Gauge theory diary

Ethan Lake

September 14, 2020

1 *The $\mathbb{C}P^N$ model and emergent electromagnetism* ✓

The $\mathbb{C}P^N$ model is defined by the Lagrangian

$$\mathcal{L} = \frac{1}{g^2}\left(|\partial_\mu z|^2 - |z^\dagger \partial_\mu z|^2\right), \tag{1}$$

where z is a $N+1$ component field, subject to the relations

$$|z|^2 = 1, \qquad (z_1, \dots, z_{N+1}) \sim (e^{i\alpha(x)} z_1, \dots, e^{i\alpha(x)} z_{N+1}). \tag{2}$$

Today we will do the following: first, we will make some brief mathematical comments on $\mathbb{C}P^N$ which didn't feel important enough to put as a standalone diary entry. Next we will go to two spacetime dimensions and do the standard mean field analysis at large N to show that in the mean field approximation, the theory is gapped with mass

$$m = \Lambda e^{-\pi/(g^2 N)}. \tag{3}$$

We will then show that at long distances, the effective action for the theory is Maxwell electrodynamics.

Edit: after writing this, I realized that the large-N analysis is essentially textbook material. The nontrivial parts of this diary are thus the discussion on kinematic issues and the expression for the effective charge in terms of N and m.

1.1 Mathematical preliminaries

First, a few general comments on $\mathbb{C}P^N$, which is the space of complex lines through the origin in $\mathbb{C}^{N+1}$. More often we see $\mathbb{C}P^N$ written as the space $(z_1, \dots, z_{N+1})/\sim$, where $(z_1, \dots, z_{N+1}) \sim (\lambda z_1, \dots, \lambda z_{N+1})$, $\lambda \in \mathbb{C}$, and $(z_1, \dots, z_{N+1}) \neq 0$. If we take

$$\lambda = \frac{1}{\sum_{i=1}^{N+1} |z_i|^2} \tag{4}$$

and multiply through, we get the conditions written in the problem statement, which shows us that[1]

$$\mathbb{C}P^N = S^{2N+1}/U(1). \tag{5}$$

[1] A similar thing occurs for $\mathbb{R}P^N$: each line through the origin in $\mathbb{R}^{N+1}$ can be identified with a pair of points on the unit sphere S^N, and so $\mathbb{R}P^N \cong S^N/\mathbb{Z}_2$. Random stream-of-conciousness fact: $\mathbb{R}P^N$ is non-orientable when N is even, and orientable when N is odd. To see this, realize that the antipodal identification in the $\mathbb{Z}_2$ quotient of S^N is performed by sending a vector $\mathbb{R}^{N+1} \ni v \mapsto -v$. For N odd the map $v \mapsto -v$ is equivalent to a rotation, since it has determinant 1, while for even N it is an inversion, since it has determinant -1. Thus the S^N in the definition of $\mathbb{R}P^N$ is identified with itself in an even / odd way depending on N, and the claim follows.

For example, $\mathbb{C}P^1 = S^3/U(1)$, as realized via the Hopf map.

A way to construct $\mathbb{C}P^N$ as a quotient which is often more useful is

$$\mathbb{C}P^N \cong \frac{SU(N+1)}{U(N)}, \tag{6}$$

where the $U(N)$ factor is embedded via the map

$$U(N) \ni u \mapsto u \oplus \frac{1}{\det u} \in SU(N+1), \tag{7}$$

which ensures that the image of $U(N)$ in $SU(N+1)$ has determinant 1. Alternatively, one might write this as

$$\mathbb{C}P^N \cong \frac{SU(N+1)}{S[U(N) \times U(1)]}, \tag{8}$$

where $S[\dots]$ denotes the part of $\dots$ with unit determinant.

To prove this, we note that $SU(N+1)$ acts transitively on $\mathbb{C}P^N$: it acts transitively on the unit sphere in $S^{2N+1} \subset \mathbb{C}^{N+1}$ and therefore acts transitively on $\mathbb{C}P^N$, which as we saw is a quotient of the aforementioned sphere by $U(1)$. Thus we can identify

$$\mathbb{C}P^N \cong SU(N+1)/\mathrm{Stab}_x, \tag{9}$$

where $\mathrm{Stab}_x \subset SU(N+1)$ is the stabilizer subgroup of an arbitrary point $x \in S^N_{\mathbb{C}}$. This is basically the first isomorphism theorem: the image of a point x under the group action is isomorphic to the group modulo the kernel of the action, which is Stab_x. Anyway, to figure out Stab_x, we can just pick a convenient $x \in \mathbb{C}P^N$, e.g. $x = \{(e^{i\theta}, 0, \dots, 0) \,|\, \theta \in \mathbb{R}/2\pi\mathbb{Z}\}$. This is clearly fixed under the $U(N)$ action on the last N coordinates, and also fixed under the $U(1)$ action on the first coordinate. We can combine these two actions to get something in $SU(N+1)$ if we shift the first coordinate by $\det^{-1}$ of the matrix acting on the last N coordinates, and so $\mathrm{Stab}_x = U(N)$, with the embedding into $SU(N+1)$ as written above. This then proves the claim.

1.2 Symmetries and alternate path integral representation

Showing that the Lagrangian $\mathcal{L}$ is invariant under gauge transformations $z \mapsto e^{i\alpha(x)} z$ (which it must be, since these generate the $U(1)$ in the quotient) is straightforward after making use of $|z|^2 = 1$ and $\partial |z|^2 = 0 \implies z^\dagger \partial z = -z \partial z^\dagger$. One can emphasize this point by noting that $|z^2| = 1$ means that the linear combination

$$A_\mu \equiv \frac{1}{2i}(z^\dagger \partial_\mu z - z \partial_\mu z^\dagger) \tag{10}$$

transforms under a gauge transformation as $A \mapsto A + d\alpha$, meaning that $\partial_\mu - iA_\mu$ is an appropriate covariant derivative. One is then led to write down an alternate representation of the theory as (now in $i\mathbb{R}$ time)

$$Z = \int \mathcal{D}A \mathcal{D}\lambda \mathcal{D}z \exp\left(-\frac{1}{g^2} \int \left((D_A z)^\dagger D_A z - \lambda(|z^2| - 1) \right) \right), \tag{11}$$

where now A is an independent field (which just gets set to the definition given above when it's integrated out, which can be done exactly since we haven't included any kinetic terms for it) and λ enforces the $|z|^2 = 1$ constraint. One can confirm that this representation works by integrating out λ (since we're in $i\mathbb{R}$ time, the contour is along $i\mathbb{R}$) to get the sphere constraint on z, and then shifting A by

$$A \mapsto A + iz\partial z^\dagger, \tag{12}$$

which eliminates the coupling between A and z, and results in an action like $|A|^2 + |\partial z|^2 - |z\partial z^\dagger|^2$. Integrating out A then produces the original $\mathbb{C}P^1$ Lagrangian.

<digression>

A brief digression on the $N = 1$ case and its relation to the $O(3)$ non-linear sigma model: the claim is that these two models are related through the Hopf map $S^3 \to S^2$:

$$n^i = z^\dagger \sigma^i z. \tag{13}$$

Here the RHS is a vector since the zs are in the fundamental of $SU(2)$, and $1/2 \otimes (1/2)^* \cong 1/2 \otimes 1/2 = 0 \oplus 1$: choosing $\mathbf{1}$ instead of σ^i would project onto 0, while choosing σ^i gives us the vector (adjoint) rep. n^i is also normalized properly.[2]

Now the z spinors live in $S^3 \cong SU(2)$, but because of the $U(1)$ redundancy the target space for the $\mathbb{C}P^1$ model is actually $SU(2)/U(1) = S^2$, the same as the $O(3)$ nlsm. As a sanity check, we can check the global symmetries on each side. The $\mathbb{C}P^1$ has a global $U(2)$ symmetry but a local $U(1)$ symmetry, so the actual global symmetry is[3]

$$PU(2) = U(2)/U(1) = PSU(2) = SU(2)/\mathbb{Z}_2 = SO(3). \tag{15}$$

This matches the global symmetry of the nlsm if we pretend the global symmetry is $SO(3)$ instead of $O(3)$. We can get the full symmetry by taking into account the reflection that extends $SO(3)$ to $O(3)$, which is represented by the matrix $-\mathbf{1}$ and hence sends $n \mapsto -n$. On the $\mathbb{C}P^1$ side, this reflection is implemented by charge conjugation, which is unitary and acts through the invariant antisymmetric form of $SU(2)$:

$$\mathbb{Z}_2^C : z \mapsto Jz^*, i \mapsto i, \qquad J = -iY. \tag{16}$$

Indeed, n is odd under this symmetry:

$$\mathbb{Z}_2^C : z^\dagger \sigma^j z \mapsto z^\dagger (J^T \sigma^j J)^T z = -z^\dagger (J[\sigma^j]^T J) z = z^\dagger (\sigma^j J^2) z = -z^\dagger \sigma^j z. \tag{17}$$

[2] The matrix $\sigma^\mu |z\rangle\langle z|\sigma^\mu$ is Hermitian and has trace 3 (since $\mathrm{Tr}[|z\rangle\langle z|] = 1$), so we can write it as $3\mathbf{1}/2 + c_i\sigma^i$. Also, since $n^i n_i$ is rotation-invariant, we can perform an $SU(2)$ rotation on the zs to any convenient spinor: we will choose $|z\rangle = (1, 0)^T$. Then $|z\rangle\langle z| = \mathbf{1}/2 + Z/2$, and conjugating by σ^j and summing over j tells us that $c_z = -1/2$, and so

$$n^i n_i = \langle z|(3\mathbf{1}/2 - Z/2)|z\rangle = 1 \tag{14}$$

as required.

[3] To get this, we have used the second isomorphism theorem: if H, N are subgroups of G with N normal, then $(NH)/N \cong H/(H \cap N)$. This just says that one can't "cancel the Ns" in $(NH)/N$ since if $N \cap H \neq 0$ then some elements of H will also be killed by the quotient. We have used the special case where $N = SU(n), H = Z(U(n)) = U(1)$ to show the obvious fact that $PU(n) = (SU(n)U(1))/U(1) \cong SU(n)/(U(1) \cap SU(n)) = PSU(n)$.

One similarly checks that the $\mathbb{C}P^1$ Lagrangian is invariant under $\mathbb{Z}_2^C$, and that $A \mapsto -A$, which is how a gauge field is expected to transform. Now $O(3) = SO(3) \times \mathbb{Z}_2^C$ is a simple direct product, since $\mathbb{Z}_2^C$ is represented by the central element $-\mathbf{1}$. This means that $\mathbb{Z}_2^C$ should commute with the $PSU(2)$ symmetry on the $\mathbb{C}P^1$ side, which it does:

$$JU^* = UJ \implies J(Uz)^* = UJz^*, \quad \forall U \in PSU(2), \tag{18}$$

where the first equality holds since $J\mathcal{K}$ with $\mathcal{K}$ complex conjugation is the antilinear automorphism from the spin 1/2 representation of $SU(2)$ to itself, which is the manifestation of that representation's pseudoreality.

Let us now show the equivalence between the nlsm and $\mathbb{C}P^1$ actions explicitly. The kinetic term for the nlsm maps as

$$\frac{\rho}{2}|\nabla \mathbf{n}|^2 = \rho\left(\delta_{il}\delta_{jk} - \frac{1}{2}\delta_{ij}\delta_{kl}\right)(z_i^* \overset{\leftrightarrow}{\nabla} z_j)(z_k^* \overset{\leftrightarrow}{\nabla} z_l) = 2\rho\left((z^\dagger \nabla z)^2 + |\nabla z|^2\right). \tag{19}$$

Other derivatives map similarly.

Depending on the context, we will also need to find a map for the WZW term. We will show that the WZW term at level S for the nlsm goes to $2Sz^\dagger \partial_\tau z$. This is rather obvious if one simply identifies the generators of $H_2(S^2) = \mathbb{Z}$ on both sides. However I've never seen the algebra actually worked out, so here we will give an explicit proof. It first helps to write the action density of the WZW term as

$$s_{WZW} = 2S \int \mathrm{Tr}[\widehat{n}\, d\widehat{n} \wedge d\widehat{n}], \tag{20}$$

where $\widehat{n} \equiv \mathbf{n}\cdot\boldsymbol{\sigma}/2$ so that $n^a = \mathrm{Tr}[\sigma^a \widehat{n}]$, and the integral is over a two-manifold which bounds the thermal circle. Note that another way of writing $\widehat{n}$ is as $\widehat{n} = |z\rangle\langle z| - \mathbf{1}/2$, which follows from "$1 = 1/2 \otimes 1/2 - 0$". Using this representation, we have

$$\begin{aligned}
s_{WZW} &= m_0 \int \mathrm{Tr}\left[(2|z\rangle\langle z| - \mathbf{1})\left(|dz\rangle\langle z| + |z\rangle\langle dz|\right) \wedge \left(|dz\rangle\langle z| + |z\rangle\langle dz|\right)\right] \\
&= m_0 \int \mathrm{Tr}\left[(2|z\rangle\langle z|dz\rangle\langle z| + |z\rangle\langle dz| - |dz\rangle\langle z|) \wedge (|dz\rangle\langle z| + |z\rangle\langle dz|)\right] \\
&= m_0 \int \left(2\langle z|dz\rangle \wedge \langle z|dz\rangle + 2\langle z|dz\rangle \wedge \langle dz|z\rangle + 2\langle dz| \wedge |dz\rangle + \langle dz|z|\rangle \wedge \langle dz|z\rangle - \langle z|dz\rangle \wedge \langle z|dz\rangle\right) \\
&= 2m_0 \int \langle dz| \wedge |dz\rangle \\
&= 2im_0 \int da,
\end{aligned} \tag{21}$$

where $a = -i\langle z|dz\rangle$ and we have used the supercommutativity of $\wedge$ and the fact that $\langle dz|z\rangle = -\langle z|dz\rangle$. Hence the WZW term becomes

$$S_{WZW} = 2im_0 \int d^dx\, d\tau\, a_\tau = 2m_0 \int d^dx\, d\tau\, z^\dagger \partial_\tau z, \tag{22}$$

as claimed.

</digression>

1.3 Effective action at large N and emergent electromagnetism

We now integrate out z to get an effective action ($\mathbb{R}$ time[4])

$$S_{eff}[A,\lambda] = iN \ln\det(-D_A^2+\lambda) - \frac{1}{g^2}\int \lambda. \tag{23}$$

Note that the D_A^2 is really D_A^2, and not $|D_A|^2$. As a first step, consider the case where $A=0$ and where we approximate λ by a constant. In the large N limit, we can figure out what this constant is by taking the saddle point of the effective action with respect to λ. This produces

$$\frac{1}{Ng^2} = i\mathrm{Tr}\frac{1}{-\partial^2+\lambda}, \tag{24}$$

where the trace is to be carried out in momentum space (the factor of the spacetime volume has canceled on both sides). We rewrite this as

$$\frac{1}{Ng^2} = \int_{\mathbb{R}^2}\frac{1}{k^2+\lambda}, \tag{25}$$

where the integral is over Euclidean momenta. Doing the integral up to a cutoff Λ gives a saddle point value of

$$\frac{4\pi}{Ng^2} = \ln(\Lambda^2/\lambda_*) \implies m_* = \sqrt{\lambda_*} = \Lambda e^{-4\pi/(g^2N)}, \tag{26}$$

which is the usual dimensional transmutation result: even though the theory was conformally invariant at the classical level, quantum effects produce a non-zero beta function through the Λ dependence of λ. The fact that we get a massive theory is also interesting because there is no obvious symmetric mass term we can write down in the UV (e.g. $m^2\sum_i |z_i|^2 = m^2$ doesn't work because of the constraint), so that this mass must be dynamically generated.

Note that if the dimension of spacetime were greater than two, then the presence of the cutoff would make the integral on the RHS of (25) bounded for arbitrarily small λ, and so if the t' Hooft coupling Ng^2 were small enough, there would be no nonzero solution to the mean-field equation for λ. This would indicate spontaneous symmetry breaking[5] and a Higgsing of the gauge field, and we'd have to go back and expand about the correct vacuum (however, since we're focusing on two dimensions, this need not concern us). Spontaneous symmetry breaking here can be argued by looking at the mean-field equations (which are exact as $N\to\infty$) for a configuration where z is constant. The mean-field equation for z is just $z\lambda = 0$. Hence if $\lambda_*\neq 0$ then we must be in the symmetric phase (which is the case in two dimensions), while if $\lambda_* = 0$ then we can have $\langle z\rangle \neq 0$, giving SSB. Hence the lower critical dimension for this theory is two.

Now we will take λ to be a constant and expand about small A, deriving the effective action for A at one-loop order. We write the $\ln\det$ as, now in $i\mathbb{R}$ time (note to self: the sign

[4] It's usually better to do this kind of stuff in $\mathbb{R}$ time since then the integration contour for λ runs along the $\mathbb{R}$ axis, which makes keeping track of signs and is eaiser.

[5] Santiy check: SSB occurs at small t' Hooft coupling since small Ng^2 means small fluctuations and hence increased tendency to order.

in the $-\lambda$s is annoying; should see if this can be fixed),

$$-N\mathrm{Tr}\ln(-D_A^2-\lambda) = -N\mathrm{Tr}\ln\left[(-\partial^2-\lambda)\left(1+\frac{A^2+2iA\partial+i(\partial A)}{-\partial^2-\lambda}\right)\right], \tag{27}$$

where the ∂ is understood to act on the z's. The overall factor of $(-\partial^2-\lambda)$ in the logarithm is an unimportant constant, so we drop it. We now expand the logarithm to second order. The first order contribution yields a term like G_zA^2 (G_z is the z propagator), plus things which vanish after integration; this first term doesn't depend on the momentum of A and hence can be dropped. To $O(A^2)$, the next diagram that contributes is a polarization bubble for the A propagator. Reading the Feynman rules off from the interaction vertex $(2iA\partial+i(\partial A))^2$, we get (still in Minkowski space)

$$-N\mathrm{Tr}\ln(-D_A^2-\lambda) \approx -\frac{N}{2}\int_{p,q} A^\mu(p)A^\nu(-p)\frac{(2q+p)_\mu(2(q+p)-p)_\nu}{(q^2-\lambda)((q+p)^2-\lambda)}. \tag{28}$$

We are interested in the IR properties of this action, so we will take the small p limit. By gauge invariant we already know the form of the effective action, but we need to go through the details to figure out what the effective electric charge is. Since the first term in the integrand has no p-dependence, we focus on the second term, since this is the term that will produce the Maxwell term. The integral is done with the usual Feynman parameters: we use the Feynman parameters to simplify the denominator, shift the q momentum to eliminate the $q\cdot p$ term in the denominator, and then drop integrals which vanish because their integrands are odd. This yields (here $\Delta = -((x^2+x)p^2-\lambda)$)

$$\begin{aligned}\int_{p,q} A^\mu(p)A^\nu(-p)\frac{1}{2}\frac{(2q+p)_\mu(2q+p)_\nu}{(q^2-\lambda)((q+p)^2-\lambda)} &= \frac{1}{2}\int_{p,q} A^\mu(p)A^\nu(-p)\frac{(2(q-xp)+p)_\mu(2(q-xp)+p)_\nu}{(q^2-\Delta)^2} \\ &= \frac{1}{2}\int_{p,q} A^\mu(p)A^\nu(-p)\frac{4q_\mu q_\nu + p_\mu p_\nu(1-2x)^2}{(q^2-\Delta)^2}.\end{aligned} \tag{29}$$

We can now look up the integrals. To get a nice answer, we need to focus on small p (i.e. $O(p^2)$), so that e.g.

$$\frac{p_\mu p_\nu}{\Delta} \approx \frac{p_\mu p_\nu}{\lambda}, \qquad \ln\Delta \approx \ln\lambda - \frac{(x^2+x)p^2}{\lambda}. \tag{30}$$

Working in this approximation, the p-dependent parts of the effective action for A become

$$-iN\frac{1}{12\pi\lambda}\int_p A^\mu(p)A^\nu(-p)(g_{\mu\nu}p^2 - p_\mu p_\nu), \tag{31}$$

and so we have generated an emergent Maxwell theory with effective charge

$$\frac{1}{e^2_{\text{eff}}} = \frac{N}{6\pi\lambda}. \tag{32}$$

Here the g^2 dependence of the charge is contained within the saddle-point value for λ. Just like the Schwinger model this theory confines, with the spectrum consisting of z-$\bar{z}$ bound states. We can argue that this theory is asymptotically free by computing the β function for

g; if this is asympotically free then from the mean field solution for λ we know that e^2_{eff} will be as well. The beta function for g can be found by requiring that the effective potential for λ be independent of Λ. The calculation of the effective potential is essentially contained in the diary entry on the Gross-Nevau model, and so we won't repeat it in detail here. The gist is that one basically does the integral $\int dk\, k \ln(\lambda + k^2)$ by use of the replica trick and dim reg; doing so produces an effective potential like

$$V_{\text{eff}}(\lambda) \sim \lambda/g^2 + N\lambda \ln(\lambda/\Lambda^2), \tag{33}$$

which should remind us of the characteristic log found in the effective potentials for the Higgs fields in four-dimensional gauge theories (see the entries on the CW potential and fluctuation-induced first-order Higgs transitions). Requiring this to be independent of Λ gives the β function for the t' Hooft coupling as

$$\beta_\gamma \sim \gamma^2, \qquad \gamma \equiv g^2 N, \tag{34}$$

which indeed demonstrates asympotitic freedom.

2 Dynamic generation of topological photon mass and domain wall anomalies ✓

Consider QED$_3$, namely

$$S = \int d^3x \left(i\bar{\psi} \not{D}_A \psi + m(x)\bar{\psi}\psi - \frac{1}{2} F \wedge \star F \right). \tag{35}$$

Today's diary entry has two parts. First, we will show that radiative corrections from the fermions induce a Chern-Simons term and hence a topological mass for the photon (at momentum scales smaller than the fermion mass). Secondly, we will consider a domain wall where $m(x)$ changes sign and will show that such an object hosts chiral fermions with a gauge anomaly, with the anomaly being canceled by the dynamically generated CS terms away from the domain wall.

Chern-Simons terms: To find the CS term induced for the gauge field, we just have to compute the one-loop contribution to the effective action for the gauge fields after integrating out the fermions.[6] The relevant integral is

$$\text{bubble}^{\mu\nu} = (-1)i^2(-ie)^2 \int_p \text{Tr}\left[\gamma^\mu \frac{\not{p}+\not{q}+m}{(p+q)^2-m^2}\gamma^\nu \frac{\not{p}+m}{p^2-m^2}\right]. \tag{36}$$

After doing Feynman parameters to simplify the denominator there are two contributions to the integral in the large m limit: one proportional to $m\gamma^\mu \not{q}\gamma^\nu$, and another which contains terms like $g^{\mu\nu}m^2$ which do not depend on $\text{sgn}(m)$ (the others vanish under $\not{p} \to -\not{p}$ or under the trace). The latter terms will get renormalized away when we regularize e.g. a la PV, so we will ignore them in what follows. In mostly-negative signature $\mathbb{R}$-time, our γ matrices are

$$\gamma^0 = X, \qquad \gamma^1 = iY, \qquad \gamma^2 = iZ. \tag{37}$$

Thus they satisfy

$$\text{Tr}[\gamma^\mu\gamma^\nu\gamma^\lambda] = -2i\epsilon^{\mu\nu\lambda}. \tag{38}$$

Using this, we get (there are terms in the numerator involving the Feynman parameter and $\not{q}$ coming from the shift in p, but these end up cancelling due to the spin sum)

$$\text{bubble}^{\mu\nu} = 2ie^2\epsilon^{\mu\lambda\nu}\int_{p,x}\frac{q_\lambda m}{(p^2-\Delta)^2}, \tag{39}$$

where Δ is a function of m, q and the Feynman parameter x. Doing the integral,

$$\begin{aligned}\text{bubble}^{\mu\nu} &= -2ei\epsilon^{\mu\nu\lambda}q_\lambda m\frac{i}{(4\pi)^{3/2}}\Gamma(1/2)\Delta^{-1/2}\\ &= \text{sgn}(m)\frac{e^2}{4\pi}\epsilon^{\mu\lambda\nu}q_\lambda,\end{aligned} \tag{40}$$

where we have taken the long-wavelength limit where $m^2 \gg q^2$ so that $\Delta \to m^2$. This diagram appears in the effective action for the gauge field with a coefficient of $-1/2$ since it is the quadratic term in the expansion of the $\ln\det$, and so it thus gives us the CS term at level $-\text{sgn}(m)/2$. The CS term violates parity (i.e. either reflection or time-reversal) and so our theory violates parity if $m \neq 0$—but we already knew this, since fermion mass terms are parity-odd in odd dimensions (this provides another derivation of the oddness of m: under T the CS level is odd, so that m must be odd as well). The fractional level here arises since we didn't properly regulate the theory using e.g. PV regularization. If we did and chose e.g. the mass of the PV field to be large and positive, we would get a level of $k = [\text{sgn}(m)-1]/2 \in \{0,-1\}$, which is well-defined. We could also have chosen the PV field to have a large negative mass, in which case the level would be valued in $\{0,1\}$. We have a freedom of changing the sign of m and also the sign of the mass of the PV regulators, but the relative sign between the two masses is physical and determines what CS level we get (the two choices are related by time reversal). Another comment is that as we mentioned

[6] Recall that when the chiral anomaly is derived using the Feynman diagram approach, the one-loop calculation is exact, due to the fact that the anomaly is computed using the index theorem, and so the coefficient of the $F \wedge F$ term is quantized, implying one-loop exactness. The quantization of the CS term means that similar one-loop exactness applies in this setting.

in an earlier footnote, since the IR CS level is quantized, our calculation must be one-loop exact—diagrams with l loops scale with the coupling constant as $(e^2)^{l-1}$, so if $l \neq 1$ made a contribution we could tune e^2 continuously and get a non-integral level.

Generalizing slightly to N_f fermions of masses m_i, we have

$$k_{IR} = k_{UV} - \sum_i \frac{\mathrm{sgn}(m_i)}{2}, \tag{41}$$

where k_{UV} is the level of the sum of the PV regulator fields. In particular if $N_f \in (2\mathbb{Z}+1)$, we always have a fractional CS level in the UV. Also note that when $m_i = 0$, if $N_f \in (2\mathbb{Z}+1)$ the UV theory always breaks parity symmetry, since we have an odd number of PV fields—this is the parity anomaly (if $N_f \in 2\mathbb{Z}$, we could choose $N_f/2$ positive-mass PV fields and $N_f/2$ negative-mass PV fields, and the effective CS level would be zero). Note that for odd N_f parity is broken in the UV; it is not an infrared effect associated with the CS level generated by integrating out the fermions and instead indicates our inability to regularize the theory in a symmetry-preserving way.

Domain wall and anomaly cancellation: Now we consider a domain wall where $m(x)$ changes sign. For concreteness, let $x^3 = z$ be the direction normal to the domain wall. Then the CS terms generated by the fermions are by themselves not gauge-invariant, since under $A \mapsto A - d\alpha$ the action changes as

$$S \mapsto S + \frac{2e^2}{4\pi} \int_w \alpha F, \tag{42}$$

where $\int_w$ is an integral over the domain wall and we have chosen the mass to be positive on the $z > 0$ side of the domain wall wolog. This gauge-non-invariance must be canceled by something living on the wall.

Indeed it is; let's solve the Dirac equation to get the relevant anomaly-cancelling zero modes. We have, say on the $z > 0$ side of the domain wall,

$$i(\not{D}^w_A + \gamma^2(\partial_z - ieA_z))\psi = -m(z)\psi, \tag{43}$$

where $\not{D}^w_A$ denotes the Dirac operator restricted to the wall. We choose an ansatz where $\psi = \eta(x^0, x^1) f(z = x^2)$, with η a zero mode of the Dirac operator restricted to the wall. We can choose it to have definite chirality under $\bar{\gamma}_w = i\gamma^0\gamma^1$ since the wall is two-dimensional. Let $\bar{\gamma}_w \eta = c_\eta \eta$. We can match the $m(z)$ on the RHS with the usual exponential factor, so we take

$$\psi = i\eta \exp\left(- \int_0^z dz'(m(z') - ieA_z) \right). \tag{44}$$

For this to work, we need

$$i\gamma^2 \eta = \eta. \tag{45}$$

But we have

$$i\gamma^2\eta = -i\gamma^0\gamma^1\eta = -\bar{\gamma}_w \eta = -c_\eta \eta, \tag{46}$$

and so if we choose η to be of negative chirality on the wall, we'll get a solution to the Dirac equation. A similar story gets told if we focus on $z < 0$: the two minus signs from

$m(-z) = -m(z)$ and $\partial_{-z} = -\partial_z$ cancel out, and in the end we get two chiral modes on the domain wall, propagating in the *same* direction, with action

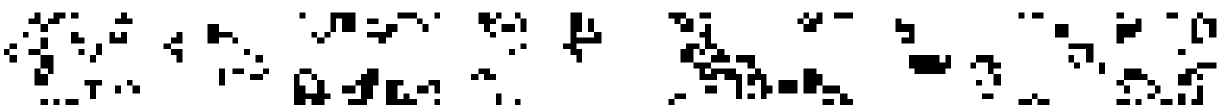
$$S_w = 2 \int_w i\bar{\eta} \not{D}_A \eta. \tag{47}$$

The fact that the two modes propagate in the same direction along the wall (and so propagate with opposite handedness in their respective half-planes) is because a region with $m(z) > 0$ is essentially the time-reversed version of a region with $m(z) < 0$, since the Dirac mass is odd under time reversal in three dimensions.

Since the wall modes have a net chirality and coupled to the gauge field, they will have a gauge anomaly. Under $A \mapsto A - d\alpha$, the action thus shifts as

$$S \mapsto S - \frac{e^2}{2\pi} \int_w \alpha F. \tag{48}$$

This cancels the gauge anomaly coming from the bulk CS terms, and so the full action is well-defined.



3 *Checking the chiral anomaly for non-Abelian gauge theory with operator relations* ✓

Today we will see how the chiral anomaly in YM theory can be derived in a UV way by explicitly computing the divergence of the axial current by way of a point-splitting approach used to define the current operator (the simpler $U(1)$ case is in P&S). In cond-mat language, this way of computing the anomaly with point-splitting is analogous to determining the anomaly by testing whether or not the action of the symmetry can be realized in an on-site manner.

To this end, consider a non-Abelian gauge theory coupled to Dirac fermions in four dimensions. We will first find the terms in the divergence of the chiral vector current which are quadratic and cubic in the gauge field, and show that they match with the usual $F \wedge F$ form for the anomaly. We will do the calculation by explicitly computing the divergence of j_5^μ and splitting the fermion two point function as

$$\bar{\psi}(x - \epsilon/2) P \exp\left(i \int_{x-\epsilon/2}^{x+\epsilon/2} A_\mu \, dx^\mu \right) \psi(x + \epsilon/2), \tag{49}$$

and taking $\epsilon \to 0$ in a symmetric limit.

The chiral vector current comes from the symmetry $\psi \mapsto \exp(-\gamma^5\alpha)\psi$, where by γ^5 we really mean $\gamma^5 \otimes \mathbf{1}_G$, where $\mathbf{1}_G$ is the identity matrix for the representation of the gauge group that the fermions live in. From the $i\bar{\psi}\not{D}\psi$ part of the action (with the convention $D = \partial + ieA$), we get

$$\partial_\mu \langle j_5^\mu \rangle = \partial_\mu \langle \bar{\psi}(x-\epsilon/2)\gamma^\mu\gamma^5 W \psi(x+\epsilon/2)\rangle, \tag{50}$$

where we've written W for the Wilson line connecting the two fermions. We take the derivatives and use the equations of motion

$$\not{\partial}\psi = ie\not{A}^a t^a \psi, \qquad \partial_\mu \bar{\psi}\gamma^\mu = -ie\bar{\psi}\not{A}^a t^a \tag{51}$$

on the terms containing $\partial_\mu\psi$ and $\partial_\mu\bar{\psi}$. This produces

$$\partial_\mu \langle j_5^\mu \rangle = ie\langle \bar{\psi}(x+\epsilon/2)\gamma^\mu\gamma^5 \left(\partial_\mu A^a_\nu t^a \epsilon^\nu W - A^a_\mu(x+\epsilon/2)t^a W + W A^a_\mu(x-\epsilon/2)t^a\right)\psi(x-\epsilon/2)\rangle, \tag{52}$$

where the first term comes from the derivative of the Wilson line. We now need to move all the Wilson lines to stand to the right of all the t^a matrices. To this end, we expand the Wilson line to first order in ϵ and write

$$W t^a \approx \left(1 + ie\int A^b t^b\right) t^a = t^a + ie\int A^b(t^a t^b - [t^a, t^b]) \approx t^a W + eA^b_\nu \epsilon^\nu f^{abc} t^c. \tag{53}$$

Then to first order in ϵ,

$$\partial_\mu \langle j_5^\mu \rangle = ie\langle \bar{\psi}(x+\epsilon/2)\gamma^\mu\gamma^5 \left((\partial_\mu A^a_\nu - \partial_\nu A_\mu)t^a\epsilon^\nu + A^a_\mu A^b_\nu \epsilon^\nu f^{cab} t^c\right)\psi(x-\epsilon/2)\rangle, \tag{54}$$

which we obtained by doing some re-arranging of the gamma matrices and expanding the A's about x (two terms coming from expanding the Wilson line cancel). We will see momentarily that the leading singularity in the fermion contraction will go as $1/\epsilon$, which will justify our first-order expansion of the terms in the previous equations.

Now we need to contract the fermions. To leading order, the fermion contraction is just a propagator connecting $x - \epsilon/2$ with $x + \epsilon/2$. However, this gives us something like $\mathrm{Tr}[\gamma^5\gamma^\mu\gamma^\nu]$, which is always zero. In order to get something nonzero, we are going to need four gamma matrices to give a nonzero trace when they get hit by the γ^5. The required four gamma matrices come from a process in which the background A field interacts once with the fermion line connecting the two points, forming a T-shaped Feynman diagram (see the picture in chapter 19 of P&S). The Feynman diagram for the $\partial_{[\mu}A_{\nu]}$ term gives

$$\partial_\mu \langle j_5^\mu \rangle = (ie)^2 \int_{p,q} \mathrm{Tr}\left[\frac{\not{p}+\not{q}}{(p+q)^2}\gamma^5\gamma^\mu \partial_{[\mu}A^a_{\nu]} t^a \epsilon^\nu A^b_\sigma(p) t^b \gamma^\sigma \frac{\not{q}}{q^2} e^{iq\epsilon - ip(x-\epsilon/2)}\right] + \dots. \tag{55}$$

Note that here that the brackets simply mean $\partial_{[\mu}A_{\nu]} = \partial_\mu A_\nu - \partial_\nu A_\mu$ (there is no 1/2! normalization). In our signature $\mathrm{Tr}[\gamma^5\gamma^\mu\gamma^\nu\gamma^\sigma\gamma^\rho] = -4i\epsilon^{\mu\nu\sigma\rho}$ and $\mathrm{Tr}[t^a t^b] = C(r)\delta^{ab}$ (where r

is the representation the fermions are in, e.g. $C(N) = 1/2$ for the fundamental of $SU(N)$), so

$$\text{relevant diagram} = 4e^2 C(r) i\epsilon^\nu \epsilon^{\mu\lambda\sigma\rho} \partial_{[\mu} A^a_{\nu]} \int_{p,q} \frac{(q+p)_\lambda A^a_\sigma(p) q_\rho}{(p+q)^2 q^2} e^{iq\epsilon - ip(x-\epsilon/2)}. \tag{56}$$

Since we are sending $\epsilon \to 0$, the large q limit is what will be relevant, so we can take the denominator to just be q^4. This is essentially because we are computing the divergences coming from splitting the product $\bar{\psi}\psi$, which come from high-momentum UV physics. Then the two integrals factor as a product: one is just $i\partial_\lambda A^a_\sigma$, while the other goes to

$$\int_q q_\rho \frac{e^{iq\epsilon}}{q^4} = -i\partial_{\epsilon_\rho}\left(\frac{i}{16\pi^2}\ln(1/\epsilon^2)\right) = -\frac{1}{8\pi^2}\frac{\epsilon_\rho}{\epsilon^2}, \tag{57}$$

where the other factor of i came from Wick rotation. This has a $1/\epsilon$ divergence, as promised earlier. The integral over p gave us a $\partial_\lambda A_\sigma$ term, and due to the antisymmetry of the ϵ tensor up front we can turn this into a $\partial_{[\lambda} A_{\sigma]}$ at the cost of a factor of $1/2$. This means that after contracting the fermions, the first term in (54) gives us

$$-\frac{C(r)}{16\pi^2}\epsilon^{\mu\nu\lambda\sigma}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(\partial_\lambda A^a_\sigma - \partial_\sigma A^a_\lambda). \tag{58}$$

Now we need to contract out the fermions in the other term that came from commuting the t^a through the Wilson line. Tracing out the spin and gauge indices using the rules described earlier, we get (the Feynman diagram looks the same, just with a different interaction vertex in between the fermions)

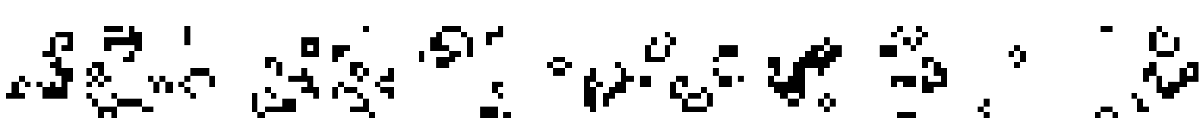
$$\text{relevant diagram} = 4iC(r)e^3 \epsilon^{\lambda\mu\sigma\rho} A^a_\mu A^b_\nu f^{cab} \epsilon^\nu \int_{p,q} \frac{(q+p)_\lambda}{(q+p)^2} A^c_\sigma(p) \frac{q_\rho}{q^2} e^{iq\epsilon - ip(x-\epsilon/2)}. \tag{59}$$

We do the integrals in the same way as before, and this term adds to the previous one to give us the expression for the divergence of the chiral current up to terms cubic in the gauge field.

Recapitulating, we have

$$\partial_\mu \langle j^\mu_5 \rangle = -\frac{C(r)e^2}{16\pi^2}\epsilon^{\mu\nu\lambda\sigma}\left((\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(\partial_\lambda A^a_\sigma - \partial_\sigma A^a_\lambda) + e(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(f^{abc} A^b_\lambda A^c_\sigma)\right) + \dots, \tag{60}$$

where $\dots$ is the second term with $\mu,\nu \leftrightarrow \lambda,\sigma$ and also the A^4 terms that we didn't write out due to laziness. This is exactly what we get from writing down the familiar $F^a \wedge F^a$ formula obtained with e.g. Fujikawa's method. In the common case where the gauge group is $SU(N)$ and the fermions are in the fundamental, the $C(r)$ becomes a $1/2$ and we get the $F^a \wedge F^a$ term with a $1/(32\pi^2)$ in front.



4 *BF theory basics*

Today we will carefully go through sections 1, 2, and 3 of "Coupling QFTs to TQFTs" [**?**] , explain how the various presentations of BF theory are constructed, and explain what the global symmetries are.

First, a note on notation: I will largely only write down dA in an integrand when A is actually a well-defined form, so that dA is exact and d actually acts as the exterior derivative. For example, given a connection A on a nontrivial $U(1)$-bundle, I will try to only write integrals over F_A, the curvature of A, and will try to avoid writing dA. Also, as a warning, I will generally be sloppy about keeping track of various minus signs coming from the supercommutativity of the exterior derivative and from interchanging the order of wedge products; in any case these are easy to figure out post facto. Another thing to keep in mind vis-a-vis factors of 2π: if $B \in H^p(X;\mathbb{Z})$ and $A \in H^q(X;\mathbb{Z})$, then $B \wedge A \in H^{p+q}(X;\mathbb{Z})$, i.e. the wedge product is a product operation on cohomology, meaning that the wedge product of two forms with integer periods is also a form with integer periods. We will thus alwaye divide by 2π-s in such a way that we are only wedging together forms in $H^*(X;\mathbb{Z})$, so as to avoid writing cohomology groups like $H^*(X;2\pi\mathbb{Z})$.[7]

First formulation: The first way of writing BF theory is

$$\frac{i}{2\pi}\int_X H \wedge (F_a + nA), \tag{61}$$

where A is a q-form gauge field and a is a $q-1$ form gauge field. Both are defined up to large gauge transformations[8]. For example, if $q = 1$ then a is a scalar, and $a \sim a + 2\pi\mathbb{Z}$. If $q = 2$, then a is a 1-form, with $a \sim a + \alpha$ for all α such that $\int_{M_1} \alpha \in 2\pi\mathbb{Z}$ for all closed 1-manifolds M_1. In the above action, H is a $(D-1)$-form Lagrange multiplier field.[9] If $\partial X \neq \emptyset$, we

[7] This is because $2\pi\mathbb{Z}$ is not a ring with the usual multiplication (e.g. $4\pi^2 \notin 2\pi\mathbb{Z}$), and so the cup product / wedge product operation no longer works.

[8] Large gauge transformations on a q-form gauge field are transformations $A \mapsto A + \alpha$, where α integrates to an element of $2\pi\mathbb{Z}$ on every closed q-manifold (and as such is a closed form). Note that α is a globally well-defined q form, although it may not be exact. Large gauge transformations of course never change the field strength and do not change the global well-definedness of A, despite some statements in the literature along the lines of "large gauge transformations take you between different magnetic flux sectors" when discussing th.

[9] In [**?**], H is taken to be quantized as

$$\int_{M_{D-q}} H \in 2\pi\mathbb{Z}, \tag{62}$$

for any closed $D-q$ submanifold M_{D-q}. Thus in this case, we can think of H as the field strength for some $U(1)$ $(D-q-1)$-form gauge field. I think we actually don't want to put this restriction on H though—if we do then the a field serves no purpose in the action (its job is to turn H into a field strength), and if we make this restriction H cannot come from a field arising from manipulating a Higgs theory (more on this in a sec).

take $H|_{\partial X} = 0$. This type of action comes from a generalized $U(1) \to \mathbb{Z}_n$ Higgs transition. Indeed, start with the Lagrangian

$$\mathcal{L}_{Higgs} = \frac{\rho^2}{2}|F_a + nA|^2, \qquad \rho \to \infty \tag{63}$$

which describes the deep IR theory for a charge n $(q-1)$-form field which has been Higgsed. Because we are interested in the $\rho \to \infty$ limit, we can freely add a term $\frac{\rho^*}{2}|H|^2$ to $\mathcal{L}$, where $\rho^* = \frac{1}{2\pi\rho}$. Now make the shift $\delta H = (\rho^*)^{-1} i \star (F_a + nA)$ (notice that we couldn't make such a shift if we were imposing a quantization condition on H's periods). This gives

$$\mathcal{L} = \frac{\rho^*}{2}|H|^2 + \frac{i}{2\pi} H \wedge (F_a + nA), \tag{64}$$

which goes over to the action we wrote above for $\rho \to \infty$.

Anyway, using this action, we can first do the integral over a: this tells us that H must be closed, and it sets the periods of H to be in $2\pi\mathbb{Z}$. Since the field strength of H vanishes (and we are assuming the spacetime manifold is torsion-free), we can globally Hodge-decompose it as

$$H = d\alpha_H + \omega_H, \tag{65}$$

where ω_H is harmonic and where the absence of a co-exact part comes from the fact that H must be closed. Up to constants relating $\mathcal{D}H$ to $\mathcal{D}\alpha_H$ and prefactors depending on $\dim H^{D-q}(X;\mathbb{Z})$, the path integral is

$$Z = \int \mathcal{D}A\mathcal{D}\alpha_H \sum_{\omega_H \in 2\pi H^{D-q}(X;\mathbb{Z})} \exp\left(\frac{i}{2\pi}\int_X (d\alpha_H + \omega_H) \wedge nA\right). \tag{66}$$

If $\partial X \neq 0$, the cohomology group becomes the relative cohomology $H^{D-q}(X, \partial X; \mathbb{Z})$.

The integral over α_H eliminates the local degrees of freedom (it sets the curvature of A to zero), while because the periods of ω_H over any closed manifold are in $2\pi\mathbb{Z}$, the sum over cohomology classes acts as a δ function setting

$$\int_{M_q} A \in \frac{2\pi}{n}\mathbb{Z} \tag{67}$$

for all closed M_q[10]. We can also get to this conclusion by integrating over H first: this tells us that n copies of A need to be exact; this is another way of saying that A is a $\mathbb{Z}_n$ gauge field.

[10]To see this more carefully, by Poincare duality we can write any $\omega_H \in 2\pi H^{D-q}(X;\mathbb{Z})$ as

$$\omega_H = \sum_{c \in H_q(X;\mathbb{Z})} 2\pi m_c \widehat{c}, \tag{68}$$

where $m_c \in \mathbb{Z}$ and the hat indicates the Poincare dual. When we put this in the path integral, we get something like (assuming the homology of X is torsion-free for simplicity)

$$\sum_{\{m_c\} \in \mathbb{Z}^{\dim H_q(X;\mathbb{Z})}} \prod_c \exp\left(i\int_c nA\right), \tag{69}$$

where the product is over all homology classes in $H_q(X;\mathbb{Z})$. This acts as a bunch of δ functions which set $n\int_c A \in 2\pi\mathbb{Z}$ for all closed q-manifolds c.

