int \partial_\mu A^{\mathrm{a}}_\nu\,\partial^\mu A^{\nu\,\mathrm{a}} - \partial_\mu A^{\mathrm{a}}_\nu\,\partial^\nu A^{\mu\,\mathrm{a}} + \frac{1}{\xi}\,\partial^\mu A^{\mathrm{a}}_\mu\,\partial^\nu A^{\mathrm{a}}_\nu \\
&= -\frac{1}{2}\int A^{\mathrm{a}}_\nu\left[\delta^\nu_\mu\,\partial^2 - \left(1-\frac{1}{\xi}\right)\partial_\mu\partial^\nu\right]A^{\mu\,\mathrm{a}}\,\mathrm{d}^d x\,.
\end{aligned}
\tag{8.76}
$$

Inverting this kinetic operator, one finds the gluon propagator is

$$
D^{\mathrm{ab}}_{\mu\nu}(p) = -\frac{\hbar\delta^{\mathrm{ab}}}{p^2}\left[\delta_{\mu\nu} - (1-\xi)\frac{p_\mu p_\nu}{p^2}\right] \tag{8.77}
$$

in momentum space. This propagator is often said to be in R_ξ gauge. Since it originally appeared in front of a BRST exact term, the value of ξ can be chosen freely; common choices are $\xi = 0$ (Landau's choice – which recovers the original Lorenz gauge as for electromagnetism) and $\xi = 1$ (Feynman and 't Hooft's choice). We'll usually take $\xi = 1$ below, so that

$$
D^{\mathrm{ab}}_{\mu\nu}(p) = -\delta^{\mathrm{ab}}\,\delta_{\mu\nu}\frac{\hbar}{p^2}\,, \tag{8.78}
$$

an especially simple form for the gluon propagator that is diagonal in both space-time and colour indices.

We must also consider ghost fields which can run around loops even if they do not appear externally. The action

$$
S_{\mathrm{gh}}[c,\bar{c},A] = \int \partial^\mu \bar{c}^{\mathrm{a}}\partial_\mu c + \mathrm{g}_{\mathrm{YM}} f^{\mathrm{a}}_{\mathrm{bc}}\partial^\mu\bar{c}^{\mathrm{a}} A^{\mathrm{b}}_\mu c^{\mathrm{c}}\;\mathrm{d}^d x \tag{8.79}
$$

leads to a momentum space propagator

$$
C^{\mathrm{ab}}(p) = \delta^{\mathrm{ab}}\frac{\hbar}{p^2} \tag{8.80}
$$

for the ghosts. This is the standard form expected for a massless complex scalar. Finally, in Lorenz gauge we have an antighost–gluon–ghost interaction of the form

$$
\Gamma^{\mathrm{abc}}(p) = -\frac{\mathrm{g}_{\mathrm{YM}}}{\hbar} f^{\mathrm{abc}} p_\mu \tag{8.81}
$$

where the component p_μ of the momentum of the antighost couples to the component A_μ of the gauge field. It's a really good exercise — and a standard exam question — to check you can derive all these terms from the action above.

The most likely feeling at this point is panic, while the most appropriate one is disgust. It doesn't take much imagination to see that any attempt to using these vertices to construct

Feynman diagrams will quickly run into a huge proliferation of terms. In fact, counting each term in the vertices (8.72) & (8.73) separately, even a fairly simple process like $2 \to 3$ gluon scattering receives contributions from $\sim$ 10,000 terms, already just at *tree* level! In theories with charged matter, such as QCD and the Standard Model, there are further interactions coming from the gluons in the covariant derivatives, and at loop level there are further contributions from the ghosts.

On the one hand, perhaps this is just the way it is. After all, Yang–Mills theory is a complicated, non–linear theory. If you come along and prod it in a more or less arbitrary way (*i.e.* do perturbation theory), you should expect that the consequences will indeed be messy and complicated. But another possible response to the above is a slight feeling of nausea. The whole point of our treatment of Yang–Mills theory in terms of bundles, connections and curvature was to show how tremendously natural this theory is from a geometric perspective. Yet this naturality is badly violated by our splitting of the Yang–Mills action into $(dA)^2$, A^2dA and A^4 pieces, none of which separately have any geometric meaning. Surely there must be a different way to treat this theory — one that is less ugly, and treats the underlying geometric structure with more respect?

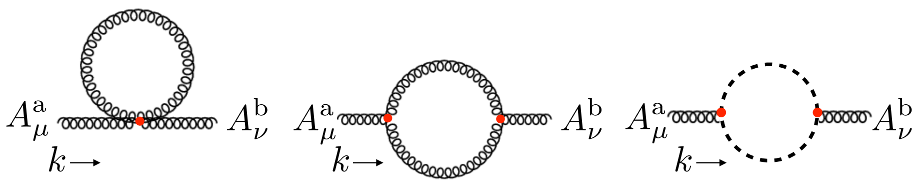
Many physicists sympathize with this view (me included). In fact, over the years various different ways to think about Yang–Mills theory have been proposed, ranging from viewing Yang–Mills theory as a type of string theory, to writing it in twistor space instead of space-time, to putting it on a computer. Some of these approaches have been highly successful, others only partially so. For now though, we must soldier on and do our best to understand the theory perturbatively in the neighbourhood of the trivial connection. To do otherwise would be somewhat akin to trying to understand differential geometry without first knowing what a vector is.

8.3.2 Vacuum polarization diagrams in Yang–Mills theory

Following what we did in QED, we can extract the Yang–Mills β-function from the form of the gluon kinetic term in the 1PI quantum effective action. We'll work in d dimensions, with $\mathrm{g}^2_{\mathrm{YM}} = \mu^{4-d}\, g^2(\mu)$ in terms of a dimensionless coupling $g^2(\mu)$ and arbitrary scale μ, chosen to be the typical scale of our experiments.