Thus the Lagrangian of the theory is zero after integrating out H and a, and A is set to be a flat $\mathbb{Z}_n$ connection. If the Lagrangian is just zero, why are we going through all this trouble of writing down 0 in a bunch of different ways? The point is that as physicists we like to work with continuum fields and like to draw intuition from actions, and so hence prefer to re-write things like $Z \sim \sum_{\alpha \in H^p(X;G)} 1$ in terms of path integrals (additionally, writing things in terms of actions lets us more easily generalize away from the topological limit, where the gauge fields are allowed to have nonzero field strength).

Symmetries: What are the gauge transformations and the symmetries of the action? We obviously have the gauge transformation $a \mapsto a + d\lambda_{q-2}$. We also need to have gauge transformations on A as well, but in order for these to leave the action invariant, we need F_a to shift as well. So, the fields are tied together in the way they transform, and gauge transformations act as

$$a \mapsto a + d\lambda_{q-2} - n\lambda_{q-1}, \qquad A \mapsto A + d\lambda_{q-1}. \tag{70}$$

This is a local symmetry of the action, and so it really is a gauge symmetry. But notice that it shifts a by something which is not an exact form! This has consequences for what the global symmetries are.

Now for the global symmetries. First, we see that we can shift A by

$$A \mapsto A + \frac{1}{n}\epsilon_q, \qquad \epsilon_q \in 2\pi H^q\left(X;\mathbb{Z}\right), \tag{71}$$

so that the integral of ϵ_q over any q-manifold is valued in $2\pi\mathbb{Z}$ (and hence ϵ_q is closed). Due to the n in the action, such a shift leaves the action invariant modulo $2\pi\mathbb{Z}$ provided we also shift the cohomology class of F_a appropriately, which is allowed since we are summing over all $[F_a]$ in the path integral.[11] This means that we have a global $\mathbb{Z}_n$ q-form symmetry.

It also looks like we have a global $(q-1)$-form symmetry, since if we shift $a \mapsto a + \epsilon_{q-1}$, where ϵ_{q-1} is any flat form (with periods equal to arbitrary elements of $\mathbb{R}/2\pi\mathbb{Z}$), then S is left invariant, since it only depends on F_a. What would be the charged objects under this symmetry? Of course, they would be the Wilson loops for a, namely $W_a(M_{q-1}) = \exp(i\int_{M_{q-1}} a)$. Normally, Wilson loop operators are gauge invariant, since they contain integrals over closed surfaces and since gauge transformations shift $U(1)$ gauge fields by exact forms. This is no longer true however, since we have a gauge transformation $a \mapsto a + n\lambda_{q-1}$, where λ_{q-1} is not exact and has no quantization conditions on its periods. Thus the $W_a(M_{q-1})$ Wilson loops are actually not gauge invariant, and the $U(1)$ higher symmetry actually does not exist.

A brief aside: one might be tempted to make $W_a(M_{q-1})$ gauge invariant by attaching a q-manifold M_q to it (with $\partial M_q = M_{q-1}$), and adding an integral of A over M_q to the Wilson operator. There are two problems with this: first, this is only possible if M_{q-1} is homologically trivial, in which case $W_a(M_{q-1})$ can not be charged under a higher symmetry in the first place. Second, the upgraded Wilson operator would be a trivial operator since it would be constructed using an integral of $F_a + nA$ over M_q, but $F_a + nA$ is trivial (which one sees by integrating out H).

[11] Note that we wouldn't also have to shift F_a if we had originally taken the periods of H to be quantized in $2\pi\mathbb{Z}$.

So far we have only identified a single $\mathbb{Z}_n$ q-form symmetry, but it turns out that there is another hidden higher $\mathbb{Z}_n$ symmetry. It's easiest to see in the second formulation, so we'll come back to it after we've discussed the second formulation.

Second formulation: To get the second presentation of the action, we can "dualize" a from a $q-1$ form gauge field to a $D-(q-1)-2 = D-q-1$ form gauge field B.

Verbose way: This way is longer, but is what usually goes down during dualization. We'll go through it here just to make sure that the second formulation is actually obtained by dualization.

We do the dualization by adding in a new q form gauge field G and a $D-q-1$ form gauge field B, so that we get the hard-to-look-at expression

$$Z = \int \mathcal{D}A\mathcal{D}H\mathcal{D}G\mathcal{D}B\mathcal{D}a \exp\left(\frac{i}{2\pi}\int_X (H \wedge (F_a - G + nA) + F_B \wedge G)\right), \tag{72}$$

where in the path integral we are summing over all possible bundles for all the fields[12] except for G, which is a globally well-defined form (i.e. dG really is trivial in cohomology). Adding all these fields hasn't actually done anything, which we can see by integrating out B: the globally well-defined part of B gives a δ function setting $dG = 0$, while the sum over cohomology classes of $F_B \in 2\pi H^{D-q}(X;\mathbb{Z})$ sets $\int_{M_q} G \in 2\pi\mathbb{Z}$ for all M_q. This means that G is the exterior derivative of some $q-1$ form gauge field, and so we can send $F_a - G \mapsto F_a$ by a field redefinition on a, recovering the original action.

Now we Hodge decompose G as

$$G = d\alpha_G + d^\dagger\beta_G + \omega_G. \tag{73}$$

We gauge-fix G by setting the exact component of the Hodge decomposition to be the exact part of F_a. Since we are summing over all cohomology classes for G in the path integral, the cohomologically nontrivial part of F_a can be absorbed by shifting ω_G. [13] This eliminates a from the theory entirely.

We then do the path integral over H, which sets $G = nA$. So finally we get the BF action in its usual form, namely

$$S = \frac{in}{2\pi}\int_X B \wedge F_A. \tag{74}$$

Again, in this presentation, both F_A and F_B have periods in $2\pi\mathbb{Z}$. Technically, to write it like this we have integrated by parts, trading the integral over $\frac{i}{2\pi}F_B \wedge A$ for one over $\frac{i}{2\pi}B \wedge F_A$. This can be done since although the two integrals are not equal, they differ by an element of $2\pi\mathbb{Z}$ (this is best thought about with DB cocycles — more on this later).

Fast way: Starting from the first formulation (61), just integrate out a directly! As we have explained, the integral over the globally well-defined part of a sets $dH = 0$, while the sum over cohomology classes for F_a enforces the quantization of the periods of H. Thus after

[12] e.g. all $U(1)$-bundles for A if A is a 1-form, all $U(1)$ gerbes for A if A is a 2-form, etc.

[13] Such a shift is fine since although the shift changes the $F_B \wedge G$ term, it changes it by something in $2\pi\mathbb{Z}$ due to the quantization condition on F_B.

integrating out a we can write H as the field strength of a $U(1)$ gauge field, $H = F_B$, which after integrating by parts gives (74).

Symmetries: Let's now check the symmetries in this formulation. The gauge symmetries are just shifts in B and A by exact forms, and we have no gauge symmetries that act on both fields as we had in the first formulation.

As for the global symmetries, we still have the $\mathbb{Z}_n$ q-form symmetry coming from shifting A, as we must. Note that naively looking at the action, we might conclude that we in fact have a $U(1)$ q-form symmetry, since the action only contains F_A which doesn't change under shifting A by a flat q-form, no matter what the holonomy of the flat q-form is. This isn't true though, and the actual symmetry is discrete: one of the ways to see this is to integrate by parts and write the integrand as $F_B \wedge A$, in which the $\mathbb{Z}_n$ character of the shift symmetry on A is manifest. But really, one should formulate the integral using DB cohomology (more on this to come). The advantage of this presentation is that the other global symmetry manifests itself as a $(D-q-1)$-form $\mathbb{Z}_n$ symmetry coming from shifting B as

$$B \mapsto B + \frac{1}{n}\epsilon_{D-q-1}, \qquad \epsilon_{D-q-1} \in 2\pi H^{D-q-1}(X; 2\mathbb{Z}). \tag{75}$$

This is a symmetry because of the quantization on F_A.

Third formulation: The final presentation of the action is a "magnetic" one, in which we treat F_B as an independent field (i.e. not necessarily flat), and enforce the fact that is actually the curvature of B with a Lagrange multiplier $\widetilde{B}$. So we write the partition function as

$$Z = \int \mathcal{D}B\mathcal{D}F\mathcal{D}\widetilde{A} \exp\left(\frac{i}{2\pi}\int_X F \wedge (F_{\widetilde{A}} + nB)\right). \tag{76}$$

Hopefully the notation here is clear: $F_{\widetilde{A}}$ is the field strength of $\widetilde{A}$, which is locally $d\widetilde{A}$, while F is its own independent field (not necessarily the field strength of any q-connection). As we have seen several times already, the integral over the globally-defined part of $\widetilde{A}$ sets $dF = 0$, while the summation over the cohomology classes for $F_{\widetilde{A}}$ enforce Dirac quantization on F. This means that integrating out $\widetilde{A}$ sets F to be the curvature of a q-form gauge field, and so indeed this presentation is equivalent to the usual BF action (note that this is similar in form to, but not exactly the same as, the first presentation).

Symmetries: Let's now check the symmetries in this formulation. The gauge symmetries, like in the first formulation, act on two fields simultaneously. First, we can shift $\widetilde{A}$ by an exact form. Second, we can shift B by $d\lambda_{D-q-2}$ while also shifting $\widetilde{A}$ by $-n\lambda_{D-q-2}$.

Similarly to our analysis of the first formulation, we see that B has a $(D-q-1)$-form $\mathbb{Z}_n$ global symmetry. One might think that we have a higher symmetry corresponding to constant shifts in $\widetilde{A}$, but since $\exp(i\int \widetilde{A})$ is not gauge-invariant, this is not so. We also have the $\mathbb{Z}_n$ q-form symmetry identified earlier, but it is "hidden" in this presentation.

5 *Careful explanation of CS level quantization* ✓

Today we will explain why the Chern-Simons level is quantized in $U(1)$ gauge theory. Our argument will hold on any manifold (e.g. even $\mathbb{R}^3$), and in particular will work even when $\int_{M_2} F = 0$ for all 2-submanifolds M_2 of spacetime. Thus the usual story about manifolds like $S^2 \times S^1$ and large gauge transformations isn't the whole story. I learned about this approach to CS terms from Alvarez's nice paper [**?**]; here we will follow these ideas and spell everything out in detail.

The usual explanations which everybody always repeats for why the CS coefficient is quantized are either "place the theory on $S^2 \times S^1$ with a unit of flux through the S^2 and look at large gauge transformations" or "realize it by integrating over a bulk 4-manifold and require independence of the extension". These are both somewhat unsatisfying to me since I want to know why even on an open manifold (like e.g. all the ones we are interested in the real wold; at least time will be $\mathbb{R}$), the CS term is quantized. That is, what is the reason that CS theory on say $\mathbb{R}^3$ only makes sense when the level is quantized? In what follows we will answer this question for the case when the gauge group is $U(1)$ for simplicity; the extension to non-Abelian gauge groups actually looks to be nontrivial, at least as far as notation is concerned, and may be revisted again in a future diary entry.

We can answer this by looking more carefully at what $\int_X A \wedge dA$ really means. It is often stated that the integrand is only well-defined up to a total derivative, but in fact the ambiguity in the integrand is much more serious than that.

The correct way to think about things is by using DB cohomology (differential characters), which is essentially a way of defining gauge fields within the framework of the Cech-de-Rham bicomplex. Recall that a $U(1)$ gauge field actually consists of three pieces of data: for a given decomposition of X into patches U_α, this data includes A_α (1-forms on each patch), $\Lambda_{\alpha\beta}$ ($\mathbb{R}$-valued 0-forms on each double overlap), and $n_{\alpha\beta\gamma}$ ($\mathbb{Z}$-valued 0-forms on each triple overlap). They relate with one another by

$$\delta_0 A_{\alpha\beta} = A_\alpha - A_\beta = d_0 \Lambda_{\alpha\beta}, \qquad \delta_1 \Lambda_{\alpha\beta} = 2\pi d_{-1} n_{\alpha\beta\gamma}, \tag{77}$$

where d_{-1} is just a suggestive way of writing the inclusion $\mathbb{Z} \hookrightarrow \mathbb{R}$ and the δ's are the Cech differentials. The transition functions in the bundle are $g_{\alpha\beta} = \exp(i\Lambda_{\alpha\beta})$, so that sending $\Lambda_{\alpha\beta} \mapsto \Lambda_{\alpha\beta} + 2\pi m_{\alpha\beta}$ for $m_{\alpha\beta}$ valued in $\mathbb{Z}$ does nothing (one can also check that this changes $n_{\alpha\beta\gamma}$ by a coboundary, and so doesn't affect the cohomology class of $n \in H^2_C(X;\mathbb{Z})$ (the C is for Cech cohomology). We can sum this up by writing A as the triple

$$A = (A_\alpha, \Lambda_{\alpha\beta}, n_{\alpha\beta\gamma}), \tag{78}$$

with

$$\delta_{-1} A = A_\alpha, \quad \delta_0 A = d\Lambda_{\alpha\beta}, \quad \delta_1 \Lambda_{\alpha\beta} = 2\pi d_{-1} n_{\alpha\beta\gamma}, \tag{79}$$

where $(\delta_{-1}A)_\alpha = A|_{U_\alpha} = A_\alpha$ is the restriction.

Morally speaking, this is kind of like doing a Hodge decomposition. The A_α part (a $(1,0)$ form, i.e. de Rham degree 1 and Cech degree 0) keeps track of the local curvature (the field strength), while the $\Lambda_{\alpha\beta}$ part (a $(0,1)$ form) keeps track of the holonomy of the gauge field around non-contractible loops. We can see this because the holonomy is captured by a flat 1-form, i.e. an element $\lambda \in H^1(X;\mathbb{R})$. Since λ is globally well-defined, we can write it simply as

$$\lambda = (\delta_{-1}\lambda, 0, 0). \tag{80}$$

Alternately, since we are only interested in its holonomy, we can just as well write it as

$$\lambda = (0, f_{\alpha\beta}, 0), \tag{81}$$

where the $f_{\alpha\beta}$ are real-valued functions. The holonomy of λ around a given loop can be computed by summing up the $f_{\alpha\beta}$'s along 2-fold intersections of patches along the loop, as we will see later. Note that there is no $n_{\alpha\beta\gamma}$ term here because $\delta_1(f_{\alpha\beta}) = \delta_1(\delta_0\lambda)_{\alpha\beta} = 0$.

The $n_{\alpha\beta\gamma}$ part (a $(-1,2)$ form) in the decomposition of A keeps track of the topology of the bundle (i.e. the Chern class). This is because, as we will see, the integral of F_A (the curvature of A) over a closed 2-manifold is given by a sum of the $n_{\alpha\beta\gamma}$'s.

Why is all this data needed in order to be able to do integrals? The philosophy is basically "we want the integrals we write down to be independent of the way in which we choose to decompose X into coordinate patches". With that in mind, consider integrating the gauge field along a 1-cycle that starts at a point $a \in U_\alpha \setminus (U_\alpha \cap U_\beta)$ and ends at $b \in U_\beta \setminus (U_\alpha \cap U_\beta)$, with $U_\alpha \cap U_\beta$ non-empty. To define the integral, we need to integrate part of the way with A_α, and then the rest of the way with A_β. Suppose the transition point between these two is at $p \in U_\alpha \cap U_\beta$. Then tentatively our integral is

$$I(a,b;p) = \int_a^p A_\alpha + \int_p^b A_\beta. \tag{82}$$

The problem is that I is not independent of p! Indeed, one can check that, for $q \in U_\alpha \cap U_\beta$, we have

$$I(a,b;p) - I(a,b;q) = \Lambda_{\alpha\beta}(p) - \Lambda_{\alpha\beta}(q). \tag{83}$$

The fix is to just add this transition function term into the integral. Thus, the following integral is independent of p:

$$I(a,b) = \int_a^p A_\alpha - \Lambda_{\alpha\beta}(p) + \int_p^b A_\beta. \tag{84}$$

However, recall that we need the shift $\Lambda_{\alpha\beta} \mapsto \Lambda_{\alpha\beta} + 2\pi m_{\alpha\beta}$ to not do anything. But, this shift changes the value of $I(a,b)$ by something in $2\pi\mathbb{Z}$! The only way to fix this is to ensure that the only time we write $I(a,b)$ is in exponentials as $\exp(iqI(a,b))$, where $q \in \mathbb{Z}$ (really we should be taking the integration to be over a closed cycle, but of course the same $2\pi\mathbb{Z}$ ambiguity still occurs). This is just another way of saying that the Wilson loop operators must be taken with integer charge. We know that if the Wilson loop wraps a nontrivial cycle then $q \in \mathbb{Z}$ is required by invariance under large gauge transformations, but here we are saying that *even for topologically trivial cycles*, the charge in the Wilson loop must be taken to be in $\mathbb{Z}$, a fact dictated only by the topology of the gauge group and not by the topology of any particular Wilson loop.

On a related note, this formulation lets us see why flat connections can be either specified as collections $(\lambda_\alpha, 0, 0)$, or entirely in terms of transition functions $(0, f_{\alpha\beta}, 0)$. Indeed, for the first formulation, we write the integral $\int_C \lambda$ for some cycle C as

$$\int_C \lambda = \sum_\alpha \int_{C_\alpha} \lambda_\alpha, \tag{85}$$

since the transition functions vanish. Hopefully the notation is clear: the C_α are the segments of C that lie in the patch U_α. On the other hand, since λ is closed and each U_α is contractible, we can write $\lambda_\alpha = d\omega_\alpha$ for some 0-forms ω_α, and so

$$\int_C \lambda = \sum_{p \in U_\alpha \cap U_\beta} (\omega_\beta(p) - \omega_\alpha(p))). \tag{86}$$

Thus if we define the transition functions $f_{\alpha\beta} = \omega_\alpha - \omega_\beta$, we see that if we only care about the holonomy of λ, we can just as well replace it will the collection $(0, f_{\alpha\beta}, 0)$ — the transition functions entirely determine the holonomy of closed forms.

The next step up in complexity comes from integrating F_A over a surface S. We need to do so in a way that doesn't depend on what sort of way we choose to cover the spacetime manifold with patches. Consider at first the case where S is closed. F_A is closed, and so on each patch it is exact (we can and will always take our patches, as well as their n-fold intersections, to be topologically trivial). So then we have

$$\int_S F_A = \sum_\alpha \int_{C_\alpha} dA_\alpha = \sum_\alpha \int_{\partial C_\alpha} A_\alpha. \tag{87}$$

Here, the $C_\alpha \subset U_\alpha$ are non-overlapping 2-chains contained in each of the patches, such that $\cup_\alpha C_\alpha = S$. Note that there are many ways of choosing the C_α, but different choices do not affect the integral, since they differ by integrals of the form $\int d(A_\alpha - A_\beta) = \int d^2 \Lambda_{\alpha\beta} = 0$. Additionally, we can always choose the C_α so that at most C_α meet at any given point (we can always choose the boundaries of the C_α to be a triangulation of S).

Returning to the integral of F_A over S, and assuming that S is orientable, we see that

$$\int_S F_A = \sum_{\alpha\beta} \int_{\partial C_\alpha \cap \partial C_\beta} (A_\alpha - A_\beta) = \sum_{\alpha\beta} \int_{\partial C_\alpha \cap \partial C_\beta} d\Lambda_{\alpha\beta}. \tag{88}$$

Each of the integrals in the above sum is over a line segment, and so each integral contributes a term like $\Lambda_{\alpha\beta}(b) - \Lambda_{\alpha\beta}(a)$. When we sum over all such line segments, we get three $\Lambda_{\alpha\beta}$ terms at each vertex (where three C_α's meet), and they appear in the form $\delta_1 \Lambda_{\alpha\beta\gamma} = 2\pi n_{\alpha\beta\gamma}$. Thus we have

$$\int_S F_A = \sum_{\alpha\beta\gamma} 2\pi n_{\alpha\beta\gamma}. \tag{89}$$

This is why we said that the $n_{\alpha\beta\gamma}$ determine the topology of the bundle (if $\partial S \neq \emptyset$ then the only thing that changes is that we get an additional integral of A over the boundary of S).

Note that in order to get a non-zero Chern class, the transition functions $\Lambda_{\alpha\beta}$ could not all be constant. Thus in order to create bundles which are twisted, it is not enough to just twist the transition functions by constants: we have to have "twisting" inside of double overlaps as well[14].

Let us finally now turn to Chern-Simons theory. Our naive guess of what the relevant integral is would be

$$\sum_\alpha \int_{C_\alpha} A_\alpha \wedge dA_\alpha, \tag{90}$$

where each C_α is now a 3-cycle, the C_α are all non-overlapping, and their union is the full spacetime X. This is not invariant under moving around the boundaries of the C_α though, which is a problem. When we wiggle one of the C_α boundaries, the difference in the integral as written above is an integral like

$$\int_{\delta C_{\alpha\beta}} (A_\alpha \wedge dA_\alpha - A_\beta \wedge dA_\beta) = \int_{\partial\delta C_{\alpha\beta}} \Lambda_{\alpha\beta} \wedge dA_\beta, \tag{91}$$

where $\delta C_{\alpha\beta}$ is the volume enclosed by the two different choices of the boundary between C_α and C_β (I may add pictures at some point to make this clearer). Thus, we can take care of this ambiguity by adding in this term to the definition of the Chern-Simons integral, like like how we added $\Lambda_{\alpha\beta}(p)$ into the definition of the Wilson line integral. So, our improved integral for the CS action now looks like

$$\sum_\alpha \int_{C_\alpha} A_\alpha \wedge dA_\alpha - \sum_{\alpha\beta} \int_{C_{\alpha\beta}} \Lambda_{\alpha\beta} \wedge dA_\beta, \tag{92}$$

where the $C_{\alpha\beta}$ are the 2-cells where the C_α 3-cells meet. Notice that to find the correction term to the naive $A_\alpha \wedge dA_\alpha$ term, we computed $\delta(A_\alpha \wedge dA_\alpha) = d\Lambda_{\alpha\beta} \wedge dA_\beta$, which we found to be a total derivative (we also used that $dA_\alpha = dA_\beta$). Thus when we took the Cech differential, we got something that was exact in de Rham cohomology. This is in keeping with the general Cech-de-Rham bicomplex structure of this whole construction.

Sadly, even the improved integral is not invariant under re-arranging the patches. Now we have to consider what happens when we wiggle one of the 1-cells $C_{\alpha\beta\gamma}$, which is a common boundary of three of the C_α's. Drawing some pictures, one can convince oneself that for two choices of $C_{\alpha\beta\gamma}$ that differ by the surface $\delta C_{\alpha\beta\gamma}$, the term that we added to the naive CS integral changes by the term

$$-\int_{\delta C_{\alpha\beta\gamma}} (\Lambda_{\alpha\beta} + \Lambda_{\beta\gamma} + \Lambda_{\gamma\alpha}) \wedge dA_\gamma = -2\pi \int_{\partial\delta C_{\alpha\beta\gamma}} n_{\alpha\beta\gamma} \wedge A_\gamma, \tag{93}$$

where we used that dA is the same on all three of the patches. Again, we see that the change in the integral is computed by taking a Cech differential, and that taking the differential

[14] Another way to say this is that if the transition functions are constants, we can choose a gauge in which the connection is flat: $A_\alpha = 0$ is a connection which obeys $A_\alpha = g_{\alpha\beta}^{-1}(A_\beta - d)g_{\alpha\beta}$. Flat connections can't have non-zero Chern class, and so we conclude that the transition functions need to not be constant if we are to get $\int F_A \neq 0$.

gives us something exact in de Rham cohomology. Thus to cancel out *this* variation, we modify the CS action to

$$\sum_\alpha \int_{C_\alpha} A_\alpha \wedge dA_\alpha - \sum_{\alpha\beta} \int_{C_{\alpha\beta}} \Lambda_{\alpha\beta} \wedge dA_\beta + 2\pi \sum_{\alpha\beta\gamma} \int_{C_{\alpha\beta\gamma}} n_{\alpha\beta\gamma} \wedge A_\gamma. \tag{94}$$

Even this isn't good enough, since we haven't looked at what happens when we wiggle around the 0-cells where four 3-cells meet. At this point, the pattern about how to fix the ambiguity should be clear: we take the Cech differential of the last term we added to the integral, find that we get an exact form, and then add that term back to the integral, but with opposite sign. Doing so gives the final form of the Lagrangian, and so the correct CS action is

$$\begin{aligned} S &= \frac{k}{4\pi} I_{CS}; \\ I_{CS} &= \sum_\alpha \int_{C_\alpha} A_\alpha \wedge dA_\alpha - \sum_{\alpha\beta} \int_{C_{\alpha\beta}} \Lambda_{\alpha\beta} \wedge dA_\beta + 2\pi \sum_{\alpha\beta\gamma} \int n_{\alpha\beta\gamma} \wedge A_\gamma - 2\pi \sum_{\alpha\beta\gamma\sigma} \int n_{\alpha\beta\gamma} \wedge \Lambda_{\gamma\sigma}. \end{aligned} \tag{95}$$

The integrals are evaluated on 3, 2, 1, and 0 cells, in turn.

Now we can note several things about this expression that aren't clear if we were to think of it as $A \wedge dA$. First, notice that I_{CS} is ambiguous up to elements of $(2\pi)^2\mathbb{Z}$, even if the spacetime manifold is completely trivial and F_A is globally exact. This is because we have an equivalence $\Lambda_{\alpha\beta} \sim \Lambda_{\alpha\beta} + 2\pi m_{\alpha\beta}$ where $m_{\alpha\beta}$ takes values in $\mathbb{Z}$, which e.g. shifts the $2\pi \int n \wedge \Lambda$ term by something in $(2\pi)^2\mathbb{Z}$ (the $\int n \wedge \Lambda$ term is really a sum over discrete points in the manifold). If we want S to be invariant modulo $2\pi\mathbb{Z}$, this forces the quantizaton of k even in topologically trivial scenarios. In fact, it forces $k \in 2\mathbb{Z}$ to be an *even* integer. We know that even levels describe bosonic systems, and this construction is only able to directly handle this case. For the fermionic case where we have a genuine spin TQFT and k is odd, we need a little bit more data: the spin structure needs to be introduced explicitly into the procedure described above, with minus signs coming from the spin structure cancelling out the extra minus signs that come from the above presentation of the action not being completely invariant under re-arrangements of the patches when k is odd. This is essentially the framing anomaly: the theory looks topological, but it actually retains a hidden dependence on the spin structure.

Also, note that changing A by a flat connection (a flat connection can be captured purely by transition functions, i.e. it can be written in the form $(0, f_{\alpha\beta}, 0)$) does *not* leave I_{CS} invariant (even if the spacetime is closed), contrary to what we would expect from writing the action as $A \wedge dA$. Indeed, we see that shifting A by a flat connection only leaves the action invariant if that flat connection is in $\Omega^1_{2\pi\mathbb{Z}/k}(X)$, i.e. only if the holonomy of the action is quantized in units of $2\pi/k$ for all 1-cycles of the spacetime. This is why CS theory has a $\mathbb{Z}_k$ 1-form symmetry, and not a $U(1)$ 1-form symmetry. One again we stress that this argument works on topologically trivial spacetimes with globally exact field strengths.

The same argument shows that e.g. in the three-dimensional BF theory with action

$$\frac{in}{2\pi} \int B \wedge F_A, \tag{96}$$

the action is *not* invariant under shifting A by a flat form λ, because of the presence of the $\int n_{\alpha\beta\gamma} \wedge \Lambda_{\gamma\sigma}$ correction term. If λ has holonomy $\exp(i\int_C \lambda) = \exp(2\pi i k/n)$ for all 1-cycles C and for $k \in \mathbb{Z}$ then the action changes only by an element in $2\pi i\mathbb{Z}$, and so we have a global $\mathbb{Z}_n$ 1-form symmetry, but not a global $U(1)$ 1-form symmetry.

6 *Quantization for Chern-Simons* ✓

The goal of today's problem is to try to understand the connection between CS theory and the WZW action through the quantization of the former (motivated by reading the paper by Jackiw and others back in 1989 on the quantization of Chern-Simons theory). We will also try to work out the details of some of the results in [**?**].

We will be working with CS theory / WZW actions for $SU(2)_k$ for concreteness:

$$S_{CS} = \frac{k}{4\pi}\int \mathrm{Tr}(A\wedge dA + \frac{2}{3}A\wedge A\wedge A). \tag{97}$$

We will quantize the theory using holomorphic quantization, by first quantizing the fields and then imposing the constraint of gauge-invariance. We will then check that the generators of gauge transformations form a linear representation of the gauge group, and find the wavefunctional for the quantized fields. Our answer will involve the WZW action.

When dealing with a system with constraints (for us, Gauss' law), we can either solve the constraint and then quantize, or the other way around.[15] In what follows we will adopt the latter strategy: this means we will quantize the fields in the normal way, and solve the constraint afterwards by requiring that the generator of the constraint act trivially on the wavefunction. As usual with these types of quantization problems, it will be helpful to do everything in complex coordinates.

From the $A\wedge\partial_t A$ term in the CS action, we can write $A = t^a A^a$ and take the trace and directly read off the commutator[16]

$$[A_i^a(z), A_j^b(w)] = \epsilon_{ij}\frac{4\pi i}{k}\delta_{ab}\delta(z-w). \tag{98}$$

[15] However it is actually not always clear that these two procedures give the same results; I remember hearing from Greg Moore that in fact there are some examples where they indeed give different results—they may be rather pathological examples, though. Should come back to this.

[16] In the commutator, the δ function is normalized to have unit integral with the measure $dx\wedge dy$, not with $dz\wedge d\bar{z}$ (sorry)

To get the right coefficient, we need to remember that $t^a = \sigma^a/2$, so that $\mathrm{Tr}(t^a t^b) = \delta^{ab}/2$, turning the part of the action that survives in the $A_0 = 0$ gauge into $(k/8\pi)\int \epsilon^{ij}(\partial_t A_i^a)A_j^a$.

Slightly more carefully, we can find the symplectic form by varying the action. We get, for a spacetime X,

$$\delta S = \frac{k}{2\pi}\int \mathrm{Tr}[\delta A \wedge (dA + A\wedge A)] - \frac{k}{4\pi}\int_{\partial X} \mathrm{Tr}[A \wedge \delta A] \tag{99}$$

where the sign of the last term comes from the signs in the product rule for d (we take the two differentials d and δ to commute). The boundary term gives the symplectic current, and so varying it again gives

$$\Omega = -\frac{k}{4\pi}\int_{\partial X} \mathrm{Tr}[\delta A \wedge \delta A]. \tag{100}$$

The fact that the symplectic form is a wedge product of δA with itself means e.g. that Wilson loops in A don't commute with themselves and is responsible for all the stuff we know and love about CS theory. Also note that $\delta A \wedge \delta A \neq 0$ even in the $U(1)$ case—this may look at first glance like the wedge product of two one-forms, but it is actually two $(1,1)$ forms (one degree in de Rham and one variational degree), and is non-vanishing (the degree relevant for the graded commutation rules is the total degree in the $d\delta$ bicomplex).

As a sanity check, this gives a commutator of $[A_i(z), A_j(w)] = \varepsilon_{ij}\frac{2\pi i}{k}\delta(z-w)$ in the $U(1)_k$ case (note that it's 2π and not 4π because of the absence of the $1/2$ factor from the trace). This means that linked wilson lines with charges p, q can be unlinked at the cost of a phase factor $e^{2\pi i pq/k}$. This gives us the correct result that $q = k$ lines have trivial braiding, and that for $k \in 2\mathbb{Z}+1$ the $q = k$ line is a fermion.

Anyway, we see from the symplectic form that we can choose e.g. either A_x or A_y as the canonical momentum. However as usual when doing geometric quantization it will be better to work with a holomorphic polarization, choosing A_z as the coordinate and $\bar{A} \equiv A_{\bar{z}}$ as the momentum. So then we have

$$[A^a(z), \bar{A}^b(w)] = \frac{4\pi i}{k}\delta_{ab}\delta(z-w). \tag{101}$$

As in Maxwell theory, A_0 has no momentum and imposes a Gauss' law constraint. Varying the action with respect to A_0^a, the first $A \wedge dA$ part gives us two copies of $\partial_{[i}A_{j]}^a/2$ (the $1/2$ from the trace), while the second A^3 part gives us three copies of $\epsilon^{ij}A_i^b A_j^c i f^{abc}/4$, where various factors of 2 from the matrix structure and trace have cancelled. Thus after a bit more algebra, we see that the constraint from A_0 means that for any physical state $|\Psi\rangle$,

$$(\partial_z \bar{A}^a - \partial_{\bar{z}} A^a + i f^{abc} A^b \bar{A}^c)|\Psi\rangle = 0. \tag{102}$$

Thus the field strength will be the generator of gauge transformations on a given Cauchy slice (for us, the $z, \bar{z}$ plane). This is in contrast with Maxwell theory, where the generator of gauge transformations is the electric field $\star F$, rather than the magnetic field F. The fact that the momentum is magetic rather than electric is responsible for flux attachment and at the technical level just comes from the fact that the action has a $\wedge F$ rather than a $\wedge \star F$.

To check that the field strength generates gauge transformations when it acts on the gauge field, define the operator

$$U(\lambda) = \exp\left(-i\frac{k}{4\pi}\int \lambda^a \wedge F^a\right). \tag{103}$$

Using the commutation relations we have

$$\bar{A}^a = -\frac{4\pi i}{k}\frac{\delta}{\delta A^a}, \tag{104}$$

and so we can write $U(\lambda^a)$ as (after integrating by parts)

$$U(\lambda) = \exp\left(\int\left[\partial_z\lambda^a\frac{\delta}{\delta A^a} - i\frac{k}{4\pi}\partial_{\bar{z}}\lambda^a A^a - if^{abc}\lambda^a A^b\frac{\delta}{\delta A^c}\right]\right). \tag{105}$$

Then expanding the exponentials and a little bit of algebra shows that

$$U(\lambda)A^a[U(\lambda)]^\dagger = A^a + \partial_z\lambda^a + if^{abc}A^b\lambda^c, \tag{106}$$

which is exactly what we want.

When deriving the action of $U(\lambda)$ on A^a, the part proportional to $\partial_{\bar{z}}\lambda^a$ canceled out. It is still essential to keep though, since it generates gauge transformations for the antiholomorphic part. Indeed, a little bit of algebra gives

$$U(\lambda)\frac{\delta}{\delta A^a}[U(\lambda)]^\dagger = \frac{\delta}{\delta A^a} + i\frac{k}{4\pi}\partial_{\bar{z}}\lambda^a + if^{abc}\lambda^c\frac{\delta}{\delta A^b}, \tag{107}$$

so that

$$U(\lambda)\bar{A}^a[U(\lambda)]^\dagger = \bar{A}^a + \partial_{\bar{z}}\lambda^a + if^{abc}\bar{A}^b\lambda^c, \tag{108}$$

as required.

Now let's find the wavefunctionals, which will be holomorphic functionals of A. In what follows, it will be convenient to be able to work with group elements $g = e^\lambda, h = e^\gamma$, as well as Lie algebra elements. We find the wave functionals by requiring that they are invariant under the action of $U(g)$. This means we have to know how the exponentiation of $\int \mathrm{Tr}[\lambda\wedge F]$ acts on arbitrary functionals of A. This is slightly tricky, since the exponentiation of F is difficult—it contains operators that don't commute among themselves.

Let us break up $U(g)$ into two parts: the part which implements the gauge transformation on the holomorphic part (the part with the $\delta/\delta A$'s), and the part containing $\mathrm{Tr}[\partial_{\bar{z}}\lambda A]$ needed for the gauge transformations of the antiholomorphic part. Hence we will write

$$-\frac{ik}{2\pi}\int \mathrm{Tr}[\lambda\wedge F] = \mathcal{G}(g) + \frac{ik}{2\pi}\int \mathrm{Tr}[g^\dagger\partial_{\bar{z}}gA], \tag{109}$$

where

$$\mathcal{G}(g) \equiv \int\left[\partial_z\lambda^a\frac{\delta}{\delta A^a} - if^{abc}\lambda^a A^b\frac{\delta}{\delta A^c}\right] \tag{110}$$

is the logarithm of the part of $U(g)$ which does the gauge transformation.

Let $|\Psi[A]\rangle$ be a candidate gauge-invariant wavefunctional, and write the action of $U(g)$ on it craftily as

$$U(g)|\Psi[A]\rangle = U(g)e^{-\mathcal{G}(g)}|\Psi[{}^gA]\rangle. \tag{111}$$

If $U(g)$ only performed gauge transformations, then we would have $U(g)e^{-\mathcal{G}(g)} = \mathbf{1}$. Because of the extra part in $U(g)$ though, this operator is nontrivial.

Now we need to find out what $U(g)e^{-\mathcal{G}(g)}$ is. Since manipulating stuff in the exponentials is difficult, let us bring down the stuff in the exponentials using the "fake one-parameter evolution" trick that comes up often when e.g. talking about anomalies. Namely, introduce a homotopy parameter $\phi \in [0,1]$ and consider $U(g^\phi)e^{-\mathcal{G}(g^\phi)}$. At $\phi = 0$ the g field is just the identity on all of spacetime, while it becomes equal to the value of interest at $\phi = 1$. Geometrically, what we are doing is extending the spatial manifold to be realized as the boundary of a three-manifold, where the added direction is the ϕ direction. Since $g^\phi = \mathbf{1}$ at $\phi = 0$ we can compactify space to a point at $\phi = 0$, and so this three-manifold looks like a cone. Anyway, we use $g^\phi = e^{\phi\lambda}$ to compute the derivative

$$\partial_\phi(U(g^\phi)e^{-\mathcal{G}(g^\phi)}) = U(g^\phi)\left(\frac{ik}{2\pi}\int \mathrm{Tr}[g^\dagger\partial_{\bar{z}}gA] + \mathcal{G}(g)\right)e^{-\mathcal{G}(g^\phi)} - U(g^\phi)e^{-\mathcal{G}(g^\phi)}\mathcal{G}(g), \tag{112}$$

since the exponents are just linear in ϕ. When we bring the integral through the $e^{-\mathcal{H}(g^\phi)}$ it gets gauge-transformed, and so since the $\mathcal{G}(g)$'s cancel,

$$\partial_\phi(U(g^\phi)e^{-\mathcal{G}(g^\phi)}) = U(g^\phi)e^{-\mathcal{G}(g^\phi)}\frac{ik}{2\pi}\int \mathrm{Tr}[g^\dagger\partial_{\bar{z}}g(g_\phi^\dagger A g_\phi + g_\phi^\dagger\partial_z g_\phi)], \tag{113}$$

where g_ϕ is the same as g^ϕ but is used since $(g^\phi)^\dagger$ looks uglier.

Now to simplify this mess. The first term on the RHS is a is a total derivative since

$$\frac{d}{d\phi}\mathrm{Tr}[(\partial_{\bar{z}}g_\phi)g_\phi^\dagger A] = \mathrm{Tr}[\partial_{\bar{z}}(g_\phi\lambda)g_\phi^\dagger A - (\partial_{\bar{z}}g_\phi)\lambda g_\phi^{-1}] = \mathrm{Tr}[(\partial_{\bar{z}}\lambda)g_\phi^\dagger A g_\phi], \tag{114}$$

as $\partial_\phi g_\phi = g_\phi\lambda$. We break the second term up as

$$\mathrm{Tr}[g^\dagger\partial_{\bar{z}}g(g_\phi^\dagger\partial_z g_\phi)]dz\wedge d\bar{z} = \frac{1}{2}\mathrm{Tr}[g_\phi^\dagger\partial_z g_\phi\partial_{\bar{z}}\lambda + g_\phi^\dagger\partial_{\bar{z}}g_\phi\partial_z\lambda]dz\wedge d\bar{z} - \frac{1}{2}\mathrm{Tr}[g_\phi^\dagger dg_\phi\wedge g^\dagger dg], \tag{115}$$

where the wedge product (taken only on the spatial slice; not involving the time coordinate) treats ∂_z as coming first and $\partial_{\bar{z}}$ as coming second. The second term on the RHS is

$$\frac{1}{2}\mathrm{Tr}[g_\phi^\dagger\partial_z g_\phi\partial_{\bar{z}}\lambda + g_\phi^\dagger\partial_{\bar{z}}g_\phi\partial_z\lambda] = \frac{d}{d\phi}\mathrm{Tr}[g_\phi^\dagger\partial_z g_\phi g_\phi^\dagger\partial_{\bar{z}}g_\phi] \tag{116}$$

since two of the terms after taking the derivative on the RHS cancel. Finally, in the last term with the wedge products, we can use $\lambda = g_\phi^\dagger\partial_\phi g_\phi$ to plug in for λ. Then we can antisymmetrize the three derivatives and divide by a factor of 3 to get

$$\frac{1}{2}\int \mathrm{Tr}[g_\phi^\dagger dg_\phi\wedge g^\dagger dg] = \frac{1}{2\cdot 3}\int d^2z\,\epsilon^{\alpha\beta\gamma}\mathrm{Tr}[g_\phi^\dagger\partial_\alpha g_\phi g_\phi^\dagger\partial_\beta g_\phi g_\phi^\dagger\partial_\gamma g_\phi], \tag{117}$$

where on the RHS α, β, γ run over $z, \bar{z}$, and ϕ. Look at how WZW-like this is! Since the integral on the RHS is only over space, the RHS is also a total ϕ derivative, and it equals

$$\frac{d}{d\phi}\frac{1}{6}\int_{B_{3,\phi}} \mathrm{Tr}[g^\dagger_{\phi'}dg_{\phi'} \wedge g^\dagger_{\phi'}dg_{\phi'} \wedge g^\dagger_{\phi'}dg_{\phi'}], \tag{118}$$

where $B_{3,\phi}$ is a bounding 3-ball extending from $\phi' = 0$ to $\phi' = \phi$.

Recapitulating, the whole integral on the RHS of (113) is a total derivative. This means that $U(g^\phi)e^{-\mathcal{G}(g^\phi)}$ is actually the exponential of the argument of the total derivative, which means that after setting $\phi = 1$, we have found $U(g)e^{-\mathcal{G}(g)}$. Keeping track of the various factors of $k/2\pi$, we get

$$U(g)e^{-\mathcal{G}(g)} = \frac{ik}{4\pi}\int \mathrm{Tr}[g^\dagger(\partial_{\bar{z}}g)A + g^\dagger\partial_z g g^\dagger\partial_{\bar{z}}g] + \frac{ik}{24\pi}\int_{B^3}\mathrm{Tr}[g^\dagger dg \wedge g^\dagger dg \wedge g^\dagger dg]. \tag{119}$$

Thus when acting on wavefunctionals, $U(g)$ both implements gauge transformations and multiplies the wavefunctionals by this exponential factor. We write it as

$$U(g)|\Psi[A]\rangle = e^{i\Omega[g,A]}|\Psi[{}^gA]\rangle \equiv \exp\left(\frac{ik}{4\pi}\int \mathrm{Tr}[A\bar{J}_g]\right)e^{iS[g]}|\Psi[{}^gA]\rangle, \tag{120}$$

where we have suggestively written the current as $\bar{J}_g = g^\dagger\partial_{\bar{z}}g$ and where $S[g]$ is the WZW action (both kinetic and topological terms).

Since $U(g)$ is a representation of the gauge group, we need it to satisfy

$$U(h)U(g) = U(gh). \tag{121}$$

Note the perverse ordering of the group elements on the RHS. Such a perversion is needed since gauge transformations conventially act adjointly as

$$A \mapsto g^{-1}Ag + g^{-1}\partial_z g, \tag{122}$$

so that for the product gh, it is g which acts first, and h which acts second.

Do the $U(g)$ form a linear representation? Consider the product $U(h)U(g)$ acting on $|\Psi[A]\rangle$. If this is equal to the action of $U(gh)$, then we need

$$\Omega[g, A] + \Omega[h, {}^g A] = \Omega[gh, A] \qquad \mathrm{mod}\ 2\pi\mathbb{Z}. \tag{123}$$

That is, we need Ω to have a coboundary which vanishes mod $2\pi\mathbb{Z}$. After some algebra, this condition means that we need

$$\frac{ik}{4\pi}\int \mathrm{Tr}[g^\dagger\partial_z g h^\dagger\partial_{\bar{z}}h] + S[g] + S[h] = S[gh]. \tag{124}$$

In a previous diary entry on the WZW term, we showed that this is indeed true. The coboundary of the WZW action is precisely equal to the term needed to ensure that the $U(g)$ form a linear representation of the gauge group.

Now that we know how the $U(g)$'s act on candidate wavefunctionals, we need to actually find a particular solution for $|\Psi[A]\rangle$. But this is easy: we want it to be invariant under the action of the gauge group, and so we can project onto the trivial representation of the gauge group by integrating over all gauge transformations, that is, by acting with $\int \mathcal{D}gU(g)$ on any candidate wavefunctional. For example, we may just take

$$\Psi[A] = \int \mathcal{D}g \exp\left(iS[g] + \frac{ik}{4\pi}\int \mathrm{Tr}[A\bar{J}_g]\right). \tag{125}$$

We see that the wavefunctional for the CS gauge fields is obtained just by plugging them in as sources for the WZW theory.

Finally we briefly touch on some stuff we have brushed over. What happens if the spatial manifold X has a boundary? Recall that

$$\delta S = \frac{k}{2\pi}\int \mathrm{Tr}[\delta A \wedge F_A] - \frac{k}{4\pi}\int_{\partial X} \mathrm{Tr}[A \wedge \delta A], \tag{126}$$

We will fix our boundary conditions by fixing $A|_{\partial X}$ to be some specific (not for sure zero) function. Then the boundary term in δS is an integral over $\mathrm{Tr}[A\delta\bar{A}]$. In order to receive no boundary corrections to the equations of motion, we have to add the counterterm

$$S_\partial = \frac{1}{4\pi}\int_{\partial X} \mathrm{Tr}[A\bar{A}]. \tag{127}$$

This counterterm then also shows up in the action of $U(g)$ that we use to find the wavefunctional, in order to ensure that gauge invariance is maintained.

Secondly, what if we are on a spatial manifold with nontrivial 1-cycles? In this case, since the constraint from A_0 merely fixes the gauge field to be flat, we can have fields with nontrivial holonomy. I think the best way to deal with the problem in this case is to adopt the "solve the constraint and then quantize" approach, whereby we first decompose $A = U^{-1}\alpha U + U^{-1}dU$, and then quantize. Here U is single-valued and α keeps track of the holonomy.

7 *Chern-Simons Propagator* ✓

Today is a quickie: a simple calculation I realized I'd never seen done before. We will find the propagator for Abelian CS + Maxwell theory

$$S = \frac{k}{4\pi}\int A \wedge F_A + \frac{1}{2e^2}\int F_A \wedge \star F_A \tag{128}$$

and show that the theory describes massive excitations. We will also explain how a long-ranged statistical interaction between particles is possible in a massive theory.

Let's find the answer for the mass in an easy way first so we can check our work. First we find the equations of motion, which we write as

$$\partial_\mu F^{\mu\nu} - \frac{\kappa}{2}\epsilon^{\nu\mu\lambda}F_{\mu\lambda} = 0, \qquad \kappa \equiv \frac{e^2 k}{2\pi}. \tag{129}$$

This is equivalently written as[17]

$$(d^\dagger - \kappa\star)F = 0. \tag{130}$$

Now we act on both sides with the operator $\star d + \kappa$, so that

$$(\star d + \kappa)(\star d - \kappa)\star F = (d^\dagger d - \kappa^2)\star F = 0. \tag{131}$$

Now since $F = dA$, $d^\dagger \star F = 0$. Thus

$$(d^\dagger d + dd^\dagger - \kappa^2)\star F = 0 \implies (\partial_\mu\partial^\mu - \kappa^2)(\star F)^\nu = 0, \tag{132}$$

indicating that $\star F$ is a massive vector field with mass

$$m = \kappa = \frac{e^2 k}{2\pi}. \tag{133}$$

When $e^2 \to \infty$ we have $m \to \infty$, which means that at strong coupling (in the deep IR), we have an infinite mass (we are "projecting onto the LLL") and we can focus only on the Chern-Simons term.

Now let's find the progagator and check this. We will choose to use Feynman gauge, by adding the gauge-fixing term

$$S_{gf} = \frac{1}{2e^2}\int d^\dagger A \wedge \star d^\dagger A \tag{134}$$

to the action. This turns the Maxwell term into $\int A \wedge \star(d^\dagger d + dd^\dagger)A$. Since the thing in the parenthesis is the Hodge Laplacian, we just get a k^2 term. So then in momentum space, we need to find

$$G_{\mu\nu} = \left(iC\epsilon^{\mu\lambda\nu}k_\lambda + Dg^{\mu\nu}k^2\right)^{-1}, \qquad C \equiv \frac{k}{4\pi}, \qquad D \equiv \frac{1}{2e^2}. \tag{135}$$

The strategy for inverting this guy is to try to break it up into projectors, as usual. We let

$$\Pi_L^{\mu\nu} = \frac{k^\mu k^\nu}{k^2}, \qquad \Pi_T = \frac{k^\mu k^\nu}{k^2} - g^{\mu\nu}. \tag{136}$$

[17] $d^\dagger = - \star d \star$ in Euclidean signature in three dimensions when acting on 1-forms while $d^\dagger = + \star d \star$ when acting on 2-forms. The hodge star satisfies $\star^2 = 1$ when acting on any degree form.

The extra piece $iC\epsilon^{\mu\lambda\nu}k_\lambda$ is in the image of Π_T (since it's orthogonal to Π_L), but it is not itself a projector. We look for a solution of the form

$$G^{\mu\nu} = X\Pi_T^{\mu\nu} + Y\Pi_L^{\mu\nu} + Z\epsilon^{\mu\lambda\nu}k_\lambda. \tag{137}$$

We find the constants X, Y, Z just by multiplying this ansatz by $iC\epsilon^{\mu\lambda\nu}k_\lambda + Dg^{\mu\nu}k^2$ and setting the result equal to $g^{\mu\nu}$. This gives three equations (the coefficients of the three types of terms in G) in three unknowns, which we solve for in terms of C, D. The algebra is kind of boring and gives

$$G^{\mu\nu} = \frac{1}{k^2(k^2 - C^2/D^2)}\left(g^{\mu\nu}\frac{k^2}{D} - \frac{1}{D}k^\mu k^\nu - \frac{iC}{D^2}\epsilon^{\mu\lambda\nu}k_\lambda\right). \tag{138}$$

This tells us that we have something with mass

$$m = \frac{C}{D} = \frac{ke^2}{2\pi}, \tag{139}$$

which agrees with our previous result. As a sanity check, we note that we get the pure Maxwell answer when $C = 0$. Also, note that D^{-1} has mass dimension 1, so that all the dimensions of the various terms agree.

We can now resolve the question posed at the beginning: if the theory is massive, how can it lead to nontrivial statistical interactions between particles which are infinitely long ranged? We see that the answer to this is that although we have a pole at $k^2 = C^2/D^2$ rendering the propagator massive, we also have a pole at $k^2 = 0$, since the last iC/D term has just a single power of momenta. This zero-momentum pole is what allows us to transmit the information needed to transmute the statistics of the particles.

8 *Maxwell in two dimensions* ✓

Today is a little brain-warmer that I realized I'd never done. We will be working though and elaborating on the appendix D of [?]. That is, we will be considering pure $U(1)$ gauge theory in two dimensions, with theta angle θ. We will put it on $\mathbb{R} \times S^1$, where the S^1 has circumference L and the $\mathbb{R}$ is time. We will find the spectrum as a function of θ and discuss issues related to superselection sectors and the role of Wilson lines in this theory.

First let us fix a gauge. We will choose a gauge in which A_1 is constant. This means we need to find an α such that

$$\begin{aligned}\partial_x^2\alpha(x,t) =&\partial_x A_1(x,t) \implies \partial_x\alpha(x,t) = A_1(x,t) - A_1(0,t) + c \\ &\implies \alpha(x,t) = \int_0^x dx' \left(A_1(x',t) - A_1(0,t)\right) + cx + d.\end{aligned} \tag{140}$$

This ensures that $A' = A - d\alpha$ will satisfy $\partial_x A_1' = 0$ for all t. Note that α is time-dependent! We can get the constants by requiring that α be well-defined, at least modulo 2π. This doesn't allow us to fix d (as d parametrizes the global $U(1)$ symmetry, it cannot be fixed), but it allows us to fix c so that (setting $d = 0$)

$$\alpha(x,t) = \int_0^x dx' \left(A_1(x',t) - A_1(0,t)\right) + x\frac{LA_1(0,t) - \int_0^L dx'\, A_1(x',t)}{L}. \tag{141}$$

One sees that at both $x = 0$ and $x = L$ we have $\alpha = 0$. This means that in the following, we can work with the variable

$$\phi(t) \equiv \int_0^L dx\, A_1(x,t), \tag{142}$$

so that the gauge-fixed $A_1(t)$ is $\phi(t)/L$.

Now we integrate out A_0 to enforce

$$\partial_x \star F = 0, \tag{143}$$

so that the electric field is a constant. We are left with the action (we have done the dx integral since everything is independent of x)

$$S = \frac{1}{2g^2L}\int dt\, (\partial_t\phi)^2 + \frac{\theta}{2\pi}\int dt\, \partial_t\phi. \tag{144}$$

The canonical momentum is

$$p = \frac{1}{g^2L}\partial_t\phi + \frac{\theta}{2\pi}. \tag{145}$$

Dimensionality check: $[g] = 1$, so p is dimensionless as it should be. After some aglebra we then find the Hamiltonian

$$H = \frac{g^2L}{2}\left(p - \frac{\theta}{2\pi}\right)^2. \tag{146}$$

Now we are in $U(1)$ gauge theory, not $\mathbb{R}$ gauge theory (this is slightly artifical since there are no charges), so that $\phi \sim \phi + 2\pi$ as a result of large gauge transformations being gauged (that they are gauged is what we mean by a $U(1)$ gauge theory). So since ϕ is periodic, the eigenfunctions p are just $e^{in\phi}$ for $n \in \mathbb{Z}$, and thus the spectrum is

$$E_n = \frac{g^2L}{2}\left(n - \frac{\theta}{2\pi}\right)^2, \qquad n \in \mathbb{Z}. \tag{147}$$

This has the expected twofold ground state degeneracy at $\theta = \pi$, exhibits the periodicity $\theta \sim \theta + 2\pi$, and so on. Note that the energy levels are liniearly proportional to the circumference of the circle, so that all the energy in the different states comes from the energy density of the vacuum. That is, there are no particles, which of course we know must be the case for gauge theory in two dimensions. Also note that the spectrum is dependent on θ, even though the added term $\theta \int F$ is topological, and hence it is independent of the metric and doesn't contribute to $T_{\mu\nu}$. In particular, it doesn't contribute to T_{00}, the Hamiltonian. But through its modification of the canonical momentum, it still has an effect on the spectrum.

What do the different n levels represent? They essentially represent the different values that the quantized electric flux $\star F$ can take on. We see that the electric flux is determined via

$$\star F = g^2(n - \theta/2\pi). \tag{148}$$

The θ term contributes to the electric flux in the usual way, with $\theta \mapsto \theta + 2\pi$ equivalent to changing the electric flux by one unit (this is one of the reasons why θ acts as a background electric field). Thus the different n levels are distinguished by the value of the electric flux which is threaded around the circle. We can go between different n by appling the Wilson line operator, since

$$[e^{i\int dx\, A_1(t)}, (\star F)(t)] = -g^2 e^{i\int dx\, A_1(t)}, \tag{149}$$

which means that the operator $e^{i\star F/g^2}$ generates the 1-form symmetry which shifts the holonomy. Since $\star F$ is constant, the charge operator $e^{i\star F/g^2}$ is independent of position, which is the statement of "current conservation $\implies$ topological charge operator" for a point charge operator. Since of course this 1-form symmetry cannot be broken in two dimensions, we know we will be able to label the states by their 1-form charges, which is just the obvious statement that we can label states by their electric fluxes. Anyway, if $W = e^{i\phi}$ is the Wilson loop, at equal times we have

$$W(\star F/g^2 + 1) = \star F W. \tag{150}$$

Thus we get the obvious statement that acting with the Wilson line increases the electric flux by g^2. This takes $n \to n+1$, and so the Wilson loop takes us between the different E_n. We can also phrase this in terms of the easily-checked similarity transform (here $H(p;\theta)$ is the Hamiltonian)

$$WH(p;\theta)W^\dagger = H(p;\theta + 2\pi), \tag{151}$$

which again demonstrates $\theta \sim \theta + 2\pi$. However, since there are no charges in the theory, the Wilson loop is kind of a pathological operator, since there is no way to apply it "gradually". In order to actually have it at our disposal, we would need to have charges (very massive ones would be fine) that we could pair-create and use to make W. Since this option is not available to us (pure gauge theory in two dimensions has no particles, as we have said), the different E_n levels are actually disconnected from one another, regardless of L. Also note that if we were to put our theory on T^2 instead of $\mathbb{R} \times S^1$, we would have to be more careful with the quantization procedure, since the operator $e^{i\int A_1}$ would not make sense: the electric flux needs to jump by 1 when crossing the Wilson line, but the Wilson line does not divide T^2 into disjoint pieces, so we get a contradiction (for a similar reason, $e^{i\int A_0}$ would not make sense). This is a trivial example of a higher-symmetry-enforced selection rule.

9 *Alternate approach to Wilson line expectation values in Chern-Simons* ✓

Today we will go over a functionally-flavored method for computing expectation values of Wilson line correlators in Abelian CS theories, focusing on the correlators $\langle W^q(C)\rangle$ and $\langle W^q(C)W^p(C')\rangle$, where $W(C) = \exp(i\int_C A)$ and where C, C' are homologically trivial but may have nontrivial linking. The usual way to do this is to solve the classical equations of motion in the presence of a source of charge q so that e.g. $F = (2\pi q/k) \star j$, where j is the source worldline.

The strategy is the same routine of using shifts of integration variables and Poincare duality that we know and love so well by now. For the insertion of two Wilson loops, for $U(1)_k$ we have

$$\langle W^q(C)W^p(C')\rangle = \frac{1}{Z}\int \mathcal{D}A \, \exp\left(i\frac{k}{4\pi}\int A\wedge dA + i\int A\wedge (q\widehat{C} + p\widehat{C}')\right). \tag{152}$$

To get rid of the Wilson loop insertion, we perform the shift

$$A \mapsto A - \frac{2\pi}{k} d^{-1}(q\widehat{C} + p\widehat{C}'). \tag{153}$$

Note that $\widehat{C}$ and $\widehat{C}'$ are both 2 forms, but taking the d^{-1} turns them into 1 forms. One can check using some integrations by parts that this shift kills the term in the exponent that is linear in A. This produces

$$\begin{aligned}\langle W^q(C)W^p(C')\rangle = \frac{1}{Z}\int \mathcal{D}A \, \exp\Big(i\frac{k}{4\pi}\int A\wedge dA - i\frac{2\pi}{k}pq\int \widehat{C}\wedge \frac{1}{d}\widehat{C}' - i\frac{\pi}{k}q^2\int \widehat{C}\wedge\frac{1}{d}\widehat{C} \\ - i\frac{\pi}{k}p^2\widehat{C}'\wedge\frac{1}{d}\widehat{C}'\Big).\end{aligned} \tag{154}$$

Now writing $C = \partial D$ and $C' = \partial D'$ so that $\widehat{C} = dD$ and $\widehat{C}' = dD'$, we have

$$\langle W^q(C)W^p(C')\rangle = \exp\left(-i\frac{2\pi}{k}pq\int \widehat{D}\wedge d\widehat{D}' - i\frac{\pi}{k}q^2\int \widehat{D}\wedge d\widehat{D} - i\frac{\pi}{k}p^2 D'\wedge d\widehat{D}'\right). \tag{155}$$

The terms with the self-CS interaction of $\widehat{D}$ and $\widehat{D}'$ are the expectation values of single Wilson loops, which we see by setting e.g. $p = 0$:

$$\langle W^q(C)\rangle = \exp\left(-i\frac{q^2\pi}{k}\int \widehat{D}\wedge d\widehat{D}\right). \tag{156}$$

This is ill-defined since the integral computes the intersection of ∂D with D, which doesn't make sense. We can regulate it by using a framing of the curve C, following the strategy in Witten's original Jones Polynomial paper. Given such a framing, we replace $d\widehat{D}$ in the above integral with $d\widetilde{D}'$, where $\partial D' = C'$ is a copy of C displaced infinitesimally along the vector field defined by the framing. The integral $\int \widehat{D}\wedge d\widehat{D}'$ then becomes the linking number of C and C', which depends only on the topological class of the framing (how many times the framing winds as it travels around C). Thus this framing-assisted regularization is just a choice of how to do point-splitting regularization for the Wilson operator. In $\mathbb{R}^3$ or S^3 we can always choose the framing so that the linking number of C and C' is zero, but for more general manifolds this may not be possible. In what follows we will actually choose a framing that winds by 2π along C if C is homologically trivial in the ambient spacetime (which we will assume to be the case). The reason for doing this will become clear in a second. Note that we are not loosing much by doing this, since we have a controlled way of determining how the answer changes upon changing the framing.

Anyway, doing the framing regularization so that the self-intersection number is equal to 1, we obtain

$$\langle W^q(C)\rangle = (-1)^{q^2/k}. \tag{157}$$

This means that with this convention, a lone Wilson loop computes the topological spin $s = q^2/(2k) \mod 1$ of the relevant anyon[18]. This comes from the fact that with our convention, the ribbon formed by C and its deformed copy has a 2π twist in it, so that unlinked loops compute the topological spin. In another convention where the framing is topologically trivial, unlinked loops would simply have expectation value 1. Now we have

$$\langle W^q(C)W^p(C')\rangle = \langle W^q(C)\rangle\langle W^p(C')\rangle \exp\left(-2\pi i\frac{pq}{k}\mathcal{L}(C,C')\right), \tag{158}$$

where $\mathcal{L}(C,C') = \int \widehat{D}\wedge d\widehat{D}'$ is the linking number of C and C'. In particular, note that a line with charge k is transparent with respect to all other lines. If k is odd this transparent line has spin $(-1)^k = -1$, and so odd k theories contain a transparent fermion—this is why they are spin TQFTs.

If we were to repeat this exercise with e.g. the Abelian CS theory with K matrix kX, then we would start with

$$\langle W^q(C)W^p(C')\rangle = \frac{1}{Z}\int \mathcal{D}A\, \exp\left(i\frac{k}{4\pi}\int(A\wedge dB + B\wedge dA) + i\int(qA\wedge C + pB\wedge C')\right), \tag{159}$$

where we have assumed that C is an A line and C' is a B line. If they were both A lines or both B lines, then we see we could perform a shift on just one of the fields so that $\langle W^q(C)W^p(C')\rangle = 1$. This is a check that the A and B fields are bosons (there is no self-interaction to change their statistics). Also note here that the lack of a self-interaction in the

[18]The topological spin is defined only modulo 1 since a Maxwell term (which we always imagine to be included in the action; it's just less relevant than the CS term in the IR using the standard scaling) leads to massive spin-1 particles (photons) that don't have any braiding phase with the sources. Hence by computing Wilson line vevs we can't distinguish a given anyon from the same anyon with a massive photon attached to it, and so the spin of the anyons is only well-defined modulo 1.

action means that the single loop expectation values $\langle W^p(C)\rangle$ don't need to be renormalized: they are equal to 1 identically. Anyway, we can perform the shifts

$$A \mapsto A - \frac{2\pi}{k} q \frac{1}{d} \widehat{C}', \qquad B \mapsto B - \frac{2\pi}{k} p \frac{1}{d} \widehat{C}', \tag{160}$$

which produces the familiar formula

$$\langle W^q(C) W^p(C')\rangle = \exp\left(-2\pi i \frac{pq}{k} \mathcal{L}(C, C')\right). \tag{161}$$

10 *The Witten effect* ✓

For some reason I found Witten's original paper on θ terms and monopole / dyon statistics rather hard to understand, and so the task for today is to explain the Witten effect in detail / with a slightly more modern presentation. We will be interested both in pure $U(1)$ gauge theory and in a situation where some larger non-Abelian gauge group is Higgsed down to $U(1)$.

We'll first do the easy part of looking at $U(1)$ gauge theory in four dimensions. We write the action as (in Minkowski signature)

$$S = \frac{1}{2e^2} \int F \wedge \star F - \frac{\theta}{8\pi^2} \int F \wedge F. \tag{162}$$

Here $\frac{1}{8\pi^2} \int F \wedge F = \frac{1}{2} \int (F/2\pi) \wedge (F/2\pi)$, which is in $\mathbb{Z}$ if the spacetime X is spin, so that we have the correct normalization of the θ term, with $\theta \sim \theta + 2\pi$. This is checked by remembering that the "instanton number" for $U(1)$ gauge theory is the second Chern character, which is

$$\mathrm{ch}_2 = \frac{1}{2}(c_1 \wedge c_1 - 2c_2), \tag{163}$$

where the c_i's are the Chern classes. Since $c_2 = 0$ for Abelian theories and $c_1 = F/(2\pi i)$, we see that the θ term is $\theta \int \mathrm{ch}_2$, which is the proper normalization.

Anyway, back to the problem. One way to motivate the Witten effect is to look at the equation of motion near a domain wall where θ jumps by some amount $\Delta\theta$. Not paying attention to getting the numbers right, this gives

$$d^\dagger F \propto d^\dagger \star (F \wedge \theta) = \star(dF \wedge \theta) + \star(F \wedge d\theta), \tag{164}$$

which means that the effective magnetic ($j_m = \star dF$) and electric currents ($j_e = d^\dagger F$) look like

$$j_e^\mu \sim j_m^\mu \theta + \Delta\theta \epsilon^{\mu\nu\lambda z} F_{\nu\lambda} \delta(z), \tag{165}$$

where we have taken the domain wall to lie in the xy plane. In particular, the $\mu = 0$ component says that

$$\rho_e \sim \theta \rho_m + \Delta\theta B^z \delta(z). \tag{166}$$

From the first term on the LHS, we see that monopoles in a $\theta \neq 0$ medium get electric charge attached to them. Alternatively, we can consider a spherical shell of material at θ, surrounded by a vacuum at $\theta = 0$. If there is magnetic flux B^r leaving the surface of the shell (this could be due to a magnetic monopole inside the shell or could come from a nontrivial 1st Chern class), then although $\rho_m = 0$ at the interface of the outer part of the shell with the vacuum, the second term on the RHS means that after integration we have $Q_e = \Delta\theta Q_m$ (where we have used Gauss' law for the magnetic field). This again shows how sources of magnetic field (be they monopoles or Chern classes) pick up electric charge when θ is turned on.

Now for a more precise justification. Consider the 1-form symmetry

$$A \mapsto A + \lambda\epsilon, \qquad \epsilon \in H^1(X; \mathbb{Z}), \tag{167}$$

where λ is a constant. Here large gauge transformations mean we identify $\lambda \sim \lambda + 2\pi$. The charge operator which generates the symmetry in the Hamiltonian formalism is

$$Q^{(1)}_{[\epsilon]} = \int_\Sigma \epsilon \wedge \frac{\delta}{\delta A}, \tag{168}$$

where Σ is space. Since $e^{2\pi Q^{(1)}_{[\epsilon]}}$ acts as the identity, the charges of this symmetry are quantized in $\mathbb{Z}$. To get the canonical momentum of A, we just differentiate the action and get

$$\frac{\delta}{\delta A} = i\left(\frac{\star F}{e^2} - \frac{\theta}{2\pi} F\right). \tag{169}$$

Now the electric charge operator is

$$Q_e(M) = \frac{1}{e^2} \int_M \star F \tag{170}$$

while the magnetic one is

$$Q_m(M) = \frac{1}{2\pi} \int_M F, \tag{171}$$

where the $1/2\pi$ normalization comes from normalizing the eigenvalues of Q_m to be integers. So we get

$$\begin{aligned} Q^{(1)}_{[\epsilon]} &= \int_\Sigma \epsilon \wedge \left(\frac{\star F}{e^2} - \frac{\theta}{2\pi} F\right) = \int_{\widehat{\epsilon} \subset \Sigma} \left(\frac{\star F}{e^2} - \frac{\theta}{2\pi} F\right) \\ &= Q_e([\widehat{\epsilon}]) - \frac{\theta}{2\pi} Q_m([\widehat{\epsilon}]). \end{aligned} \tag{172}$$

In particular, we see that the electric charge of a dyon is

$$Q_e([\widehat{\epsilon}]) = Q^{(1)}_{[\widehat{\epsilon}]} + \frac{\theta}{2\pi} Q_m([\widehat{\epsilon}]). \tag{173}$$

Now $Q_{[\widehat{\epsilon}]}$ and $Q_m([\widehat{\epsilon}])$ are both quantized in $\mathbb{Z}$, so for $\theta \notin 2\pi\mathbb{Z}$ the electric charge is not integral. We also see that $T : \theta \mapsto \theta + 2\pi$ acts on the charge lattice (q, m) by $T : (q, m) \mapsto (q + m, m)$. Of course, in order to actually have nonzero charges we need $H_2(X; \mathbb{Z})$ to be nontrivial so that we can have nontrival choices for $\widehat{\epsilon}$ (although this homology group can always be made nontrivial but excising small balls from spacetime).

Now for the non-Abelian version. We will do the usual example where we have $SU(2)$ broken down to $U(1)$ by giving a scalar ϕ in the fundamental of $SO(3)$ a vev (if we gave a vev to a scalar in the fundamental of $SU(2)$, the gauge group would be broken down completely). The symmetry breaking allows us to get dyons with charge assignements as in the $U(1)$ case but on spacetimes that have $H_2(X; \mathbb{Z}) = 0$, e.g. $\mathbb{R}^4$.

We will find it helpful to work with ϕ in the adjoint of $SU(2)$, rather than the fundamental of $SO(3)$ (of course they are the same, only the notation is different). This means we will write ϕ as a matrix in $\mathfrak{su}(2)$ as $\phi = \sigma^a \phi^a$. This is slightly more convinent compared to writing ϕ as a three-vector and having it transform under three-dimensional matrices.

The action is (Minkowski signature)

$$S = \int \left(-\frac{1}{2g^2} \mathrm{Tr}[F \wedge \star F] - \frac{\theta}{8\pi^2} \int \mathrm{Tr}[F \wedge F] + \mathrm{Tr}[d_A\phi \wedge \star d_A \phi] + \lambda \left(\mathrm{Tr}[\phi^2] - \frac{v^2}{2} \right)^2 \right), \tag{174}$$

where the covariant derivative is $d_A\phi = d\phi - i[A, \phi]$. Locally then the potential makes ϕ want to go like e.g. $v\sigma^3/2$, provided $\lambda \neq 0$. We will look for a monopole solution where at infinity ϕ goes to $\phi = \frac{r^a \sigma^a v}{2r}$ at infinity. If $\lambda = 0$ then we can have $\phi = \frac{r^a \sigma^a v}{2r} + O(1/r)$, but if $\lambda \neq 0$ then this leads to an infinite potential energy. We also need to choose asymptotic falloff conditions on the gauge field so that the kinetic term $|d_A\phi|^2$ is finite when integrated over space: i.e., we need $d_A\phi \sim O(1/r^2)$ as $r \to \infty$. We can ensure that this is the case provided that we choose A as follows:

$$A = 2\frac{i}{v^2}[\phi, d\phi] + \frac{2}{v}\mathcal{A}\phi, \tag{175}$$

where $\mathcal{A}$ is a $U(1)$ gauge field. This works since

$$\begin{aligned} d_A\phi &= d\phi + \frac{2}{v^2}[[\phi, d\phi], \phi] = d\phi - \frac{2}{v^2}\epsilon^{abc}\epsilon^{cde}\phi^a(d\phi)^b\phi^d\sigma^e \\ &= d\phi - \frac{2}{2v^2}(\delta_{ad}\delta_{be} - \delta_{ae}\delta_{bd})\phi^a(d\phi)^b\phi^d\sigma^e = d\phi - \frac{2}{v^2}(\mathrm{Tr}[\phi^2]d\phi - \phi d\mathrm{Tr}[\phi^2]) \\ &\sim \frac{1}{v^2 r^2}(\phi/r - d\phi), \end{aligned} \tag{176}$$

where in the last step we have kept the leading terms as $r \to \infty$. Since this goes as $1/r^2$, the kinetic term $|d_A\phi|^2$ has a finite integral over $\mathbb{R}^3$. Note that we don't actually need to know the functional form of ϕ for this to work (we can only solve for ϕ analytically when $\lambda = 0$).

The important part here is that the abelian gauge field $\mathcal{A}$ that we tacked on doesn't contribute to $d_A\phi$. It represents the gauge freedom in the inbroken $U(1)$ subgroup at infinity (rotations about the radial axis). When we multiply A by ϕ and take the trace the $[\phi, d\phi]$ part gets killed, and so $\mathcal{A}$ is proportional to the projection of A onto the radial direction:

$$\mathcal{A} = \frac{1}{v}\mathrm{Tr}[\phi A](1+\dots), \tag{177}$$

where if $\lambda = 0$ we can have $\dots \sim O(r^{-1})$; otherwise since we need $\mathrm{Tr}[\phi^2] \to v^2/2 + O(r^{-2})$ we have $\dots \sim O(r^{-2})$. The identification of the $U(1)$ gauge field makes sense since projecting onto the $\hat{r}$ direction in $SU(2)$ by tracing with ϕ selects out the generator of rotations about the radial direction, which leave $\langle\phi\rangle$ invariant (the structure group is reduced to $U(1)$ at infinity). Similarly, the $U(1)$ field strength is defined asymptotically as

$$\mathcal{F} = \frac{1}{v}\mathrm{Tr}[\phi F]. \tag{178}$$

Since $F \sim O(1/r^2)$ at infinity in order for the gauge field kinetic term to have finite energy, the parts of ϕ which go as negative powers of r can be ignored (thus we have not written any $(1+O(r^{-1}))$ factor on the RHS of the above equation), and the $U(1)$ field strength becomes exactly the projection of F onto the radial direction (locally we can write $\phi \to v\sigma^3$ and then $\mathcal{F} \to F^3$).

Now we need to examine the residual "zero mode" gauge transformations that act on the $U(1)$ gauge field $\mathcal{A}$. We need to look for "gauge transformations" which act nontrivially on the fields at infinity (and hence are not gauged). We want to leave the scalar field configuration invariant, and since it transforms in the adjoint, our gauge transformation parameter should be something built out of ϕ so that its action on ϕ is trivial. We also need the gauge transformation parameter to be purely radial, so that it only affects $\mathcal{A}$. It also needs to not mess with the falloff conditions we've imposed on e.g. $d_A\phi$ so that the energetics are unchanged. A transformation which fits the bill is given by $U_\alpha = \exp(i\alpha\phi/v)$, where $\alpha \in \mathbb{R}$. Actually, the fact that this is the generator is kind of obvious: it performs rotations about the $\hat{r}$ direction, which is exactly what the unbroken $U(1)$ does. This maps $A \mapsto A + \alpha d_A\phi/v$ (this has to be nontrivial since we can't have $d_A\phi = 0$ identically: if we did, we would have a global $U(1)$ symmetry [global symmetries are parametrized by covariantly constant things]. But we know this can't happen, since the structure group does not globally reduce to $U(1)$). Under the transformation, $\mathcal{A}$ changes by

$$\delta\mathcal{A} = \frac{1}{v}\mathrm{Tr}[\phi d_A\phi] = \frac{1}{2v}d\mathrm{Tr}[\phi^2]. \tag{179}$$

If $\lambda = 0$ this can have an $O(r^{-1})$ contribution, and we get the familiar $\mathcal{A} \sim r^{-1}$ falloff behaviour of a gauge field in the Coulomb phase. Now since when $\alpha \in 2\pi\mathbb{Z}$ we have $U_\alpha \to \mathbf{1}$, the operators $U_{2\pi k}, k \in \mathbb{Z}$ act as gauged gauge tranformations. Therefore the "physical gauge transformations" (I know, this terminology is awful) are parametrized by $\alpha \in [0, 2\pi)$ and thus give a $U(1)$ symmetry as expected.

Since the symmetry is $U(1)$, the charges associated to the asymptotic $U(1)$ symmetry will be integral. The charge operator for the symmetry is

$$\mathcal{U}(\alpha) = \exp\left(-i\frac{\alpha}{v}\int_{\mathbb{R}^3}\mathrm{Tr}[d_A\phi \wedge \delta_A]\right). \tag{180}$$

From the Lagrangian we read off

$$\delta_A = \frac{1}{g^2} \star F - \frac{\theta}{4\pi^2} F, \tag{181}$$

so that $\mathcal{U}(\alpha) = e^{-i\alpha Q_{\mathcal{A}}}$, where

$$\begin{aligned} Q_{\mathcal{A}} &= \frac{1}{v} \int_{\mathbb{R}^3} \mathrm{Tr}\left[d_A\phi \wedge \left(\frac{1}{g^2} \star F - \frac{\theta}{4\pi^2} F \right) \right] \\ &= \frac{1}{v} \int_{S^2_\infty} \left(\frac{1}{g^2} \mathrm{Tr}[\phi \wedge \star F] - \frac{\theta}{4\pi^2} \mathrm{Tr}[\phi \wedge F] \right) - \frac{1}{v} \int_{\mathbb{R}^3} \mathrm{Tr}\left[\phi \wedge \left(\frac{1}{g^2} d_A \star F - \frac{\theta}{4\pi^2} d_A F \right) \right], \end{aligned} \tag{182}$$

where we've used the fact that ϕ being in the adjoint means that e.g. $\mathrm{Tr}[\phi \wedge F]$ is $SU(2)$-neutral, so that we may write $\mathrm{Tr}[d_A\phi \wedge F] = d\mathrm{Tr}[\phi \wedge F] - \mathrm{Tr}[\phi \wedge d_A F]$. Also in the above, the hodge star is taken with respect to the full spacetime. The Bianchi identity means $d_A F = 0$, while we have

$$\mathrm{Tr}[\phi \wedge d_A \star F] \to \mathrm{Tr}[\phi (d_A)_i F^{0i}] \propto \mathrm{Tr}[\phi \sigma^a \epsilon_{abc} \phi^b (d_A \phi)^c_0] = 0, \tag{183}$$

where we have assumed our monopole solution is such that ϕ is covariantly constant in time (of course, we see from this formula that a moving monopole will produce an electric field, just like a moving electric charge produces a magnetic field). Thus only the surface integrals at infinity contribute, and we have

$$\begin{aligned} Q_{\mathcal{A}} &= \frac{1}{v} \int_{S^2_\infty} \left(\frac{1}{g^2} \mathrm{Tr}[\phi \wedge \star F] - \frac{\theta}{4\pi^2} \mathrm{Tr}[\phi \wedge F] \right) \\ &= Q_e - \frac{\theta}{2\pi} Q_m. \end{aligned} \tag{184}$$

Now $Q_{\mathcal{A}}$ is quantized in $\mathbb{Z}$ and so is Q_m (it is the first Chern class of the $U(1)$ bundle at infinity, and in fact is valued in $2\mathbb{Z}$, because of the factor of 2 from the trace or alternatively because the thing getting the vev was in the adjoint of $SU(2)$ i.e. the fundamental of $SO(3)$, instead of the fundamental of $SU(2)$. This is in fact the minimal possible magnetic charge since in these conventions, if we introduced a field charged in the fundamental of $SU(2)$ to do a Dirac string experiment, the field would have electric charge $1/2$). This implies, as in the Abelian case, that the electric charge of a monopole is non-integral, and dependent on the value of θ.

11 *More on the Schwinger model, its phases, and the theta term* ✓

Today's problem came from wanting to understand a statement made by Zohar Komargodski in his lecture on 1-form symmetries at a Stony Brook workshop; a good reference looks to be [**?**].

Consider the massive Schwinger model on the spacetime $S^1 \times \mathbb{R}$, where the fermions carry charge $q \in \mathbb{Z}$ and have a real mass $m \in \mathbb{R}$ (just for simplicity—axial rotations allow us to interchange the phase of the fermion mass with the θ angle, so this is done without loss of generality):

$$S = \int_{S^1\times\mathbb{R}} \left(\frac{1}{2\pi}\bar\psi(\partial\!\!\!/ - iqA\!\!\!/)\psi + \frac{m}{2\pi}\bar\psi\psi + \frac{1}{2e^2}F\wedge\star F - i\frac{\theta}{2\pi}F \right). \tag{185}$$

Because the fermion has charge q, this theory has a $\mathbb{Z}_q$ 1-form symmetry.

We will explain what happens in the limits $m \ll e$ and $m \gg e$ and discuss the differences between $q = 1$ and e.g. $q = 2$. As usual, most of the action happens at $\theta = \pi$.

$m \gg e$. First we do the case of large mass. It is then reasonable to throw the fermions away, and look at the pure gauge theory. We already analyzed QED_2 with a θ term in a previous diary entry, where we showed that the spectrum was labeled by the different quantized values of the electric flux:

$$E_n = \frac{e^2}{2}\left(n - \frac{\theta}{2\pi}\right)^2 = \frac{e^2}{2}F_{01}^2, \tag{186}$$

where we have set the circumeference of the circle to 1 for simplicity. When $\theta = \pi$ we have a degeneracy corresponding to the choice of $F_{01} = \pm 1/2$. This degeneracy can get lifted by a mxing between the two ground states of order e^{-mL} if $q = 1$, but is exact if $q > 1$.

We can also look at this from the boson side. The action bosonizes to (using the quantum fields and strings part II conventions and working in Euclidean time)

$$S = \int \left(\frac{1}{8\pi}d\phi\wedge\star d\phi - \frac{m}{\pi}\cos\phi + \frac{1}{2e^2}F\wedge\star F - i\frac{\theta + q\phi}{2\pi}F \right). \tag{187}$$

Note that changing $m \mapsto -m$ is the same as changing $\phi \mapsto \phi + \pi$, which from the fact that ϕ appears in the θ term reminds us of why TIs occur inside regions of spacetime where the fermion mass has a sign opposite to the sign it has in vacuum. Anyway, sending $m \to \infty$ freezes out the boson and we get pure QED at $\theta = \pi$, which as we have said has two orthogonal degenerate ground states (and as we said the orthogonality is exact even in finite systems when $q > 1$).

$m \ll e$. Now we look at small mass. In fact, we start with $m = 0$. Here the θ dependence can be completely removed by a shift in ϕ: this is the chiral anomaly, since a shift in ϕ is generated by the vector current for ϕ, which is dual to the axial current for the fermions. The action for $m = 0$ is quadratic and thus easy to solve. We can integrate out the gauge field and use the results of previous diary entries to write the effective action as

$$S = \int \left(\frac{1}{8\pi} d\phi \wedge \star d\phi + \frac{e^2}{2} \min_{k\in\mathbb{Z}} (k - q\phi/2\pi)^2 \right), \tag{188}$$

which holds as long as ϕ is slowly varying (if ϕ were non-compact, we would just have a quadratic mass term). Thus we see that we end up with a massive scalar (the soliton), and so the massless Schwinger model ("massless" in the sense that the action doesn't contain an explicit mass term for the fermions) is in fact actually massive. Since ϕ is valued in $[0, 2\pi)$, there is only one minumum of the potential if $q = 1$, but if $q > 1$ then we have q distinct minima.

Now we turn on a small mass. After shifting ϕ to kill the θF coupling, the potential for ϕ is

$$V(\phi) = -\frac{m}{\pi} \cos(\phi - \theta/q) + \frac{e^2}{2} \min_{k\in\mathbb{Z}} (k - q\phi/2\pi)^2. \tag{189}$$

Now the effect of θ is more important. Let's take $\theta = \pi$. If $q = 1$ then θ has the effect of shifting the minimum of the oscillating part of the potential to π. Superimposing the cosine on top of the quadratic potential has the effect of creating two distinct minima if the mass is large enough. We thus get the attractive picture of an Ising-type phase transition where two minima merge into one as m is varied, although where exactly this happens is hard to say, since it relies on us trusting this form of the effective potential for ϕ beyond the regime of parameters for which it was derived. Since the Ising transition is described by free fermions, we see that at some value of $m \sim e$ the confinement disspaears. The picture here is a line of alternating ± 1 charges, which since they change the flux by $\Delta F_{01} = 1$ are the domain walls for the Ising order parameter. When the domain walls start to proliferate confinement goes away at the massless Ising point since the electric fields of the domain walls cancel the background electric field coming from the θ term. Thus the physics is that of a tradeoff between the energy needed to create the screening charges (coming from the explicit $m\bar{\psi}\psi$ mass term) and the electrostatic energy saved when the screening fields are set up. One is led to think about this confinement transition as some sort of a toy model for the confinement transition occurring in $D = 4$ YM as the fermion masses are tuned.

If $q = 2$ then the quadratic part of $V(\phi)$ has two minima. Upon adding the mass term, there are still two minima. They are not equally spaced in ϕ because the cosine part of the potential is not symmetric about $\phi = 0$ while the quadratic part is. A similar picture holds for $q > 2$: the q distinct minima have their positions shifted, but for small m there remain q different minima.

The different minima are distinguished by the value of the electric flux around the ring, which comes spaced in integer units. Shifting the flux by q units is the same as shifting ϕ by 2π, and since we identify ϕ with $\phi + 2\pi$, different vacua that differ in their electric fluxes by q are connected. This is of course due to the fact that we have charge q particles which can propagate around the circle and change the flux by q. However, there is no process which can change the electric flux by $p < q$ units. Thus, even though we are on a circle, different vacua related by $\Delta F < q$ do not mix—they do not even mix in a way which is exponentially small in the particle mass / size of the circle. They only mix through a Wilson operator $e^{ip\oint A}$, which is nonlocal. Thus we genuinely have q distinct superselection sectors, even when space is compact.

So, at $\theta = \pi, q = 1$ we have the following picture: for $m \gg e$ we have two degenerate ground states distinguished by the electric flux, with an exponentially small mixing between them. Then at $m \ll e$ we only have one ground state, and so at some finite $m \sim e$ we have an Ising transition. If we e.g. change to $q = 2$, then the $m \gg e$ story is the same (although the two ground states do not mix), but for $m \ll e$ we still have two ground ground states (again with no mixing). Thus it is natural to guess that for $q = 2$ the Ising transition is eliminated. If $q > 2$, then it seems like as m is increased, there will be phase transitions where pairs of distinct minima merge into each other, eventually pairing down to leave behind the two minima of the $m \gg e$ limit (note to self: come back and think about this).

12 *Allowed spectrum of charges in Abelian and non-Abelian gauge theory and generalizations of the Dirac quantization condition* ✓

Today the goal is to figure out what types of electric and magnetic charges are allowed to be possessed by line operators in non-Abelian gauge theory.

For two dyons of electric and magnetic charges (q, m), (q', m'), we will show that for $U(1)$ gauge theory the quantization condition is

$$qg' - q'g \in \mathbb{Z}, \tag{190}$$

and will determine the non-Abelian analogue of this formula.

We will then describe the allowed line operators in a non-Abelian gauge theory based on a general compact Lie group G (here by "line operators", we mean operators that are literally supported on a line, and do not come with any surface operator [topological or otherwise] attached).

Let's first do $U(1)$ gauge theory, which is the easiest to understand. The charge quantization condition we want to derive is

$$qg' - q'g \in \mathbb{Z}. \tag{191}$$

The minus sign is the tricky part, which is somewhat unintuitive if one just looks at a given braiding process and thinks about AB phases. Basically, the minus sign is the minus sign in $\star\star = -1$, which holds on 2-forms in four-dimensional spacetime in Lorentzian signature ($\star^2 = (-1)^{p(D-p)+1}$ on a Lorentzian-signature D manifold when acting on p-forms, while for Euclidean signature the exponent is shifted by one).

From the Lorentz force law and EM duality $E \mapsto B, B \mapsto -E$, the force on a dyon (q, g) moving at velocty v in the field of another (motionless) dyon (q', g') is (ignoring factors of $1/4\pi$)

$$m\frac{dv^i}{dt} = (qq' + gg')\frac{r^i}{r^3} + (qg' - q'g)\epsilon^{ijk}v_j\frac{r_k}{r^3}, \tag{192}$$

where the all-important minus sign comes from doing $B \mapsto -E$ in the Lorentz force term. We want to get an angular momentum out of this so that we can find something which is quantized, and so we cross both sides with r: $r \times d_t(v) = d_t(r \times v)$ since $v \times v = 0$, so

$$m\frac{d(r \times v)}{dt} = (qg' - q'g)\frac{r \times (r \times v)}{r^3}. \tag{193}$$

Using the identity for two epsilon symbols with one index contracted between them, the cross product goes to $[r \times (r \times v)]^i = r^2 v^i - (r \cdot v)r^i$ (maybe the sign is wrong). When we divide by r^3 we get $v^i/r - (r \cdot v)r^i/r$, which is exactly the time derivative of r^i/r^2. So then we have total time derivatives on both sides, and we conclude that

$$(r \times p)^i = (qg' - q'g)\hat{r}^i. \tag{194}$$

Thus from quantization of angular momentum, we see that $qg' - q'g$, with the minus sign, is the correct thing to put a quantization condition on.

Now we go to looking at the spectrum of line operators in a general non-Abelian gauge theory. We want to examine the quantization condition on the magnetic charge carried by a given dyonic line operator $\mathcal{O}$. Consider a (small) S^2 linking $\mathcal{O}$. We can take the magnetic field on this S^2 to be uniform, with the gauge field on the S^2 being

$$A_\pm = \frac{B}{2}(\pm 1 - \cos\theta)d\phi, \tag{195}$$

where B is a covaraintly constant matrix (it should be covariantly constant since F is covariantly constant; it can't be an actual constant since F isn't gauge-invariant) determined by the magnetic field[19]. A proof of why the field is convariantly constant is in the following footnote[20]. Here the coordinates A_+ are used for the northern hemisphere, and A_- is used for the southern hemisphere. The two expressions differ at the equator by $d(B\phi)$, which

[19] More precisely, it is a covaraintly constant section of the adjoint bundle on the S^2. Recall that sections of the adjoint bundle $\mathrm{Ad}P$ are gauge-invariant things (like field strengths). The adoint bundle is given by taking the product of a principal G bundle P over the relevant spacetime (sub)manifold with $\mathfrak{g}$, and then quotienting by the adjoint action Ad, so $\mathrm{Ad}P = (P \times \mathfrak{g})/\sim_{\mathrm{Ad}}$. The identification here is $(g \cdot \phi, F) \sim (\phi, \mathrm{Ad}_{g^{-1}}F)$, or $(g \cdot \phi, \mathrm{Ad}_g F) \sim (\phi, F)$, which is telling us to mod out by gauge transformations.