We let $\Pi^{\mathrm{ab}}_{\mu\nu}(k)$ denote the sum of all 1PI 1-loop diagrams with exactly two external gluons, constructed using the gluon propagator (8.77), 3-gluon vertex (8.72), and 4-gluon vertex (8.73) from the classical Yang–Mills action, together with the ghost propagator (8.80) and ghost–gluon vertex (8.81) from the ghost action. To 1–loop accuracy, there are three diagrams that contribute (not including counterterms):

$$\Pi^{\mathrm{ab}}_{\mu\nu}(k)_{1\,\mathrm{loop}} =$$



where, as always, the external legs are amputated in computing the 1PI contribution. Note that each of these graphs contributes an amount $\propto g^2(\mu)$, just reflecting the fact that in the original action, $g_{\rm YM}^2$ was indistinguishable from $\hbar$.

From our experience with $\lambda\phi^4$ theory, we do not expect the first of these diagrams to play any role in renormalizing the kinetic term, because the external momentum does not flow around the loop. Indeed – the only term it can contribute to would be a mass term for the gauge field. Likewise, there is a k-independent contribution to the second diagram involving the 3-pt gluon interaction that again must contribute to the mass of the gauge field in the quantum effective action. Neither of these diagrams, nor their sum, is compatible with BRST invariance which we recall acts on functions of the gauge field alone – such as the quadratic term in the effective action – just like a gauge transformation. However, we'll see that this would-be mass contribution is precisely cancelled by the k-independent part of the diagram involving the ghosts. The sum of all three diagrams indeed has a tensor structure compatible with $F^{\mu\nu}F_{\mu\nu}$.

Using the Feynman choice $\xi = 1$ of our gauge smearing parameter, the first diagram corresponds to the (dimensionless) momentum space expression

$$-\frac{g^2(\mu)\,\mu^{2-d}}{2}\int\frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\delta^{\rho\sigma}\delta^{\rm cd}}{p^2}\Gamma^{\rm abcd}_{\mu\nu\rho\sigma} = -g^2(\mu)\,\mu^{2-d}\,C_2(G)\,\delta^{\rm ab}\int\frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\delta_{\mu\nu}}{p^2}\,(d-1) \tag{8.82}$$

built from the Feynman gauge propagator and the four–point vertex (8.73). Here, $C_2(G)$ is the quadratic Casimir of the gauge group G in the adjoint representation, defined by $\mathrm{tr}_{\rm adj}(t^{\rm a}t^{\rm b}) = \delta^{\rm ab}C_2(G)$. It arises from accounting for the different possible 'colours' of the gauge field running around the loop (*i.e.*, from considering the possible structure of the Lie algebra indices) and in the case $G = \mathrm{SU}(N)$, we have $C_2(G) = N$. Of course, this integral diverges for any $d \in \mathbb{N}$, but to understand the full effect we should combine it with the remaining two diagrams before aiming to cancel the divergence with counterterms. To do this, for ease of comparison to the remaining diagrams, it will be helpful to write the integral in (8.82) as

$$\begin{aligned}\int\frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\delta_{\mu\nu}}{p^2} &= \int\frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\delta_{\mu\nu}}{p^2}\frac{(p+k)^2}{(p+k)^2} = \int_0^1\mathrm{d}x\int\frac{\mathrm{d}^dP}{(2\pi)^d}\frac{\delta_{\mu\nu}}{(P+\Delta)^2}\left[P^2+(1-x)^2k^2\right]\\ &= -\frac{\delta_{\mu\nu}\,k^2}{(4\pi)^{d/2}}\left[\frac{d}{2}\Gamma(1-d/2)\int_0^1\frac{x(x-1)}{\Delta^{2-d/2}}\,\mathrm{d}x + \Gamma(2-d/2)\int_0^1\frac{(1-x)^2}{\Delta^{2-d/2}}\,\mathrm{d}x\right]\end{aligned} \tag{8.83}$$

where we have introduced a factor of $1 = (p+k)^2/(p+k)^2$, combined the two denominators using a Feynman parameter x and defined $P = p + xk$ and $\Delta = -x(1-x)k^2$.

We now turn to the second diagram, involving two copies of the three-gluon vertex (8.72). This diagram gives a contribution

$$\begin{aligned}&\frac{g^2(\mu)\mu^{2-d}}{2}\int\frac{\mathrm{d}^dp}{(2\pi)^d}D^{\rm ad}_{\mu\sigma}(p+k)\,D^{\rm bc}_{\nu\rho}(p)\,\Gamma^{\rm dbc}_{\sigma\nu\rho}(p+k,-k,p)\,\Gamma^{\rm cad}_{\rho\mu\sigma}(p,k,p+k)\\ &= -\frac{g^2(\mu)\mu^{2-d}}{2}\,C_2(G)\,\delta^{\rm ab}\int_0^1\mathrm{d}x\int\frac{\mathrm{d}^dP}{(2\pi)^d}\frac{N_{\mu\nu}}{(P^2-\Delta)^2}\,,\end{aligned} \tag{8.84}$$

where again we have combined the two propagators using a Feynman parameter, with $P = p + xk$ and $\Delta = -x(1-x)k^2$ as before. The tensor structure of this integral is the rather unappealing expression

$$\begin{aligned} N_{\mu\nu} &= [\delta_{\mu\nu}(k-p)_\sigma + \delta_{\rho\sigma}(2p+k)_\mu - \delta_{\sigma\mu}(p+2k)_\rho]\,[\delta_\nu{}^\rho(p-k)^\sigma - \delta^{\rho\sigma}(2p+k)_\nu + \delta_\nu{}^\sigma(p+2k)_\rho] \\ &= -\delta_{\mu\nu}\left[(2k+p)^2 + (p-k)^2\right] - d(k+2p)_\mu(k+2p)_\nu \\ &\qquad + (2k+p)_{(\mu}(k+2p)_{\nu)} + (k-p)_{(\mu}(2k+p)_{\nu)} - (k+2p)_{(\mu}(k-p)_{\nu)}\,, \end{aligned} \tag{8.85}$$

where $a_{(\mu}b_{\nu)} = a_\mu b_\nu + a_\nu b_\mu$. Writing this in terms of $P = p + xk$, discarding linear terms in P and replacing $P_\mu P_\nu \to \delta_{\mu\nu}P^2/d$, we obtain the simpler form