[20] An S^2 linking the line operator in question sees a magnetic field

$$F_{ij} = \epsilon_{ijk}\frac{x^k}{4\pi|x|^3}B(x) = \mathrm{vol}_{S^2} \wedge B(x), \tag{196}$$

needs to be $-ig^{-1}dg$ for some well-defined g since the gauge fields on different patches are glued together with exterior derivatives of transition functions. We see that $g = e^{Bi\phi}$ on the equator, and so in order for g to be well-defined we need

$$e^{2\pi i B} = \mathbf{1} \tag{201}$$

to hold as a matrix equation.

Another essentially equivalent way to get the quantization is as follows. Gauge transformations act on B adjointly by conjugation. We now need a math fact: for any $B \in \mathfrak{g}$, we can always find a $B' = S^\dagger B S$ such that $B' \in \mathfrak{t}$, where $\mathfrak{t}$ is a maximal torus in $\mathfrak{g}$. Basically, we can always make a rotation in the gauge group to diagonalize the uniform field B (all elements in $\mathfrak{g}$ lie in *some* maximal torus). Since the components of the gauge field now commute with eachother, we can use stokes' theorem and write the Wilson loop as an integral over either the southern or northern hemispheres of the S^2. Demanding that these two integrals given consistent Wilson loops gives us the same quantization condition as before.

Now since $B \in \mathfrak{t}$, we can write $B = \beta^i H_i$ for some coefficient vector $\beta \in \mathbb{R}^r$, with $r = \dim \mathfrak{t}$ and where H_i are the generators of $\mathfrak{t}$. When we take the Wilson line to be in some representation, since the H_i can be simultaneously diagonalized, we can replace them with their eigenvalues μ_i, which are the weights of the given representation. Thus the quantization condition $e^{2\pi i B} = \mathbf{1}$ reads

$$\beta_i \mu^i \in \mathbb{Z}, \qquad \mu \in \Lambda_w(G). \tag{202}$$

Looking at the diary entry on weights and root lattices, we see that we can satisfy this condition by taking

$$\beta^i = 2\frac{\alpha_i}{\alpha^2} = \alpha_i^\vee, \qquad \alpha \in \Lambda_r(G),\, \alpha^\vee \in \Lambda_r^\vee(G) \tag{203}$$

---

since in this configuration there is a monopole at the center of the S^2. Note that B has x dependence since setting B to be a constant would not be a gauge-invariant thing to do in the non-Abelian case. However, B is covariantly constant.

To show this, consider the equations of motion on the S^2, namely $D_i F^{ij} = 0$. This reads

$$0 = (d^\dagger \mathrm{vol}_{S^2})^j B(x) + (\mathrm{vol}_{S^2})^{ij} D_i B(x). \tag{197}$$

Since the volume form on S^2 is co-closed (it is harmonic), we get that

$$\epsilon_{ijk} x^j D_k B(x) = 0. \tag{198}$$

Now consider the Bianchi identity, $D_{(i}F_{jk)} = 0$. This reads

$$0 = d\mathrm{vol}_{S^2} \wedge B(x) + \mathrm{vol}_{S^2} \wedge DB(x). \tag{199}$$

Again the first term vanishes, and this implies that

$$x^k D_k B(x) = 0, \tag{200}$$

since the antisymmetrizations from the wedge product and the definition of the volume form cancel out.

So together with the previous equation derived from the eom, we see that the vector $D_k B(x)$ is orthongonal to both the radial direction and to the directions tangent to the S^2. Thus it vanishes identically, and so $D_k B(x) = 0$ as claimed.

so that the quantization condition is satisfied by virtue of quantization of angular momentum in $\mathfrak{su}(2)_\alpha$. This is why we often think of the allowed t'Hooft lines as coming from representations that are created with $\otimes$s of the adjoint representation, in contrast to Wilson lines which can be in any representation: if the β^i can always be written in terms of a root in $\Lambda_r(G)$, then since $\Lambda_r(G)$ is the root lattice for the adjoint representation, all allowed magnetic charges must come from $\otimes$s of adjoint representations (the $\otimes$ operation corresponding to the fusion of t'Hooft lines).

This is not strictly true though, since while $\beta \in \Lambda_r^\vee(G)$ is sufficient for satisfying the quantization condition, it is not always neccessary. The allowed values for β actually come from a sublattice of the co-root lattice. The most general choice of β would be to take $\beta \in (\Lambda_w(G))^*$, where the dual indicates functions into $\mathbb{Z}$. Looking back at the previous problem, we see that this is precisely what it means for β to be a weight of the dual group $G^\vee$. Thus $\Lambda_w(G^\vee)$ parametrizes the allowed values of magnetic charge.

The diagonalization we made to rotate the magnetic field so that $B \in \mathfrak{t}$ was made with a gauge transformation that was constant on the S^2 which we were using to study the quantization condition. This doesn't completely fix B though, since there are still rotations we can do within the Cartan subalgebra which represent residual gauge redundancies. These redundancies are precisely given by the action of the Weyl group for the dual Lie algebra $\mathfrak{g}^\vee$ (remember that the Weyl group is given by reflections about the roots, so that it only depends on the Lie algebra, and not on the choice of Lie group). A math fact is that if $WBW^\dagger \in \mathfrak{t}$ for $B \in \mathfrak{t}$, then W impliments a Weyl transformation. So, the Weyl group contains all the residual gauge transformations not fixed by our choice of magnetic field $B \in \mathfrak{t}$. Now the roots for the dual group are the co-roots of the original group, which means that the Weyl group acts in the same way on both the lattice of G and the lattices of $G^\vee$, since

$$\text{Weyl} : \mu \mapsto \mu - \alpha(\alpha_i^\vee \mu^i) = \mu - \alpha^\vee(\alpha_i \mu^i), \tag{204}$$

so that Weyl and Weyl$^\vee$ act in the same way. Thus our tentative classification scheme for magnetic charges is to label them by elements of the quotient $\Lambda_w(G^\vee)/\text{Weyl}$.

A slighty better way to classify the charges is to realize that no matter what the exact Lie group and dual Lie group are (given a particular $\mathfrak{g}$), magnetic lines in the dual root lattice $\Lambda_r(\mathfrak{g}^\vee) = \Lambda_r^\vee(\mathfrak{g})$ will always be allowed. Indeed, for $\mu \in \Lambda_w(G)$ and $H \in \Lambda_r^\vee(\mathfrak{g})$ we have $\mu(H) \in \mathbb{Z}$ regardless of the exact choice of $G = \widetilde{G}/\Gamma_G$, so that magnetic charges in $\Lambda_r(\mathfrak{g}^\vee)$ are always allowed. Thus to obtain a classification which distinguishes between the line operators that are allowed for different choices of Γ_G, we can quotient the lattice of all possible magnetic charges (viz. $\Lambda_w(G^\vee)$) by the lattice of those that will be there no matter what (viz. $\Lambda_r(G^\vee)$. Thus we propose to classify magnetic charges by the quotient

$$\Lambda_w(G^\vee)/\Lambda_r(G^\vee) = \Lambda_w^*(G)/\Lambda_r^\vee(G) = \Lambda_w(\widetilde{G})/\Lambda_w(G), \tag{205}$$

where we have used various manipulations derived in the last diary entry. Using the last section of that diary entry, we see that

$$\Lambda_w(G^\vee)/\Lambda_r(G^\vee) = \Gamma_G = \pi_1(G), \tag{206}$$

where again $\widetilde{G}/\Gamma_G = G$. That we get such a classification for the magnetic charges of 't Hooft operators in terms of $\pi_1(G)$ makes perfect sense, since the 't Hooft operators can be defined by the holonomy (valued in $\pi_1(G)$) they induce in gauge field configurations. So the full class of allowed magnetic charges is $\Lambda_w(G^\vee)$ (up to Weyl invariance, more on this in a sec), and once we mod out by the lines which always appear regardless of the Lie group, we see that they are classified by $\Gamma_G = \pi_1(G)$.

A similar statement can be made for the electric operators. Since the allowed electric operators are determined by the allowed representations that we can take the trace in, they are classified by $\Lambda_w(G)$ (up to Weyl invariance). But the lines in the adjoint representation, corresponding to points in the $\Lambda_r(\mathfrak{g})$ lattice, always appear (they are the "worldlines of the gauge fields", since the gauge fields transform adjointly under constant gauge transformations) regardless of the choice of Lie group ($\Lambda_r(\mathfrak{g}) \subset \Lambda_w(G)$ for all G with Lie algebra $\mathfrak{g}$). Thus we propose to classify electric operators by the quotient

$$\Lambda_w(G)/\Lambda_r(G) = \Lambda_w^*(G^\vee)/\Lambda_w^*(\widetilde{G}) = \Lambda_w(\widetilde{G}^\vee)/\Lambda_w(G^\vee) = \Gamma_G^\vee = \pi_1(G^\vee), \tag{207}$$

where we used that e.g. $\Lambda_r^\vee(G) = \Lambda_r^\vee(G^\vee)$ and $\Lambda_w(\widetilde{G}) = (\Lambda_r^\vee(\widetilde{G}))^*$. Note the nice symmetry of this quotient with the quotient for the magnetic charges!

So far we have been considering lines that were either purely electric or purely magnetic. What happens if we have dyonic lines? Some of the details for this are in [**?**], so we will be brief. Basically, you can only have a consistent dynonic line if the electric and magnetic fields commute with one another (so that they can be fused together on a line in an unambiguous way). This is already ensured if we take the charges to be given by the classification scheme above, since everything done above involved only the Cartan subalgebras of $\mathfrak{g}$ and $\mathfrak{g}^\vee$. In more detail, one first fixes a magnetic field B, and then chooses an electric field in the centralizer of B in G. The centralizer G_B gives rise to a Lie algebra $\mathfrak{g}_B$, whose Cartan algebra is still the Cartan algebra of G since the elements in the Cartan algebra commute with one another. The Weyl group acts on B, and we also have a redundancy coming from Weyl_B acting on the electric sector, where Weyl_B is the subgroup which fixes B. But the combined action of Weyl_B and $\text{Weyl}/\text{Weyl}_B$ on the electric sector is just the same as having the full Weyl act, and so we just get a single (diagonal) action of Weyl on the electric and magnetic sectors. Thus the whole analysis goes through unchanged for dyonic operators.

If we propose to classify line operators by $\Gamma_G^\vee \times \Gamma_G$, what about Weyl invariance? It turns out that this is already accounted for, since the Weyl group acts trivially on $Z(\widetilde{G})$, and hence on the above assignments of both magnetic and electric charges ($\Gamma_G, \Gamma_G^\vee \subset Z(\widetilde{G})$). Proof: recall that a weight vector μ under the action of the Weyl group changes by $\delta\mu = \alpha(\alpha_i^\vee \mu^i)$, where $\alpha^\vee$ is some co-root. The inner product here is just $\mu(H_\alpha)$, which since H_α is an $\mathfrak{sl}(2,\mathbb{C})$ generator, is an integer. Thus $\delta\mu \in \Lambda_r(G)$, and so the Weyl group acts on vectors by adding integer multiples of roots to them (see the $SU(3)$ figure from the last diary entry for a nontrivial example of how this plays out). In the quotient $Z(\widetilde{G}) = \Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G})$ this action is trivial, and so Weyl acts trivially on $Z(\widetilde{G})$.

Summarizing, we can classify the lattice L of line operators by

$$L = (\Lambda_w(G)/\Lambda_r(G)) \times (\Lambda_w(G^\vee)/\Lambda_r(G^\vee)) = \Gamma_G^\vee \times \Gamma_G = \pi_1(G^\vee) \times \pi_1(G) \subset Z(\widetilde{G})^2. \tag{208}$$

Since $\Gamma_G^\vee \times \Gamma_G$ is an Abelian group, multiplying equivalence classes of line operators is done easily by using addition in the group. The simplest cases are when one of Γ_G, $\Gamma_G^\vee$ is the center of $\widetilde{G}$. If $\Gamma_G = Z(\widetilde{G})$ then $\Lambda_w(G) = \Lambda_r(G)$ and $\Gamma_G^\vee = \mathbb{Z}_1$, so that the spectrum of line operators modulo gauge field world lines only includes magnetic line operators. Likewise, if the gauge group is simply connected so that $\Gamma_G = \mathbb{Z}_1$ then we have no 't Hooft lines (since $\pi_1(G) = 0$), and the spectrum has only electric line operators. Also note that this implies that no matter what the gauge group is, the number of line operators is always equal to the order of the center:

$$|L| = |\Gamma_G||\Gamma_G^\vee| = |Z(\widetilde{G})|. \tag{209}$$

Now we take another slightly approach to identifying L by looking at the analogue of the quantization condition on $qm' - q'm$ for the non-Abelian case. The same angular momentum argument goes through unmodified (I think—the argument for the $U(1)$ unfortunately took S-duality for granted, though), but the charges involved are now matrices and need to get turned into numbers with the help of an inner product. Since the electric (magnetic) charges are in $\mathfrak{t}^*$ ($\mathfrak{t}$), the inner product is just the evaluation map $\langle , \rangle : \mathfrak{t}^* \otimes \mathfrak{t} \to \mathbb{C}$. For two dyons with electric / magnetic field strengths (Q, B) and (Q', B'), we thus require

$$\langle Q, B' \rangle - \langle Q', B \rangle \in \mathbb{Z}. \tag{210}$$

Let us pretend we didn't know about the $\Gamma_G, \Gamma_G^\vee$ groups calculated previously, and just wanted to go about solving the quantization condition directly. By the physical arguments given earlier, we know that the lattice of line operators has to sit inside $Z(\widetilde{G})^2$:

$$L \subset (\Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G})) \times \Lambda_w(\widetilde{G}^\vee)/\Lambda_r(\widetilde{G}^\vee) = Z(\widetilde{G})^2, \tag{211}$$

where the group $\widetilde{G}^\vee$ is defined so that its roots are the co-roots of $\widetilde{G}$.

Solving the quantization condition by calculating the inner product is simple in our reduced classification scheme in terms of $\Gamma_G^\vee \times \Gamma_G \subset Z(\widetilde{G})^* \times Z(\widetilde{G})$ (we have been ignorning the difference between $Z(\widetilde{G})$ and $Z(\widetilde{G})^*$ since $Z(\widetilde{G})$ is Abelian), since we can use the group law in $Z(\widetilde{G})$. For simplicity, let us first assume that $Z(\widetilde{G}) = \mathbb{Z}_N$. We can thus write a given electric line in $\Lambda_w(\widetilde{G})/\Lambda_r(\mathfrak{g})$ as $\frac{q}{N}R$, where $R \in \Lambda_r(\mathfrak{g})$ is some root that we fix and $q \in \mathbb{Z}_N$. Similarly, we can choose a given magnetic line to be mH, where $H \in \Lambda_w(\widetilde{G}^\vee) = \Lambda_r^*(\widetilde{G})$ is a fundamental magnetic weight that has inner product 1 with R (this is possible since the fundamental weights are the dual of the roots), and $m \in \mathbb{Z}_N$. Thus the quantization condition for two dyons $((q/N)R, mH), ((q'/N)R, m'H)$ is

$$\frac{1}{N}(qm' - q'm)\langle R, H \rangle = \frac{qm' - q'm}{N} \in \mathbb{Z}. \tag{212}$$

More generally, if $Z(\widetilde{G})$ is an Abelian group with n $\mathbb{Z}_k$ factors, then we can proceed as above but assign electric lines as $q_1 R_1 + \cdots + q_n R_n$, and likewise for magnetic lines. One then ends up with the same quantization condition as above in each $\mathbb{Z}_k$ factor. Thus if we specify two dyons by $(q, m), (q', m') \in Z(\widetilde{G})^2$, the quatization condition is

$$qm' - q'm = 0 \text{ in } Z(\widetilde{G}). \tag{213}$$

Let's now again specialize to the case where $Z(\widetilde{G}) = \mathbb{Z}_N$, which is basically the only case of interest (all simple Lie groups have $Z(\widetilde{G})$ cyclic expect $\mathrm{Spin}(4n)$, where it is $\mathbb{Z}_2^2$). How do the solutions to the quantization condition (213) relate to the subgroups $\Gamma_G, \Gamma_G^\vee$ that we identified earlier? Hopefully, both the solutions to (213) and the choices of Γ subgroups enumerate the same list of sublattices of the "full" charge lattice $\mathbb{Z}_N^2$. Let's now see why this is indeed the case.

Without loss of generality we can write

$$\Gamma_G = \mathbb{Z}_a, \quad \Gamma_G^\vee = \mathbb{Z}_b, \tag{214}$$

for some $a, b \in \mathbb{Z}$ such that $ab = N$. The most obvious way of embedding these groups into the full lattice $\mathbb{Z}_N^2$ is to take $\Gamma_G = (0, \frac{N}{a}k) \subset \mathbb{Z}_N^2$ for $k \in \mathbb{Z}_a$, and likewise to take $\Gamma_G^\vee = (\frac{N}{b}l, 0) \subset \mathbb{Z}_N$ for $l \in \mathbb{Z}_b$ (here the first factors are electric charges and the second factors are magnetic charges). Then the lattice of allowed operators is

$$L = \{(N/b)l, (N/a)k \,|\, (l,k) \in \mathbb{Z}_b \times \mathbb{Z}_a\} \tag{215}$$

This of course satisfies the quantization condition (213) for all l, k. However, depending on the choice of groups involved, this lattice will not be the only lattice allowed, as there may be multiple ways of embedding the $\Gamma_G, \Gamma_G^\vee$ groups into the full lattice $\mathbb{Z}_N^2$. This is related to the Witten effect.

In our conventions, the emedding of the group $\Gamma_G^\vee$ into $\mathbb{Z}_N^2$, which determines the allowed electric lines, will always be uniquely defined as

$$\Gamma_G^\vee = \mathbb{Z}_b = \{((N/b)k, 0) \,|\, k \in \mathbb{Z}_b\} \subset \mathbb{Z}_N^2. \tag{216}$$

That is, the group $\Gamma_G^\vee$ determines the purely electric operators in the charge lattice. This is because $\Gamma_G^\vee = \Lambda_w(G)/\Lambda_r(G)$ classifies the allowed representations of the gauge group (modulo those constructed from $\otimes$s of the adjoint). Each representation R always defines a purely electric line operator via $W_C = \mathrm{Tr}_R[\exp(\int_C A)]$.

By contrast, we have some freedom when it comes to the magnetic operators and the embedding of Γ_G. This freedom essentially comes from our ability to make a re-definition of what we mean by a magnetic charge. For example, when we were deriving the constraints on allowed magnetic charges for line operators, we probed a magnetic line operator $\mathcal{O}$ with a purely electric line, and then constrained the allowed magnetic lines as a function of the different representations the electric line could be taken in. But since electric lines don't have statistical phases with other electric lines, the quantization conditions on the magnetic charge of $\mathcal{O}$ would be unchanged if we replaced $\mathcal{O}$ with $\mathcal{O} \otimes W$, where W is a purely electric line. Thus we could define our magnetic operators to come attached with electric lines, and the whole story would go through unchanged. Theories with different types of W's attached to the magnetic operators $\mathcal{O}$ are related by the Witten effect (i.e. usually by a shift of some θ angle), and correspond to different ways of embedding Γ_G in the full $\mathbb{Z}_N^2$ charge lattice. For example, if $\Gamma_G = \mathbb{Z}_N$, we may choose to embed it in $\mathbb{Z}_N$ as any of the N distinct subgroups $\{(nl, l)\}$, with $n, l \in \mathbb{Z}_N$.

In general, we can say that the lattice L of line operators fits into an exact sequence

$$0 \to \Gamma_G^\vee \to L \to \Gamma_G \to 0. \tag{217}$$

The injection is unique and the choice of $\Gamma_G^\vee$ completely determines the purely electric operators we have access to. There are usually mutliple ways of projecting onto the magnetic group Γ_G though, each of which gives us a distinct solution to the quantization condition. Note that this sequence will not always be split, which means that there may be no subset of lines in the full charge lattice whose magnetic charges fuse in a Γ_G subalgebra (if there were, there would exist a splitting homonorphism $\Gamma_G \to L$ with that subalgebra as its image).

Again returning to $Z(\widetilde{G}) = \mathbb{Z}_N$, we see that if $\Gamma_G = \mathbb{Z}_1$, there is only one theory, whose charge lattice is the purely electric $\mathbb{Z}_N \times 0$ sublattice of $\mathbb{Z}_N^2$. As already mentioned, if $\Gamma_G = \mathbb{Z}_N$, there are N distinct theories, which differ in the charge assigned to the fundamental magnetic line and which are permuted by the modular T operation. Also note that regardless of the choice of Γ_G, if there is a single line (q, m) with $m = 1$, then there are no purely electric lines in the $\mathbb{Z}_N^2$ charge lattice (this follows from $|L| = N$). Likewise, if there is a line $(1, 0)$ then there are no magnetically charged lines (and as previously mentioned we must have $\Gamma_G = \mathbb{Z}_1$).

For $Z(\widetilde{G}) = \mathbb{Z}_N$, we can take $\Gamma_G^\vee = \mathbb{Z}_b$, $\Gamma_G = \mathbb{Z}_a$ for $N = ab$. The different equivalence classes of extensions are given by

$$\mathrm{Ext}^1_{\mathbb{Z}}(\mathbb{Z}_a, \mathbb{Z}_b) = \mathbb{Z}_a \otimes_{\mathbb{Z}} \mathbb{Z}_b = \mathbb{Z}_{\gcd(a,b)}. \tag{218}$$

In this classification, the split extensions are trivial. The number of different split extensions is classified by different ways of taking semidirect products, i.e. different maps $\mathbb{Z}_b \to \mathrm{Aut}(\mathbb{Z}_a) = \mathbb{Z}_a^*$, where $\mathbb{Z}_a^*$ is the multiplicative group (of order $|\mathbb{Z}_a^*| = \phi(a)$). Thus the number of split extensions is $|\mathrm{Hom}(\mathbb{Z}_b \to \mathbb{Z}_a^*)|$ (note to self: where was I going with this paragraph?)

Two last things worth re-iterating before wrapping up. First, we have classifed the allowed line operators in the theory given $\widetilde{G}$ and Γ_G by picking out a certain subset of the charge lattice. Operators with charges running over all values of the charge lattice exist no matter what Γ_G is, but unless they are part of the given subset, they must come with surfaces (which may be topological depending on the phase) attached to them, and cannot be defined on homologically non-trivial cycles. Also, it is good to remember the distinction between magnetic charge (i.e. "GNO magnetic charge") and t' Hooft flux. The former is basically a lattice point in $\Lambda_w(G^\vee)$, while the latter is an element in $\pi_1(G)$. In particular, we have no operators with t Hooft flux in theories where the gauge group is simply connected, while we still have operators with nonzero GNO magnetic charge.

13 *Higher symmetries in non-Abelian gauge theories* ✓

Today we will consider non-Abelian gauge theory for some gauge group G, obtained from a simply-connected covering space $\widetilde{G}$ by $G = \widetilde{G}/\Gamma_G$. We will determine the higher symmetries possessed by this theory, and show how theories with different choices of Γ_G are related from a higher-symmetry point of view. We will be basically be trynig to understand the comments in [**?**, **?**] and explain what these papers are doing in detail. Throughout this diary entry we will be using notation as if we were dealing with differential forms even when we are really dealing with cochains with coefficients in a discrete group. Therefore things like $\wedge$ will always implicitly stand for their appropriate discrete analogues. Pointryagin squares and the like will also be implicitly understood but will be written with $\wedge$.

The difference between gauge theories based on different quotients $G = \widetilde{G}/\Gamma_G$ for finite Γ_G comes in the line operators that are allowed in the theory, as we saw in an earlier diary entry. The difference between the different Gs is the types of transition functions that they allow: a given set of transition functions may satisfy the cocycle condition (closed under the Cech differential) for one choice of Γ_G, but not another.

How does this affect the allowed line operators? Naively the Wilson lines operators do not care about the transition functions and the different choices of Γ_G (at least if their topology is trivial), since they are integrals of a Lie-algebra-valued quantity. But this is too hasty. To see why, we write the Wilson line by splitting it up into a bunch of patches. We take the path C to lie in the union of a collection of patches $\{U_\alpha\}$, with the segment of the Wilson line W_C lying in U_α denoted by W_α. The naive formula for $W(C)$ is then

$$W(C) = \mathrm{Tr}\left[\prod_\alpha W_\alpha\right], \tag{219}$$

but this is not quite correct. Indeed, it is not invariant under changing the local trivializations on each U_α. Under a change in trivialization g_α which is constant on each patch, we have $W_\alpha \mapsto g_\alpha^\dagger W_\alpha g_\alpha$, and so if we e.g. change the trivialization on a single patch, our fomula for $W(C)$ is not invariant. The correct thing to do is to glue each W_α together with transition functions $t_{\alpha\beta}$:

$$W(C) = \mathrm{Tr}\left[W_\alpha t_{\alpha\beta} W_\beta t_{\beta\gamma} W_\gamma \cdots\right]. \tag{220}$$

Under a change in transition functions of $\{g_\alpha\}$, we have $t_{\alpha\beta} \mapsto g_\alpha^\dagger t_{\alpha\beta} g_\beta$, and $W(C)$ is left invariant. This construction is related to the DB cohomology approach for integrating gauge fields, which is easier to spell out in more detail in the Abelian case.

Anyway, we now see how the choice of transition functions affects the Wilson loop. For example take $SU(N)$, and consider twisting the transition functions such that a single transition function appearing in $W(C)$ gets twisted by $t_{\alpha\beta} \mapsto e^{2\pi i/N} t_{\alpha\beta}$ (we are not changing any of the trivializations, just a transition function—this can be done without creating extra field strength only if C is homologically nontrivial). Then under this change we see that $W(C) \mapsto e^{2\pi i/N} W(C)$, desipte the fact that we haven't actually changed the coordinate-patch realizations of the gauge field $\{A_\alpha\}$ at all. Thus it is good to keep in mind that although $W(C)$ only involves the gauge field, the gauge field is really a principal bundle, which carries more information (viz. information about the transition functions) than just the 1-forms A_α.

The example we will be focusing on primarily is $\widetilde{G} = SU(N)$. The fundamental Wilson lines depend on the transition functions, and may not be good line operators when Γ_G is taken to be nontrivial. However, Wilson lines in the adjoint are always good line operators, regardless of Γ_G. We know this from the previous diary entries since the root lattice always gives legit line operators, but now we can see it in a different way.

For N the fundamental of $SU(N)$, we have

$$N \otimes \bar{N} = 1 \oplus A, \tag{221}$$

where A is the adjoint. Thus for all $SU(N)$ we can write a matrix in the adjoint as follows:

$$[U_A]^{ij}_{kl} = [U_N]^i_j [U_N^\dagger]^k_l - \frac{1}{N}\delta^i_l \delta^j_k. \tag{222}$$

In this expression, i, j are fundamental indices and k, l are antifundamental indices. Here the second factor projects out the $\mathbf{1}$ in the direct sum, and the $1/N$ is so that when we put in $U_N = \mathbf{1}$ then we get $\mathrm{Tr}[\mathbf{1}_A] = N^2 - 1 = \dim(A)$. The index structure is fixed to be $\delta^l_i \delta^j_k$, since this is the $\otimes$ of invariant symbols (δ functions) for the two pairs of $N, \bar{N}$ indices.

This means that the Wilson loop in the adjoint can be computed from the Wilson loop in the fundamental by

$$W_A(C) = \mathrm{Tr}_A e^{i\oint_C A} = \mathrm{Tr}\left\{ \left[e^{i\oint_C A}\right]^i_k \left[e^{-i\oint_C A}\right]^j_l - \frac{1}{N^2}\delta^i_k \delta^j_l \right\}. \tag{223}$$

The trace sets $i = k$ and $j = l$, so

$$W_A(C) = \left|\mathrm{Tr}_f e^{i\oint_C A}\right|^2 - 1. \tag{224}$$

Thus we see that since the center symmetry changes the fundamental Wilson line by a phase, it leaves $W_A(C)$ invariant. Thus no matter what choice of $\Gamma_G \subset \mathbb{Z}_N = Z(SU(N))$ we make, the adjoint Wilson lines will always be well-defined well defined operators, blind to the allowed t' Hooft line operators.

With these introductory comments out of the way, let's now see how this works in a more detailed way. We'll focus on the simple example of the relation between $SU(N)$ and $PSU(N)$ gauge theory, although quotienting $SU(N)$ by other subgroups of $\mathbb{Z}_N$ can be done analogously.

Of course, the real difference between $SU(N)$ and $PSU(N)$ gauge theories is in the transition functions, and passing from $SU(N)$ to $PSU(N)$ means changing the transition functions. This isn't something that's easy to do in a transparent way as far as the variables naturally appearing in a QFT are concerned, so we will try to encode the changed transition functions into a field that appears in the action.

We want to "gauge" the $\mathbb{Z}_N$ part of the transition functions to obtain $PSU(N)$ gauge theory, which we will do by coupling the $SU(N)$ theory to a $\mathbb{Z}_N$ gauge field. Recall that discrete gauge fields are basically just $\mathbb{Z}_N$ transition functions: they have no local degrees of freedom on patches; all of their physical content is in transition functions between patches. Coupling $SU(N)$ to a $\mathbb{Z}_N$ gauge field will then allow us to identify two states that differ by twisting the transition functions by something in $\mathbb{Z}_N$, and we will have obtained a $PSU(N)$ theory. Since we are in the continuum, we will write down the $\mathbb{Z}_N$ gauge field using the BF theory approach, where we only deal with $U(1)$ fields. Since the $\mathbb{Z}_N$ symmetry we want to gauge is a 1-form symmetry, the gauge field will be a $\mathbb{Z}_N$ 2-form field B. we use the presentation of the BF action where NB gets set to be the field strength of a $U(1)$ gauge field though a Lagrange multipler coupling:

$$S \supset \frac{i}{2\pi} \int H \wedge (F_{\mathcal{A}} - NB), \tag{225}$$

where $F_{\mathcal{A}}$ is the field strength for a $U(1)$ gauge field $\mathcal{A}$ (we are avoiding writing it as $d\mathcal{A}$ since $\mathcal{A}$ may not be a 1-form). Here H is quantized to have $2\pi\mathbb{Z}$ periods around all closed 2-manifolds.

If F_A is the $SU(N)$ field strength, the naive thing to do would be to take $F_A \mapsto F_A - B\mathbf{1}$ in the action. This is what we would do if we were trying to gauge a $U(1)$ 1-form symmetry, since gauging the 1-form symmetry that acts on A means that F_A changes by total derivates under gauge transformations. But this isn't right: locally we can replace B with $\frac{1}{N}F_{\mathcal{A}}$, which means that the action $||F_A - \frac{1}{N}F_{\mathcal{A}}||^2$ now has more local degrees of freedom than it did when we started, which is definitely not what we want. The correct thing to do is to cancel out these extra local degrees of freedom with the $\mathcal{A}$ field. We will do this by taking a $U(N) = [SU(N) \times U(1)]/\mathbb{Z}_N$ gauge theory, and killing off the $U(1)$ factor to produce an $SU(N)/\mathbb{Z}_N = PSU(N)$ theory.

To this end, consider the $U(N)$ field

$$\mathscr{A} = A + \frac{1}{N}\mathcal{A}\mathbf{1}. \tag{226}$$

Here A is traceless and $\mathcal{A}$ is the $U(1)$ part of the $U(N)$, with $\mathrm{Tr}\mathscr{A} = \mathcal{A}$. By saying that this is a $U(N)$ field, we really just mean that the transition functions of A and those of $\mathcal{A}$ are correlated in a way such that their combination gives legit $U(N)$ transition functions. In particular, we can let the transition functions for the $SU(N)$ part fail to be closed in $\check{C}^1(M;\mathbb{Z})$, provided that the transition functions for $\mathcal{A}/N$ compensate this ($\mathcal{A}$ can do this while still remaining a legit $U(1)$ gauge field because of the $1/N$ in front of it in the definition for $\mathscr{A}$: $\mathcal{A}$ is a well-defined $U(1)$ gauge field, but $\mathcal{A}/N$ is not). This identification is the quotient by $\mathbb{Z}_N$ in $[SU(N) \times U(1)]/\mathbb{Z}_N$.

Since we now can shift the transition functions of the A bundle by Nth roots of unity through a gauge transformation, we almost have a $PSU(N)$ gauge field. The only problem is that it has an extra $U(1)$ local degree of freedom that the $PSU(N)$ theory doesn't have. This extra local degree of freedom is eliminated by adding the background field B to the action though, via

$$S \supset \frac{1}{2g^2} \int \mathrm{Tr}\left[(F_{\mathscr{A}} - B\mathbf{1}) \wedge \star(F_{\mathscr{A}} - B\mathbf{1})\right]. \tag{227}$$

This is the proper way to couple the $SU(N)$ theory to a $\mathbb{Z}_N$ gauge field: neither B nor $\mathcal{A}$ by themselves is the $\mathbb{Z}_N$ gauge field; the $\mathbb{Z}_N$ gauge field involves both of them. This is the price we pay for wanting to work in the continuum. Note that A is not really the $PSU(N)$ field: while the bundle for A is allowed to have transition functions which fail the cocycle condition by something in $\mathbb{Z}_N$, we need the gauge fields $\mathcal{A}, B$ to make sure that the places where the cocycle condition fails can be removed by gauge transformations, so that such places do not contribute to a physical field strength. So it's really the whole package that constitutes the $PSU(N)$ field.

Now it may seem like we've actually just gotten back to our starting point by adding B, since it appears in the combination $\frac{1}{N}F_{\mathcal{A}}\mathbf{1} - B\mathbf{1}$, which seems to vanish upon integrating out H. But integrating out H only says that $F_{\mathcal{A}} = NB$ and does *not* imply that $\frac{1}{N}F_{\mathcal{A}} = B$. Locally it does, and so we have indeed added no new local degrees of freedom. But globally, knowing NB does not let you know B. Knowing NB means that you know B locally, and means that you know $(e^{i\int_\Sigma B})^N$ for all closed 2-manifolds Σ. In fact since $NB = F_{\mathcal{A}}$, we have $(e^{i\int_\Sigma B})^N = 1$ for all Σ, and so knowing NB doesn't give you any information about the holonomies of B around closed 2-manifolds (which are always Nth roots of unity). It may help to again explain exactly what happens when we integrate out H. Doing a Hodge decomposition $F_{\mathcal{A}} = d\alpha + \omega_{\mathcal{A}}$, $B = d\lambda + d^\dagger\epsilon + \omega_B$, $H = d\gamma + \omega_H$, $\omega_{\mathcal{A}} \in 2\pi H^2(M;\mathbb{Z})$, $\omega_H \in H^2(M;\mathbb{R}/2\pi\mathbb{Z})$, $\omega_H \in 2\pi H^2(M;\mathbb{Z})$ (recall that $\int_\Sigma H \in 2\pi\mathbb{Z}$ for all closed 2-manifolds Σ), we get a δ function setting $\epsilon = 0$, so that the non-cohomological degrees of freedom are all pure gauge and thus dissapear. Upon summing over $\omega_H \in 2\pi H^2(M;\mathbb{Z})$ we get the term

$$\sum_{\omega_H \in 2\pi H^2(M;\mathbb{Z})} \exp\left(\frac{i}{2\pi}\int \omega_H \wedge (\omega_{\mathcal{A}} - N\omega_B)\right) = \sum_{\widehat{\omega}_H \in H_2(M;\mathbb{Z})} \exp\left(i\int_{\widehat{\omega}_H} (\omega_{\mathcal{A}} - N\omega_B)\right). \quad (228)$$

Given the quantization of $\omega_{\mathcal{A}}$, this means that we can constrain

$$\omega_B \in \frac{2\pi}{N} H^2(M;\mathbb{Z}), \quad (229)$$

but ω_H is free to take on any cohomology class in this cohomology group (classes in $H^2(M;\mathbb{Z})$ are gauged under the 1-form gauge symmetry though, so only the classes in $(2\pi)/NH^2(M;\mathbb{Z}_N)$ are physically distinct). In particular, integrating out H does not actually set $B = \frac{1}{N}F_{\mathcal{A}}$, since the cohomology class of B is not fixed after integrating out H. This class is the sole degree of freedom carried by the $\mathbb{Z}_N$ gauge field.

Thus, the combination $\frac{1}{N}F_{\mathcal{A}} - B$ carries no local degrees of freedom, but carries global $\mathbb{Z}_N$ degrees of freedom: it is the $\mathbb{Z}_N$ 2-form gauge field that want to couple to the $SU(N)$ fields. As a sanity check, if we have a situation where $\int_\Sigma B \in 2\pi\mathbb{Z}$ for all Σ, then B is pure gauge and we can make a gauge transformation to eliminate $\mathcal{A}$ and B from the theory entirely. This gives us back the pure $SU(N)$ theory as required.

We can now look at the operators in the theory. The Wilson line in the fundamental of $SU(N)$ now is not gauge invariant under 1-form gauge transformations $\mathcal{A} \mapsto \mathcal{A} + N\lambda$, and so we must write it as a surface operator by attching a B surface:

$$W_f(C) \mapsto \operatorname{Tr}\left[\exp\left(i\oint_C \mathcal{A}\right)e^{-i\int_\Sigma B\mathbf{1}}\right] = \operatorname{Tr}\left[\exp\left(i\oint_C A\right)\right]e^{\frac{i}{N}\oint_C \mathcal{A}}e^{-i\int_\Sigma B}, \quad (230)$$

where $\partial\Sigma = C$ and we used $e^{x\mathbf{1}} = \mathbf{1}e^x$ for $x \in \mathbb{C}$. Because of the attached surface operator the fundamental Wilson lines are no longer part of the lines operators in the theory. The adjoint lines are the new "smallest charge" electric line operators: as we saw earlier they depend on the fundamental Wilson lines through the product $|W_f(C)|^2$. Since the Abelian part relating to the holonomy of $\mathcal{A}$ cancels when the square is taken, the adjoint Wilson loops don't see the $\mathcal{A}$ field and are consequently gauge-invariant well-defined line operators. Note that similar things like this are discussed in Tong's gauge theory notes, but unlike in his notes we are saying that the line operators in the $PSU(N)$ theory are the adjoint representation Wilson lines, not N-fold powers of the fundamental line. N-fold powers don't get you anything since you still need a surface operator attached to the line to ensure gauge invariance[21]. Also e.g. for $SU(3)$ we have

$$3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1. \tag{231}$$

Since 10 cannot be built from $\otimes$s of the adjoint 8, the RHS is not invariant under the $\mathbb{Z}_N$ 1-form symmetry and thus taking three fundamental Wilson lines is not quite the right way to get something invariant under the 1-form symmetry.

Now, are the global symmetries in the $PSU(N)$ theory right? As we've seen before, a pure $\mathbb{Z}_N$ 2-form gauge field comes with a $\mathbb{Z}_N$ 2-form electric symmetry and a $\mathbb{Z}_N$ $(D-2-1)$-form magnetic symmetry. For us $D = 4$ so the magnetic symmetry is a 1-form symmetry. The 2-form electric symmetry sends $B \mapsto B+\gamma$ for a flat 2-form γ with periods in $\frac{2\pi}{N}\mathbb{Z}$. This symmetry is broken by the coupling to the $U(N)$ field $\mathcal{A}$, which is good since the $PSU(N)$ theory shouldn't have any 2-form symmetries. However, in keeping with our discussion the other day about $\widetilde{G}$ and $\widetilde{G}/\Gamma_G$ gauge theories, we know that the $PSU(N)$ theory should have a $\mathbb{Z}_N$'s worth of t' Hooft operators, which are the charged objects for a $\mathbb{Z}_N$ 1-form symmetry. This is precisely the 1-form magnetic symmetry of the B field. The t' Hooft loop is constructed with $\widetilde{A}$, where $F_{\widetilde{A}} = H$ (writing H like this is possible since we took H to be a closed 2-form with periods in $2\pi\mathbb{Z}$). So the t' Hooft operator is

$$T(C) = \exp\left(i \oint_C \widetilde{A}\right). \tag{232}$$

The last type of operator we have is

$$\mathcal{W}(\Sigma) = \exp\left(i \int_\Sigma B\right), \tag{233}$$

where Σ is closed. From the commutation relation between B and $\widetilde{A}$ (roughly $[\widetilde{A}, B] = i/N$), we see that $\mathcal{W}(\Sigma)$ is the charge operator for the magnetic 1-form symmetry, and that it has the correct commutation relations with the fundamental Wilson line $W_f(C)$ (the nontrivial commutation relation is due to the B surface operator attached to $W_f(C)$). It thus computes the integral of ω_2 over Σ, where $\omega_2 \in H^2(M;\mathbb{Z}_N)$ is the 2nd Stiefel-Whitney class for the $PSU(N)$ bundle restricted to Σ.

Now given this, how would we get back to the $SU(N)$ theory? We would need to gauge the $\mathbb{Z}_N$ 1-form magnetic symmetry with a $\mathbb{Z}_N$ 2-form gauge field, whose magnetic 1-form

[21] The surface operator is now $e^{-iN\int_\Sigma B}$. If Σ is closed then this is always equal to 1, and so the attached surface operator is topological (independent of the exact choice of Σ. But the surface operator still needs to be there, and if C does not bound a surface then the N-fold power of $W_f(C)$ can't be defined in a gauge-invariant way.

symmetry would become the electric symmetry of the $SU(N)$ theory. This new gauge field needs to have the effect of forcing $F_\mathcal{A}$ to be quantized in periods of $2\pi N\mathbb{Z}$, since as we saw this turns the $PSU(N)$ transition functions into $SU(N)$ transition functions. So we can add some fields with the action

$$\frac{i}{2\pi}\int H'\wedge(F_\mathcal{A}-NF_{\mathcal{A}'}), \tag{234}$$

where H' is yet another Lagrange multiplier 2-form gauge field and $\mathcal{A}'$ is a $U(1)$ gauge field. The periods of H are not quantized, so that integrating out H sets the cohomology classes of $F_\mathcal{A}$ and $NF_{\mathcal{A}'}$ to be equal. This makes $F_\mathcal{A}$ pure gauge under the original 1-form $\mathbb{Z}_N$ gauge symmetry and so $\mathcal{A}$ and B can be eliminated from the action, leaving us with an $SU(N)$ field. The electric symmetry of $SU(N)$ in this presentation is the 1-form symmetry that shifts $\mathcal{A}'$ by a closed form that has periods in $\frac{2\pi}{N}\mathbb{Z}$ (again, the fact that this is the symmetry can be most easily seen by integrating by parts and writing the relevent term in the action as $\frac{iN}{2\pi}\int F_{H'}\wedge\mathcal{A}'$, and recalling that H' is a $U(1)$ 2-form gauge field so that $F_{H'}$ has periods in $2\pi\mathbb{Z}$). The electric 2-form symmetry of the added $\mathbb{Z}_N$ gauge field that we used to get back to $SU(N)$ is broken by the coupling to $F_\mathcal{A}$, and so all the global symmetries are properly accounted for.

So, summarizing, we've seen that $SU(N)$ has an electric $\mathbb{Z}_N^{(1)}$ symmetry, while $PSU(N)$ as a magnetic $\mathbb{Z}_N^{(1)}$ symmetry, and that these two symmetries can be related to one another through a gauging procedure. We know that the electric symmetry in $SU(N)$ is only $\mathbb{Z}_N$ since we always have gluons in the adjoint representation, so Wilson lines in the adjoint can end and hence can't carry a 1-form charge. How do we know that we haven't missed e.g. a magnetic symmetry in $SU(N)$? One way to argue is to say that we can couple the theory to a Higgs field in the adjoint: this leads to dynamical t'Hooft Polyakov monopoles as we have seen in an earlier entry, which carry magnetic charge N in the magnetic lattice. Since adding the Higgs field doesn't break any symmetries, the symmetries of the Higgsed theory should be the same as the un-Higgsed one, and so the un-Higgsed one can't have any magnetic 1-form symmetries. Actually, apparently the pure glue theory actually does have dynamical charge N monopoles, created at the intersection of N Wilson lines (should find the reference for this).