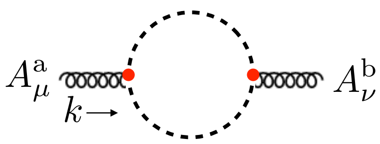
$$N_{\mu\nu} = -\delta_{\mu\nu}6P^2\left(1-\frac{1}{d}\right) - \delta_{\mu\nu}k^2\left[(2-x)^2+(1+x)^2\right] + k_\mu k_\nu\left[(2-d)(1-2x)^2 + 2(1+x)(2-x)\right] \tag{8.86}$$

which holds under the integral sign. Altogether, this second pure gauge field diagram yields a contribution

$$\begin{aligned} &-\frac{g^2(\mu)\,\mu^{d-2}\,C_2(G)\,\delta^{\rm ab}}{(4\pi)^{d/2}}\left[\frac{3(d-1)}{2}\Gamma(1-d/2)\,k^2\,\delta_{\mu\nu}\int_0^1\frac{x(1-x)}{\Delta^{2-d/2}}\,\mathrm{d}x\right. \\ &\left. + \Gamma(2-d/2)\left(\delta_{\mu\nu}\int_0^1\frac{(2-x)^2+(1+x)^2}{2\Delta^{2-d/2}}\,\mathrm{d}x - k_\mu k_\nu\int_0^1\frac{(1-d/2)(1-2x)^2+(1+x)(2-x)}{\Delta^{2-d/2}}\,\mathrm{d}x\right)\right]. \end{aligned} \tag{8.87}$$

Like the previous integral from the 4-pt vertex, this contribution just looks like pure garbage, as does the sum of the two. Indeed, this is correct – these two diagrams are completely meaningless, because their value depends on the choice of gauge we used to write down the propagator for A_μ. This is to be expected. They come from considering just the pure Yang–Mills part of the full action, together with the $(\partial^\mu A_\mu)^2$ term we obtained from integrating out the Nakanishi–Lautrup field in the gauge–fixing term. Just as in the zero–dimensional toy case (8.10), this part of the path integral is indeed gauge dependent, and knows about all the details of our choice of gauge slice.

To cancel this dependence and obtain some meaningful, gauge–invariant quantity, we must also include the Faddeev–Popov determinant, represented at $O(\hbar)$ by the 1-loop graph



involving the ghost fields. This diagram yields

$$
\begin{aligned}
& - g^2(\mu)\, \mu^{2-d} \int \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{p^2} \frac{1}{(p+k)^2} f^{\mathrm{dac}} (p+k)_\mu\, f^{\mathrm{cbd}} p_\nu \\
& = -\frac{g^2(\mu)\mu^{2-d}}{(4\pi)^{d/2}}\, C_2(G)\, \delta^{\mathrm{ab}} \left[-\frac{1}{2} \Gamma(1-d/2) \delta_{\mu\nu} k^2 + \Gamma(2-d/2) k_\mu k_\nu \right] \int_0^1 \frac{x(1-x)}{\Delta^{2-d/2}}\, \mathrm{d}x
\end{aligned}
\tag{8.88}
$$

where we recall that, as with the electron loop in our vacuum polarization calculation of QED, we obtain a minus sign from the fact that the fields running around the loop are fermionic.

We're now ready to combine these diagrams. We'll begin by looking at the coefficient of $\Gamma(1-d/2)$. This Γ-function has a pole in $d=2$. The only way our integrals could diverge in as low a dimension as $d=2$ is if they behave in the UV as $\sim \int \mathrm{d}^d p/p^2$, with only one factor of $1/p^2$. Such terms come from the original diagram (involving the 4-pt vertex), as well as the k-independent parts of the remaining two diagrams. Therefore, unless it is cancelled by the coefficient, this pole would correspond to the generation of a mass term in the quantum effective action. Combining the diagrams, the coefficient contains a factor

$$
\frac{3d - 3 - d^2 + d - 1}{2} = (1 - d/2)(d-2)
$$

and since $(1-d/2)\Gamma(1-d/2) = \Gamma(2-d/2)$ we see that indeed the pole cancels! BRST symmetry is working: the ghosts have ensured that the diagrams computed from the Yang–Mills action alone (together with a choice of gauge in which to write propagator) do not generate a gauge–violating mass term in the effective action. Even better, we now see that all the terms are in fact proportional to $\Gamma(2-d/2)$ (with a pole at $d=4$, but not at $d=2$) and that they combine to give[89]

$$
\Pi^{\mathrm{ab}}_{\mu\nu}(k) = \delta^{\mathrm{ab}}\, k^2 \left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2} \right) \pi(k^2) \tag{8.89}
$$

where

$$
\pi(k^2) = -\frac{g^2(\mu)}{(4\pi)^{d/2}}\, C_2(G)\, \Gamma(2-d/2) \int_0^1 \frac{\mu^{2-d/2}}{\Delta^{2-d/2}} \left((1-d/2)(1-2x)^2 + 2 \right) \mathrm{d}x + O(\hbar) \tag{8.90}
$$

and I remind you that $\Delta = -x(1-x)k^2$. This may be compared to the expressions (5.74a)-(5.74b) for vacuum polarization in QED. The crucial point is that it this expression is proportional to the tensor structure $\delta^{\mathrm{ab}}(\delta_{\mu\nu} - k_\mu k_\nu/k^2)$, corresponding to a quantum correction to the coefficient of the kinetic term $\partial^{[\mu} A^{\nu]}\, \partial_{[\mu} A_{\nu]}$ in the effective action, together with higher derivative terms coming from the k dependence in Δ.