We now briefly cover θ angles. After gauging to $PSU(N)$, the θ term is (remember that the $1/8\pi^2$ comes from expanding $\exp(F/2\pi)$)

$$S\supset\frac{i\theta}{8\pi^2}\int\mathrm{Tr}\left[(F_\mathcal{A}-B\mathbf{1})\wedge(F_\mathcal{A}-B\mathbf{1})\right]. \tag{235}$$

Since F_A is traceless, this is

$$S\supset\frac{i\theta}{8\pi^2}\int\mathrm{Tr}[F_\mathcal{A}\wedge F_\mathcal{A}]+\frac{i\theta N}{8\pi^2}\int B\wedge B-\frac{i\theta}{4\pi^2}\int F_\mathcal{A}\wedge B. \tag{236}$$

Since $\mathrm{Tr}[F_\mathcal{A}]=F_\mathcal{A}$, we can write this as

$$S\supset i\theta c_2+\frac{i\theta N}{8\pi^2}\int B\wedge B-\frac{i\theta}{4\pi^2}\int F_\mathcal{A}\wedge B+\frac{i\theta}{8\pi^2}\int F_\mathcal{A}\wedge F_\mathcal{A}. \tag{237}$$

Here the second Chern class is

$$c_2 = \frac{1}{2(2\pi)^2}\left(\mathrm{Tr}[F_{\mathcal{A}} \wedge F_{\mathcal{A}}] - \mathrm{Tr}[F_{\mathcal{A}}] \wedge \mathrm{Tr}[F_{\mathcal{A}}]\right), \tag{238}$$

and is always integral. Recall that we get this from the second-order contribution to $\det(\mathbf{1} + F_{\mathcal{A}}/2\pi)$, which we compute as $\exp(\mathrm{Tr}\ln(\mathbf{1} + F_{\mathcal{A}}/2\pi))$ by using the Taylor series for the log. Using the constraint from integrating out H, we can eliminate $\mathcal{A}$ and write everything in terms of B (we do this by writing $F_{\mathcal{A}} = NB$. This is not the same as writing $B = N^{-1}F_{\mathcal{A}}$, which is not a replacement that we are making):

$$S_\theta = i\theta \int c_2 - (N^2 - N)\frac{i\theta}{8\pi^2}\int B \wedge B, \tag{239}$$

where now B is constrained to have periods in $2\pi/N$ around all 2-submanifolds. We see that shifting θ by 2π is now nontrivial: where we dropped $2\pi i \int c_2 \in 2\pi\mathbb{Z}$. The N term is certainly non-trivial because of the quantization on B, while the $+1$ part is in $2\pi\mathbb{Z}$ if we are working on a spin manifold. If we are not on a spin manifold, this factor can contribute a $\pm$ sign to the path integral[22].

The $\int B \wedge B$ term here prompts us to go back to our original action and add a counter-term of this form. We have several options for counterterms to add, given by an integer p. The counter-term we add is

$$S_{ct} = \frac{ipN}{4\pi}\int B \wedge B. \tag{244}$$

[22]The claim is that on a spin manifold M,

$$\int_M \frac{F}{2\pi} \wedge \frac{F}{2\pi} \in 2\mathbb{Z}, \tag{240}$$

for any 2-form $F \in 2\pi H^2(M;\mathbb{Z})$ while on a non-spin manifold, the RHS is replaced by $\mathbb{Z}$. Why does the (non)admitance of a spin structure determine how gauge fields integrate? This is because if a four-manifold M is spin, the intersection form

$$H^2(M;\mathbb{Z}) \times H^2(M;\mathbb{Z}) \to \mathbb{Z} \tag{241}$$

is even. This is because mod 2 we have, for any $A \in H^2(M;\mathbb{Z})$,

$$A \frown A = \omega_2 \frown A \qquad \mod 2, \tag{242}$$

where ω_2 is the second Stiefel Whitney class. Thus if M is spin then ω_2 lifts to an even class in $H^2(M;\mathbb{Z})$, and so since $(F/2\pi) \in H^2(M;\mathbb{Z})$, the integral above must indeed be even if M is spin.

To prove the last equation, one notes that $\omega_2 \frown A = \omega_2(TM|_A)$, i.e. $\omega_2 \frown A$ meansures the Stiefel-Whitney class of A embedded in M (using Poincare duality to think of A as a 2-manifold). Now $TM|_A = TA \oplus TN$, where TN is the component of the tangent bundle of M normal to A. Then since A is an orientable manifold, the Whitney product formula reads

$$\omega_2(TM|_A) = \omega_2(TA) + \omega_2(TN). \tag{243}$$

The first term on the RHS is the mod 2 Euler class of A, which is trivial for A a 2-manifold, since $\chi(A) = 2 - 2g \in 2\mathbb{Z}$. The second term is the mod 2 Euler class of TN, which is precisely the self-intersection number of A mod 2. This is beacuse $\omega_2(TN)$ measures the zeros of vector fields in TN mod 2, and the zeros in TN precisely come from self-intersection points of A (the intersection at each self-intersection point can be made transverse, so at these points the tangent space of A generates the full tangent space of M; hence any vector field in TN must vanish at these points).

Under a gauge transformation that shifts B by F_λ (the curvature of a $U(1)$ gauge field), this counter-term changes by

$$\delta S_{ct} = ipN\pi \int \frac{F_\lambda}{2\pi} \wedge \frac{F_\lambda}{2\pi} + \frac{ipN}{2\pi} \int B \wedge F_\lambda. \tag{245}$$

The first term is in $2\pi\mathbb{Z}$ if we are on a spin manifold, but is non-trivial in general. There is no other term in the action that will give us a quadratic term in F_λ, so in order for S_{ct} to make sense we must have $pN \in 2\mathbb{Z}$ (if we are on a spin manifold, just $pN \in \mathbb{Z}$ is okay).

The second $B \wedge F_\lambda$ term is trivial if we use the constraint that B has periods in $\frac{2\pi}{N}\mathbb{Z}$ coming from integrating out H, since then the term is valued in $2\pi p\mathbb{Z}$. Integrating out H gets rid of the operator responsible for the t' Hooft lines though (recall $H = F_{\widetilde{A}}$), so it would be better if we could keep H around (we want to talk about what happens to the t' Hooft lines when θ is shifted, which is tricky to do if we have to integrate it out to insure gauge invariance. I admit this argument is a little shaky). Basically, we want to ensure that the action is gauge invariant without needing to integrate out any of the Lagrange multipliers. We see that we can cancel the second term directly if we take

$$H = F_{\widetilde{A}} \mapsto F_{\widetilde{A}} - pF_\lambda \tag{246}$$

under the 1-form gauge transformation. This adds an extra term $ipN \int F_{\mathcal{A}} \wedge F_\lambda/2\pi$, but this is trivial since both $F_{\mathcal{A}}$ and F_λ are quantized in $2\pi\mathbb{Z}$ (by definition of $\mathcal{A}$ and λ, not because of Lagrange multipliers). Finally, note that if we take $p \mapsto p + 2N$ then the action changes by $\frac{iN^2}{2\pi} \int B \wedge B$, which is trivial (after integrating out H). So we have the periodicity $p \sim p + 2N$ (for generic parity N). On a spin manifold, or if N is even, we can do better and write $p \sim p + N$. This is the same as the fact that in Abelian CS theory with odd level, there are $2k$ particles with the kth one a fermion, while for k even the periodicity is smaller and there are only k particles. Adding this counter-term means that doing $\theta \mapsto \theta + 2\pi$ is the same as doing $p \mapsto p + 1 - N$ (if spacetime is spin then it is just $p \mapsto p + 1$).

Let's now check that the line operators we predicted to exist in the last diary entry are indeed the ones that are realized. As expected, for $\theta \neq 0$, a pure charge-1 t' Hooft line doesn't exist, since $e^{i\oint \widetilde{A}}$ is not gauge invariant (recall that we made $F_{\widetilde{A}} \mapsto F_{\widetilde{A}} - pF_\lambda$ under the 1-form gauge symmetry). However, we see that its gauge variance is exactly canceled by the gauge variance of p copies of the antifundamental Wilson line (since $\delta F_{\mathcal{A}} = NF_\lambda$ under the 1-form gauge transformation), so that the t' Hooft line becomes

$$T(C;p) = e^{\frac{i}{N}\oint_C \widetilde{A}} \mathrm{Tr}\left[\exp\left(-ip\oint_C \mathcal{A}\right)\right], \tag{247}$$

demonstrating that shifting the θ angle by 2π attaches electric charge 1 (really, -1) to the magnetic lines. By changing θ and thus changing p, we can realize the full range of $PSU(N)$ charge sublattices that we found in the previous diary entry, generated by the lines $(k, 1)$ for $k \in \mathbb{Z}_N$ the electric charge of the minimal magnetically charged line.

Before we end, we briefly mention that some quotient theories have charge lattices that are not related by shifting θ. To find examples, we need Γ_G to be a proper subgroup of the center. The simplest example where we can quotient by a subgroup of the center is for $G = SU(4)/\mathbb{Z}_2$. Recall that the charge lattice is given by the exact sequence

$$1 \to \Gamma_G^\vee \to L \to \Gamma_G \to 1. \tag{248}$$

For this example we have $\Gamma_G = \mathbb{Z}_2$ (and hence $\Gamma_G^\vee = \mathbb{Z}_2$), and so there are two possible choices for L: the split extension $L = \mathbb{Z}_2 \rtimes \mathbb{Z}_2 \cong \mathbb{Z}_2^2$, or the non-trivial extension in $H^2(\mathbb{Z}_2; \mathbb{Z}_2)$, namely $L = \mathbb{Z}_4$. In the former case we have two generators for the charge lattice, namely $(0, 2)$ and $(2, 0)$. In the later case, we have a single generator with magnetic charge 2, namely $(1, 2)$. Changing θ can only relate theories that differ in the way in which they define the electric charge of the generator(s) of the charge lattice; it cannot relate two theories with cohomologically distinct extensions L. Indeed, we see that the higher symmetries in the problem are not given by two 1-form symmetries $\Gamma_G^\vee \times \Gamma_G$, but rather are determined by L. In the case of $L = \mathbb{Z}_4$, there is only a single mixed electromagnetic 1-form symmetry, while for $L = \mathbb{Z}_2^2$ we have both electric and magnetic 1-form $\mathbb{Z}_2$ symmetries. This gives us an understanding of why the two $SU(4)/\mathbb{Z}_2$ theories are not related by a shift in a θ angle: they have different global symmetries.

14 *All about (fractional) instanton numbers* ✓

Today's entry is a collection of things about instantons and their normalizations in different gauge groups that I thought would be handy to have around as a reference.

14.1 Mathematical prelude

We begin this diary entry with definitions and useful math facts. For us, the instanton number I for a gauge bundle E over a four-dimensional spacetime X will be defined as

$$I = \int_X \mathrm{ch}_2(E), \tag{249}$$

where $\mathrm{ch}_2(E)$ is the second chern character. Recall that the Chern characters are obtained in the context of Chern-Weil theory from the expansion of $\mathrm{Tr}[e^{F/2\pi}]$ as

$$\mathrm{ch}_k(E) = \frac{1}{k!}\mathrm{Tr}\left[(F/2\pi)^{\wedge k}\right]. \tag{250}$$

Note that the Chern characters involve only a single trace, unlike the Chern classes.

We define I as the integral of $\mathrm{ch}_2(E)$ and not of $\mathrm{c}_2(E)$ (the second chern class), since we want I to be nonzero when we choose the gauge group to be $U(1)$ (and also, since it is the chern character, not the chern class, that appears in the index formula). For gauge groups like $SU(N)$ with traceless generators, $\mathrm{c}_2(E)$ and $\mathrm{ch}_2(E)$ are *almost* the same. The difference comes from torsion phenomena: the Chern classes can have torsion contributions, so that even when $\mathrm{Tr}[F] = 0$, we can have e.g. $c_1 \neq 0$, provided that c_1 is pure torsion. This can happen when the gauge group is the quotient of some simply connected group, like in the case of $SU(N)/\mathbb{Z}_N$. These torsionful contributions are ignored by Chern-Weil theory, but are important to keep track of.

In contrast, the Chern characters are defined as classes in $H^*(X;\mathbb{Q})$[23], and as such *never* have any torsionful elements. They are calculated purely from the local curvature, and are only sensitive to data about the gauge group's Lie algebra (whereas the Chern classes care about the full Lie group).

An important difference between the Chern characters and the Chern classes is that Chern classes always integrate to integers. The example relevant to us is that the integral of c_2 is an integer on any manifold, spin or otherwise. In contrast, the integral of ch_2 is not generically integral on a non-spin manifold, since the intersection form on a non-spin manifold is not for sure even. Thus we should remember that the Chern classes are good $\mathbb{Z}$ characteristic classes, while the Chern characters are not.

The chern characters satisfy

$$\mathrm{ch}(E \otimes F) = \mathrm{ch}(E) \wedge \mathrm{ch}(F), \qquad \mathrm{ch}(E \oplus F) = \mathrm{ch}(E) + \mathrm{ch}(F). \tag{251}$$

The former can be seen by plugging in $F_{E\otimes F} = \mathbf{1}_G \otimes F_H + F_G \otimes \mathbf{1}_H$ into the formulae for the chern characters, while the latter is straightforward to see since the chern characters involve only a single trace. On the other hand, the Chern class of the direct sum is the wedge of the Chern classes, instead of the sum:

$$\mathrm{c}(E \oplus F) = \mathrm{c}(E) \wedge \mathrm{c}(F). \tag{252}$$

This is the Whitney sum formula and can be seen from the definition of the Chern classes in terms of the expansion of $\det(\mathbf{1} + F_A/2\pi)$, and the fact that $\det(A \oplus B) = \det(A)\det(B)$. I'm unaware of any simple formula for $\mathrm{c}(E \otimes F)$, unless $E \cong \bigoplus_i \mathcal{L}_i, F \cong \bigoplus_j \mathcal{L}'_j$ for line bundles $\mathcal{L}_i, \mathcal{L}'_j$. In that case, we have

$$\mathrm{c}(E \otimes F) = \mathrm{c}\left(\bigoplus_{i,j} \mathcal{L}_i \otimes \mathcal{L}'_j\right) = \bigwedge_{i,j} \left(1 + \mathrm{c}_1(\mathcal{L}_i \otimes \mathcal{L}'_j)\right) = \bigwedge_{i,j} (1 + \mathrm{c}_1(\mathcal{L}_i) + \mathrm{c}_1(\mathcal{L}'_j)), \tag{253}$$

where we used the Whitney sum formula and the fact that $\mathrm{c}_1(\mathcal{L} \otimes \mathcal{L}') = \mathrm{c}_1(\mathcal{L}) + \mathrm{c}_1(\mathcal{L}')$. This can be seen by recalling that the first Chern class can be defined by the Euler class of the underlying real bundle. Since the expression for the Euler class involves the log of

[23] The rational coefficients here are simply because the mod 1 contributions to the Chern characters come from the coefficients in the expansion of the exponential. Since the cohomology groups are isomorphic, we may equivalently just write $H^*(X;\mathbb{R})$.

the transition functions, and since the transition functions of $\mathcal{L} \otimes \mathcal{L}'$ are the product of the transition functions for $\mathcal{L}$ and $\mathcal{L}'$, the Euler class of $\mathcal{L}\otimes\mathcal{L}'$ splits as a sum of the Euler classes of each line bundle—hence $\mathrm{c}_1(\mathcal{L} \otimes \mathcal{L}') = \mathrm{c}_1(\mathcal{L}) + \mathrm{c}_1(\mathcal{L}')$.

Something else we sometimes need to do is to determine characteristic classes / instanton numbers for product bundles $E \otimes F$, where E is a principal G-bundle and F is a principal H-bundle. The answer for the instanton number is what you would expect: for theories not involving a $U(1)$ factor so that their Lie algebra generators are traceless, we have

$$I_{E\otimes F} = \mathrm{ch}_2(E \otimes F) = I_G \dim H + I_H \dim G, \tag{254}$$

where the dimension means the dimension of the defining representaiton of the associated Lie algebras. This follows from $\mathrm{ch}(E \otimes F) = \mathrm{ch}(E) \wedge \mathrm{ch}(F)$: taking the second order terms, we have

$$\mathrm{ch}_2(E \otimes F) = \mathrm{Tr}_G[\mathbf{1}_G]\frac{1}{8\pi^2}\mathrm{Tr}_H\left[F_H \wedge F_H\right] + \mathrm{Tr}_H[\mathbf{1}_H]\frac{1}{8\pi^2}\mathrm{Tr}_G\left[F_G \wedge F_G\right], \tag{255}$$

which gives us what we want.

14.2 Instantons

Normal instantons come from transitions between pure gauge field configurations in different homotopy classes of $\pi_3(G)$, where the 3 in $\pi_3(G)$ is a spatial slice (or region thereof) where the gauge fields asymptote to a constant (the elements in $\pi_3(G)$ are the glueing data for nontrivial bundles on S^4) . These instantons can live in any $\mathbb{R}^4$-like region of a given 4-manifold, regardless of its topology. Furthermore they will exist for all choices of (non-Abelian) gauge groups, since $\pi_3(G) = \mathbb{Z}$ for all simple compact non-Abelian Lie groups G. These instantons are common to all gauge groups G that descend from some simply connected group $\widetilde{G}$ by quotienting by some finite Γ_G (which may be $\mathbb{Z}_1$). To show this, one uses the long exact sequence coming from $1 \to \Gamma_G \to \widetilde{G} \to G \to 1$. This sequence contains

$$\ldots \to \pi_4(\widetilde{G}/G) \to \pi_3(G) \to \pi_3(\widetilde{G}) \to \pi_3(\widetilde{G}/G) \to \cdots . \tag{256}$$

Now $\pi_{k>0}(\widetilde{G}/G) = \pi_{k>0}(\Gamma_G) = 0$ since Γ_G is discrete and the homotopy groups are basepoint preserving (the basepoint is fixed to be a given element of the target space for the definition of the homotopy group, so we don't get a $|\Gamma_G|$'s worth of constant maps, we just get a $\mathbb{Z}_1$'s worth). Thus we have an isomorphism $\pi_3(G) \cong \pi_3(\widetilde{G})$, and so the "small" instantons associated with $\pi_3(G)$ have the same instanton number no matter what Γ_G is.

As in [**?**], we will normalize the instanton number so that the minimal "small" instanton has instanton number $I = 1$. This minimal small instanton can always be taken to be a minimal $SU(2)$ instanton, on an S^3 around which $A \sim U^\dagger dU, U \sim e^{ix^aT^a}$, for an appropriately chosen trio of generators $T^z, T^+, T^- \in \{T^a\}$, with T^z, T^+, T^- generating an $\mathfrak{su}(2)$ Lie algebra. Recall that this embedding of $\mathfrak{su}(2)$ can always be done: we pick a pair of roots T^+, T^- that are eigenvalues under the action of Ad_A where A is such that Ad_A has maximal kernel, and then from these generators we can construct a T^z in the Cartan subalgebra of $\mathfrak{g}$ that together with the $T^\pm$ generates an $\mathfrak{su}(2)$. Thus for all choices of (compact, simple) Lie group G, we can always embed an $SU(2)$ instanton through a choice of $\mathfrak{su}(2) \to \mathfrak{g}$. This induces a map

$SU(2) \to G$, and the normalization of the instanton number depends on the index of this map.

One foolproof way to find the normalization for the instanton number is to compute the instanton number by requiring that for a minimal small instanton field configuration F, we have

$$1 = \frac{1}{N_{\mathfrak{g}}} \int \mathrm{Tr}_{Ad_{\mathfrak{g}}} \left(\frac{F}{2\pi} \wedge \frac{F}{2\pi} \right). \tag{257}$$

Here $\mathrm{Tr}_{Ad_{\mathfrak{g}}}$ is taken in the adjoint representation of $\widetilde{G}$, which is always a representation for all $G = \widetilde{G}/\Gamma_G$, and $N_{\mathfrak{g}}$ is a normalization constant that fixes the equality. For example, consider $SU(2)$. We know the bundle E with a minimal $SU(2)$ instanton is such that

$$1 = \int \mathrm{c}_1(E) = \frac{1}{2} \int \mathrm{Tr}_f \left(\frac{F}{2\pi} \wedge \frac{F}{2\pi} \right). \tag{258}$$

In the fundamental, we have the normalization

$$\mathrm{Tr}(T_f^a T_f^b) = \frac{\delta^{ab}}{2}. \tag{259}$$

On the other hand, in the adjoint we have

$$\mathrm{Tr}(T_{Ad}^a T_{Ad}^b) = N\delta^{ab}, \tag{260}$$

so that if we take the minimal $SU(2)$ instanton F^a but change the representation to the adjoint, the answer changes by a factor of $2N$. Thus we have

$$N_{\mathfrak{su}(N)} = 4N. \tag{261}$$

When Γ_G is nontrivial we can have "large" instantons that contribute to the instanton number I but which make rational contributions to I instead of integral contributions. This is because if the topology of spacetime is nontrivial, we can have G bundles which are not $\widetilde{G}$ bundles. This is not as contrived a scenario as it seems, since nontrivial spacetime topologies can be created by inserting t' Hooft operators with nontrivial t' Hooft flux, which exist if $\pi_1(G) = \Gamma_G$ is nontrivial. To visualize the types of processes that give nonzero "large" instanton number, we can think about the $U(1)$ case, where our intuition is aided by the fact that the integral in $\int \mathrm{ch}_2(E)$ can be interpreted as a self-intersection number of the Poincare dual of F (such an interpretation is only possible in the non-Abelian case if $E = \bigoplus_i \mathcal{L}_i$ is a direct sum of line bundles so that F is diagonal). For example, if we consider a process in which two initially separated magnetic flux loops pass through each other to form a Hopf link and then later re-separate, then the self-intersection number of $\widehat{F}/2\pi$ is 2, and we get $I = 1$.

An important comment is that the fractional part of the instanton number is due entirely to w_2, as noted in [?] (temporarily assuming that the gauge group is semisimple; $U(1)$ factors can of course give fractional contributions not associated with a torsionful class). That is, we have small instantons, which give integral contributions to the instanton number, and large instantons caused by w_2 which can be fractional, and no other types of instantons.

We see this as follows: to examine different ways to construct a $G = \widetilde{G}/\Gamma_G$ bundle, we start from a trivialization over the 1-skeleton of X. This is always possible for orientable X and connected G. Then, we try to extend this bundle over the 2-skeleton. If $\pi_1(G) = \Gamma_G$ is nontrivial, this may not be possible: if the trivialization on the 1-skeleton winds by an element of $\pi_1(G)$ along the boundary of a given 2-cell, the trivialization is not extendable into that 2-cell (this is a global obstruction if the product of all such holonomies in $\pi_1(G)$ accross all 2-cells is nontrivial). 2-cells where there is an obstruction to extending the trivialization are determined by w_2. Now we try to extend the trivialization over the 3-cells. This may be obstructed by an element in $\pi_2(G)$. However, since G is a topological group, we have the magical fact that $\pi_2(G) = \mathbb{Z}_1$, and so there is no obstruction at this level. Finally we try to extend into the 4-skeleton: this is obstructed by $\pi_3(G)$. But as we have seen $\pi_3(G) = \pi_3(\widetilde{G})$ parametrizes the "small" instantons, which are the same for both G and its universal cover, and so the contribution of $\pi_3(G)$ elements to I is always integral in our normalization. Thus the only possible contribution to the fractional part of I is w_2.

14.3 Examples

Now we will compute examples for a few classes of groups of interest.

As we mentioned above, if our gauge group is simply connected it can have no "large" instantons, and we will never have fractional instanton numbers. Fractional values of I ("large instantons") can arise from two things:

- $U(1)$ factors in the gauge group corresponding to $\mathbb{Z}$ factors in $\pi_1(G)$ (this is no surprise; we know that $I \in \frac{1}{2}\mathbb{Z}$ for $U(1)$ theories on general 4-manifolds)
- G bundles that cannot be lifted to $\widetilde{G}$ bundles because of a discrete obstruction, corresponding to torsion factors in $\pi_1(G)$.

In what follows we will derive expressions for the fractional part of the instanton number I for various gauge groups, written in terms of the discrete characteristic class w_2 that describes the gauge bundle (as well as the $U(1)$ part of the field strength, if it exists). We emphasize that the instanton number is never actually coming from doing a calculation with any discrete objects—the instanton number is the (integral of the) second Chern character, which never has any torsionful contributions. However, as we will see, the choice of the characteristic class w_2 puts constraints on what the instanton number can be, and so that the fractional part of I is indeed a function of w_2. This function will always turn out to be the Pontryagin square $\mathcal{P}(w_2)$.

We will be interested in the question of whether or not $\mathcal{P}(w_2)$ constitutes a topological action that is linearly independent from the instanton number. Recall that for $w_2 \in H^2(X;\mathbb{Z}_n)$,[24] we have $\mathcal{P}(w_2) \in H^4(X;\mathbb{Z}_k)$, where $k = 2n$ if n is even, and $k = n$

[24] As elsewhere in the diary, we are being slightly imprecise as writing $w_2 \in H^2(X;\mathbb{Z}_n)$. In reality w_2 is a characteristic class $w_2 : b_G(X) \to H^2(X;\mathbb{Z}_n)$, i.e. a map from the isomorphism classes of G-bundles over X to the cohomology on X, in this case with $\mathbb{Z}_n$ coefficients. When we write something like $w_2 \in H^2(X;\mathbb{Z}_n)$, we are using w_2 as a standin for the cohomology class one gets when evaluating w_2 on the gauge bundle E, which is left implicit in the notation.

if n is odd. We may therefore write the topological terms

$$S_{top}/2\pi \supset \frac{p}{2n}\int P(w_2), \tag{262}$$

where $p \in \mathbb{Z}_{2n}$ if $n \in 2\mathbb{Z}$, and $p \in 2\mathbb{Z}_{2n}$ if $n \in 2\mathbb{Z}+1$.

Suppose the fractional part of the instanton number which depends on $P(w_2)$ is $(q/2n)\int P(w_2)$, so that the part of S_{top} involving w_2 is

$$S_{top} = \theta\left(\frac{q}{2n}\int P(w_2) + \dots\right) + \frac{p}{2n}\int P(w_2), \tag{263}$$

where $\dots$ are the other contributions to the instanton number. If we can equivalently write this as

$$S_{top} = (\theta + \delta\theta)\left(\frac{q}{2n}\int P(w_2) + \dots\right), \tag{264}$$

where $\delta\theta \in 2\pi\mathbb{Z}$ so as not to affect the integer part of I, then the discrete $P(w_2)$ class will not be independent from the instanton number. This is important because again, despite appearances, the stuff within the ()s in the above equation comes from integrating a local density, allowing the θ angle to take on a continuum of values.[25] We will be able to write S_{top} as (264) for any p provided that q generates all of $\mathbb{Z}_{2n}$ or $2\mathbb{Z}_{2n}$, depending on the parity of n. The condition for this to happen is that

$$\gcd(q, 2n) = \begin{cases} 1 & n \in 2\mathbb{Z} \\ 2 & n \in 2\mathbb{Z}+1 \end{cases}. \tag{265}$$

When this condition is satisfied, there is no independent torsionful characteristic class that we can add to the topological action, and hence the whole topological action will appear with a continuously tunable coefficient θ.

14.4 $SU(N)$ and $PSU(N)$

We will now specialize to the case where $G = PSU(N)$. The degree to which a given $PSU(N)$ bundle E does not lift to an $SU(N)$ bundle is determined by a class

$$w_2(E) \in H^2(X; \mathbb{Z}_N), \tag{266}$$

where X is spacetime. We can construct E by taking an $SU(2)$ bundle $\widetilde{E}$ and relaxing the cocycle condition on the transition functions to only hold modulo an Nth root of unity: $[g_{ij}g_{jk}g_{ki}]_{ab} = \delta_{ab}e^{2\pi i f_{ijk}/N}$, where the f_{ijk} are integers. The choice of f_{ijk} determines the $w_2(E)$ class, which when integrated over a given closed 2-submanifold tells us the fractional flux passing through that manifold.

A naive first guess would be that the instanton number for $PSU(N)$ bundles can be (in our normalization) an element of $\frac{1}{N}\mathbb{Z}$. This is because if E is a $PSU(N)$ bundle, then

[25] The term in ()s is well defined since a shift in $P(w_2)$ by a 4-form valued with periods in $2n\mathbb{Z}$ can be compensated for by a shift in the integer-valued part of I (the part coming from small instantons).

$E^{\otimes N}$ is a bundle whose transition functions those of an $SU(N)$ bundle, since the transition functions of $E^{\otimes N}$ are N-fold $\otimes$s of the transition functions for E, which ensures that the cocycle condition holds exactly in $E^{\otimes N}$ (i.e., not just up to an Nth root of unity). The instanton number for $E^{\otimes N}$ is found from

$$\mathrm{ch}(E^{\otimes N}) = \mathrm{ch}(E)^{\wedge N} = 1 + N\mathrm{ch}_2(E) + \dots, \tag{267}$$

where we used $\mathrm{ch}_1(E) = 0$ on account of the traclessness of the $SU(N)$ generators, and so one then might think that $\int \mathrm{ch}_2(E^{\otimes N}) \in \mathbb{Z}$ on account of its transition functions satisfying the cocycle condition. This is not quite true however, and in fact $\int \mathrm{ch}_2(E^{\otimes N}) \in \frac{1}{2}\mathbb{Z}$, a situation which is possible due to the fact that for line bundles, $I \in \frac{1}{2}\mathbb{Z}$ on non-spin manifolds (we will never be restricting to spin manifolds).[26] We'll see how this works in a second.

Now let us see how such a fractional instanton number can be realized. In what follows we will basically be working out in gory detail a computation described in [**?**] for the case of a spin manifold. The goal is to explicitly construct a $PSU(N)$ bundle that will get us the minimal possible I of $1/2N$.

First, let us fix a class w_2. Let $\mathcal{L}$ be the line bundle over X with first Chern class reducing to w_2 mod N:

$$w_2 = \mathrm{c}_1(\mathcal{L}) \mod N. \tag{268}$$

Here the LHS is viewed as an element in $H^2(X;\mathbb{Z})$, but we will usually use the correspondence between elements of $H^*_{dR}(X;\mathbb{R})$ with quantized periods and those in $H^*(X;\mathbb{Z})$ to think of it as an actual 2-form in the de Rham sense. From $\mathcal{L}$ we can form the bundle $\mathcal{L}^{-1/N}$, defined to have transition functions which are $1/N$th roots of the transition functions of $\mathcal{L}$. In particular, the cocycle conditions in $\mathcal{L}^{-1/N}$ are only satisfied up to Nth roots of unity. We can then construct a $PSU(N)$ bundle E as follows:

$$E = \mathcal{L}^{-1/N} \otimes \left(\mathcal{L} \oplus \bigoplus_{i=1}^{N-1} T_i \right), \tag{269}$$

where T_i is a trivial line bundle. The $\mathcal{L}^{-1/N}$ means that E is not an $SU(N)$ bundle. However, the $\mathcal{L}^{-1/N}$ factor does not turn the thing in parenthesis from an $SU(N)$ bundle into a $PSU(N)$ bundle, since the thing in the parenthesis is not an $SU(N)$ bundle: it has nonzero first Chern character, which precludes it from being an $SU(N)$ bundle. Indeed, (vector bundles associated to) $SU(N)$ principal bundles always have zero first Chern character, simply because $\mathrm{Tr}(F) = 0$ (the Chern characters never have torsion; they are defined totally within the context of Chern-Weil theory). Note that this doesn't necessarily mean that the first Chern *class* must vanish though, since the Chern classes can have torsionful contributions.

Anyway, if E is to be a $PSU(N)$ bundle then since at the Lie algebra level $PSU(N)$ and $SU(N)$ are identical, E must also have a first Chern character which vanishes. This indeed is true, and is the reason for the choice of powers of $\mathcal{L}$ appearing in E: we first use

$$\mathrm{ch}(E) = \mathrm{ch}(\mathcal{L}^{-1/N}) \wedge \left(\mathrm{ch}(\mathcal{L}) + \sum_{i=1}^{N-1} \mathrm{ch}(T_i) \right). \tag{270}$$

[26] A better but more mathematical way to say this would be to say that we can have instanton numbers for $E^{\otimes N}$ that are in $\frac{1}{2}\mathbb{Z}$ because of the existence of the Pontryagin square operation, which lets us consistently "divide" torsionful intersection numbers by two. See the diary entry on the Pontryagin square for details.

Taking the first degree component gives

$$\mathrm{ch}_1(E) = -\frac{1}{N}c_1(\mathcal{L}) \cdot N + 1 \cdot c_1(\mathcal{L}) = 0 \tag{271}$$

as required.

The construction of building E from "fractional" line bundles makes it clear that it is a $PSU(N)$ bundle. If $\lambda_{ij} = e^{i2\pi g_{ij}}$ are the transition functions for $\mathcal{L}$, then the transition functions for E are the matrices

$$\Lambda_{ij} = \mathrm{diag}(e^{i2\pi g_{ij}(1-\frac{1}{N})}, e^{-i2\pi g_{ij}/N}, \dots, e^{-i2\pi g_{ij}/N}). \tag{272}$$

Note that while we still have $\det(\Lambda_{ij}) = 1$, $\delta\Lambda$ is no longer trivial:

$$(\delta\Lambda)_{ijk} = e^{-2\pi i f_{ijk}/N}\mathbf{1}, \tag{273}$$

where the $f_{ijk} \in \mathbb{Z}$ are as before determined by the class w_2. This means that the Λ_{ij} are transition functions for a $PSU(N)$ bundle, but not for an $SU(N)$ bundle[27].

Now we will compute the instanton number of E, working modulo integral classes (i.e., just focusing on the fractional part). For some reason I chose to first do this by computing the second Chern class of E, which gives the instanton number since the second Chern class and second Chern character are equal in this case (the calculation of the Chern character is a little later on). We use the Whitney sum formula to write

$$\mathrm{c}(E) = (1 + \mathrm{c}_1(\mathcal{L}^{-1/N}) + \mathrm{c}_1(\mathcal{L})) \wedge \bigwedge_{i=1}^{N-1} (1 + \mathrm{c}_1(\mathcal{L}^{-1/N})). \tag{274}$$

Taking the degree-2 part, we have

$$\mathrm{c}_2(E) = \frac{N^2 - N}{2}\mathrm{c}_1(\mathcal{L}^{-1/N}) \wedge \mathrm{c}_1(\mathcal{L}^{-1/N}) + (N-1)\mathrm{c}_1(\mathcal{L}) \wedge \mathrm{c}_1(\mathcal{L}^{-1/N}). \tag{275}$$

Now the wedge product of the chern classes is, using the Pontryagin square to take the wedge product so as to properly count the self-intersections of the w_2 surface (see the diary on the Pontryagin square for more),

$$\mathrm{c}_1(\mathcal{L}^{-1/N}) \wedge \mathrm{c}_1(\mathcal{L}^{-1/N}) = \frac{1}{N^2}P(w_2). \tag{276}$$

Then we get

$$\mathrm{c}_2(E) = \frac{P(w_2)}{2}\left(1 - \frac{1}{N} - 2 + 2\frac{1}{N}\right) = -\frac{1}{2}\left(1 - \frac{1}{N}\right)P(w_2). \tag{277}$$

We can also do the computation by computing $\mathrm{ch}_2(E)$, which should agree with $\mathrm{c}_2(E)$ since $\mathrm{c}_1(E) = 0$. The calculation goes as follows:

$$\mathrm{ch}(E) = \mathrm{ch}(\mathcal{L}^{-1/N}) \wedge (\mathrm{ch}(\mathcal{L}) + N - 1). \tag{278}$$

[27]Here it is very important that $\delta\Lambda$ is a constant Nth root of unity times $\mathbf{1}$: having different Nth roots of unity along each entry of the diagonal would be no good, since only the diagonal $\mathbb{Z}_N$ is moded out by when passing to $PSU(N)$.

Since for line bundles $\mathrm{ch}_2(\mathcal{L}) = \frac{1}{2}\mathrm{c}_1(\mathcal{L}) \wedge \mathrm{c}_1(\mathcal{L})$ we have, working modulo terms that are integral classes,[28]

$$\begin{aligned}\mathrm{ch}_2(E) &= N\mathrm{ch}_2(\mathcal{L}^{-1/N}) + \mathrm{ch}_2(\mathcal{L}) + \mathrm{ch}_1(\mathcal{L}^{-1/N}) \wedge \mathrm{ch}_1(\mathcal{L}) \\ &= P(w_2)\left(\frac{N}{2N^2} + \frac{1}{2} - \frac{1}{N}\right) \\ &= \frac{(N-1)}{2N}P(w_2).\end{aligned} \tag{279}$$

These results tell us that the instanton number is valued in $\frac{1}{N}\mathbb{Z}$ on a spin manifold, and $\frac{1}{2N}\mathbb{Z}$ on a non-spin manifold. Something that's kind of interesting here is that as mentioned above, on a non-spin manifold we can get an instanton number l such that $NI \notin \mathbb{Z}$! Again, this is interesting because from the point of view of transition functions, we might be led to expect that the fractional part of the instanton number is always valued in $\frac{1}{N}\mathbb{Z}$, given that the transition functions of an $PSU(N)$ bundle can always be made into the transition functions for an $SU(N)$ bundle by raising them to their Nth powers. This becomes a little less surprising if we consider the rather dumb (since it's just an issue of normalization) example of $U(1)/\mathbb{Z}_N$: here the instanton number in the $U(1)/\mathbb{Z}_N$ theory is valued in $\frac{1}{2N^2}\mathbb{Z}$, even though the transition functions all only fail the cocycle conditions by Nth roots of unity. Looking at this example, we see that the reason for N copies of a $PSU(N)$ bundle not necessarily giving a $I \in \mathbb{Z}$ just boils down to the fact that I is nonlinear in the field strength and that the $PSU(N)$ bundles are built from line bundles, which can have fractional instanton numbers on non-spin manifolds.[29]

To be pedantically explicit, we can do the construction of the minimal $PSU(N)$ bundle for $SU(2)$. Let A_m be the $U(1)$ gauge field for a 2π monopole whose worldline wraps some nontrivial cycle in spacetime, and let

$$A_{SO(3)} = \begin{pmatrix} A_m/2 & 0 \\ 0 & -A_m/2 \end{pmatrix}. \tag{280}$$

Note that this has zero first Chern character as required,[30] and that it is constructed as a $\oplus$ of "fractional" line bundles: the total $SO(3)$ bundle is $\mathcal{L}^{1/2} \oplus \mathcal{L}^{-1/2}$, where $\mathcal{L}$ is a legit $U(1)$ line bundle. The instanton number for this field configuration is

$$I = \frac{1}{16\pi^2}\int F_{A_m} \wedge F_{A_m} \in \frac{1}{4}\mathbb{Z} \tag{281}$$

[28] We write $\mathrm{c}_1(\mathcal{L}) = w_2 + N\alpha$, where α has integral periods. We can then throw away a term $\frac{N^2-N}{2}w_2 \wedge \alpha$ in the second step since it is an integral class, on account of $(N^2 - N)/2 \in \mathbb{Z}$ for all N.

[29] It's also kind of interesting that we can get the minimal instanton number by working with a direct sum of line bundles regardless of the spin-ness of the background manifold, since for $SU(N)$ bundles this isn't the case: consider e.g. an $SU(2)$ bundle $E = \mathcal{L} \oplus \mathcal{L}^*$. Then $\mathrm{ch}_2(E) = 2\mathrm{ch}_2(\mathcal{L})$, which on a spin manifold is in $2\mathbb{Z}$, twice the minimum allowed value. This tells us that on a spin manifold the minimal instantons for $SU(N)$ are the ones that involve twisting around more than just two axes (and hence cannot be composed into a $\oplus$ of line bundles). This is then to be contrasted with the $PSU(N)$ case, where bundles formed from direct sums of line bundles can always give the minimal instanton number.

[30] However, its first Chern class is a nontrivial element in $\mathbb{Z}_2$ cohomology—the second SW class of the $SO(3)$ bundle—which is torsionful (recall that the second SW class for a complex vector bundle is the mod-2 reduction of that bundle's first Chern class).

on general manifolds, while it is in $\frac{1}{2}\mathbb{Z}$ on spin manifolds. This is in keeping with us being able to get strength $1/2N$ instantons for $PSU(N)$ gauge theories on a generic manifold, and strength $1/N$ instantons on spin manifolds. Note how this instanton picks out a particular direction in $SU(2)$ space, namely the σ^z direction, unlike the small instantons which wrap about the entire internal $SU(2)$ space.

A final way to see all of this is to write the $PSU(N)$ field in terms of a $U(N)$ field and a $\mathbb{Z}_N$ 2-form gauge field, as was done in the previous diary entry. Using the same notation as in that diary entry, we have

$$I = \frac{1}{8\pi^2}\int \mathrm{Tr}[(F_{\mathcal{A}} - B\mathbf{1}) \wedge (F_{\mathcal{A}} - B\mathbf{1})], \tag{282}$$

where $\mathcal{A} = A_{SU(N)} + \mathcal{A}\mathbf{1}/N$ is a $U(N)$ gauge field, and $\mathcal{A}$ is a properly quantized $U(1)$ gauge field, with $F_{\mathcal{A}} = NB$ enforced through a Lagrange multiplier constraint. Then

$$I = \mathrm{c}_2(E_{U(N)}) + \frac{i}{8\pi^2}\int (F_{\mathcal{A}} \wedge F_{\mathcal{A}} + NB\wedge B - 2F\wedge B) \to \mathrm{c}_2(E_{U(N)}) + \frac{i}{8\pi^2}(N^2 - N)\int B\wedge B, \tag{283}$$

where we used the Lagrange multiplier constraint. The first term, the second Chern class of the $U(N)$ bundle, is in $\mathbb{Z}$. However, the second term is in $\frac{1}{2N}\mathbb{Z}$, since $B/2\pi$ has periods in $1/N$. Thus the instanton number for $PSU(N)$ theories is valued in $\frac{1}{2N}\mathbb{Z}$. B here is the 2-form that measures w_2 of the bundle, and so as above, we see that the fractional part of the instanton number comes from the "large" instantons (the small ones are determined by the $\mathrm{c}_2(E_{U(N)})$ factor).

Finally we ask whether the instanton number I is linearly independent from the discrete torsionful term $p/2n\int P(w_2)$. To answer this we calculate

$$\gcd(N-1, 2N) = \gcd(N-1, 2), \tag{284}$$

which according to (265) means that in this case, the discrete class $P(w_2)$ not independent from the instanton number for any N.

14.5 $SU(N)/\mathbb{Z}_M$

The computation for this case is similar to the computation for $PSU(N)$, but it's not in the literature so it's worth doing (we will be very brief, though). We can assume that $M|N$ with $1 < M < N$ wolog. We form an $SU(N)/\mathbb{Z}_M$ bundle by

$$E_{SU(N)/\mathbb{Z}_M} = \mathcal{L}^{(1-N)/M} \oplus (\mathcal{L}^{1/M})^{\oplus(N-1)}, \tag{285}$$

which has zero first Chern character and has transition functions which fail the cocycle condition by powers of $e^{2\pi i/M}\mathbf{1}_N$, as required. The second Chern character is, after a little algebra,

$$\mathrm{ch}_2[E_{SU(N)/\mathbb{Z}_M}] = \frac{N(N-1)}{2M^2}\mathrm{c}_1(\mathcal{L})^{\wedge 2}, \tag{286}$$

where $\mathrm{c}_1(\mathcal{L})$ mod M reduces to w_2 of the bundle, which is a $\mathbb{Z}_M$ valued form because $\pi_1[SU(N)/\mathbb{Z}_M] = \mathbb{Z}_M$. This result is pretty easy to understand from the $PSU(N)$ case:

the Ns in the denominator come from the number of line bundles in the direct sum (which of course doesn't change when we change the group we're quotienting by), while the N^2 in the denominator changes to M^2 because the extent to which the constituent line bundles are allowed to be fractional changes when we change the quotient group to $\mathbb{Z}_M$. Therefore the instanton number is

$$I_{E_{SU(N)/\mathbb{Z}_M}} = \frac{2\pi N(N-1)}{2M^2} \int P(w_2) + \dots, \tag{287}$$

where $\dots$ are the integer parts.

The condition for the discrete class $P(w_2)$ to not be independent from the instanton number is

$$\gcd(N(N-1)/M, 2M) = \begin{cases} 1 & M \in 2\mathbb{Z} \\ 2 & M \in 2\mathbb{Z}+1 \end{cases}. \tag{288}$$

14.6 $U(N)/\mathbb{Z}_M$

In another diary entry, we show that

$$\pi_1[U(N)/\mathbb{Z}_M] = \mathbb{Z} \times \mathbb{Z}_g, \qquad g \equiv \gcd(N, M). \tag{289}$$

Therefore $Q \equiv U(N)/\mathbb{Z}_M$ bundles will be labeled by a $\mathbb{Z}_g$ characteristic class w_2, along with the regular $\mathbb{Z}$-valued class for the "small" instantons (and, since we have a $U(1)$ factor, a fractional contribution from large instantons in the $U(1)$ part of the group).