[89]It's still (!) not entirely trivial to obtain this result. Among other things, one needs to use the fact that the integral over the Feynman parameter x is invariant under $x \mapsto (1-x)$ to judiciously write $x = \frac{1}{2}x + \frac{1}{2}(1-x) = \frac{1}{2}$ in terms in the numerator of the integral that are linear in x. See *e.g.* Peskin & Schroeder, *An Introduction to Quantum Field Theory*, section 16.5 for more details.

The factor of $\Gamma(2-d/2)$ shows there is still a pole in $d=4$, and setting $d=4-\epsilon$ leads to the asymptotic expression

$$\pi(k^2) \sim -\frac{g^2(\mu)\,C_2(G)}{16\pi^2}\left[\frac{5}{3}\left(\frac{2}{\epsilon}-\gamma+\ln\frac{4\pi\mu^2}{k^2}\right)+\frac{31}{9}\right]+O(\epsilon) \tag{8.91}$$

as $\epsilon \to 0^+$. We can cancel the divergence as $\epsilon \to 0$ with a counterterm $\frac{\delta Z_3}{4}\mathrm{tr}\,(F^{\mu\nu}F_{\mu\nu})$ adjusting the coefficient of the Yang–Mills action, and it's convenient to choose

$$\delta Z_3 = \frac{g^2(\mu)\,C_2(G)}{16\pi^2}\left[\frac{5}{3}\left(\frac{2}{\epsilon}-\gamma-\ln 4\pi\right)+\frac{31}{9}\right] \tag{8.92}$$

to leave us with the quantum effective kinetic term

$$\Gamma^{(2)}[A] = \frac{1}{4}\int\left[1+\frac{5\hbar g^2}{48\pi^2}C_2(G)\,\ln\frac{k^2}{\mu^2}\right]k^{[\mu}\tilde{A}^{\nu]\,\mathrm{a}}(k)\,k_{[\mu}\tilde{A}^{\mathrm{a}}_{\nu]}(-k)\,\mathrm{d}^4k \tag{8.93}$$

in four dimensional momentum space. In particular, this shows that the gauge boson propagator (in Feynman gauge) is now

$$D^{\mathrm{ab}}_{\mu\nu} = \frac{1}{k^2}\,\frac{\delta_{\mu\nu}\,\delta^{\mathrm{ab}}}{1+\frac{5\hbar}{48\pi^2}g^2C_2(G)\,\ln\frac{k^2}{\mu^2}} \tag{8.94}$$

to this order. This gives us another interpretation of the scale μ: it is the energy of the gauge boson at which the propagator is just $1/k^2$.

Unlike in QED, the quadratic part of the effective action is not gauge (or BRST) invariant. Indeed, whilst in any gauge the self-energy is transverse and free from quadratic divergences, its coefficient may depend on the BRST trivial parameter ξ. For general values of the gauge smearing parameter, it turns out we should replace

$$\frac{5}{3} \longrightarrow \frac{13}{6}-\frac{\xi}{2}$$

in expressions such as (8.91) and (8.92), showing that the effective gluon propagator depends on ξ. This result is not surprising: the gluon propagator depended on ξ even at the classical level. However, it shows that our expression still contains BRST non-invariant pieces. We do obtain a BRST invariant answer for any physical observable, such as the $2\to 2$ gluon scattering amplitude, but if we compute this to 1-loop accuracy, then as well as the 1-loop propagator we must include vertex corrections such as

and the corresponding diagrams involving ghosts running around the loop. These diagrams also individually depend on ξ in a way that intricately cancels out in the sum of all diagrams contributing to a given physical process. Only in the sum of all such diagrams do we obtain a BRST invariant answer.

Rather than compute all these terms, in the next section we'll examine a simpler way to extract physically meaningful results, such as the running of the Yang-Mills coupling constant.

8.4 Background field method

Suppose we expand the Yang–Mills action not around the trivial connection $\nabla_0 = d$, but around some other background connection ∇_0 with $F_{\nabla_0} \neq 0$. We will also suppose that ∇_0 obeys the classical Yang–Mills equations so that $\nabla_0^* F_{\nabla_0} = 0$. Setting the full connection to be $\nabla = \nabla_0 + a$, as in (8.6) we have

$$\begin{aligned} S_{\mathrm{YM}}[\nabla] &= \frac{1}{2\mathrm{g}_{\mathrm{YM}}^2} \int_M \mathrm{tr}(F_\nabla \wedge *F_\nabla) \\ &= S_{\mathrm{YM}}[\nabla_0] + \frac{1}{2\mathrm{g}_{\mathrm{YM}}^2} \int_M \mathrm{tr}(F \wedge *[a,a]) + \frac{1}{2\mathrm{g}_{\mathrm{YM}}^2} \int_M \mathrm{tr}\left(\nabla_0 a + \frac{1}{2}[a,a]\right) \wedge * \left(\nabla_0 a + \frac{1}{2}[a,a]\right) . \end{aligned} \tag{8.95}$$

This action is invariant under gauge transformations h acting as $\nabla \mapsto h^{-1}\nabla h$ on the true connection, and hence as

$$\nabla_0 \mapsto h^{-1}\nabla_0 h \qquad \text{and} \qquad a \mapsto h^{-1} a h \tag{8.96a}$$

on the background and fluctuations. Notice that a transforms in the *adjoint* here; this is as expected, since it is the difference $a = \nabla - \nabla_0$ between the true connection and our choice of background connection ∇_0. However, we obtain the same transformation of ∇ if we instead declare

$$\nabla_0 \mapsto \nabla_0 \qquad \text{and} \qquad a \mapsto h^{-1}\nabla_0 h + h^{-1} a h\,, \tag{8.96b}$$

assigning the whole transformation to a and leaving the background connection invariant. Thus, if we choose the gauge fixing functional to be

$$f(a) = \nabla_0^\mu a_\mu\,, \tag{8.97}$$

we break invariance under the gauge transformations (8.96b) whilst preserving invariance under the background gauge transformations (8.96a). This is significant because the results of our loop calculations will give a function of ∇_0 that is *manifestly* invariant under $\nabla_0 \mapsto h^{-1}\nabla_0 h$, which is a powerful constraint on the form these loop corrections can take.

With the choice (8.97) of $f(a)$ we have the ghost action

$$S_{\mathrm{gh}} = \int_M \mathrm{tr}\left(\bar{c}\, \nabla_0 * \left(\nabla_0 c + [a,c]\right)\right). \tag{8.98}$$

while, in the presence of an R_ξ-type gauge smearing term, the path integral over the Nakanishi–Lautrup field gives

$$\int \mathcal{D}h\ \exp\left(-\int_M \mathrm{itr}\,(h \nabla_0 * a) + \frac{\xi}{4} * \mathrm{tr}(h^2)\right) = \exp\left(-\frac{1}{\xi}\int_M \mathrm{tr}(\nabla_0^\mu a_\mu\, \nabla_0^\nu a_\nu)\, \mathrm{d}^d x\right) . \tag{8.99}$$

This is essentially the same as before, except that the derivative is covariant with respect to background gauge transformations.

We'd like to obtain the effective action for ∇_0 when the fluctuations a are integrated out. As always, to one–loop accuracy it's sufficient to restrict our attention to the quadratic terms in the fluctuations, though of course the higher order terms would be important at higher loops. Consider first the fluctuations in the gauge field. Keeping just the quadratic terms in the gauge smeared action with $\xi = 1$, we have

$$\begin{aligned}
&S_{\rm YM}[\nabla_0 + g_{\rm YM}\, a] - S_{\rm YM}[\nabla_0]|_{O(a^2)} \\
&= \int \mathrm{tr}\left((\nabla_{0\mu} a_\nu)(\nabla_0^\mu a^\nu - \nabla_0^\nu a^\mu) + F_{\nabla_0}^{\mu\nu}[a_\mu, a_\nu] + \nabla_0^\mu a_\mu\, \nabla_0^\nu a_\nu\right)\, \mathrm{d}^d x \\
&= \int \mathrm{tr}\, \Big(a_\nu\left(-\delta^{\mu\nu}\,\nabla_0^2 + \nabla_0^\mu\nabla_0^\nu - \nabla_0^\nu\nabla_0^\mu\right) a_\mu + F_{\nabla_0}^{\mu\nu}[a_\mu, a_\nu]\Big)\ \mathrm{d}^d x \\
&= \int \mathrm{tr}\, \Big(-a_\nu \nabla_0^2 a^\nu + 2a_\nu[F_{\nabla_0}^{\mu\nu}, a_\mu]\Big)\ \mathrm{d}^d x\,.
\end{aligned} \tag{8.100}$$

Here, in going to the third equality we have used the fact that, since a transforms in the adjoint under background gauge transformations, $(\nabla_0^\mu\nabla_0^\nu - \nabla_0^\nu\nabla_0^\mu)a_\mu = [F_{\nabla_0}^{\mu\nu}, a_\mu]$ and also $\mathrm{tr}\, F_{\nabla_0}^{\mu\nu}[a_\mu, a_\nu] = \mathrm{tr}\, a_\nu[F_{\nabla_0}^{\mu\nu}, a_\mu]$ using cyclicity of the trace. The ghosts can only appear in loops, so to 1-loop we can ignore the $\bar{c}\,\nabla_0(*[a,c])$ vertex, treating the ghost action simply as

$$S_{\rm gh}^{\rm quad}[\bar{c}, c, a] = \int \mathrm{tr}\, \bar{c}\, \nabla_0^2 c\ \mathrm{d}^d x\,. \tag{8.101}$$

Thus, integrating out the fluctuations, to one loop accuracy we have the effective action

$$S_{\rm YM}^{\rm eff}[\nabla_0] = S_{\rm YM}[\nabla_0] + \frac{1}{2}\ln\det(\triangle) - \ln\det(-\nabla_0^2)\,, \tag{8.102}$$

where the first term is the classical contribution, while the determinants come from integrating out the bosonic fluctuations in the gauge field and the fermionic ghosts, respectively. In the first determinant, $\triangle$ is the matrix of operators

$$(\triangle_{\mu\nu})^{\rm a}{}_{\rm c} = -\delta^{\rm a}{}_{\rm c}\delta_{\mu\nu}\nabla_0^2 - 2f^{\rm a}_{\rm bc}(F_{\nabla_0})^{\rm b}_{\mu\nu} \tag{8.103}$$

acting on fluctuations $a_\mu = -\mathrm{i}a_\mu^{\rm c} t_{\rm c}$ in the adjoint representation for background transformations. The first term here is a Laplacian, covariant with respect to the background field, whilst the second term is the magnetic moment coupling between the fluctuation and the background fieldstrength.