We now need to ask how we can form Q bundles which are not $U(N)$ bundles. We claim that we can always obtain the minimal instanton number with a bundle

$$E_Q(r,q) = \mathcal{L}_d^{r/M} \otimes \begin{pmatrix} \mathcal{L}_t^{q(1-N)/g} & & & \\ & \mathcal{L}_t^{q/g} & & \\ & & \ddots & \\ & & & \mathcal{L}_t^{q/g} \end{pmatrix}, \qquad r, q \in \mathbb{Z}. \tag{290}$$

Here the line bundle $\mathcal{L}_d$ (d for "diagonal") keeps track of the fractional $U(1)$ part of the instanton number, with the transition functions in $\mathcal{L}_d^{r/M}$ failing by powers of $\zeta_M^r \mathbf{1}_N$, where $\zeta_M = e^{2\pi i/M}$. The Chern character of $\mathcal{L}_d$ does *not* reduce to anything relating to the torsionful class w_2 when modded out by some integer: the element of $\pi_1[Q]$ defined by a triple patch overlap where the transition functions of $\mathcal{L}_d^{r/M} \otimes \mathbf{1}_N$ fail the cocycle condition by $\zeta_M^r \mathbf{1}_N$ is an element of the $\mathbb{Z}$ factor in $\pi_1[Q]$ (coming from topologically nontrivial maps $\det : U(N) \to S^1$), and hence is unrelated to torsionful w_2 class.

The dependence of the instanton number on the w_2 of the bundle $E_Q(r,s)$ is instead determined by the term involving $\mathcal{L}_t$ (t for "torsion"). $\mathcal{L}_t^{q/g}$ is a line bundle that fails the cocycle conditon by $\zeta_g^q \mathbf{1}_N$ on each triple overlap; this is allowed since ζ_g is a power of ζ_M. The structure of the direct sum of the $\mathcal{L}_t$s is such that around a triple patch overlap where the cocycle condition fails, one traces out a loop in the finite $\mathbb{Z}_g$ factor of $\pi_1[Q]$ (see the diary entry on $\pi_1[U(N)/\mathbb{Z}_M]$ for details). This means that the Chern class of $\mathcal{L}_t$ can be taken to reduce to w_2 mod g. Note that the factor involving the $\mathcal{L}_t$s gives us a minimal $SU(N)/\mathbb{Z}_g$

bundle when $q = 1$, which is what we expect from general considerations of how the quotient in Q acts on the $U(1)$ and $SU(N)$ factors.

Now we should actually compute the instanton number I. The total Chern character is

$$\mathrm{ch}(E_Q(r,q)) = (N-1)\mathrm{ch}(\mathcal{L}_d^{r/M} \otimes \mathcal{L}_t^{q/g}) + \mathrm{ch}(\mathcal{L}_d^{r/M} \otimes \mathcal{L}_t^{q(1-N)/g}). \tag{291}$$

Taking the degree 2 part and simplifying modulo integer terms, we get

$$I = \int \left(\frac{Nr^2}{2M^2} \mathrm{c}_1(\mathcal{L}_d) \wedge \mathrm{c}_1(\mathcal{L}_d) + \frac{pq^2}{2g} P(w_2) \right) + \dots, \qquad p \equiv \frac{N(N-1)}{g} \in \mathbb{Z}, \tag{292}$$

again with $\dots$ representing integer contributions. Note the absence of mixed terms between the $\mathcal{L}_d$ and $\mathcal{L}_t$ factors; this is because the bundle associated with the $\mathcal{L}_t$ factors has zero first Chern class. Also note that the term $P(w_2)$ is well-defined mod $2\pi\mathbb{Z}$: its well-definedness mod $2\pi\mathbb{Z}$ for arbitrary q depends on $pg \in 2\mathbb{Z}$, but this is always the case since $pg = N(N-1) \in 2\mathbb{Z}$. The minimal instanton number is then found by choosing either r or q to be zero and the other to be 1, depending on the choices of M and N.

From the above expression, we see that the condition for the instanton number to reproduce all possible discrete $P(w_2)$ terms is that

$$\gcd\left(\frac{N(N-1)}{g}, 2g\right) = \begin{cases} 1 & g \in 2\mathbb{Z} \\ 2 & g \in 2\mathbb{Z}+1 \end{cases}. \tag{293}$$

Actually there's a small subtlety here: to "absorb" a possible discrete term into the instanton number, we have to shift θ by something in $2\pi\mathbb{Z}$. In the previous examples this hasn't been a problem, since the part of I that can be written in terms of $P(w_2)$ is the only contribution to I mod 1. However in the $U(1)$ case, we also have a fractional part of I which comes from the $U(1)$ part: the term $\mathrm{Tr}[F] \wedge \mathrm{Tr}[F] \propto \mathrm{c}_1(E)^{\wedge 2}$ appears in I, and is fractional in general. Therefore the shift of θ to cancel the discrete $P(w_2)$ term is not as innocuous as it seems, since the phase contributed by this fractional $U(1)$ part will change.

However, in the $U(N)$ case we should really be considering a more general topological action with two distinct θ angles, since in the $U(N)$ case the second Chern class and second Chern character give distinct topological terms:

$$S_{top} = \theta_1 \int \mathrm{ch}_2[E] + \theta_2 \int \frac{1}{2} \mathrm{ch}_1[E] \wedge \mathrm{ch}_1[E] + \frac{2\pi p}{2g} \int P(w_2), \tag{294}$$

where again, $\mathrm{ch}_1[E] \wedge \mathrm{ch}_1[E]$ is just a pretentious way of writing $\mathrm{Tr}[F] \wedge \mathrm{Tr}[F]/4\pi^2$. The precise statement to make is then that if (293) is satisfied, then the torsionful $P(w_2)$ term can be absorbed into the continuous θ terms by adjusting both θ_1 *and* θ_2. Just for fun, the condition that the discrete $P(w_2)$ term can be absorbed into the continuous theta term is shown as a function of N and M in figure 1.

14.7 $Sp(N)$ and $PSp(N)$

First, let's disambiguate the notation: here, by $Sp(N)$, we mean the *compact* group

$$Sp(N) \equiv U(2N) \cap Sp(2N; \mathbb{C}), \tag{295}$$

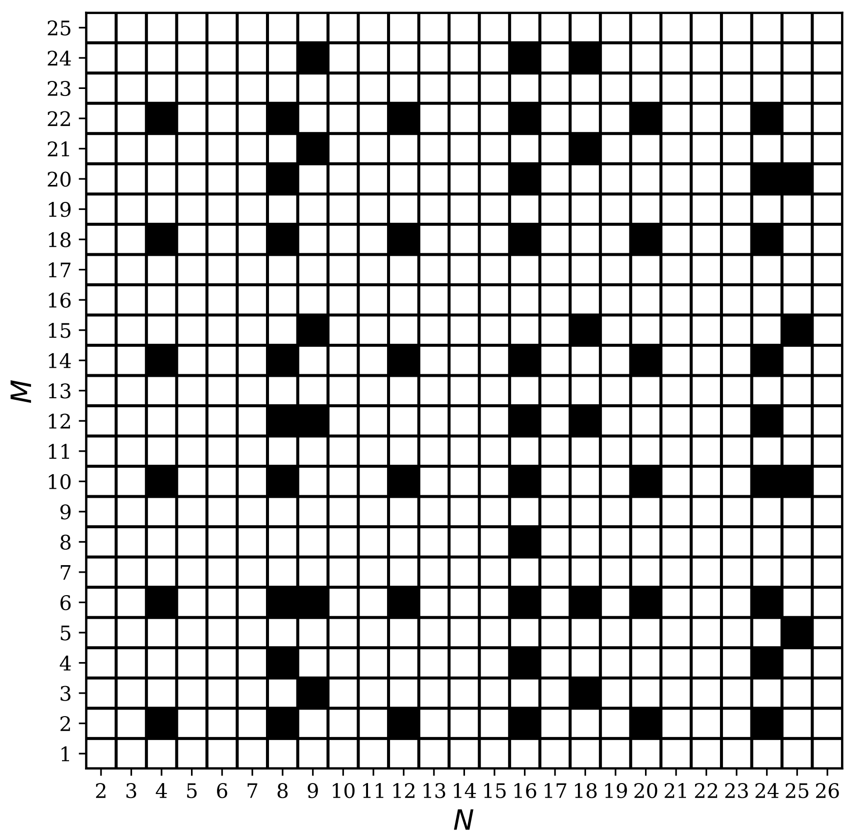


Figure 1: When the instanton number and discrete theta angle are independent for $U(N)/\mathbb{Z}_M$. A black filling means that they are independent. Going to larger values of N and M gives you something that looks like its nearly periodic, but actually isn't.

where $Sp(2N;\mathbb{C})$ is the *non-compact* group of complex $2N$-$2N$ matrices that preserve $J\otimes\mathbf{1}_{N\times N}$, in the sense that

$$U\in Sp(N)\implies U^\dagger U=1,\qquad U^TJU=J\otimes\mathbf{1}_{N\times N},\tag{296}$$

where

$$J\equiv(-iY)\otimes\mathbf{1}_{N\times N},\qquad J^2=-\mathbf{1}_{2N\times 2N}\tag{297}$$

is our choice of symplectic form. The Lie algebra for the compact symplectic group [31] can be obtained by writing a general Lie algebra element T as a linear combination

$$\mathfrak{sp}(N)\ni T=i\mathbf{1}\otimes A+X\otimes B_1+Y\otimes B_2+Z\otimes B_3,\tag{298}$$

where A is traceless and antisymmetric, and the B_i's are symmetric. Both A and the B_i's are real (they have to be (anti-)Hermitian in order for $e^{i\alpha T}$ to be unitary, and they have to be (anti-)symmetric in order for $e^{i\alpha T}$ to preserve the symplectic form). In this presentation we see clearly how $\mathfrak{su}(2)$ is embedded in $\mathfrak{sp}(N)$, viz. as the first factors in the $\otimes$. Additionally, we see that $Sp(1)=SU(2)$ and the center of $Sp(N)$ is $\mathbb{Z}_2$, as can be easily checked by looking for diagonal things that preserve $J\otimes\mathbf{1}$.

To get the normalization for the instanton number straight, we need to look at how $SU(2)$ embeds into $Sp(N)$. First, note that there can only be a single full $SU(2)$ factor in $Sp(N)$, since $Z(Sp(N))=\mathbb{Z}_2=Z(SU(2))$ means that we can't have multiple copies without having a quotient by their centers as well. We can also find such a full $SU(2)$ just by looking at matrices of the form $U\otimes\mathbf{1}_{N\times N}$, where $U\in SU(2)$. These are obviously unitary, and a quick check shows that they are also in $Sp(2N;\mathbb{C})$. Furthermore setting $U=-\mathbf{1}_{2\times 2}$ gives the center of $Sp(N)$, so we know that $Sp(N)$ really does have a full $SU(2)$ inside of it (i.e., the $SU(2)$ doesn't appear in a form where it's quotiented by $\mathbb{Z}_2$ in some way). Thus minimal $SU(2)$ instantons have instanton number 1 in $Sp(N)$, which we might write as

$$I_{Sp(N)}=\int p_1(Sp(N)).\tag{299}$$

Explicitly, we can write the gauge field $A^{Sp(N)}$ for a minimal instanton in terms of the $SU(2)$ minimal instanton gauge field $A_{SU(2)}$ as e.g.

$$A_\mu^{Sp(N)}=A_\mu^{SU(2)}\otimes E_{11},\tag{300}$$

where E_{11} is the matrix with a 1 in the upper leftmost entry, and zeros everywhere else. Since E_{11} is symmetric, $A_\mu^{Sp(N)}$ is indeed in the Lie algebra $\mathfrak{sp}(N)$. Writing it like this, it's clear that $I_{Sp(N)}=p_1(Sp(N))$.

Now for the quotient groups $PSp(N)=Sp(N)/\mathbb{Z}_2$. How might we obtain a $PSp(N)$ bundle that's not an $Sp(N)$ bundle? We consider the bundle $E_{SO(3)}=\mathcal{L}^{1/2}\oplus\mathcal{L}^{-1/2}$, which is an $SO(3)$ bundle that does not lift to an $SU(2)$ bundle. Similarly to as in our discussion of $SU(N)$, $\mathcal{L}$ is the line bundle whose first Chern class reduces mod 2 to some class $w_2\in$

[31] Why's it called symplectic? Since it preserves iY, which is the antisymmetric form used in the commutation relations for the symplectic form on phase space: if $v=(x,p)^T$, then $v^TJv=i$ is the CCR, and we can send $v\mapsto Rv$ for any $R\in Sp(N)$ preserving the CCR.

$H^2(X;\mathbb{Z}_2)$. Since roughly the transition functions fail the cocycle condition on triple overlaps by the value of w_2 on the triple overlap, the cocycle conditions of $\mathcal{L}^{1/2}$ fail by an amount controlled by $w_2/2$. We then use the diagonal embedding $SU(2) \to Sp(N)$ to use $\mathcal{L}^{1/2} \oplus \mathcal{L}^{-1/2}$ to create a $PSp(N)$ bundle that doesn't lift to an $Sp(N)$ bundle. Since the diagonal $SU(2) \to Sp(N)$ embedding sends

$$SU(2) \ni U \mapsto U \otimes \mathbf{1}_{N\times N} \in Sp(N), \tag{301}$$

the $PSp(N)$ bundle we get is a direct sum of N copies of $E_{SO(3)}$[32]:

$$E_{PSp(N)} = E_{SO(3)}^{\oplus N} = (\mathcal{L}^{1/2} \oplus \mathcal{L}^{-1/2})^{\oplus N}. \tag{302}$$

Using manipulations like the ones used for looking at $PSU(N)$ bundles, we see that the instanton number mod 1 (i.e. the part of the instanton number that doesn't come from small instantons) is

$$[\mathrm{ch}_2(E_{PSp(N)})]_1 = \frac{N}{2}[\mathrm{ch}_2(\mathcal{L})]_2 = \frac{N}{2}\frac{P(w_2)}{2}, \tag{303}$$

where $P(w_2)$ is again the Pontryagin square (a $\mathbb{Z}_4$ class, since it's acting on $\mathbb{Z}_2$ cochains), and where $[]_k$ denotes the mod k reduction. Since $P(w_2)/2$ is an integer class on a spin manifold by the even-ness of the intersection form, on spin manifolds we can have fractional instantons for $PSp(N)$ if N is odd, but not if N is even. From this we see that the discrete class $P(w_2)$ is independent from I provided that $N \in 2\mathbb{Z}+1$.

Before moving on, let's just clarify why we needed to choose the diagonal embedding of $SU(2)$ into $Sp(N)$, instead of e.g. the embedding $U \mapsto E_{11} \otimes U$ used to compute the normalization of $I_{Sp(N)}$ (I'm writing the tensor product in the opposite order since I find it slightly easier to visualize). Indeed, suppose we chose this embedding for the $SO(3)$ bundle. Then we would end up with a bundle whose transition functions could fail the cocycle condition by the matrix $-\mathbf{1}_{2\times 2} \oplus \mathbf{1}_{2N-2\times 2N-2}$. In a $PSp(N)$ bundle, the transition functions are only allowed to fail the cocycle condition by the matrix $-\mathbf{1}_{2N\times 2N}$, since this is the thing that gets quotiented out by upon passing to $PSp(N)$. In contrast, if we choose the diagonal embedding $U \mapsto \mathbf{1} \otimes U$, then we get a bundle whose transition functions fail the cocycle condition by $-\mathbf{1}_{2N\times 2N}$, which is what we want. Thus, we must choose the diagonal embedding.

14.8 $SO(N)$

We will now briefly look at the normalization of the instanton number for $SO(N)$. Some time in the future I may come back and discuss Spin(N) and quotients of $SO(N)$.

First for $SO(3)$, which we've already mentioned above. To find the normalization, we compute the value that a minimal $SU(2)$ instanton has when lifted to the adjoint representation. This is easy: we can take the same $U^\dagger dU$ with $U \sim e^{ir^aT^a}$ type of instanton, we just have to change the T^a's. Now for $SU(N)$ we have $\mathrm{Tr}[T^aT^b] = \frac{1}{2}\delta^{ab}$, while for $SO(3)$ we have

[32]This is more obvious if we write the embedding as $\mathbf{1}_{N\times N} \otimes U$, and change our definition so that the elements in $Sp(N)$ preserve $\mathbf{1}_{N\times N} \otimes J$.

$\mathrm{Tr}[T^aT^b] = 2\delta^{ab}$. So $\int p_1(E_{SO(3)})$ for an $SO(3)$ bundle with a minimal $SU(2)$ instanton is $4\int p_1(E_{SU(2)})$. Thus for $SO(3)$,

$$I_{SO(3)} = \frac{1}{4}\int p_1(SO(3)). \tag{304}$$

The notation $p_1(SO(3))$ has the hopefully obvious meaning "$p_1(E)$ for some $SO(3)$ bundle E". Note that this conclusion was reached for an arbitrary manifold, spin or not spin. If we restrict ourselves to spin manifolds, the Pontryagin class is even, so that $I_{SO(3)} \in \frac{1}{2}\mathbb{Z}$ on spin manifolds. This can be proved decomposing the $SO(3)$ bundle as $\mathcal{L}^{1/2} \oplus \mathcal{L}^{-1/2}$ and realizing that the second chern character depends on the even-ness of the intersection form, or by using the relation

$$p_1(E) \mod 2 = P(w_2), \tag{305}$$

where P is the Pointryagin square. Again we see an example of the fact that the quantization of the instanton number for simply connected Lie groups doesn't depend on whether the base manifold is spin (since there the instanton number is also the Chern class, which is integral on any manifold), but that for quotients of simply connected Lie groups, the quantiation of the instanton number does depend on whether the base manifold is spin.

Now for $SO(N \geq 4)$. We use

$$SO(4) = [SU(2) \times SU(2)]/\mathbb{Z}_2, \tag{306}$$

where the quotient is the diagonal $\mathbb{Z}_2$ (some people write this with $\otimes$ instead of $\times$, which I don't like: the tensor unit is $\mathbb{C}$, which means that we would already be making the $/\mathbb{Z}_2$ identification!). Note that since $Z(SU(2) \times SU(2)) = \mathbb{Z}_2^2$, taking the quotient leaves behind a factor of $\mathbb{Z}_2$ in the center, which is just right to match with $Z(SO(4)) = \mathbb{Z}_2$.

Regular (non-fractional) instantons are created in $SO(N > 3)$ through embedding a minimal $SU(2)$ instanton into one of the $SU(2)$ factors in the decomposition for the subgroup $SO(4) \subset SO(N)$. Now, we can form fractional instantons in $SO(N)$ by embedding an $SO(3)$ instanton inside of $SO(N)$. The way this embedding works is also through the $SO(4)$ subgroup (note to self: can we show there are no other ways to do the embedding?), but it is the embedding into the diagonal subgroup of $[SU(2) \times SU(2)]/\mathbb{Z}_2$. The reason that the embedding must be done though the diagonal subgroup is because $SO(3)$ has trivial center, and so we need to embed $SO(3)$ in the diagonal subgroup so that the quotient by $\mathbb{Z}_2$ gives us something without a $-\mathbf{1}$ central element. Anyway, the point of this is that the minimal fractional instanton number in $SO(N)$ will be *twice* that in $SO(3)$, since both $SU(2)$ factors contribute. So

$$I_{SO(N)} = \frac{1}{2}\int p_1(SO(N)), \qquad N \geq 4. \tag{307}$$

Again, this holds over arbitrary manifolds, be they spin or not spin. If the manifold is spin, we can conclude that $I_{SO(N)} \in \mathbb{Z}$ since in that case $p_1(SO(N))$ is an even class, as discussed earlier.

15 *Dyon spin, statistics, and statistical transmutation from θ angles* ✓

Today's problem statement is as follows: consider $U(1)$ gauge theory in four dimensions. Explain why, if the 2π monopoles are bosonic at $\theta = 0$, they become fermionic when $\theta = 2\pi$. What if the theory is coupled to fermionic matter, so that the $q = 1$ electric charges are fermions? Derive the spin and statistics for the dyons in the charge lattice for all values of θ, and explain why the spectrum depends on the value of θ, even though $\theta \int F \wedge F$ is a topological term which doesn't contain the metric and hence doesn't contribute to $T_{\mu\nu}$. Why does the induced electric charge created by the θ term not contribute to the statistics of the monopoles? Also, what is the periodicity of θ on (non)-spin manifolds?

Now consider a discrete analogue, namely $\mathbb{Z}_N$ BF theory. What is the analog of the θ term, and how does it affect the statistics of the line / surface operators in the theory?

We start with yet another derivation of the Witten effect, prompted by reading an old paper by Wilczek that seemed cool but which made absolutely no sense to me when I read it. Consider a single monopole in spatial $\mathbb{R}^3$ of unit magnetic charge, with the magnetic field being set up by a vector potential A_0. Consider a change of the gauge field A which takes place over a time Δt:

$$A(t) = A_0 + \left[\frac{t - t_0}{\Delta t}\Theta(t - t_0)\Theta(t_0 + \Delta t - t) + \Theta(t - (t_0 + \Delta t))\right](0,\, \alpha f'(r),\, 0,\, 0), \quad (308)$$

where we are in (t, r, θ, ϕ) coordinates. Here α is any real number and $f(r)$ is a smoothened step function with $f(\infty) = 1, f(0) = 0$. Since $A(t) - A_0 = 0$ for r near the origin, adding this change to the gauge field can be thought of as a change only of the gauge field on the coordinate patch that does not encompass the origin—thus it is well-defined even though the full $A(t)$ must be constructed by gluing patches. This $A(t)$ leads to a radial electric field

$$E_r(t) = \frac{\alpha f'(r)}{\Delta t}\Theta(t - t_0)\Theta(t_0 + \Delta t - t). \quad (309)$$

This electric field gives zero contribution to the energy if we take $\Delta t \to \infty$, since $\int d^3r\, dt\, E_r(t)^2 \propto \frac{1}{\Delta t} \to 0$. However, it does contribute to the θ term:

$$\frac{\theta}{8\pi^2}\int F \wedge F = \frac{\theta}{8\pi^2}\int \frac{1}{4} \cdot 4 \cdot 2B_i E^i = \frac{\theta\alpha}{4\pi^2} 4\pi \int_0^\infty dr\, r^2 B_r f'(r) = \frac{\alpha\theta\Phi_B}{4\pi^2}, \quad (310)$$

where in the last step we integrated by parts and used $f(0) = 0$ to kill the $\nabla \cdot B$ term. Here the magnetic flux Φ_B is measured in units where a unit monopole has 2π flux. Thus the θ term contributes a phase to the path integral.

What is the physical interpretation of this phase? The final and initial gauge configurations differ by

$$A(t > t_0 + \Delta t) - A(t < t_0) = \alpha \partial_r f. \quad (311)$$

This is a function that goes to the constant α at spatial infinity. Thus it is a "large" gauge transformation, better called an asymptotic symmetry, which rotates the boundary conditions of the sections of the $U(1)$ bundle in question by a phase $e^{i\alpha}$ (if $\alpha = 2\pi$ then this is a legit gauge transformation). The electric charge of a system is defined as the representation that the system transforms under when acted on by asymptotic symmetries like this, and so we identify the relative phase between these two configurations (the phase picked up by the θ term in the path integral) with $e^{i\alpha q}$, where q is the electric charge of the monopole. Thus since this holds for all α, we have

$$e^{i\theta\Phi_B\alpha/2\pi} = e^{i\alpha q}, \implies q = \frac{\theta\Phi_B}{4\pi^2}. \tag{312}$$

In particular, a unit monopole comes attached with an electric charge of $\theta/2\pi$.

Since for any current j such that $\langle e^{i\int A\wedge\star j}\rangle \neq 0$ we have that $\star j \in d\Omega^1(X\ \mathbb{R})$[33], we can invert the d and write $d^{-1}\star j = D$ for some $D \in \Omega^2(X;\mathbb{R})$. In the case where $\partial X \neq 0$, D may have support on ∂X. D is basically just the worldsheet swept out by the electric fluxes that link the two points created when we take the intersection of the Poincare dual of $\star j$ with any constant time slice.

Let's remind ourselves of why dyons can have fermionic statistics, thinking classically without a θ term. The angular momentum for a configuration with electromagnetic fields E, B is (I derived this by finding $T_{\mu\nu} \sim g_{\mu\nu}F\wedge\star F - F_{\mu\sigma}F^{\sigma}{}_{\nu}$ and taking T_{0i} to get the

[33]This is just because there does not exist a solution to the classical eom otherwise: $d\star F = \star j$ tells us that $\star j$ is exact, since $\star F$ is always a globally well-defined 2-form. Define the Poincare dual $\widehat{\mathcal{J}}$ of $\mathcal{J} = \star j$. Then

$$\star j \in d\Omega^3(X;\mathbb{R}) \implies \widehat{\mathcal{J}} \in B_1(X, \partial X;\mathbb{Z}). \tag{313}$$

Here $B_1(X, \partial X;\mathbb{Z})$ are the 1-chains which are relative boundaries, i.e. for each element $C \in B_1(X, \partial X;\mathbb{Z})$ there is a 2-manifold M such that ∂M consists of C and 1-submanifolds of ∂X. This follows from Poincare duality applied to relative cohomology groups (we are being a bit caviler about switching between $\mathbb{R}$ and $\mathbb{Z}$ coefficients: we usually want to think of j as being in $C^1(X;\mathbb{Z})$, but we usually want to think of F as being in $\mathbb{R}$-valued dR cohomology, so we are sloppily mixing the two).

What sorts of current loops $\mathcal{J}$ are allowed by this condition? All contractible current loops are of course allowed. Non-contractible loops are only allowed on non-compact manifolds: this is because on non-compact manifolds we can have non-contractible loops that are in $B_1(X, \partial X;\mathbb{Z})$ (think of the loop on a cylinder). This is just another way of saying that we can't have a single charge on a compact manifold (think of a current line wrapping a temporal circle), since the flux can't be well-defined everywhere. On a non-compact manifold, we can have a single charge since flux can end on the boundary: thus if we are on a non-compact manifold we can choose $\mathcal{J}$ to be a non-contractible element of $B_1(X, \partial X;\mathbb{Z})$, and this implies that $d\star F$ must be trivial in $H^3(X;\mathbb{R})$ and must be non-trivial in $H^3(X, \partial X;\mathbb{R})$ (since in this case $(d\star F)|_{\partial X} = 0$ but $(\star F)|_{\partial X} \neq 0$; flux lines are ending on the boundary). A corollary of this / another way of saying the same thing is that $\langle e^{i\oint_C A}\rangle = 0$ if the curve C is nontrivial in $H_1(X, \partial X;\mathbb{Z})$: the only Wilson loops that can have vevs are ones integrated around curves which are relative boundaries (otherwise there is a trivial shift in integration variables [one which doesn't affect the boundary conditions] which the Wilson line transforms nontrivially under). A quick comment on a common current confusion: if we add background matter through $A\wedge\star j$, it seems like the 1-form symmetry on A is unbroken, since $\star j = dD$ means $\delta(A\wedge\star j) = \lambda\wedge dD$ for a flat 1-form λ, which seems to vanish upon integration by parts. The key is that for any allowed $\star j = dD$, we have $D|_{\partial X} \neq 0$, and since λ is flat we also have $\lambda|_{\partial X} \neq 0$ (provided $\lambda \notin H^1(X, \partial X;\mathbb{R})$ in which case no Wilson lines can have vevs under the symmetry coming from shifting A by λ). Thus an integration by parts actually gives $\int_{\partial X}\lambda\wedge D \neq 0$, and so the coupling to the current indeed breaks the symmetry.

momentum, but surely there's a better way)

$$L_i = \int d^3 r \epsilon_{ijk} r^j \epsilon^{klm} E^l B^m. \tag{314}$$

Now let B^i be a monopole of strength g at the origin and E^i be sourced by an electric point charge q at position r_0^i. Using the triple product $a \times (b \times c) = b(a \cdot c) - c(a \cdot b)$, we get

$$L_i = \frac{1}{4\pi} \int d^3 r \frac{r_{0,i} r^2 - r_i (r_j r_0^j)}{r^3 |r - r_0|^3}. \tag{315}$$

Now we get crafty and use

$$\partial_i \widehat{r}^j = \frac{\delta_{ij}}{|r|} - \frac{r_j r_i}{|r|^3}. \tag{316}$$

Using this we can re-write the integrand above as $E^j \partial_j \widehat{r}_i$, and then integrate by parts. The boundary term is like $\int d^2 \Omega_{S^2_\infty} \widehat{r}^j \to 0$ since we are integrating over all angles. This gives an integral over $(\nabla \cdot E) \widehat{r}^j$. Then using $\nabla \cdot E = 4\pi \delta(r - r_0)$, we have

$$L_i = -qg \widehat{r}_i. \tag{317}$$

Note that L_i is pointed along the vector from the charge to the monopole, and that $|L|$ is *independent* of the distance between the charge and the monopole.

This electromagnetic angular momentum needs to be added on to the usual angular momentum of charged particles in order to get a conserved quantity. Using $\partial_t \widehat{r} = |r|^{-3}(\dot{r} - r(r \cdot \dot{r}))$, one can check that the conserved angular momentum for a particle of charge q moving in a monopole field of strength g is now ($m = 1$ units)

$$L_i = \epsilon_{ijk} r^j \dot{r}^k - eg \widehat{r}_i. \tag{318}$$

If these angular momentum generators are correct, they need to satisfy the correct commutation relations. When we quantize, we write the angular momentum generators as

$$L_i = \epsilon_{ijk} r^j \pi^k - eg \widehat{r}_i, \tag{319}$$

where $\pi^k = p^k - eA^k$ is the kinetic momentum and $p^k \leftrightarrow -i\partial^k$ is the canonical momentum. Since π^k parallel transports in the presence of A, its commutator picks up the field strength:

$$[\pi_i, \pi_j] = ieF_{ij}. \tag{320}$$

One can then check, with the help of the identity $-\epsilon^{inl}\epsilon^{ilm} + \epsilon^{ilm}\epsilon^{jnl} = \epsilon^{ijk}\epsilon^{knm}$, that

$$[\epsilon_{ijk} r^j \pi^k, \epsilon_{joq} r^o \pi^q] = i\epsilon_{ijl}\epsilon_{lmn} r^m \pi^n + ieF_{ij}. \tag{321}$$

The first term on the RHS is what we want if we want the usual $SU(2)$ relations for angular momentum, while the (uniform by assumption) F_{ij} term is some vaguely central-extensiony thing that screws up the commutation relations. So if our angular momentum generators are correct, the commutators with the $-eg\widehat{r}$ term need to cancel the field strength term above. This is indeed what happens: we compute

$$[\epsilon^{ilm} r_l \pi_m, -eg\widehat{r}^j] = i r_l eg \epsilon^{ilm} \frac{1}{|r^3|}(\delta_{jm} r^2 - r_j r_m). \tag{322}$$

When we subtract the $i \leftrightarrow j$ counterpart to get the full expression in the commutator, we find that these commutators contribute a total of $-2i\epsilon^{ijk} eg\widehat{r}_k$. One of these goes into the definition of L_k, and so

$$[L_i, L_j] = i\epsilon_{ijk} L^k + ieF_{ij} - ieg\epsilon_{ijk}\widehat{r}^k. \tag{323}$$

We've defined g so that $\int F = 4\pi g$ on a small sphere surrounding the monopole, and so F is given by $F_{ij} = \epsilon_{ijk} g\widehat{r}^k$. Thus the two extra terms precisely cancel each other, and we recover the correct $SU(2)$ commutation relations.

Anyway, let's go back and look at the extra $-eg\widehat{r}$ term added on to L_i. In these conventions the quantization condition on the magnetic charge is $4\pi ge \in 2\pi\mathbb{Z}$, where e is the minimal electric charge. Thus $g \in \frac{1}{2}\mathbb{Z}$. Taking $g = 1/2$ for the minimal monopole and $q = 1$, we see that the minimal dyon has $L_i = -\frac{1}{2}\widehat{r}_i$. So, if we do a rotation about the $\widehat{r}$ axis through an angle of 2π, we get a phase of $e^{2\pi i \widehat{r}_i L^i} = -1$. Thus the minimal dyon is a fermion (via spin-statistics; more on this in a moment), as we would expect.

This same argument also works for deriving the quantization condition for a pair of dipoles $(q_1, g_1), (q_2, g_2)$: we put the electric and magnetic fields generated by the two dyons into the integral for angular momentum, and use the same trick described above. This gives $L_i = -(q_1 g_2 - q_2 g_1)\widehat{r}_i$, with r_i the separation vector between the monopoles. Note the relative minus sign between the two terms! This is kind of counter-intuitive, since the two dyons are being braided around one another with the same handedness.

Let's now check that the spin-statistics argument is correct by computing the self-statistics of a dyon. We do this by moving one (e, g) dyon in a π semicircle around the other (the two-dyon system has translation invariance so we can scoot the rotated system back to the original one after the π rotation for free). The electric charge of the moving dyon picks up a phase $e^{ie\int_C A(r)}$, where A is sourced by the magnetic monopole at the origin and C is the semicircular contour. By electromagnetic duality $(e, g) \mapsto (-e, g)$ the moving magnetic monopole picks up a phase $e^{-ig\int_C \widetilde{A}(r)}$. We find $\widetilde{A}(r)$ by noting that the Lagrangian for a single dyon contains the couplings $e\dot{r}^i_e A_i(r)$ and $-g\dot{r}^i_g \widetilde{A}_i(r)$. Since the total dyon system is translationally invariant, the total canonical momentum $p_e + p_g$ is conserved, $p_e + p_g = 0$. From varying the action, we have

$$p_e + p_g = m_e \dot{r}_e + m_g \dot{r}_m + eA(r_e - r_g) - g\widetilde{A}(r_g - r_m) = 0. \tag{324}$$

Since this has to hold when e.g. the electric and magnetic charges are at rest, we require

$$eA(r) = g\widetilde{A}(-r). \tag{325}$$

Now for a monopole we can use the solution

$$A_i(r) = \frac{g}{r}\frac{\epsilon_{ijk} r^j \widehat{n}^k}{r - r_l \widehat{n}^l}, \tag{326}$$

where $\widehat{n}^j$ is some unit vector that we may choose freely. This solution is valid on a large enough patch on S^2 for our purposes, and we will take $\widehat{n} = \widehat{z}$ to be the unit vector in the plane normal to the movement of the dyon, so that $r^i\widehat{n}_i = 0$. For motion in the xy plane then, we have

$$A(r) = -\frac{g}{r}d\phi, \qquad \widetilde{A}(r) = \frac{e}{r}d\phi. \tag{327}$$

Thus the total phase accumulated during the exchange is

$$\exp\left(-ieg\int_C rd\phi\frac{1}{r} + ieg\int_C rd\phi\frac{-1}{r}\right) = e^{2\pi ieg} = (-1)^{2eg}, \tag{328}$$

which indeed gives us fermionic statistics if $eg \in \frac{1}{2}(2\mathbb{Z}+1)$, which agrees with spin-statistics.

Another way to see this is to make a "gauge transformation" to get rid of the potentials, at the expense of making the wavefunction not single-valued. I really don't like this way of doing things since wavefunctions should always be single valued and one should never do singular gauge transformations, but because a lot of other people seem to do similar things it's good to understand what the argument is. If we let r denote the vector pointing from dyon 1 to dyon 2, the Hamiltonian is (setting the mass of the dyons to 1/2 for simplicity)

$$H = (-i\partial_1 - eA(r) + g\widetilde{A}(r))^2 + (-i\partial_2 - eA(-r) + g\widetilde{A}(-r))^2 + (e^2+g^2)\frac{1}{|r|}. \tag{329}$$

Now note that $eA(r) - g\widetilde{A}(r) = 2egd\phi/|r|$. Thus we can eliminate the gauge fields in H if we make a "gauge transformation"

$$\psi(|r|,\phi) \mapsto \widetilde{\psi}(|r|,\phi) \equiv \exp\left(-2ige\phi\right)\psi(|r|,\phi), \tag{330}$$

where ϕ is the angular coordinate in the plane of the dyon's motion, with one of the dyons fixed at the origin. So the Schrodinger equation is now

$$\left(-\partial_1^2 - \partial_2^2 + \frac{e^2+g^2}{|r|}\right)\widetilde{\psi}(|r|,\phi) = E\widetilde{\psi}(|r|,\phi), \tag{331}$$

where $\widetilde{\psi}$ has a non-single-valued part $e^{-2ige\phi}$. Changing $\phi \mapsto \phi+\pi$ exchanges the dyons, and does

$$\widetilde{\psi} \mapsto (-1)^{2ge}\widetilde{\psi}. \tag{332}$$

Thus we again find that the dyons are fermions if $2ge \in 2\mathbb{Z}+1$.

Recapitulating, we have seen that bound states of a unit charge and a unit monopole are fermions, both in terms of their spin and in terms of their statistics. We have also seen that turning on a θ term changes the electric charge of the monopoles. This raises the question: are the statistics of the monopoles affected by θ? This seems reasonable because of the charge attachment, but definitely can't be true since we are in three dimensions and can only have bosons and fermions, but can tune θ continuously, implying that if the statistics did depend on θ then the statistics would vary continuously, which is a contradiction. Indeed, we will see that the electric field induced on the monopole is a "polarization effect" and makes no contribution to either the spin or the statistics of the monopole. This means that only the "microscopic" charge and monopole number are relevant for determining dyon spins and statistics. For example, suppose at $\theta = 0$ the pure minimal monopole is a boson. When we increase θ, the charge of the monopole increases, but it remains a boson. At $\theta = 2\pi$, it becomes a bosonic $(1, 1)$ dyon. This means at $\theta = 2\pi$ the new charge-neutral monopole is really a $q = -1$ "microscopic" charge bound to the $(1, 1)$ bosonic dyon. Since only the microscopic charge is relevant for determining the statistics, the new charge-neutral monopole is a fermion. More on this to follow.

Let's now verify the claim that the induced electric charge doesn't contribute to spin or statistics. For the spin, there is a very simple argument: applying Gauss' law $d(\star F/e^2 + \theta F/4\pi^2) = 0$ around a monopole of flux $\Phi = 2\pi m$ tells us that the induced electric field is

$$E^i_{ind} = \frac{m\theta}{2\pi}\frac{r^i}{r^3}. \tag{333}$$

In particular, the induced electric field is purely radial, and parallel to the magnetic field (since the θ term is an $E \cdot B$ term). Now as we recalled earlier the angular momentum goes like $L \sim \int r \times (E \times B)$, and so the contribution to the total angular momentum from the induced field vanishes. Thus the spin of the monopoles is independent of θ.

Now we look at the statistics. We could do a wavefunction approach like we did previously, but here we will do something more field-theory-centric. Write the action in the presence of sources for monopoles and electric charges as (we're in $\mathbb{R}$ time since it seemed to be easiest to keep track of signs that way—can't promise that all the signs are correct though)

$$S = \int \left[-\frac{1}{2e^2}F \wedge \star F + \frac{\theta}{8\pi^2}F \wedge F + qA \wedge \star j\right], \tag{334}$$

where $\star j$ is dual to the electric worldlines. We will keep track of the monopoles by doing the decomposition

$$F = dA + 2\pi\beta + \omega. \tag{335}$$

Here β is the magnetic monopole part such that $\star d\beta$ is the monopole current, normalized so that its Poincare dual has $\mathbb{Z}$ periods (we are *not* assuming β is coexact; this is not exactly a Hodge decomposition. The reason why is so that we can more easily deal with monopoles). Here ω is a harmonic component that will only be activated when we have flux threading 2-cycles of spacetime (the monopoles are treated as locations where $dF \neq 0$, rather than excised balls in spacetime). The harmonic component decouples from the rest of the action and gives $S_\omega \sim \int \omega \wedge \star\omega + \theta \int \omega \wedge \omega$. We will avoid talking about it any further, and will thus only be dealing with non-harmonic forms in what follows. Thus the Hodge Laplacian will always be invertible on the forms we'll be working with.

The action is then re-written as

$$S = \int \left[-\frac{1}{2e^2}A \wedge \left((\star d^\dagger d + \star dd^\dagger)A + 4\pi d\star\beta - 2e^2\frac{\theta}{2\pi}d\beta - 2e^2 q\star j\right) - \frac{2\pi^2}{e^2}\beta \wedge \star\beta + \frac{\theta}{2}\beta \wedge \beta\right], \tag{336}$$

where we have taken boundary conditions so that $A|_{\partial X} = 0$ and inserted a gauge-fixing term in Feynman gauge. Now we make the shift

$$A \mapsto A - \frac{1}{2}\Box^{-1}\star\left(4\pi d\star\beta - 2e^2\frac{\theta}{2\pi}d\beta - \frac{q}{2e^2}\star j\right). \tag{337}$$

Here the $-$ sign is needed since $\star^2 = +1$ on 1-forms in Lorentzian signature. If the worldines of the monopoles and electric sources meet ∂X transversely, which we will assume, then this shift preserves the boundary conditions on A. This renders the A part of the action to just be $\int F \wedge \star F$, which we absorb into the normalization of the measure. This leaves us with

$$S = \int \left[\frac{1}{8e^2} \left(4\pi d \star \beta - 2e^2 \star \mathcal{J}_\theta\right) \wedge \Box^{-1} \star \left(4\pi d \star \beta - 2e^2 \star \mathcal{J}_\theta\right) - \frac{2\pi^2}{e^2} \beta \wedge \star \beta + \frac{\theta}{2} \beta \wedge \beta \right], \tag{338}$$

where we've defined the current

$$\mathcal{J}_\theta \equiv qj + \frac{\theta}{2\pi} \star d\beta = qj + \frac{\theta}{2\pi} m, \tag{339}$$

which is a linear combination of the charge and monopole currents j and m, in accordance with the induced electric charges stuck onto the monopoles because of the θ term. The terms involving β but not dependent on θ are

$$S \supset -\frac{2\pi^2}{e^2} \int \left(-d^\dagger \beta \wedge \star \Box^{-1} d^\dagger \beta + \beta \wedge \star \beta\right) = -\frac{1}{2\widetilde{e}^2} \int d\beta \wedge \star \Box^{-1} d\beta, \tag{340}$$

where $\widetilde{e} = e/2\pi$ is the dual charge. Here we've used $\Box = dd^\dagger + d^\dagger d$ to write

$$\begin{aligned} \int \left(-d^\dagger \beta \wedge \star \Box^{-1} d^\dagger \beta + \beta \wedge \star \beta\right) &= \int \left(-\beta \wedge \star \Box^{-1} dd^\dagger \beta + \beta \wedge \star \Box^{-1} (dd^\dagger + d^\dagger d) \beta\right) \\ &= \int \beta \wedge \star d^\dagger \Box^{-1} d\beta = \int d\beta \wedge \star \Box^{-1} d\beta, \end{aligned} \tag{341}$$

since $d, d^\dagger$ commute with $\Box$ and hence with $\Box^{-1}$.

We can also see this by going to momentum space: ignoring constants coming from combinatorial factors, we take $\beta \wedge \star \beta \to \frac{1}{q^2} \beta_{\mu\nu} q^2 \beta^{\mu\nu}$, and then use

$$\beta \wedge \star \beta = \beta \wedge \star \Box \Box^{-1} \beta = -\partial^\sigma \beta_{\sigma\lambda} \partial^\gamma \frac{1}{-\partial^2} \beta_{\gamma\lambda} - \partial_\sigma \beta^{\rho\omega} \epsilon^{\sigma\lambda\omega\rho} \partial_\lambda \frac{1}{-\partial^2} \beta^{\lambda\alpha} \to \frac{1}{q^2} \left(q^\sigma q^\gamma \beta_{\sigma\lambda} \beta_{\gamma\lambda} + q_\sigma q_\lambda \epsilon^{\sigma\lambda\omega\rho} \beta^{\rho\omega} \beta^{\lambda\alpha}\right). \tag{342}$$

The first term cancels with

$$-d^\dagger \beta \wedge \star \Box^{-1} d^\dagger \beta \to -\frac{1}{q^2} q^\sigma q^\gamma \beta_{\sigma\lambda} \beta_{\gamma\lambda}, \tag{343}$$

which indeed leaves only the second term.

Since $d, d^\dagger$ commute with the Hodge Laplacian $dd^\dagger + d^\dagger d$, we can write the term containing β but not θ as

$$S \supset -\frac{1}{2\widetilde{e}^2} \int d\beta \wedge \star \Box^{-1} d\beta = \frac{1}{2\widetilde{e}^2} \int m \wedge \Box^{-1} \star m, \tag{344}$$

where as before $m = \star d\beta$ is the monopole current. This is the electromagnetic dual of the electric current-current Coulomb interaction for the monopoles. The fact that such an interaction was induced could also have been argued in the following way: due to the fact that $dA + 2\pi\beta$, the effective action for β needs to be invariant under the shift $\delta\beta = d\lambda$ for λ

a 1-form, since it can be compensated by a shift in A. Thus the effective action for β should involve the projector onto the coexact forms. Indeed, we have

$$\int d\beta \wedge \star \Box^{-1} d\beta = \int \beta \wedge \star \frac{d^\dagger d}{\Box} \beta, \tag{345}$$

with $d^\dagger d/\Box$ the projector onto the coexact forms: $\Box^{-1} d^\dagger d d\lambda = 0$, while

$$\frac{d^\dagger d}{\Box} d^\dagger \omega = d^\dagger \frac{\Box - d^\dagger d}{\Box} \omega = d^\dagger \omega, \tag{346}$$

so that it acts as $\mathbf{1}$ on coexact forms (it is not defined on the harmonic forms since they are in $\ker \Box$).