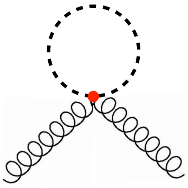
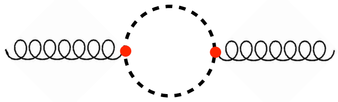
Let's compute these determinants, beginning with the ghost contribution. We have

$$\begin{aligned}
-\nabla_0^2 &= -\partial^2 + \mathrm{i}(\partial^\mu A_\mu^{\rm a} + A_\mu^{\rm a}\partial_\mu)t_{\rm a} + A^{\mu{\rm a}}A_\mu^{\rm b} t_{\rm a} t_{\rm b} \\
&= -\partial^2 + \Delta^{(1)} + \Delta^{(2)}
\end{aligned} \tag{8.104}$$

where $\Delta^{(i)}$ is the term involving i powers of the background gauge field A_μ. Thus

$$\begin{aligned}
\ln\det(-\nabla_0^2) &= \ln\det(-\partial^2) + \ln\det(1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)})) \\
&= \ln\det(-\partial^2) + \mathrm{tr}\ln(1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)})) \\
&= \ln\det(-\partial^2) + \mathrm{tr}\left(\frac{1}{-\partial^2}\Delta^{(2)}\right) - \frac{1}{2}\mathrm{tr}\left(\frac{1}{-\partial^2}\Delta^{(1)}\frac{1}{-\partial^2}\Delta^{(1)}\right) + O(A^3)\,.
\end{aligned} \tag{8.105}$$

In going to the last line we've used the fact that, for a semi-simple gauge group, $\operatorname{tr} t_{\rm a} = 0$ so the term $\operatorname{tr}(-\partial^2)^{-1}\Delta^{(1)})$ that is linear in A_μ vanishes. Note that we only need keep track of terms up to second order in A; background gauge invariance ensures that these quadratic terms will be completed to full non–Abelian curvatures. The two remaining field–dependent terms may be represented by the Feynman diagrams

where the external field is the background gauge field A. These diagrams respectively correspond to the momentum space expressions

$$\operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(2)}\right) = \mu^{4-d}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\tilde{A}^{\rm a}_\mu(-k)\tilde{A}^{\rm b}_\nu(k)\int\frac{\mathrm{d}^d p}{(2\pi)^d}\operatorname{tr}\left(\delta^{\mu\nu}\, t_{\rm a}t_{\rm b}\right)\frac{1}{p^2} \tag{8.106}$$

and

$$\frac{1}{2}\operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(1)}\frac{1}{-\partial^2}\Delta^{(1)}\right) = \frac{\mu^{4-d}}{2}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\tilde{A}^{\rm a}_\mu(-k)\tilde{A}^{\rm b}_\nu(k)\int\frac{\mathrm{d}^d p}{(2\pi)^d}\operatorname{tr}\left(\frac{(2p+k)^\mu t_{\rm a}(2p+k)^\nu t_{\rm b}}{p^2\,(p+k)^2}\right). \tag{8.107}$$

Individually, these expressions are not meaningful, but they combine (using the same manipulations as in the previous section) to give

$$\ln\det(-\nabla_0^2) = \frac{C_2(G)}{3(4\pi)^{d/2}}\Gamma(2-d/2)\,\frac{1}{2}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\tilde{A}^{\rm a}_\mu(-k)\tilde{A}^{\rm a}_\nu(k)(k^2\delta^{\mu\nu}-k^\mu k^\nu)\left(\frac{k^2}{\mu^2}\right)^{d/2-2} + O(A^3) \tag{8.108}$$

which is the contribution to the background field effective action from integrating out the ghosts to one loop.

We now turn to the determinant $\ln\det(\triangle)$ coming from integrating out fluctuations in the gauge field. Again, we write

$$\begin{aligned}\ln\det\triangle &= \ln\det(-\nabla_0^2 - 2[F_{\nabla_0},\,\cdot\,])\\ &= \ln\det(-\partial^2) + \operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(2)}\right) - \frac{1}{2}\operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(1)}\frac{1}{-\partial^2}\Delta^{(1)}\right) - \frac{1}{2}\operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(F)}\frac{1}{-\partial^2}\Delta^{(F)}\right) + \cdots\end{aligned} \tag{8.109}$$

where $\Delta^{(F)} = -2f^{\rm a}_{\rm bc}(F_{\nabla_0})^{\rm b}_{\mu\nu}$ is the magnetic moment term. The terms $\Delta^{(1)}$ and $\Delta^{(2)}$ are exactly the same as in the ghost calculation, except that here they act on all d space-time components of the fluctuating field a_μ, since each of these components can run around the loop. Thus these terms give $d\times$(8.108) so, bearing in mind the weighting of the ghost and gauge determinants in (8.102), they combine with the ghost contribution to give a term

$$\frac{d-2}{2}\,\frac{C_2(G)}{6(4\pi)^{d/2}}\Gamma(2-d/2)\int\frac{\mathrm{d}^d k}{(2\pi)^d}\tilde{A}^{\rm a}_\mu(-k)\tilde{A}^{\rm a}_\nu(k)(k^2\delta^{\mu\nu}-k^\mu k^\nu)\left(\frac{k^2}{\mu^2}\right)^{d/2-2} \tag{8.110}$$

in the effective action.

The remaining contribution is the magnetic moment term. As above, it suffices to work to quadratic order in the background field, so we can treat the coupling F_{∇_0} as its Abelian part dA. We obtain

$$
\begin{aligned}
\operatorname{tr}\left(\frac{1}{-\partial^2}\Delta^{(F)}\frac{1}{-\partial^2}\Delta^{(F)}\right) &= -4C_2(G)\mu^{4-d}\int\frac{\mathrm{d}^dk}{(2\pi)^d}\frac{\mathrm{d}^dp}{(2\pi)^d}\operatorname{tr}(\tilde{A}_\mu(k)\tilde{A}_\nu(-k))\,\frac{k^2\delta^{\mu\nu}-k^\mu k^\nu}{p^2(p+k)^2} \\
&= \frac{4C_2(G)}{(4\pi)^{d/2}}\Gamma(2-d/2)\int\frac{\mathrm{d}^dk}{(2\pi)^d}\left(k^2\delta^{\mu\nu}-k^\mu k^\nu\right)\left(\frac{k^2}{\mu^2}\right)^{d/2-2}\operatorname{tr}(\tilde{A}_\mu(k)\tilde{A}_\nu(-k))
\end{aligned}
\tag{8.111}
$$

(For further details of this calculation, see *e.g.* section 16.6 of Peskin & Schroeder.)

Combining all the pieces, we obtain the effective action for the background field as

$$
\begin{aligned}
S^{\text{eff}}_{\text{YM}}[\nabla_0] &= S_{\text{YM}}[\nabla_0] - \frac{\Gamma(2-d/2)}{(4\pi)^{d/2}}\frac{11C_2(G)}{3}\frac{1}{2}\int\mathrm{d}^dx\left(\frac{-\partial^2}{\mu^2}\right)^{d/2-2}\operatorname{tr}(F_{\nabla_0}\wedge *F_{\nabla_0}) \\
&= \frac{1}{2}\int\mathrm{d}^dx\left[\frac{1}{g^2(\mu)} - \frac{\Gamma(2-d/2)}{(4\pi)^{d/2}}\frac{11C_2(G)}{3}\left(\frac{-\partial^2}{\mu^2}\right)^{d/2-2}\right]\operatorname{tr}(F_{\nabla_0}\wedge *F_{\nabla_0})
\end{aligned}
\tag{8.112}
$$

where the first term is the classical piece and the remaining terms come from expanding the 1-loop determinants. Note that we only computed the quadratic terms in A above: in momentum space these involved the characteristic factor $k^2\delta_{\mu\nu} - k_\mu k_\nu$ corresponding to $(\partial_\mu A_\nu - \partial_\nu A_\mu)^2$ in position space, and we used background gauge invariance to conclude that these terms must be completed to the full, non-Abelian curvature $F_{\nabla_0\mu\nu}F^{\mu\nu}_{\nabla_0}$ involving cubic and quartic powers of the background field that we neglected above. As always, the full quantum effective action will also involve an infinite series of higher powers of F_{∇_0} and its derivatives, though these will be irrelevant in $d = 4$.

The Γ-function has a pole in $d = 4$ which we can remove using the $\overline{\text{MS}}$ counterterm

$$
\delta Z_3 = \frac{11}{3}\frac{C_2(G)}{(4\pi)^2}\left[\frac{2}{\epsilon} - \gamma + \ln 4\pi\right] \tag{8.113}
$$

leaving us with the 1-loop effective action

$$
S^{\text{eff}}_{\text{YM}}[\nabla_0] = \frac{1}{2}\int\mathrm{d}^4x\left[\frac{1}{g^2(\mu)} + \frac{1}{(4\pi)^2}\frac{11C_2(G)}{3}\ln\left(\frac{-\partial^2}{\mu^2}\right)\right]\operatorname{tr}(F_{\nabla_0}\wedge *F_{\nabla_0}) \tag{8.114}
$$

for the background Yang-Mills field in $d = 4$.

8.4.1 The β-function and asymptotic freedom

Arguing as before that the effective coupling must be independent of the renormalization scale μ, we have

$$
0 = \mu\frac{\partial}{\partial\mu}g^2_{\text{eff}} = \mu\frac{\partial}{\partial\mu}\left[\frac{1}{g^2(\mu)} - \frac{1}{(4\pi)^2}\frac{11}{3}C_2(G)\ln\mu^2\right] \tag{8.115}
$$

Therefore the β-function of pure Yang–Mills theory is

$$
\beta(g) = -\frac{g^3(\mu)}{(4\pi)^2}\frac{11}{3}C_2(G)\,, \tag{8.116}
$$

corresponding to the running coupling

$$g^2(\mu') = \frac{g^2(\mu)}{1 + \frac{g^2(\mu)}{16\pi^2}\,\frac{11}{3}C_2(G)\ln(\mu'/\mu)^2}\,. \tag{8.117}$$

The β-function is *negative*, and we see that if $\mu' > \mu$ then $g(\mu') < g(\mu)$. Thus the Yang–Mills coupling is marginally relevant: Yang–Mills theory approaches a free theory in the ultraviolet, and becomes strongly coupled in the infra-red. This, finally, is our paradigmatic example of a continuum QFT. Yang–Mills theory begins arbitrarily close to a Gaussian fixed point and moves out along the renormalized trajectory in a (marginally) relevant direction as we probe at lower and lower energies. The fact that it is strongly coupled at large distances means perturbation theory around the classical action is a poor guide to the low–energy physics, which is how it avoids Pauli's initial criticism that we don't observe any long range forces beyond gravity and electromagnetism.

More generally, as you'll explore in the final problem sheet, if we couple scalar fields in representations r_i of G and Dirac spinors in representations r_j, then the β-function is modified to

$$\beta(g) = -\frac{g^3}{16\pi^2}\left[\frac{11}{3}C_2(G) - \frac{1}{3}\sum_{i\in\text{scalars}} C(r_i) - \frac{4}{3}\sum_{j\in\text{fermions}} C(r_j)\right] \tag{8.118}$$

where $\mathrm{tr}_r(t^{\mathrm{a}}t^{\mathrm{b}}) = C(r)\,\delta^{\mathrm{ab}}$. For example, if $G = \mathrm{SU}(N)$ then $C_2(G) = N$ and $C(\text{fund}) = 1/2$. Provided we do not have too much matter, the β-function of Yang–Mills theory is *negative*, so this theory approaches a free theory in the UV, with

$$g^2(\mu') = \frac{g^2(\mu)}{1 + \frac{g^2(\mu)}{16\pi^2}\left(\frac{11}{3}C_2(G) - \frac{1}{3}\sum_i C(r_i) - \frac{4}{3}\sum_j C(r_j)\right)\ln \mu'^2/\mu^2} \tag{8.119}$$

as the running coupling. In fact, a theorem of Gross and Coleman states that non–Abelian gauge theories are the *only* possible non–trivial, asymptotically free QFTs in four dimensions.