The θ dependence of the action is

$$S \supset \int \left[\frac{e^2}{2} \mathcal{J}_\theta \wedge \Box^{-1} \star \mathcal{J}_\theta - 2\pi d^\dagger \beta \wedge \Box^{-1} \star \mathcal{J}_\theta + \frac{\theta}{2} \beta \wedge \beta \right], \tag{347}$$

where we've used $\star^2 = 1$ on 1-forms in real time in four dimensions. The first term is the usual current-current interaction for electrically charged sources: here the current is upgraded to include a term proportional to θ and the monopole current, reflecting the fact that the electric fields of the monopoles contribute to the usual electric interaction between Wilson lines. The second two terms are

$$\int \left[-2\pi d^\dagger \beta \wedge \Box^{-1} \star \mathcal{J}_\theta + \frac{\theta}{2} \beta \wedge \beta \right] = \int \left[-2\pi q d^\dagger \beta \wedge \Box^{-1} \star j - \theta d^\dagger \beta \wedge \Box^{-1} d\beta + \frac{\theta}{2} \beta \wedge \beta \right] \tag{348}$$

The last two terms actually cancel since (I think the signs are correct) in momentum space

$$0 = A[q_\alpha \beta_{\mu\nu} \beta_{\lambda\sigma}] q^\alpha \epsilon^{\mu\nu\lambda\sigma} \implies \beta_{\mu\nu} \beta_{\lambda\sigma} \epsilon^{\mu\nu\lambda\sigma} = \frac{1}{q^2} 4 q^\alpha \beta_{\alpha\mu} q_\nu \beta_{\lambda\sigma} \epsilon^{\mu\nu\lambda\sigma}. \tag{349}$$

Here the first equality simply follows from being in four dimensions (A represents antisymmetrization on the indices) and the factor of 4 comes from our ability to contract the q outside the antisymmetrizer with any four β indices. In the final equality, the left term corresponds to the $\beta \wedge \beta$ piece, while the right term corresponds to the $d^\dagger \beta \wedge \Box^{-1} d\beta$ piece. Another way to see this is (okay, I'm really just having fun at this point) to Hodge-decompose β as $\beta = d\alpha + d^\dagger \gamma$ (recall β had no harmonic component), and then write

$$\int (d\alpha + d^\dagger \gamma) \wedge (d\alpha + d^\dagger \gamma) = 2 \int d\alpha \wedge d^\dagger \gamma, \tag{350}$$

and

$$-\int d^\dagger (d\alpha + d^\dagger \gamma) \wedge \Box^{-1} d(d\alpha + d^\dagger \beta) = \int d \star d\alpha \wedge \Box^{-1} \star d d^\dagger \gamma = \int d\alpha \wedge \Box^{-1} (d^\dagger d + d d^\dagger) d^\dagger \gamma = \int d\alpha \wedge d^\dagger \gamma. \tag{351}$$

The θ-dependent β terms are the last expression minus half the second-to-last expression, and so they indeed cancel.

Recapitulating, the full action is

$$S = \int \left[\frac{e^2}{2} \mathcal{J}_\theta \wedge \Box^{-1} \star \mathcal{J}_\theta + \frac{1}{2\tilde{e}^2} m \wedge \Box^{-1} \star m - 2\pi q \beta \wedge \Box^{-1} \star dj \right]. \tag{352}$$

The first two terms are the electric and magnetic current-current interactions, respectively, while the last bit is the AB term. Why is it an AB term? We can write it as

$$S_{AB} = -2\pi q \int \Box^{-1} dj \wedge \star\beta = -2\pi q \oint dx^\mu \int d^4y \, \epsilon_{\mu\nu\lambda\sigma} \frac{x^\nu - y^\nu}{|x-y|^4} (\star\beta)^{\lambda\sigma}(y), \tag{353}$$

where the $\oint$ is over the Poincare dual of $\star j$. This is a linking number between the current loop and the surface dual to $\star\beta$, which turns out to be the AB phase. This is seen a bit more explicitly by writing $\star j = dD$ (if j is not of this form, $\langle e^{i\int A \wedge \star j} \rangle = 0$), and using current conservation to write

$$\Box j = d^\dagger dj \implies \Box^{-1} dj = d(d^\dagger d)^{-1} \star dD = dd^{-1} \star d^{-1} dD = \star D, \tag{354}$$

where we haven't bothered to keep track of potential minus signs. Thus S_{AB} is

$$S_{AB} = -2\pi q \int D \wedge \beta = -2\pi q \int_{\widehat{D}} \beta. \tag{355}$$

Here the integral is over a disk bound by the current loop. For a geometry where a spatial current loop is drawn in spatial $\mathbb{R}^3$ with a monopole sitting at the origin, this just becomes

$$S_{AB} = -q \int_{\widehat{D}} dA_{mag} = -qm\Omega(\widehat{D}), \tag{356}$$

where m is the monopole strength, A_{mag} the monopole part of the vector potential, and $\Omega(\widehat{D})$ the solid angle enclosed by the loop ($d\beta = 0$ away from the origin of $\mathbb{R}^3$ since the monopole current is $m \propto \delta(\vec{x})dt$, and so the dependence on $\widehat{D}$ is topological). This is precisely the AB phase we expect.

The important thing here is that the calculation of the AB phase involved only j, and not $\mathcal{J}_\theta$. This means that the monopoles have statistical interactions only with genuine microscopic electric currents, and they do not have any statistical interactions with the electric charge bound to them by the θ term. Thus the statistics of a dyon is calculated solely through its microscopic charge assignments, and its statistics are unchanged as θ is varied.

Now we will discuss the periodicity of θ, which is a bit subtle. We can see the difference in the periodicity of θ for fermionic / bosonic theories even without talking about spin structures and the evenness of the intersection form. For bosonic theories (both $(1,0)$ and $(0,1)$ are bosons, with notation (e,m)), the $(1,1)$ dyon is a fermion, as explained above. More generally, (e,m) is a fermion whenever e and m are both odd. Thus the $m=0$ row of the charge lattice is totally bosonic, the $m=1$ has alternating bosons / fermions, the $m=2$ row has all bosons, etc. Now when $\theta \mapsto \theta + 2\pi$, the $m=0$ row is invariant, the $m=1$ row moves to the right by one unit, the $m=2$ row moves to the right by two units, and so on (increasing θ changes the charge, but not the statistics, of a given dyon, since the induced charge doesn't enter in to the statistical calculation). Thus the $m \in (2\mathbb{Z}+1)$e rows are not invariant under the shift in θ, but are invariant under a 4π shift. So when e, m are bosons, $\theta \sim \theta + 4\pi$.

This is equivalent to the statement that on a non-spin manifold, the quantization of $\int (F/2\pi) \wedge (F/2\pi)$ is in $\mathbb{Z}$, not in $2\mathbb{Z}$. On a spin manifold, we have transparent fermions that we can bind to any of the particles in the charge lattice, changing their statistics. Thus on a spin manifold, the statement that the statistics of the $m = 1$ row gets changed when shifting θ by 2π is not meaningful. On a non-spin manifold, (on which the theory for which both e, m are bosons can be defined) we don't have these transparent fermions, and so the shift of the $m = 1$ row is meaningful.

Hold on, one might say, if the $(1,1)$ dyon is a fermion, how can we define the theory on a non-spin manifold? We can't use a spinc connection since the charge $(1,0)$ object is a boson, and therefore we have no spin-statistics relation. So what's going on? The answer is[34] that even though $(1,1)$ is a fermion, it doesn't need a spin structure on the spacetime manifold to have a well-defined framing. The point is that there is no choice of fundamental fields for which $(1,1)$ is a local operator: it is always a non-local object. We can give it a framing (a spin structure on its worldline) by using e.g. the vector that points from the electric charge to the magnetic charge. If it were a local operator we couldn't use its internal structure to give it a framing, so we'd have to give it a framing by using the one induced by the framing of the tangent space of the ambient manifold. If the ambient manifold is non-spin then this would be impossible. So non-spin manifolds preclude defining neutral *local* fermions, but not nonlocal (emergent) ones.

By contrast if e is a fermion and m is a boson, things change. Now the $m \in 2\mathbb{Z}$ rows of the lattice are alternating fermion-boson, while the $m \in (2\mathbb{Z}+1)$ rows are all bosonic. Shifting θ by 2π only shifts the $m \in (2\mathbb{Z}+1)$ rows of the lattice, which is trivial in this case since the statistics of all the dyons in the odd rows is bosonic. Thus $\theta \sim \theta + 2\pi$ when e is a fermion (regardless of whether we are on a spin manifold, or a non-spin manifold with a spinc connection). The case where both e and m are fermions is anomalous and probably warrants its own diary entry, to be written sometime in the future.

15.1 Discrete case

Now to briefly discuss the discrete case of $\mathbb{Z}_N$ BF theory. The action with sources included is

$$S = \int \left(\frac{n}{2\pi} F_A \wedge B + \frac{kn}{4\pi} B \wedge B + j \wedge \star A + \Sigma \wedge \star B \right), \tag{357}$$

where the second term is the discrete θ term (of course, the appropriate squaring operation is really the Pontryagin square since $\wedge$ is not dual to the proper intersection product, but we will continue to abuse notation by writing $\wedge$). Now we normally have a gauge transformation on B where $\delta B = F_\lambda$ for F_λ the field strength of a $U(1)$ gauge field. We want the action to be gauge invariant without having to integrate any fields out to impose quantization conditions (i.e. we want the action to be invariant without using our knowledge that the periods of B

[34] Thanks to Ryan Thorgren to helping me understand this! :D

will be quantized in $2\pi\mathbb{Z}/n$), and so we require that $\delta A = -k\lambda$ under this shift. Imposing this invariance in the presence of the currents means

$$k\lambda \wedge \star j - \lambda \wedge d \star \Sigma = 0 \implies j = \frac{1}{k} d^\dagger \Sigma. \tag{358}$$

This forces $\Sigma = \Sigma_j + \Sigma_c$, where $\partial \widehat{\star\Sigma_j} = \widehat{\star j}$ and $\partial \widehat{\star\Sigma_c} = 0$. Here $\star\Sigma_j$ is dual to the surfaces attached to the Wilson lines to render them gauge invariant, while $\star\Sigma_c$ is dual to the worldsheets of closed strings that are charged under B.

Integrating out A says that

$$\star j = -\frac{n}{2\pi} F_B. \tag{359}$$

Since the current is conserved, working in a gauge where $d^\dagger B = 0$ we can take $d^\dagger$ of both sides and write (we are in Euclidean time in four dimensions, so $d^\dagger = -\star d \star$, and $\star^2 = (-1)^{p^2}$ on p-forms)

$$\star dj = -\frac{n}{2\pi} \Box B \implies B = -\frac{2\pi}{n} \Box^{-1} \star dj. \tag{360}$$

This is only a solution for B up to elements of $H^2(X) = \ker \Box$. We will assume for simplicity that this cohomology group vanishes, since what we really care about are the correlation functions of the various operators, which we can study on e.g. $\mathbb{R}^4$. Relaxing this assumption is no big deal, since the harmonic parts of B decouple from most other things in the action. Also, since we always take $\star j = dD$ for $\widehat{D}$ some disk (or more precisely, any 2-manifold with $\partial D \setminus \partial X = \widehat{\star j}$, where X is spacetime—recall that if j is not of this form, the path integral vanishes [also, in our case, the associated Wilson line would not be gauge invariant]), one may also write $dD = -\frac{n}{2\pi} F_B \implies B = -\frac{2\pi}{n} D$. This solution for B is only defined up to elements in $\ker(d)$. If we impose $d^\dagger B = 0$ then the exact part drops out, and we get a solution for B up to elements of $H^2(X)$. As before we assume this vanishes, and so we get the same result. In BF theory, we know that the periods of B are quantized, Is this reproduced by our solution? Let M_2 be any closed 2-chain in $H_2(X;\mathbb{Z})$. Then

$$\frac{2\pi}{n} \int_{M_2} \Box^{-1} \star dj = \frac{2\pi}{n} \int \star \widehat{M_2} \wedge \Box^{-1} dj = \frac{2\pi}{n} \int M_2 \wedge D \in \frac{2\pi}{n} \mathbb{Z}, \tag{361}$$

since we take $\star j$, and hence D, to be in the image of $H^*(X;\mathbb{Z})$ under the inclusion into de Rham cohomology. So indeed, B has the periods we expect.

Now we put this solution back into the action:

$$S = \frac{1}{2\pi} \int \left(\frac{4\pi^2 k}{2n} \Box^{-1} dj \wedge \Box^{-1} dj - \Sigma \wedge \frac{2\pi}{n} \Box^{-1} dj \right). \tag{362}$$

We see that the discrete θ parameter k determines the self-statistical properties of the electric matter (sources for A), while the statistics between the magnetic (sources for B) and electric matter is determined just by n and is unaffected by the θ term. The fact that a θ term involving the "magnetic" field B affects the statistics of the matter coupled to the opposite field A is just like what happens in $U(1)$ gauge theory, whose action we might write as $d\widetilde{A} \wedge F/2e^2 + \theta F \wedge F/8\pi^2$, where $d\widetilde{A} = \star F$. Here the θ term involving the field F affects the charges of the sources coupled to the opposite field $\widetilde{A}$ ($d\widetilde{A} \wedge F$ is supposed to remind one of $F_A \wedge B$).

16 *Basics of large N Yang-Mills* ✓

Toady we will consider $SU(N)$ YM theory at large N. We will explain why the t' Hooft limit is essential for getting analytic control on the theory, rather than just the $N \to \infty$ limit, and derive the matrix structure of the propagator and the Feynman rules. Finally, we relate the Wilson line in the adjoint to the Wilson line in the fundamental and explain how this relationship simplifies in the large N limit. This material is of course all in the literature; I just wanted to understand things myself and make a reference I could refer to later.

First, we claim that $SU(N)$ YM theory we never want to take $N \to \infty$ limit without also taking $g \to 0$. To see why, recall that the $SU(N)$ beta function is

$$\frac{dg^2}{d\ln\mu} = -\frac{11g^4}{24\pi^2}C(SU(N)) = -\alpha g^4 N, \tag{363}$$

where α is a positive number and we've used the fact that the quadratic casimir of the adjoint representation (aka twice the dual coxeter number) is $f_{abc}f_{dbc} = \delta_{ad}C(SU(N)) \implies C(SU(N)) = N$. If we integrate the β function, then we get

$$g^2(\mu) = \frac{Ng_0^2}{1 - \alpha N g_0^2 \ln(\Lambda_0/\Lambda_\mu)}, \tag{364}$$

where the 0 subscripts are at some reference scale (like the UV). Λ_{QCD} is the scale at which perturbation theory breaks down, found by setting the denominator to zero. This gives

$$\Lambda_{QCD} = \Lambda_0 e^{-(\alpha N g_0^2)^{-1}}. \tag{365}$$

Thus we see that if we take $N \to \infty$ without also keeping Ng^2 fixed, we get $\Lambda_{QCD} \to \Lambda_0$, and so the strongly-coupled scale becomes equal to the UV cutoff scale, leaving us unable to say anything useful.

So with this in mind, we define the finite coupling constant of interest as $\lambda = g^2 N$. Thus the action is

$$S = \frac{N}{\lambda}\int \mathrm{Tr}[F \wedge \star F]. \tag{366}$$

Despite the N in front, we cannot just take the classical saddle point. This is essentially because while a non-classical field configuration will be suppressed by the N out front, it will be amplified by the fact that as N gets large there are many many more non-classical field configurations to have the fields in. So the theory is still very quantum.

To determine the Feynman rules, we will deal with the matrices A directly, rather than their components A^a in some representation. Now A transforms adjointly under the (global part of the) gauge group:

$$A^a_b \mapsto [U^\dagger]^{ad} A^d_e U_{eb} = [U^*]^{da} U_{eb} A^d_e. \tag{367}$$

Thus the upper index of A transforms in $\bar{N}$ while the lower index transforms in N, reflecting the fact that $N \otimes \bar{N} = Ad \oplus \mathbf{1}$.

Now just as how we can write something transforming with spin 1/2 under $SU(2)$ as $a|\uparrow\rangle + b|\downarrow\rangle$, we can write

$$A^a_b = \sum_{\mathcal{A}=1\dots N^2} A^{\mathcal{A}} [T^{\mathcal{A}}]^a_b. \tag{368}$$

Here we are thinking of A as a vector transforming under the adjoint of $SU(N)$: the tuple (a, b) is a composite vector index. The $T^{\mathcal{A}}$'s are basis vectors, and the $[T^{\mathcal{A}}]^a_b$'s are their components. Maybe the notation $v^{\mathcal{A}}_{a,b}$ would be slightly better to emphasize that we are thinking of the $T^{\mathcal{A}}$ as vectors, rather than generator matrices. Indeed, $\mathcal{A}$ runs over N^2 different values, since the vectors being represented live in an N^2 dimensional space (the space of $N \times N$ unitary matrices).

Now since the adjoint representation has dimension $N^2 - 1$, we need to remove one of the $T^{\mathcal{A}}$'s from the basis, in order to get a vector space of the right dimensionality, as $\dim \mathfrak{su}(N) = N^2 - 1$. We want to remove the generator corresponding to the trivial representation, which is proportional to the identity matrix. Let us work in the normalization

$$\langle T^{\mathcal{A}} | T^{\mathcal{B}} \rangle = \mathrm{Tr}[T^{\mathcal{A}} T^{\mathcal{B}}] = \delta^{\mathcal{A}\mathcal{B}}. \tag{369}$$

Then the generator proportional to $\mathbf{1}$ that we need to remove is $[T^{N^2}]^a_b = \frac{1}{\sqrt{N}} \delta^a_b$. With this generator taken out, the completeness relation is now

$$\sum_{\mathcal{A}=1}^{N^2-1} [|T^{\mathcal{A}}\rangle\langle T^{\mathcal{A}}|]^{ac}_{db} = \sum_{\mathcal{A}=1}^{N^2-1} [T^{\mathcal{A}}]^a_b [T^{\mathcal{A}}]^c_d = \left(\delta^a_d \delta^c_b - \frac{1}{N} \delta^a_b \delta^c_d \right). \tag{370}$$

If we were working with $U(N)$ so that we could have generators with nonzero trace, we would have the full N^2 generators and we would have $\mathbf{1} = \delta^a_d \delta^c_b$ on the RHS. Taking out the generator for the $SU(N)$ case means that $\mathrm{Tr}[\sum_{\mathcal{A}} |T^{\mathcal{A}}\rangle\langle T^{\mathcal{A}}|] = \frac{1}{2}(N^2 - 1)$.

Anyway, the point of writing A like this is that it allows us to figure out what the index structure of the propagator is. The kinetic term in the action looks like

$$\mathrm{Tr}[dA \wedge \star dA] = \sum_{\mathcal{A},\mathcal{B}} \langle T^{\mathcal{A}} | T^{\mathcal{B}} \rangle dA^{\mathcal{A}} \wedge \star dA^{\mathcal{B}} = \sum_{\mathcal{A}} dA^{\mathcal{A}} \wedge \star dA^{\mathcal{A}}, \tag{371}$$

so that the propagator is only non-zero when it connects two A's with the same generator $T^{\mathcal{A}}$. Thus

$$\begin{aligned}\langle A^a_{\mu b}(x) A^c_{\nu d}\rangle &= \sum_{\mathcal{A},\mathcal{B}} [T^{\mathcal{A}}]^a_b [T^{\mathcal{B}}]^c_d \langle A^{\mathcal{A}}_\mu(x) A^{\mathcal{B}}_\nu(y)\rangle = \sum_{\mathcal{A}} [|T^{\mathcal{A}}\rangle\langle T^{\mathcal{A}}|]^{ac}_{db} \langle A^{\mathcal{A}}_\mu(x) A^{\mathcal{A}}_\nu(y)\rangle \\ &= D_{\mu\nu}(x-y)\left(\delta^a_d \delta^c_b - \frac{1}{N}\delta^a_b \delta^c_d\right).\end{aligned} \tag{372}$$

Here $D_{\mu\nu}(x-y)$ is the regular vector propagator in whichever gauge fixing condition we feel like adopting. Together with this propagator, the Feynman rules are easy to write down in double-line notation.

Now we can look at general correlation functions. All operators of interest will be gauge invariant and hence will involve traces (they will have no free indices). We can focus on operators with a single trace, since operators with more traces can be built form single-trace ones. We find connected correlation functions for single-trace operators $\mathcal{O}_i$ by adding $\sum_i N \int J_i \mathcal{O}_i$ to the action, and then finding $\prod_j (N^{-1}\delta_{J_j}) W[J]$, where $Z[J] = e^{-W[J]}$ as usual. The factors of N here are just so that $W[J]$ is the same order in N as the vacuum partition function. This is order $O(N^2)$, which can be seem from evaluating the simplest planar vacuum-to-vacuum graphs. In general, we see from the action that the amplitude of a given graph is determined by

$$\mathcal{A} \sim \left(\frac{N}{\lambda}\right)^{v-e} N^{f+s} = \lambda^{e-v} N^{\chi+s}, \tag{373}$$

where v, e, f, s are the vertices, edges, faces, and sources of the Feynman diagram. We can then use $\chi = 2 - 2g - s$, where g is the genus and s is the number of holes (sources), to write

$$\mathcal{A} \sim \lambda^{e-v} N^{2-2g}. \tag{374}$$

Thus no matter how many (pure glue) sources we insert, the leading-order in N diagrams that contribute to $W[J]$ will be planar and go as N^2 (drawing some pictures to "experimentally" test this is fun).

Now from the way we are computing connected correlation functions, we see that every functional differentiation with respect to a source that we need to perform multiplies the correlation function by $1/N$. Thus $\langle \mathbf{1}\rangle$ goes as $O(N^2)$, $\langle \mathcal{O}_1\rangle$ goes as N^1, and in general an n-point connected correlation function goes as $\langle \mathcal{O}_1 \cdots \mathcal{O}_n\rangle_c \sim N^{2-n}$. This means that as $N \to \infty$ all the 2-point functions of single-trace operators factorize:

$$\langle \mathcal{O}_1 \mathcal{O}_2\rangle = \langle \mathcal{O}_1 \mathcal{O}_2\rangle_c + \langle \mathcal{O}_1\rangle\langle \mathcal{O}_2\rangle \to (\langle \mathcal{O}_1\rangle\langle \mathcal{O}_2\rangle \sim O(N^2)) + O(1). \tag{375}$$

In particular, $\langle(\mathcal{O} - \langle\mathcal{O}\rangle)^2\rangle / \langle\mathcal{O}\rangle^2 = \langle\mathcal{O}\mathcal{O}\rangle_c / \langle\mathcal{O}\rangle^2 \sim N^{-2}$, so that as $N \to \infty$ the fluctuations become small.

Now we turn to the computation of the Wilson loop. We are interested in computing the Wilson loop in the adjoint representation. In general, we have

$$W_{R_1 \otimes R_2}(C) = P \exp\left(i \oint_C dx^\mu A^\alpha_\mu (T^\alpha_{R_1} \otimes \mathbf{1} + \mathbf{1} \otimes T^\alpha_{R_2})\right) = W_{R_1}(C) \otimes W_{R_2}(C). \tag{376}$$

Here we have used that $T^{\alpha}_{R_1\otimes R_2} = T^{\alpha}_{R_1}\otimes \mathbf{1} + \mathbf{1}\otimes T^{\alpha}_{R_2}$ (it's an easy check to see that the Lie bracket holds, and that $e^{i\theta^{\alpha}T^{\alpha}_{R_1\otimes R_2}} = e^{i\theta^{\alpha}T^{\alpha}_{R_1}}\otimes e^{i\theta^{\alpha}T^{\alpha}_{R_2}}$), and then used that things which commute can be separately path-ordered.

We are most interested in the case of $SU(N)$, for which $N\otimes\bar{N} = Ad\oplus\mathbf{1}$. We need to project out the trivial representation from $W_{N\otimes\bar{N}}(C)$. We do this by (using the notation $a\bar{a}\in\mathbb{Z}_N^2$ for an adjoint index)

$$[W_{Ad}(C)]^{a\bar{a}}_{b\bar{b}} = [W_N(C)]^a_b[W_{\bar{N}}(C)]^{\bar{a}}_{\bar{b}} - \frac{1}{N}\delta^a_{\bar{b}}\delta^b_{\bar{a}}. \tag{377}$$

The index structure on the last term is such that it only activates for indices that are diagonal in both the N and $\bar{N}$ factors, and is determined by taking the $\otimes$ of the intertwiners (the δ functions) for the indices in the N and $\bar{N}$ representations (delta functions like δ^b_a that connect two indices transforming in the same representation are not invariant symbols unless $N\cong\bar{N}$ which only happens if $N=2$, so we need to have δ functions with an index structure that connects N and $\bar{N}$ indices). The $1/N$ normalization is to ensure that when we take the gauge field to vanish, we get $\mathrm{Tr}[W_{Ad}(C)|_{A=0}] = \dim(Ad) = N^2-1$.

Taking the trace and using the Hermiticity of the generators in the fundamental representation to write

$$W_{\bar{N}}(C) = P\exp\left(i\oint_C A^{\alpha}(-T^{\alpha}_N)^*\right) = P\exp\left(i\oint_C A^{\alpha}T^{\alpha}_N\right)^* = W_N(C)^*, \tag{378}$$

we see that

$$\langle \mathrm{Tr}\, W_{Ad}(C)\rangle = \langle|\mathrm{Tr}\, W_N(C)|^2\rangle - 1. \tag{379}$$

Since $W_N(C)$ is a single-trace operator, in the $N\to\infty$ limit the two point function is dominated by the disconnected part[35] , Thus at large N, we can move the square outside of the expectation value:

$$\langle \mathrm{Tr}\, W_{Ad}(C)\rangle \approx_{N\to\infty} |\langle \mathrm{Tr}\, W_N(C)]\rangle|^2. \tag{380}$$

Thus the coefficient of the adjoint area law is twice that of the fundamental line area law. This is actually kind of crazy, since we know that for small N the adjoint Wilson line always

[35]Since the Wilson lines are nonlocal, this might be a little bit subtle to see. Consider first the 1-point function for the fundamental Wilson line. To zeroth order in the t' Hooft coupling, it just looks like a single fundamental line, drawn in the shape of C. This is $O(N)$, since there is one trace. To next order, we have to integrate over all ways for a propagator to connect two points on the fundamental line together (the Abelian Wilson line is the exponential of this double integral). The propagator is a double line, and so we are integrating over diagrams that look like two loops, which are parallel along the propagator line. This diagram thus has N dependence of λN: two gluon-quark-quark vertices that go as 1 in our choice of coupling, one propagator that goes as λ/N, and two sums over N for the two loops. Higher order terms have more propagators connecting the loop to itself, but adding a propagator in a planar way increases the number of propagators by one and the number of faces by one, resulting in an extra power of λ but the same $O(N)$ N-dependence (as usual non-planar diagrams are suppressed). Thus the 1-point function for the Wilson line is $O(N)$.

The connected correlation for two fundamental Wilson lines is $O(1)$ however. Indeed, consider the $O(\lambda)$ contribution to the connected part: it looks like two single (fundamental) lines, with a single double-line propagator connecting them. This diagram has one propagator and one loop, so it goes as λ^1N^0. Adding further propagators cannot increase the N-dependence, and only increases the λ dependence. So $|\langle W_N(C)\rangle|^2$ is larger than $\langle|W_N(C)^2\rangle_c$ by a factor of N^2, in line with what we expect for two-point functions of single-trace operators.

has perimeter law, since adjoint strings can break and end on gluons (adjoint sources can be screened by gluons). Looking through the old QCD literature, apparently the adjoint line goes as (schematically)

$$\langle \mathrm{Tr}\, W_{Ad}(C)\rangle \approx N^2 e^{-\sigma A} + e^{-\sigma P}, \tag{381}$$

where σ is a string tension, A is the area and P is the perimeter. Since normally $A \gg P$, the perimeter law piece dominates. However if we make $N \to \infty$, the area-law-scaling piece can actually win out—and it does if the fundamental lines are confined, as we saw above. Note however that for large enough Wilson lines, there is always a cross over to perimeter law for any finite N.

17 *When are CS theories spin TQFTs?* ✓

Today's mission is straightforward: answering the question in the title by working through a few representative examples.

One way to examine whether a CS theory is spin or not is to carefully define the CS action by breaking up the manifold into patches and defining the action in the style of DB cohomology; see a previous diary entry on this. This approach is kind of subtle for non-Abelian gauge groups though, so we will take the bounding 4-manifold approach, which is computationally simpler.

17.1 $U(1)_k$

As usual, define the CS action on a closed 3-manifold X by integrating an $F \wedge F$ term over some 4-manifold Y with $\partial Y = X$. The exponential of the action is independent of the choice of bounding 4-manifold Y provided that

$$\frac{k}{8\pi^2}\int_X F \wedge F \equiv \frac{k}{2} I \in \mathbb{Z} \tag{382}$$

for all closed 4-manifolds M. Now, $F/2\pi \in H^2(M;\mathbb{Z})$, so we know for sure that $I \in \mathbb{Z}$ since the cup product of $F/2\pi$ with itself is then in $H^4(M;\mathbb{Z})$. Now if $k \in 2\mathbb{Z}$ then the (exponential of the) above integral is independent of M, regardless of whether M is spin or not. Thus if $k \in 2\mathbb{Z}$, the CS theory is insensitive to the spin structure and hence is bosonic. However, suppose $k \in 2\mathbb{Z}+1$. Then the CS action is only well-defined if $I \in 2\mathbb{Z}$. The constraint $I \in 2\mathbb{Z}\ \forall M$ can only be satisfied if we restrict our attention to M such that M is spin. If M is spin then $\omega_2(TM) = 0 \mod 2$ and the intersection form is even, meaning that I is always even. So, for odd k, the theory can only be defined using spin bounding 4-manifolds, and hence the original 3-manifold needs to come equipped with a spin structure as well. Thus odd k theories are spin TQFTs.

17.2 $SU(N)_k$

Now consider $SU(N)$. Now the relevant integral over a closed 4-manifold is

$$\frac{k}{8\pi^2}\int_M \mathrm{Tr}[F\wedge F] = k\mathrm{ch}_2(F) \in k\mathbb{Z}, \tag{383}$$

since the second Chern character is the second second Chern class for $SU(N)$ on account of the tracelessness of the $SU(N)$ generators, it is quantized on account of the second Chern class being a $\mathbb{Z}$ characteristic class (for $U(1)$, the integral is just the second Chern character, which is not a class in $\mathbb{Z}$ cohomology). Note that the quantization of the integral does not depend on whether M is spin or not: the second Chern class's integrality doesn't depend on the spin nature of M, since it does not (in general) compute an intersection form. Indeed, the minimal $\mathrm{ch}_2(F) = 1$ instantons are the "small" instantons that can exist on any manifold, regardless of its topology. They are constructed from bundles which are not tensor products of line bundles (if they were their quantization would be sensitive to $\omega_2(TM)$), and since they are "small" they can exist equally happily on spin- and non-spin manifolds. So, all the $SU(N)$ CS theories are bosonic.

17.3 $U(N)_{k,q}$

Now for $U(N)_{k,q}$, which is defined though

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[\mathcal{A}\wedge d\mathcal{A} - \frac{2i}{3}\mathcal{A}\wedge\mathcal{A}\wedge\mathcal{A}\right] + \frac{q-k}{4\pi N}\mathrm{Tr}[\mathcal{A}]\wedge d\mathrm{Tr}[\mathcal{A}]. \tag{384}$$

As explained before, the notation is done like this because q is ($1/N$ times) the effective $U(1)$ level, while k is the effective $SU(N)$ level.

Now we use the decomposition $U(N) = [SU(N)\times U(1)]/\mathbb{Z}_N$. At the level of actions, we simply write $\mathcal{A} = A + \boldsymbol{\mathcal{A}}\mathbf{1}$, where A is an $SU(N)$ field (whose transition functions may fail by Nth roots of unity), $\boldsymbol{\mathcal{A}}$ is a $U(1)$ field (with transition functions failing in the inverse way). The quotient comes from the correlation of the transition functions between A and $\boldsymbol{\mathcal{A}}$. In terms of these fields, we have

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[A\wedge dA - \frac{2i}{3}A\wedge A\wedge A\right] + \frac{qN}{4\pi}\boldsymbol{\mathcal{A}}\wedge d\boldsymbol{\mathcal{A}}, \tag{385}$$

so that qN is indeed the "effective $U(1)$ level". The scare quotes here are because $\boldsymbol{\mathcal{A}}$ isn't really a $U(1)$ field, because of the quotient: only $N\boldsymbol{\mathcal{A}}$ is a legit $U(1)$ field. So the legit $U(1)$ part is really

$$S \supset \frac{2\pi q/N}{8\pi^2}\int_Y d(N\boldsymbol{\mathcal{A}})\wedge d(N\boldsymbol{\mathcal{A}}), \tag{386}$$

where Y is a bounding 4-manifold. This would seem to indicate that we require $q \in N\mathbb{Z}$ in order for the action to be well-defined (independent of Y). But this is not quite the case, since the term in $\mathcal{L}$ involving A also stands a chance of being ill-defined on its own, due to the $\mathbb{Z}_N$ quotient. Indeed, from our previous diary entry on instanton numbers in $PSU(N)$ gauge theory, we saw that $\frac{k}{2}\int \mathrm{Tr}(dA/2\pi \wedge dA/2\pi)$ was quantized in $k/N\mathbb{Z}$. Thus the ill-defined-ness of the A part of the action alone is captured by $k/N \mod 1$. Since the transition functions of A and $\mathcal{A}$ fail the cocycle condition in opposite senses at each triple overlap of patches, the fractional part of the instanton number for the A field is the negative of that for the $\mathcal{A}$ field. Thus the total parameter measuring the ill-defined-ness of the action is actually $(k-q)/N \mod 1$. So, for a consistent theory, we need

$$k - q \in N\mathbb{Z}. \tag{387}$$

Another way to say this is that since $\mathrm{Tr}[\mathcal{A}]$ is a well-defined $U(1)$ gauge field (but not $\mathcal{A}$ itself), the appearance of the term $(k-q)\mathrm{Tr}\mathcal{A} \wedge d\mathrm{Tr}\mathcal{A}/4\pi N$ in the action means that in order for this to be well-defined we need to have $(k-q)/N \in \mathbb{Z}$.

Yet another way to say it is that the theory needs to be invariant under simultaneous shifts in the transition functions of A and $\mathcal{A}$ by elements in $\mathbb{Z}_N$, which is realized on $\mathcal{A}$ through the shift $\delta\mathcal{A} = \frac{1}{N}d\phi$ for some 2π-periodic scalar ϕ. Since we are shifting both A and $\mathcal{A}$, A is invariant, and the action changes by

$$\delta S = \frac{(q-k)}{2\pi}\int d\phi \wedge F_{\mathcal{A}} \tag{388}$$

(for the derivation of the fact that the prefactor is $1/2\pi$ and not $1/4\pi$, see the previous diary entry). Now since only $N\mathcal{A}$ is a $U(1)$ gauge field, the flux of $F_{\mathcal{A}}$ is quantized in $\overline{\mathbb{Z}}/N$. Thus in order for $\delta S \in \overline{\mathbb{Z}}$, we need $(q-k) \in N\mathbb{Z}$.

Anyway, when are these theories spin? Returning to the original formulation in terms of the $\mathcal{A}$ field, the appropriate four-dimensional integral to compute is

$$I = \frac{1}{8\pi^2}\int \left(k\mathrm{Tr}[F_{\mathcal{A}} \wedge F_{\mathcal{A}}] + \frac{q-k}{N}\mathrm{Tr}[F_{\mathcal{A}}] \wedge \mathrm{Tr}[F_{\mathcal{A}}] \right). \tag{389}$$

Using the definition of the second Chern class,

$$I = 2\pi \int \mathrm{c}_2(E) + 2\pi\frac{k + (q-k)/N}{8\pi^2}\int d\mathrm{Tr}\mathcal{A} \wedge d\mathrm{Tr}\mathcal{A}, \tag{390}$$

where E is the total $U(N)$ bundle. Since $\int \mathrm{ch}_2(E) \in \mathbb{Z}$ on any closed 4-manifold (spin or not), whether or not the theory is spin is determined by the second term. In particular, we get

$$k + \frac{q-k}{N} \in \begin{cases} 2\mathbb{Z} & \implies \text{not spin} \\ (2\mathbb{Z}+1) & \implies \text{spin} \end{cases}, \tag{391}$$

where these are the only two options since as we said before, $(q-k) \in N\mathbb{Z}$.

17.4 $PSU(N)_k$

As we saw in a previous diary entry, on spin manifolds, minimal $PSU(N)$ bundles have instanton numbers that are in $\frac{1}{N}\mathbb{Z}$, and thus they are only defined when the level satisfies $k \in N\mathbb{Z}$. Since the fractional part of the instanton number came from the intersection number $\int B \wedge B$ of a 2-form $\mathbb{Z}_N$ gauge field, the fractional part of the instanton number will indeed depend on the existence of a spin structure: on non-spin manifolds we only have $I \in \frac{1}{2\mathbb{Z}}$. Thus $PSU(N)_k$ is spin if the level is an odd multiple of N ($k \in 2N\mathbb{Z} + N$), and non-spin if the level is an even multiple of N ($k \in 2N\mathbb{Z}$).

For example, take $PSU(2)_2 = SO(3)_2$: we obtain this from $SU(2)_2$ by identifying the representation 1 with the trivial representation. Now $SU(2)_2$ is the Ising theory, and 1 is the fermion. So, in order to identify 1 with 0, we need a spin structure. Thus $PSU(2)_2$ is a spin CS theory.

More generally, we know that the spin j line in $SU(2)_k$ has spin

$$\theta_j = \frac{j(j+1)}{k+2}. \tag{392}$$

When $k \in 2\mathbb{Z}$, we can take the quotient to $PSU(2)_k$. The maximal spin line with $j = k/2$ is the generator of the $\mathbb{Z}_2^{(1)}$ symmetry we need to quotient by, and from the above we see that it has spin $\theta_{k/2} = k/4$. Therefore for $k \in 4\mathbb{Z}+2$ the generator is a fermion, and so $PSU(2)_k$ is spin for $k \in 4\mathbb{Z}+2$. On the other hand, when $k \in 4\mathbb{Z}$ the generator is a boson, and so for such values of k, $PSU(2)_k$ is not spin.

17.5 $DW_{p,q}$ theory

In the notation of last time, the $DW_{p,q}$ theory is

$$\mathcal{L} = \frac{p}{4\pi} a \wedge da + \frac{q}{2\pi} a \wedge db. \tag{393}$$

Writing the action as an integral over a bounding 4-manifold tells us that these theories are spin when p is odd, and non-spin when p is even. This matches with the discussion of the 1-form symmetries of the theory in the previous diary entry: the generator for the $\mathbb{Z}_q^{(1)}$ symmetry shifting b is a boson and not anomalous, while the generator U_a for the $\mathbb{Z}_l^{(1)}$, $l \equiv \gcd(p,q)$ symmetry shifting a has spin

$$s[U_a] = \frac{p}{2l^2} \mod 1. \tag{394}$$

This means that the spin of l copies of the charge operator is $s[U_b^l] = p/2 \mod 1$. Since l copies of the charge operator gives a line that has trivial statistics with everything, we see that if $p \in 2\mathbb{Z}$ we have no problem, while if $p \in 2\mathbb{Z}+1$ then the theory has a transparent fermion. However since the theory is spin if $p \in 2\mathbb{Z}+1$ the transparent fermion is trivial, and so U_b^l is a trivial line, as required.

17.6 $SO(N)_K$

The CS action for $SO(N)_K$ is written as

$$S = \frac{k}{8\pi}\int_M \mathrm{Tr}[F_A \wedge F_A], \tag{395}$$

where the trace is taken in the vector representation. Note the factor of $1/8\pi$ in front, which differs from the usual $1/4\pi$ we've seen so far—the reason for this is ultimately that the reality of the $SO(N)$ representations ensures a doubling of the index of the Dirac operator on M, which by the index theorem lets us relate the η invariant and the CS action with an extra factor of $1/2$ compared to the normal definition—more on this in another diary entry.

Anyway, requiring that the integral be independent of the bounding 4-manifold means that for all closed M, we need

$$2\pi k\frac{1}{2\cdot 8\pi^2}\int_M \mathrm{Tr}[F_A\wedge F_A] = \pi k\int p_1(A), \tag{396}$$

where $p_1(A)$ is the first Pontryagin class. Now this is a legit $\mathbb{Z}$ characteristic class, but unlike the second Chern class, its quantization *does* depend on the type of manifold that it's on. In particular, the relation

$$p_1(A) = P(w_2(A)) + 2w_4(A) \mod 4 \tag{397}$$

tells us that $\int p_1(A) \in 2\mathbb{Z}$ on spin manifolds. Thus $k \in 2\mathbb{Z}$ theories make sense on any manifold and are not spin, while $k \in 2\mathbb{Z}+1$ theories are spin.

So in general, the coefficient in front of the CS Lagrangian ($k/4\pi$, $k/8\pi$, etc.) can be determined by looking at how the relevant characteristic class (Chern or Pontryagin) is quantized on different types of manifolds. We should pick it so that for all k, the 2+1D CS action is well-defined on spin manifolds (but may require special choices for k to be defined on non-spin manifolds).

Before wrapping up, note how we never needed to compute the spectrum of line operators to make these statements, although that's certainly one way to figure our whether they are spin or not. However, just knowing whether they are spin already tells us a nonzero amount about their spectrum: we already know that e.g. a transparent fermion cannot appear in the spectrum for $SU(N)_k$ or $U(1)_{2k}$, but that one must appear in $U(1)_{2k+1}$.

18 *Gauge (in)variance of non-abelian CS action and building instantons* ✓

Today is something simple that I hadn't done before. We will compute the gauge variation of the non-Abelian CS action explicitly, and use the result to show how one can build instantons on S^4.

We will work in math conventions where the gauge transformations act as

$$A \mapsto g^{-1}(A+d)g = g^{-1}Ag + \omega. \tag{398}$$

The Lagrangian in these conventions is then

$$\mathcal{L} = \frac{ik}{4\pi}\mathrm{Tr}\left(A \wedge dA + \frac{2}{3}A \wedge A \wedge A\right). \tag{399}$$

If we had $A \mapsto g^{-1}(A+id)g$ instead, we'd need to tack an i onto the $2/3$ (which can be seen by tracking the i through the following manipulations). The gauge variation of the first part is

$$\mathrm{Tr}[A \wedge dA] \mapsto \mathrm{Tr}\left[(A^g + \omega) \wedge (-\omega A^g + (dA)^g - A^g\omega - \omega \wedge \omega)\right], \tag{400}$$

where $A^g \equiv g^{-1}Ag$ and we've used $d\omega = -\omega \wedge \omega$. Now we use

$$\mathrm{Tr}[X \wedge Y] = (-1)^{|X||Y|}\mathrm{Tr}[Y \wedge X] \tag{401}$$

to write this term as

$$\mathrm{Tr}[A \wedge dA] \mapsto \mathrm{Tr}\left[A \wedge dA - 3\omega \wedge \omega \wedge A^g - \omega^{\wedge 3} + \omega(dA)^g - 2A^g \wedge A^g \wedge \omega\right]. \tag{402}$$

Clearly, the $A^{\wedge 3}$ term is going to be needed if we want to get something gauge invariant. This term changes as

$$\frac{2}{3}\mathrm{Tr}[A^{\wedge 3}] \mapsto \frac{2}{3}\mathrm{Tr}\left[A^{\wedge 3} + \omega^{\wedge 3} + 3(A^g \wedge \omega \wedge \omega + A^g \wedge A^g \wedge \omega\right]. \tag{403}$$

Adding these two contributions, we see that

$$\delta\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[A^g \wedge \omega \wedge \omega(2-3) + \omega^{\wedge 3}(-1+2/3) + A^g \wedge A^g \wedge \omega(2-2) + \omega \wedge (dA)^g\right]. \tag{404}$$

We can collect two of the surviving terms into a total derivative, so that

$$\delta\mathcal{L} = -\frac{k}{4\pi}d\mathrm{Tr}[\omega \wedge A^g] - \frac{k}{12\pi}\mathrm{Tr}[\omega^{\wedge 3}]. \tag{405}$$

Now the first term doesn't contribute to δS, since $\omega|_{\partial X} = 0$ if g is a gauge transformation and we fix ∂ conds on A ($X =$ spacetime). Since $\omega|_{\partial X} = 0$, the second term in $\delta\mathcal{L}$ is 2π times the winding number density for a map from the (compactification of) X to the target Lie group. Together, these terms tell us what kind of WZW needs to live on ∂X in order for gauge invariance to be manifest with free boundary conditions on A. The winding number (okay, "winding" is probably best reserved for $S^1 \to S^1$ situations—maybe "wrapping" would be pedantically better) term integrates to something in $\overline{\mathbb{Z}}$ (using (401) it's straightforward to show that $\mathrm{Tr}[\omega^{\wedge 4}] = 0$, so that the winding number density term is closed. Showing that the $1/12\pi$ coefficient is the correct normalization can be done by computing the integral for a fixed example field configuration; see one of the diary entries on WZW models for more detail. The winding number is $W = \frac{1}{24\pi^2}\int \mathrm{Tr}[\omega^{\wedge 3}]$.) Anyway, using this quantization on the integral of the $\omega^{\wedge 3}$ term, we see that the whole CS action is indeed gauge invariant modulo elements of $\overline{\mathbb{Z}}$.