8.5 Topological terms and the vacuum angle

In the case $\dim(M) = 4$, there's a further term we could add to the action:

$$S_{\text{top}}[\nabla] = \frac{\mathrm{i}\theta}{8\pi^2}\int_M \mathrm{tr}(F\wedge F) = \frac{\mathrm{i}\theta}{8\pi^2}\int_M \epsilon^{\mu\nu\rho\sigma}\,\mathrm{tr}(F_{\mu\nu}\,F_{\rho\sigma})\,\mathrm{d}^4x \tag{8.120}$$

where θ is a new coupling constant, sometimes known as the *vacuum angle*.

The integral in (8.120) is a topological invariant: we defined it without picking any metric on M, instead using the totally antisymmetric ϵ symbol $\epsilon^{\mu\nu\rho\sigma}$ to join up the indices on the two F's. Furthermore, when M is compact, $S_{\text{top}}[\nabla]$ is completely independent of the connection ∇ on our bundle $P \to M$! To see this, note that after changing the connection

$\nabla \to \nabla + A$, the difference[90]

$$\mathrm{tr}(F_{(\nabla+A)} \wedge F_{(\nabla+A)}) - \mathrm{tr}(F_\nabla \wedge F_\nabla) = \mathrm{d}\left[\mathrm{tr}\left(A \wedge \nabla A + \frac{2}{3} A \wedge A \wedge A\right)\right] \tag{8.121}$$

is a total derivative. Thus the integral over this term vanishes if $\partial M = \emptyset$, or more generally if we require $A|_{\partial M} = 0$. Consequently, classical physics is completely unaffected if we include this term in the action; the Euler–Lagrange equations of

$$S[\nabla] = S_{\mathrm{YM}}[\nabla] + S_{\mathrm{top}}[\nabla] \tag{8.122}$$

are identical to those of $S_{\mathrm{YM}}[\nabla]$.

In particular, when $P \to M$ is a topologically trivial bundle (*i.e.* $P = M \times G$ globally), evaluating S_{top} on the trivial connection $\nabla = \partial$, we find

$$S_{\mathrm{top}}[\partial] = 0 \tag{8.123}$$

since $F_\partial = 0$ obviously. Our argument that the value of $S_{\mathrm{top}}[\nabla]$ is the same for *all* connections on a fixed topological type of bundle now shows that $S_{\mathrm{top}}[\nabla] = 0$ for any connection on the trivial bundle $P = M \times G$. However, we can't use this argument to conclude that $S_{\mathrm{top}}[\nabla]$ is always vanishes because for a topologically non–trivial bundle P (like a higher dimensional version of our example of the Mobius strip), it is *not* possible to express the connection ∇ as $\partial + A$ for any A defined globally over M and indeed for topologically non–trivial bundles, $S_{\mathrm{top}}[\nabla] \neq 0$. Though I won't prove it here, the factor of $1/8\pi^2$ in (8.120) ensures that when the gauge group $G = SU(N)$, in fact

$$\frac{1}{8\pi^2} \int_M \mathrm{tr}(F \wedge F) \in \mathbb{Z} \tag{8.124}$$

so that $S_{\mathrm{top}}[\nabla]$ is just $\mathrm{i}\theta$ times an *integer*. The particular integer we get is known as the **instanton number** in physics, or the **second Chern class** in maths, and gives us information about how 'topologically twisted' our bundle is.

If we also include the topological term (8.120) in the action and sum over all bundle topologies — allowing for instantons — then the partition function becomes

$$Z_{\mathrm{YM}}[\mathrm{g}^2_{\mathrm{YM}}, \theta, g] = \sum_{\substack{\text{topologically} \\ \text{distinct } P}} \left\{ \int_{\mathcal{A}/\mathcal{G}} DA\, \mathrm{e}^{-S[\nabla]} \right\} = \sum_{n \in \mathbb{Z}} \mathrm{e}^{-\mathrm{i}\theta n} \int DA\, \mathrm{e}^{-S_{\mathrm{YM}}[\nabla]} \tag{8.125}$$

where in the second equality I've used without proof the fact that, when G is semi–simple, the instanton number n gives a complete specification of the topological type of the bundle $P \to S^4$. We see that using the action (8.122) weights topologically distinct bundles by the factor $\mathrm{e}^{\mathrm{i}\theta n}$, where n is the instanton number. The value of the coupling θ allows us to vary the relative importance of different instanton contributions.

[90] It's a good exercise to derive this result for yourselves, particularly if you're taking the Applications of Differential Geometry to Physics course. If not, just treat what I'm writing here as a schematic way to keep track of all the derivatives and gauge fields.

The presence of the $\epsilon^{\mu\nu\rho\sigma}$ in the definition of $S_{\rm top}$ implies that this term violates CP symmetry, with the size of CP violation being related to θ (at least classically). Physically, such CP violation would lead to effects including the generation electric dipole moment for the neutron. Thus, the observed *absence* of any such dipole moment provides a constraint on the size of θ. In fact, experiments can test for neutron dipole moments very sensitively, and they tell us that in QCD, $|\theta| \ll 10^{-9}$. The small size of θ is known as the *strong CP problem* and calls for an explanation. The favoured theoretical solution (proposed by Roberto Peccei and Helen Quinn in 1977) involves promoting the coupling constant θ to a new *field* $\theta(x)$. While there's as yet no experimental evidence for this field, quite remarkably t it turns out that the theoretical properties $\theta(x)$ must have to enable it to solve the strong CP problem also automatically make it a good candidate to be an important constituent of dark matter. This is just one of many striking examples where the apparently different arenas of High Energy Theory and Cosmology come together.