18.1 Building instantonson S^4

Now we use this result to construct $SU(N)$ instantons on S^4. We will cover S^4 with two patches U_N and U_S, each homeomorphic to a 3-ball, with $U_N \cap U_S = S^3_{eq}$, the equatorial 3-sphere. We want to compute $I = \frac{1}{8\pi^2}\int \mathrm{Tr}[F \wedge F]$. Now on each patch U_N, U_S, the gauge field A is a well-defined 1-form, and so we can use Stoke's theorem. Thus

$$I = \frac{1}{8\pi^2}\left(\int_{U_S} \mathrm{Tr}[F_{A_S} \wedge F_{A_S}] + \int_{U_N} \mathrm{Tr}[F_{A_N} \wedge F_{A_N}]\right) = \frac{1}{2\pi}\int_{S^3_{eq}} (\mathcal{L}_{CS_1}[A_N] - \mathcal{L}_{CS_1}[A_S]), \tag{406}$$

where $\mathcal{L}_{CS_1}[A]$ is the CS action at level 1.

Now to create an instanton we glue up the sections of the gauge bundle on U_N to those on U_S through a large gauge transformation[36]. The existence of nontrivial gauge transformations in this case is guaranteed from $\pi_3(SU(N)) = \mathbb{Z}$. So, we choose the transition function g_{NS} such that g_{NS} is a nontrivial homotopy class in $\pi_3(SU(N))$. Then the gauge fields get glued together as $A_N = g^\dagger_{NS}(A_S + d)g_{NS}$. So

$$I = \frac{1}{2\pi}\int_{S^3_{eq}} (\mathcal{L}_{CS_1}[g^\dagger_{NS}(A_S + d)g_{NS}] - \mathcal{L}_{CS_1}[A_S]). \tag{409}$$

Now we can use our result for the gauge variation of $\mathcal{L}_{CS_1}$ to write

$$I = \frac{1}{2\pi}\int_{S^3_{eq}}\left(-\frac{1}{4\pi} d\mathrm{Tr}[\omega_{NS} \wedge g^\dagger_{NS} A_N g_{NS}] - \frac{1}{12\pi}\mathrm{Tr}[\omega^{\wedge 3}_{NS}]\right), \qquad \omega_{NS} = g^\dagger_{NS} dg_{NS}. \tag{410}$$

The first term dies, and so we get

$$I = -\frac{1}{24\pi^2}\int \mathrm{Tr}[\omega^{\wedge 3}_{NS}] = -W \in \mathbb{Z}, \tag{411}$$

which is (the negative of; sorry for the dumb sign choice) the winding number of g_{NS}.

Let's remind ourselves why the $1/24\pi^2$ coefficient is there, just for fun. We'll do the calculation for $SU(2)$ for simplicity. The winding number 1 map in $\pi_3(SU(N))$ is

$$g_{NS} = x_\mu \widetilde{\sigma}^\mu, \tag{412}$$

[36]This is exactly the same as how we build e.g. mangetic monopoles on S^2 for $U(1)$ gauge theory: we take the gauge field on the northern / southern hemispheres to be e.g.

$$A_N = \frac{1-\cos\theta}{2} d\phi, \quad A_S = \frac{-1-\cos\theta}{2} d\phi, \tag{407}$$

so that on the equator, $A_N - A_S = d\phi$, which means that on the equator, A_N and A_S differ by a large gauge transformation on the S^1 (also note how A_N is not well-defined at the south pole $\theta = \pi$, and A_S is not well-defined at the north pole $\theta = 0$). The "instanton" number is then

$$\frac{1}{2\pi}\left(\int_{U_N} F_{A_N} + \int_{U_S} F_{A_S}\right) = \frac{1}{2\pi}\int_{S^1_{eq}} (A_N - A_S) = 1, \tag{408}$$

where $U_{N/S}$ is the northern / southern hemisphere.

with $x_\mu \in S^3$ a unit vector and $\widetilde{\sigma}^\mu = (\mathbf{1}, iX, iY, iZ)$. Note that as required, $g_{NS}g_{NS}^\dagger = x_\mu x^\mu \mathbf{1} = \mathbf{1}$, and $\det g_{NS} = x_\mu x^\mu = 1$.

To evaluate the winding number integral we can either go to spherical coordinates and do lots of algebra, or use a clever trick. The clever trick is as follows: since g_{NS} is uniform on the S^3, we just need to compute the winding number density at a particular point on the 3-sphere, and then multiply the result by $2\pi^2 = \mathrm{vol}\,(S^3)$. Let us choose the north pole, where the field points in the $\mathbf{1}$ direction. Now ω is

$$\omega = (x_\mu \widetilde{\sigma}^\mu)^\dagger d(x_\mu \widetilde{\sigma}^\mu) = (x_\mu \widetilde{\sigma}^\mu)^\dagger (\sigma_\nu - x^\lambda \sigma_\lambda x_\nu)\, dx^\nu. \tag{413}$$

Evaluating this at the point $x_\nu = (1,0,0,0)$, the only derivatives that enter are ∂_i, where $i \in \{x,y,z\}$, since these are the coordinates in the tangent space at $(1,0,0,0)$. Thus ω becomes just $\sigma_i dx^i$, and the integrand is

$$\frac{i^3}{24\pi^2}\mathrm{Tr}[\sigma^i\sigma^j\sigma^k] dx^i \wedge dx^j \wedge dx^k = \frac{i^3}{4\pi^2}\mathrm{Tr}[XYZ] d^3x = \frac{1}{2\pi^2} d^3x, \tag{414}$$

which is just the pullback of the volume form on S^3. Multiplying this by the volume of S^3 we get 1, and so the $1/24\pi^2$ normalization was indeed correct.

What should we do if we want a winding number $W > 1$ map? To get winding number 1 we pulled back the volume form, so to get winding number W we should pullback W times the volume form. There are some physics books which say that for $\omega = g^{-1}dg$ we should keep the g configuration for $W = 1$ but replace g by g^W: this is wrong since then W appears in I as W^3 as I is cubic in the ω's. Instead, break up the spacetime S^3 as SS^2, where SS^{d-1} denotes the suspension of S^{d-1} by S^1. Let the coordinate on the S^1 that's doing the suspension be θ, and the coordinates on the S^2 be ϕ, ψ. Then if $g_1(\theta,\phi,\psi)$ is the winding number 1 configuration, the winding number W configuration is $g_W = g_1(W\theta, \phi, \psi)$. What's going on here is that the map completes winding number 1 during $\theta \in [0, \pi/W)$, and so the total winding number for $\theta \in [0,\pi]$ is W (we imagine composing W identity maps $S^3 \to S^3$ in a row, with the basepoint [$\theta = 0$ point] of each map being the terminal point [$\theta = \pi$ point] of the previous one). This is also clear from the integral formula: the wedge product means that the integrand contains one derivative for each of the coordinates on the S^3, so multiplying one of the coordinates by W will increase the integrand by a factor of W.

How do we make instantons for other gauge groups G? We do this by using a map $SU(2) \to G$ induced from a map $\mathfrak{su}(2) \to \mathfrak{g}$, which always exists because of roots. Specifically we can always get a winding number $W = 1$ map by taking the gauge configuration

$$g_{NS}^{SU(N)} = \mathbf{1}_{N-2} \oplus g_{NS}^{SU(2)}, \tag{415}$$

which is in $SU(N)$ as required.

We can also get winding number $W > 1$ maps by embedding the $SU(2)$ instanton inside of $SU(N)$ with a different representation. For example, for the map $SU(2) \to SU(3)$, we can choose to embed the $SU(2)$ either in the fundamental, or in the adjoint. The difference in the instanton number just comes from the difference in the trace of the generators; in this case we can get winding number $W = \pm 4$.

19 *Chirality of instanton-induced zero modes in four dimensions* ✓

Consider some massless fermions coupled a background gauge field. The index theorem tells us the net chirality ind $(\not{D}_A) = \nu_+ - \nu_-$ of the zero modes of the Dirac operator is determined by the instanton number (we are ignoring the gravitational contribution). However, it only tells us the difference in the left- and right-chirality zero modes; it does not tell us how many zero modes there are. Today we will argue however that in an instanton field such that $\nu_+ - \nu_- = n$, we actually generically have $\nu_+ = n, \nu_- = 0$.

Consider a zero mode of the Dirac operator with chirality $\pm$:

$$\not{D}_A(1 \pm \bar{\gamma})\psi_\pm = 0. \tag{416}$$

Now hit this with $\not{D}_A$, and use (the $\circ$ notation here is meant to emphasize that the derivatives in the left $\not{D}$ act on the A in the right $\not{D}$)

$$\begin{aligned}
\not{D}_A \circ \not{D}_A &= \partial_\mu \partial_\nu \gamma^\mu \gamma^\nu - A^a_\mu A^b_\nu \gamma^\mu \gamma^\nu T^a T^b - i A_\mu \partial_\nu \{\gamma^\mu, \gamma^\nu\} - i(\partial_\mu A_\nu)\gamma^\mu \gamma^\nu \\
&= \partial_\mu \partial^\mu - A_\mu A^\mu - i\partial_\mu A^\mu - \frac{1}{2} A^a_\mu A^b_\nu [\gamma^\mu, \gamma^\nu] T^a T^b - 2iA_\mu \partial^\mu - \frac{i}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu)\gamma^\mu \gamma^\nu \\
&= \partial_\mu \partial^\mu - A_\mu A^\mu - i\partial_\mu A^\mu - \frac{1}{2} i f^{abc} A^b_\mu A^c_\nu T^a - \frac{i}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu)\gamma^\mu \gamma^\nu \\
&= (\partial_\mu - iA_\mu)^2 - \frac{i}{2} F_{\mu\nu} \gamma^\mu \gamma^\nu,
\end{aligned} \tag{417}$$

where $(\partial_\mu - iA_\mu)^2$ means that the ∂_μ acts on the A_μ as well.

Using this, we have

$$0 = \not{D}_A \circ \not{D}_A (1 \pm \bar{\gamma})\psi_\pm = \left[(\partial_\mu - iA_\mu)^2 - \frac{i}{2} F_{\mu\nu}\gamma^\mu \gamma^\nu\right](1 \pm \bar{\gamma})\psi_\pm. \tag{418}$$

Now in Euclidean signature, $\bar{\gamma} = \prod_\mu \gamma^\mu$ (all of the γs, including $\bar{\gamma}$, are Hermitian). Thus we have

$$\gamma^\mu \gamma^\nu \bar{\gamma} = -\frac{1}{2} \epsilon^{\mu\nu\lambda\sigma} \gamma^\lambda \gamma^\sigma. \tag{419}$$

We now multiply the field strength term in (418) by $(1 \pm \bar{\gamma})/2$, which is allowable since it's a projector. Thus the putative zero mode satisfies

$$\left[(\partial_\mu - iA_\mu)^2 - \frac{i}{2}F_{\mu\nu}\Sigma_\pm^{\mu\nu}\right](1 \pm \bar{\gamma})\psi_\pm = 0, \tag{420}$$

where we have defined

$$\Sigma_\pm^{\mu\nu} = \gamma^\mu\gamma^\nu \mp \frac{1}{2}\epsilon^{\mu\nu\lambda\sigma}\gamma^\lambda\gamma^\sigma. \tag{421}$$

Note that Σ_+ is anti-self-dual while Σ_- is self-dual (note to self: missed a sign?):

$$(\star\Sigma_\pm)^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\sigma}\Sigma_\pm^{\lambda\sigma} = \mp\Sigma_\pm^{\mu\nu}. \tag{422}$$

Let us write $F = \mathcal{F}_+ + \mathcal{F}_-$, where $\mathcal{F}_+ = (F + \star F)/2$ is self-dual and $\mathcal{F}_- = (F - \star F)/2$ is anti-self-dual (we are in Euclidean signature, with $\star^2 = (-1)^{p(4-p)}$ on p-forms). Now, the contraction of a SD with an ASD form vanishes, since $A \wedge \star B = B \wedge \star A$ means that $A \wedge \star B = -A \wedge \star B$ if only one of A, B is ASD. Thus we can write

$$\left[(\partial_\mu - iA_\mu)^2 - \frac{i}{2}\mathcal{F}_\mp^{\mu\nu}\Sigma_\pm^{\mu\nu}\right](1 \pm \bar{\gamma})\psi_\pm = 0. \tag{423}$$

Now we consider a field for which

$$\frac{1}{8\pi^2}\int \mathrm{Tr}[F \wedge F] = n. \tag{424}$$

Then

$$n = \frac{1}{8\pi^2}\int \left(\mathrm{Tr}[\mathcal{F}_+ \wedge \mathcal{F}_+] - |\mathrm{Tr}[\mathcal{F}_- \wedge \mathcal{F}_-]|\right). \tag{425}$$

Here we have used $\int \mathcal{F}_+ \wedge \mathcal{F}_- = \int \star\mathcal{F}_+ \wedge \star\mathcal{F}_- = -\int \mathcal{F}_+ \wedge \mathcal{F}_- = 0$, and the fact that $0 < \int \mathcal{F}_- \wedge \star\mathcal{F}_- = -\int \mathcal{F}_- \wedge \mathcal{F}_-$. Thus we see that the self-dual part of the field strength contributes positively to the instanton number, while the anti-self-dual part contributes negatively. Both SD and ASD parts contribute positively to the $\int \mathrm{Tr}[F \wedge \star F]$ YM action. This means that if we want to look for a minimal-action configuration with a given instanton number, we can restrict ourselves to purely SD or purely ASD fields[37].

Let us suppose $n > 0$, so that the minimal action configuration has $\mathcal{F}_+ \neq 0, \mathcal{F}_- = 0$. Then we see that a putative ψ_+ zero-mode obeys

$$(\partial_\mu - iA_\mu)^2\psi_+ = 0, \tag{427}$$

[37] This is being a bit glib, since there may be instantons with $n \neq 0$ and field strengths which are not purely SD or ASD, but which are still solutions to the equations of motion (just not minimal action ones). For example, consider $SU(N \geq 4)$ gauge theory. Then we can consider the field configuration

$$A^{SU(N)} = 0_{N-4\times N-4} \oplus A^{SU(2)}_{SD,k} \oplus A^{SU(2)}_{ASD,l}, \tag{426}$$

where $A^{SU(2)}_{SD,k}$ is a configuration with self-dual field strength and $SU(2)$ instanton number k, and similarly for $A^{SU(2)}_{ASD,l}$. This configuration has instanton number $k - l$ and is a solution to the equations of motion, but is not purely SD or ASD.

since there is no anti-self-dual field strength contribution. Now $(\partial_\mu - iA_\mu)$ is anti-Hermitian, so $(\partial_\mu - iA_\mu)^2$ is Hermitian with $\mathbb{R}$ eigenvalues. Furthermore, it is negative-definite, since the eigenvalue of an eigenspinor of $(\partial_\mu - iA_\mu)$ is purely imaginary (by anti-Hermitian-ness). Thus since all the eigenvalues of $(\partial_\mu - iA_\mu)$ have the same sign and only the non-normalizable choice $\psi_+ = 0$ has a zero eigenvalue, there are no normalizable solutions to the above equation, and we conclude that there are no $+$ zero modes. Similarly, if we were to choose $n < 0$ so that the minimal action configuration for the gauge fields resulted in a purely ASD field strength, we would find $(\partial_\mu - iA_\mu)^2\psi_- = 0$, meaning that there are no $-$ zero modes.

So, at least for minimal-action purely SD / ASD field configurations, not only does the instanton number determine the net difference in $+$ and $-$ chirality zero modes, but it also tells us that $\nu_- = 0$ if the instanton number is positive, while $\nu_+ = 0$ if the instanton number is positive, and so the chiral difference in zero modes is actually equal to the (signed) total number of zero modes. Now we can imagine slowly deforming the background fields away from the minimal action purely SD / ASD configuration, while keeping the instanton number fixed. Since the number of $\pm$ chirality zero modes cannot change continuously, we expect that all configurations with a given instanton number, not just the purely SD / ASD ones, have a total number of zero modes equal to the chiral difference in zero modes.

Finally, a miscellaneous comment on reflections that I didn't know where to put. We know that reflections take left-handed fermions to right-handed ones—therefore in order for the zero mode situation to be invariant under reflections, we must have that a reflection takes a SD 2-form to an ASD 2-form. Indeed, this is true: one can see this by realizing that if we consider a reflection of the α coordinate, then

$$F_{\mu\nu} \mapsto [I_\alpha]_\mu^\lambda [I_\alpha]_\nu^\rho F_{\lambda\rho} \qquad [I_\alpha]_\mu^\lambda \equiv \delta_\mu^\lambda - 2\delta_\mu^\alpha \delta_\alpha^\lambda. \tag{428}$$

The result then follows after a little bit of algebra.

20 *GSD for K matrix CS theory from phase space* ✓

Today is a quickie: we show a cool way that I hadn't seen in the literature (I'm sure it exists somewhere though) for how to get the $|\det K|^g$ GSD on a Riemann surface of genus G for a CS theory with K-matrix K.

The strategy we will take will be to compute the volume of phase space. First we need the symplectic form. We get this by taking a variation of $K(a, da) = a_i \wedge da_j K^{ij}$, integrating by parts, and looking at the boundary term. Choosing a Cauchy slice Σ_g on which to quantize, the sympletic potential is

$$\omega = \frac{1}{4\pi}\int_{\Sigma_g} K(a,\delta a). \tag{429}$$

This gives us the symplectic potential as

$$\Omega = \frac{1}{4\pi}\int_{\Sigma_g} K(\delta a,\delta a) = \frac{1}{4\pi}\int_{\Sigma_g} K_{ij}\delta a^i \wedge \delta a^j. \tag{430}$$

Here the wedge product takes place in both actual space and in variational space. Thus e.g.

$$\delta a^i \wedge \delta a^j = \delta_1 a^i_x \delta_2 a^j_y - \delta_2 a^i_x \delta_1 a^j_y - \delta_1 a^i_y \delta_2 a^j_x + \dots \tag{431}$$

where δ_1, δ_2 are two (orthogonal) variations in variational space.

The space of solutions to the equations of motion is the space of flat connections on Σ_g. We can thus write

$$\delta a^i = \sum_{C_\mu \in H_1(\Sigma_g;\mathbb{Z})} \delta_\alpha \theta_\mu \widehat{C}_\mu, \tag{432}$$

where the Poincare dual is taken in Σ_g, so that $\widehat{C}_\mu$ is a flat 1-form. Here the coefficients $\theta_\mu \in [0,2\pi)$, since when $\theta_\mu \in \overline{\mathbb{Z}}$, $\theta_\mu \widehat{C}_\mu$ (no sum) is a large gauge transformation with $\overline{\mathbb{Z}}$ holonomy around the cycle C_μ.

Before plugging this in to the symplectic form, we note that for Σ_g of genus g, $H_1(\Sigma_g;\mathbb{Z}) = \mathbb{Z}^{2g}$, with generators $C_{0,\rho}, C_{1,\rho}$ for $\rho \in \mathbb{Z}_g$, such that

$$C_{\alpha,\rho} \cap C_{\beta,\sigma} = \delta_{\rho,\sigma}\delta_{\alpha,\beta+1}(-1)^\alpha, \tag{433}$$

with $\alpha, \beta \in \mathbb{Z}_2$ and consequently where $\beta+1$ is taken mod 2. The minus sign here is indeed because the $\widehat{C}_\mu$'s anticommute when wedged together.

Anyway, the point is that the homology of Σ_g is just g powers of the homology of the torus (since Σ_g is a connected sum). Thus, doing the integral, we can write Ω as

$$\Omega = \frac{1}{4\pi}K_{ij}\sum_{\rho=1,\dots,g}\sum_{\alpha=0,1}(-1)^\alpha \delta\theta^i_{\rho,\alpha}\wedge\delta\theta^j_{\rho,\alpha+1}, \tag{434}$$

where now $\wedge$ only takes place in variational space. Since the K matrix is symmetric (it has to be so that the off-diagonal parts add pairwise to give mutual CS terms that are properly quantized as $\sum_{i<j} a^i da^j/2\pi$), the antisymmetry of the sum on α cancels the antisymmetry of the wedge product in variational space, and we can write

$$\Omega = \frac{1}{2\pi}K_{ij}\sum_{\rho=1,\dots,g}\delta\theta^i_{\rho,0}\wedge\delta\theta^j_{\rho,1}. \tag{435}$$

Thus, for each torus ρ in the connected sum and for each flavor index i, the holonomy around the longitudinal cycle of the ρth torus, namely $\theta^i_{\rho,0}$, will constitute a position variable in the phase space. Its corresponding canonically conjugate momentum variable is then a linear combination (in flavor space) of the holonomies around the other cycle on the ρth torus, namely $\sum_j K_{ij}\theta^j_{\rho,1}$.

To find the GSD, we need to look at the symplectic volume of the ground state subspace. From the sum over ρ, we see that this factors into a product over each torus in the connected sum, each of which have the same phase space volume. Thus the GSD will be $GSD_{\Sigma_g} = (GSD_{T^2})^g$, where T^2 is the torus.

So, we just have to compute GSD_{T^2}. This is evidently

$$GSD_{T^2} = \int \bigwedge_{i=1,\dots,\dim K} \frac{K_{ij}}{4\pi^2}\delta\theta^i_0 \wedge \delta\theta^j_1, \tag{436}$$

where the integral is in variational space. Here we have remembered to divide by a further factor of 2π since the phase space volume form for a single degree of freedom is $dq \wedge dp/h$, and in our units $h = 2\pi$.

To see how $\det K$ arises from this, we just have to use the antisymmetry of the variational wedge product. Since $\delta\theta^i_\alpha \wedge \delta\theta^i_\alpha = 0$, the only terms which survive the product are those which contain the full volume form

$$V = \bigwedge_{i\in 1,\dots,\dim K} \delta\theta^i_0 \wedge \bigwedge_{j\in 1,\dots,\dim K} \delta\theta^j_1. \tag{437}$$

So, bringing the $\delta\theta$'s in our expression for GSD_{T^2} into this form,

$$GSD_{T^2} = \frac{1}{(4\pi^2)^{\dim K}} \int \bigwedge_{i=1,\dots,\dim K} \delta\theta^i_0 \wedge \sum_{\{j_\lambda\}\in\mathbb{Z}^{\dim K}_{\dim K}} K_{1j_1}\delta\theta^{j_1}_1 \wedge K_{2j_2}\delta\theta^{j_2}_1 \wedge\cdots\wedge K_{\dim K j_{\dim K}}\delta\theta^{j_{\dim K}}_1. \tag{438}$$

Moving all of the $\delta\theta_1$'s into order, which we do at the cost of an ϵ symbol, we have

$$GSD_{T^2} = \frac{1}{(4\pi^2)^{\dim K}} \int V \sum_{\{j_\lambda\}\in\mathbb{Z}^{\dim K}_{\dim K}} \epsilon^{j_1,\dots,j_{\dim K}} K_{1j_1}K_{2j_2}\cdots K_{\dim K j_{\dim K}} = \frac{\det K}{(4\pi^2)^{\dim K}} \int V. \tag{439}$$

Now since each of δ^i_0, δ^i_1 can be varied from 0 to 2π, the integral over V exactly cancels the factor in the denominator. Thus we get $GSD_{T^2} = |\det K|$, and hence $GSD_{\Sigma_g} = |\det K|^g$, as required.

21 *Topological terms from integrating out fermions in four dimensions and some characteristic class relations for vector bundles* ✓

Today's diary entry is a small compendium of results about what kind of θ terms are produced when integrating out massive fermions in four dimensions. A good reference for this diary entry is Witten's review article [?].

21.1 Complex / Dirac fermions

First we will look at the case where the fermions transform in a complex representation of the full symmetry group ("full" here meaning including spacetime symmetries) that we will assume to include a $U(1)$ fermion number conservation factor. In this case, there is no antisymmetric bilinear form we can use to construct a symmetric action involving a single fermion field, and so any symmetry-preserving Dirac operator appearing in the action will have to pair two independent fields $\bar{\psi}$ and ψ, with opposite charges under the $U(1)$ (this is what we mean when we say the fermion transforms in a complex representation: there is no antisymmetric bilinear form, invariant under the symmetry, that pairs a single fermion field with itself). Thus in the basis $(\psi,\, \bar{\psi})^T$, the Lagrangian will be purely off-diagonal:

$$\mathcal{L} = \begin{pmatrix} \psi & \bar{\psi} \end{pmatrix} \begin{pmatrix} 0 & i\not{D} \\ -i\not{D}^T & 0 \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix}, \tag{440}$$

where $i\not{D}$ is Hermitian. The minus sign and transpose are needed since the fermions need to be paired antisymmetrically: otherwise the action vanishes, since e.g.

$$\bar{\psi}_\alpha [i\not{D}]^{\alpha\beta} \psi_\beta = -\psi_\beta [i\not{D}^T]^{\beta\alpha} \bar{\psi}_\alpha. \tag{441}$$

When doing manipulations like this, we should remember that $\partial^T = \overleftarrow{\partial} = -\partial$. So in this notation, the derivative in the matrix $[\not{D}]^{\alpha\beta}$ always acts on the second index: thus the ψ in $[\not{D}]^{\alpha\beta}\psi_\beta$ is differentiated, while the ψ in $[\not{D}]^{\beta\alpha}\psi_\beta$ is not.

Anyway, the point of this is just to note that this structure for the Lagrangian means that integrating out the fermions ψ and $\bar{\psi}$ produces a determinant (of $i\not{D}$) rather than just a Pfaffian. Working in Euclidean signature and adding a mass m and a gauge field A produces a partition function $Z[A;m] = \det(i\not{D}_A - m)$, where m is real (in Euclidean time γ^0 is Hermitian, so the Lagrangian is $\bar{\psi}(i\not{D}_A - m)\psi$).

Since $\not{D}_A$ anticommutes with $\bar{\gamma}$ (recall that we are working in four dimensions), if ψ is an eigenspinor of $i\not{D}_A$ with non-zero eigenvalue, then $\bar{\gamma}\psi$ is a linearly independent eigenspinor with an eigenvalue of the opposite sign (they are linearly independent since they have different eigenvalues: $\langle \psi, i\not{D}_A\psi\rangle = \lambda\langle\psi,\psi\rangle \implies \langle \lambda, i\not{D}_A\bar{\gamma}\psi\rangle = -\lambda\langle\psi,\bar{\gamma}\psi\rangle = \langle\bar{\gamma}\psi, i\not{D}_A\psi\rangle = +\langle\psi,\bar{\gamma}\psi\rangle \implies \langle\psi,\bar{\gamma}\psi\rangle = 0$). Since they are linearly independent, $\psi_\pm \equiv (1\pm\bar{\gamma})\psi/2$ must be nonzero for both choices of sign: non-zero modes have support on both chirality subspaces, and so every non-zero-mode comes as a member of a positive-negative eigenvalue pair. Remember that here "zero-mode" means a mode which is annihilated by $i\not{D}_A$, not a mode which is annihilated by the Hamiltonian.

Now for the partition function: we have

$$\det(i\not{D}_A - m) = \left(\prod_{\lambda_j>0}(\lambda_j - m)(-\lambda_j - m)\right) m^{N_+ + N_-}, \tag{442}$$

where N_σ is the number of zero-modes with chirality σ. Note that when we say "number of zero-modes", we really mean "number of positive-charge zero modes": we are just computing the determinant of $i\not{D}_A$ as it acts on ψ, and not on $\bar{\psi}$. This number can be odd (and is the number relevant for computing the parition function), but the full number of zero modes, of both positive and negative charges, is always even. Indeed, if $\psi_\pm$ is a zero mode for the field ψ then its complex conjugate is a zero mode for the field $\bar{\psi}$[38] and so the full number of zero modes (for both the fields ψ and $\bar{\psi}$) is actually $2(N_+ + N_-)$.

The factor of $m^{N_+ + N_-}$ can also be understood from looking at how the zero modes get paired up by the mass term: each positively-charged zero mode ψ_+ appears in the path integral together with its negatively-charged partner as

$$\int \mathcal{D}\psi^\dagger \mathcal{D}\psi e^{-\int \psi_-^\dagger m \psi_+} = \int \mathcal{D}\psi^\dagger \mathcal{D}\psi \left(1 - \int \psi_-^\dagger m \psi_+\right) \propto m, \tag{443}$$

because of how Grassmann integration works. Thus we get a factor of m for each positively-charged zero mode, rather than for zero modes of all charges.

Anyway, note how the product in the expression for $\det(i\not{D}_A - m)$ is independent of the sign of m due to the pairing of opposite-chirality eigenmodes — the only dependence on $\mathrm{sgn}(m)$ comes from the zero modes. Thus we have

$$\frac{Z[A;m]}{Z[A;-m]} = (-1)^{N_+ + N_-} = (-1)^{N_+ - N_-} = e^{i\pi \mathrm{ind}\,(i\not{D}_A)}. \tag{444}$$

Now the index of the Dirac operator, for A a connection on a bundle E, is given by the index formula as

$$\mathrm{ind}\,(i\not{D}_A) = \int \widehat{A} \wedge \mathrm{ch}(E). \tag{445}$$

Here $\mathrm{ch}(E)$ is the Chern *character* of the bundle E, *not* the Chern class. So we can write this as

$$\mathrm{ind}\,(i\not{D}_A) = \int \widehat{A} \wedge e^{F_A/2\pi}. \tag{446}$$

Now the Dirac genus only involves Pontryagin[39] classes since it's a characteristic class in the real (involving traceless field strengths) tangent bundle. Thus only $4n$-dimensional classes contribute to $\widehat{A}$. For a 4-manifold M, we just need

$$\widehat{A} = 1 - \frac{1}{24}p_1(TM) + \dots, \qquad \mathrm{ch}(E) = \mathrm{Tr}[\mathbf{1}] + \mathrm{Tr}[F_A/2\pi] + \frac{1}{2}\mathrm{Tr}[F_A/2\pi \wedge F_A/2\pi] + \dots, \tag{447}$$

[38] The complex conjugate zero mode has opposite chirality: the associated zero mode of $\bar{\psi}$ is $\bar{\psi}_\pm = \gamma^0\psi_\pm^*$, and $\bar{\gamma}\bar{\psi}_\pm = -\gamma^0\bar{\gamma}\psi_\pm^* = \mp\bar{\psi}_\pm$. Note how here we are treating $\bar{\psi}$ as a column vector like ψ, which is a slightly more transparent thing to do since it really is an independent field.

[39] Spelling?! I can never remember, although in any case it seems like there is no standard romanization.

with the trace taken in the fundamental representation. Then

$$\text{ind }(i\not{D}_A) = -\frac{\dim(E)}{24}\int p_1(TM) + \frac{1}{8\pi^2}\int \text{Tr}[F_A \wedge F_A]. \tag{448}$$

Here $p_1(TM)$ is $\text{Tr}[R\wedge R]$ with some normalization that I can never remember. Writing the gravitational contribution in terms of the signature with $\int \widehat{A} = \sigma/8$, we have

$$\frac{Z[A;m]}{Z[A;-m]} = \exp\left(\frac{i\pi}{8\pi^2}\int \text{Tr}[F_A\wedge F_A] - i\pi\frac{\dim E}{8}\sigma\right), \tag{449}$$

where σ is the signature. Note that this is a completely non-perturbative result.[40]

On a spin manifold $\sigma \in 16\mathbb{Z}$, and so the signature part makes no contribution. On a general non-spin manifold, σ can be an arbitrary integer, and so the fact that ind $(i\not{D}_A) \in \mathbb{Z}$ tells us that for a spinc connection A,

$$\frac{1}{2}\int \frac{F_A}{2\pi}\wedge\frac{F_A}{2\pi} - \frac{\sigma}{8} \in \mathbb{Z}, \tag{450}$$

which means that for a spinc connection,

$$\frac{1}{2}\int \frac{F_A}{2\pi}\wedge\frac{F_A}{2\pi} \in \frac{1}{8}\mathbb{Z}. \tag{451}$$

Of course, this makes total sense: if A is spinc then $2F_A/2\pi$ is an integer class, and so we can write the above integral as $\frac{1}{8}\int(2F_A/2\pi)\wedge(2F_A/2\pi)$, which is then manifestly in $\frac{1}{8}\mathbb{Z}$.

21.2 Pseudoreal / Majoranna fermions

So far we've seen what topological term gets generated upon integrating out a Dirac fermion. What about a Majorana fermion? Our fermion χ will be assumed to transform in a pseudoreal representation of the full symmetry group, so that there exists an antisymmetric bilinear form J invariant under the symmetry, which allows us to construct a symmetric mass term via $\chi_\alpha J^{\alpha\beta}\chi_\beta$. Since J is an invariant form then so too is $J(i\not{D}_A)$[41] and so the pairing for the kinetic term is then

$$\mathcal{L} \supset \chi^T J(i\not{D}_A)\chi. \tag{452}$$

In what follows we will take J to be real, with $J^2 = -\mathbf{1}$, so that J can be thought of as a complex structure. I think that this can be done wolog (with this convention the Hermitian mass term is $i\chi^T J\chi$). Note that

$$\chi^T(Ji\not{D}_A)\chi = -\chi^T[Ji\not{D}_A]^T\chi \implies [Ji\not{D}_A]^T = -Ji\not{D}_A \implies [i\not{D}_A]^T J = Ji\not{D}_A. \tag{453}$$

[40] This is because locally, the integrand is a total derivative. If any Feynman diagram were to contribute to the effective action for A, it would then in momentum space contain a factor of p_{tot}, where p_{tot} is the sum of the momenta on all the external A legs attached to the diagram. Since momentum is conserved $p_{tot} = 0$, and therefore no Feynman diagram can contribute to this result. The only times when such topological terms can show up diagramatically is when there is an operator insertion (like j^μ_A) in the diagram to provide some extra momentum.

[41] Here it's best to think about $i\not{D}_A$ as being an operator rather than a bilinear form: J is used to raise / lower fermion indices, and $i\not{D}_A$ preserves the index placement. Thus J pairs two lower-index or two upper-index fermions, while $i\not{D}_A$ pairs and upper one with a lower one or vice versa.

Since $J^2 = -\mathbf{1}$, we then have

$$J[i\not{D}_A]^T J = -i\not{D}_A, \tag{454}$$

which indeed is telling us that J is a kind of complex structure. Now consider the operator JK, where K is complex conjugation. We have

$$JK[i\not{D}_A]^\dagger KJ = -i\not{D}_A. \tag{455}$$

Since $(JK)^2 = -\mathbf{1}$ (this is true even if we chose J to be Hermitian instead of anti-Hermitian, since then J would be complex) and $i\not{D}_A$ is Hermitian (in Euclidean signature), we have

$$JKi\not{D}_A = i\not{D}_A JK, \tag{456}$$

i.e. $JK\gamma_j = -\gamma_j JK$. Therefore pseudoreal fermions come equipped with an antiunitary action JK that commutes with the Dirac operator. Since $(JK)^2 = -\mathbf{1}$, we can then conclude that all eigenspinors of $i\not{D}_A$ come in pairs (related by JK) with identical eigenvalues: each eigenspinor χ comes with a linearly independent eigenspinor $JK\chi$, with the same eigenvalue λ, so then in the basis $(\chi,\ JK\chi)$, $i\not{D}_A$ has a block $\lambda\mathbf{1}_{2\times 2}$. Multiplying this by J, we see that a single eigenvalue λ of the Dirac operator then appears in the Lagrangian as

$$\mathcal{L} \supset \begin{pmatrix} \chi & JK\chi \end{pmatrix} \begin{pmatrix} 0 & \lambda \\ -\lambda & 0 \end{pmatrix} \begin{pmatrix} \chi \\ JK\chi \end{pmatrix}. \tag{457}$$

Summing over all such pairs $\chi, JK\chi$, we get a big antisymmetric matrix. If we expand $e^{-\int \mathcal{L}}$ to the order at which the Grassmann integration gives something nonzero, we see that the partition function becomes

$$\mathrm{Pf}(Ji\not{D}_A) = \pm[\det(i\not{D}_A)]^{1/2}, \tag{458}$$

since $\det J = 1$.

One other thing we will need to know is that the doubling of the spectrum because of JK also restricts to a doubling of each eigenspinor of definite helicity. So this means that all non-zero-mode eigenspinors of $i\not{D}_A$ are qudrupled (one for each chirality, and for each chirality two modes related by JK), while the zero-modes are merely doubled. To show this, we observe that since $(JK)^2 = -\mathbf{1}$ and since $\bar{\gamma}$ is a product of an even number of γ matrices,

$$JK\bar{\gamma}JK = JK\left(\prod_j \gamma_j\right)JK = -\prod_j (JK\gamma_j JK) = -\prod_j(-\gamma_j (JK)^2) = -\bar{\gamma} \implies JK\bar{\gamma} = \bar{\gamma}JK, \tag{459}$$

and so

$$[JK, \bar{\gamma}] = 0. \tag{460}$$

Thus each definite-chirality mode is doubled (note to self: do the Lorentzian-signature case as well).

Anyway, we can now compute the topological term induced by integrating out the fermions. Since the Pfaffian is the square root of the determinant, we can just naively take the square root of the partition function we found for the Dirac fermion (which did give a determinant), and write

$$\frac{Z[A;m]}{Z[A;-m]} = (-1)^{(N_+ + N_-)/2}. \tag{461}$$

Now because of the doubling of the spectrum due to the JK operator we discussed above, and because it commutes with $\bar{\gamma}$, we know that (unlike in the Dirac case), both N_+ and N_- must separately be even. Thus we have $(-1)^{N_-/2} = (-1)^{-N_-/2}$, and so

$$\frac{Z[A;m]}{Z[A;-m]} = (-1)^{\text{ind }(i\not{D}_A)/2} = \exp\left(\frac{1}{2}\left[\frac{i\pi}{8\pi^2}\int \text{Tr}[F_A \wedge F_A] - i\pi\frac{\dim E}{8}\sigma\right]\right). \tag{462}$$

This response is in keeping with the fact that a Majoranna is "half" a Dirac fermion.

For Majoranas, because the topological term involves $e^{\pi i p_1(B)/2}$ where B is either the gauge bundle or the tangent bundle, it evidently helps to have expressions for the Pontryagin classes mod 4. The most useful such relation is one derived in the entry on Pointryagin classes, namely

$$P(w_2(E)) = p_1(E) + 2w_4(E) \mod 4, \tag{463}$$

for any vector bundle (real or complex) E. For $SO(3)$ the $w_4(E)$ term is trivial, but in general it makes a contribution[42]. In any case, this means that we can write the topological response as

$$\frac{Z[A;m]}{Z[A;-m]} = \exp\left(\frac{i\pi}{2}\left[P(w_2(E)) + 2w_4(E) - \frac{\dim E}{8}\sigma\right]\right). \tag{464}$$

Here the second SW class of the gauge bundle is allowed to be non-trivial, as long as the second SW class of the tangent bundle is also non-trivial in the same way, so that $w_2(E) + w_2(TM) = 0 \mod 2$. Note that on a spin manifold, the even-ness of the intersection form means that $\int p_1(E) \in 2\mathbb{Z}$, and so the even-ness of the index of the Dirac operator means that $\frac{\dim E}{8}\sigma \in 2\mathbb{Z}$. In particular, taking just a single Majoranna fermion (not coupled to any gauge field) tells us that on a spin manifold, $\sigma \in 16\mathbb{Z}$. On the other hand, on a non-spin manifold, σ can be any integer, and this places constraints on how the Pointryagin term is quantized (although to work on a non-spin manifold, we need to be able to choose $w_2(E)$ in such a way that the full gauge- + spin-connection satisfies the cocycle condition).

The typical example for Majorana fermions is when A is a connection for an $SO(n)$ associated bundle. When n is even, $-\mathbf{1} \in SO(n)$ and $-\mathbf{1} \in \text{Spin}(4)$ act in the same way on fermions, and so our fermion is really coupled to a $[\text{Spin}(4) \times SO(n)]/\mathbb{Z}_2$ connection. Now for $n = \dim E$ even, the gravitational term is quantized in $\frac{1}{4}\mathbb{Z}$; thus in order to maintain the integrality of the index of the Dirac operator, the Pointryagin class must also be quantized

[42]Recall that the kth SW class is the obstruction to finding $\text{Rank}(E) - k + 1$ nowhere vanishing sections of E, and so they become trivial for $k > \text{Rank}(E)$. As a consequence, any $SO(3)$ bundle has $w_4 = 0$, since the SW classes $w_k(E)$ with $k > \text{Rank}(E)$ all vanish, and $\dim[SO(3)] = 3$.

An equivalent way to discuss w_k is to say that if the kth SW class is nonzero, then there is an obstruction to extending the trivialization of the bundle over the k-skeleton. But the converse is not true: there are plenty of cases where there is an obstruction to extending the trivialization, but the associated SW class vanishes. In general the obstruction to extend a G-bundle over the k-skeleton is captured by $\pi_{k-1}(G)$. This could fail to get detected by the SW classes either due to the fact that homotopy groups carry more data than cohomology groups, or because the obstructions always vanish mod 2. For example, the obstruction to extending an $SO(3)$ bundle over the 4-skeleton is non-zero as $\pi_3(SO(3)) = \mathbb{Z}$, even though $w_4 = 0$ because $4 > 3$. Moreover, no mod 2 class could detect this obstruction, since $\pi_3(SO(3))$ should really be thought of as $2\mathbb{Z}$. This is because elements in $\pi_3(SO(3))$ descend from elements in $\pi_3(S^3) = \mathbb{Z}$ from the map $S^3 \to SO(3)$, which is a double cover. Therefore a winding number 1 map in $\pi_3(S^3)$ maps onto a winding number 2 map in $\pi_3(SO(3))$; hence $\pi_3(SO(3)) = 2\mathbb{Z}$.

in $\frac{1}{4}\mathbb{Z}$, in such a way that the gauge and gravitational contributions add to give something in $2\mathbb{Z}$. This quantization makes sense, since when passing from $SO(n) \to SO(n)/\mathbb{Z}_2$ for n even, the quantization of the instanton number (alias $\int p_1(E)$) changes by a factor of 1/4 on a general non-spin manifold (see a diary entry in last year's diary for a discussion of why). By contrast when n is odd, there is no $\mathbb{Z}_2$ identification between the gauge and spin connections, and in order for our fermion to be well-defined, we need to work on a spin manifold.

22 *How spin CS theory sees the spin structure* ✓

Consider a CS theory which is spin. How does the partition function $Z[\sigma]$ depend on the spin structure σ? That is, under $\sigma \mapsto \sigma + \beta$ with $\beta \in H^1(X;\mathbb{Z}_2)$, how does $Z[\sigma]$ change? Today we will answer this question by attempting to give a physicist's interpretation of the results in the very nice math paper [?]. I haven't fully understood [?], but I think the results below are likely to be a physicist's translation of a subset of its results.

To illustrate the idea, we will first show how $U(1)_1$ (and by extension, $U(1)_k$) depends on a spin structure; later we will generalize to non-Abelian groups. Our notation will be such that X is a closed 3-manifold, Y is an open 4-manifold with $\partial Y = X$, and Z is a closed 4-manifold.

First let us recall the usual story. The action $\int_Y \mathrm{ch}_2$ (with ch_2 the second Chern character) is independent of the choice of Y mod $\overline{\mathbb{Z}}$ only if we restrict ourselves to choices of X and Y with spin structures, because of the resultant even-ness of the intersection form on any $Y \sqcup \bar{Y}'$ with $\partial Y = \partial Y'$. Making such a restriction is permissible since $\Omega_3^{\mathrm{Spin}}(pt) = 0$, so any choice of spin X always has a spin Y that it bounds.[43] In this way of thinking about things, the dependence of the action on the spin structure enters because the spin structure controls what sort of choices we are allowed to make for Y. Unfortunately though the bounding manifolds can become complicated, making explicit calculations of the spin structure rather difficult (consider e.g. the manifold with the RRR spin 3-torus as a boundary — understanding an explicit construction of this manfiold is prohibitively mathematical for most physicists).

[43] Contrast this with e.g.

$$\Omega_2^{\mathrm{Spin}}(pt) = \Omega_1^{\mathrm{Spin}}(pt) = \mathbb{Z}_2, \tag{465}$$

with the former generated by the RR torus and the later generated by the R circle.

However, I think a slightly different logic may be possible here. Namely, it may be possible to get away with only using a spin structure on X to define the action, and to still allow the extending manifolds Y to be *arbitrary* (i.e., not necessarily spin) 4-manifolds. If this is true, then when we say "spin CS theory", we mean that the theory needs a choice of spin structure on the 3-manifold X, but *not* on the various 4-manifolds Y that X bounds.

The claim is that we can *define* the spin CS action as

$$S[w_2] = \pi \int_Y \frac{F}{2\pi} \wedge \left(\frac{F}{2\pi} + w_2(TY) \right), \tag{466}$$

where $w_2(TY)$ is the second SW class on Y,[44] which may be non-zero. The spin structure $\mathcal{S}$ on X is parametrized by the trivialization

$$w_2(TY)|_{\partial Y} = d\mathcal{S}, \qquad \mathcal{S} \in C^1(X; \mathbb{Z}_2) \tag{467}$$

but as we said, the exactness (on the boundary) need not mean that Y is spin.

The reason for defining $S[w_2]$ in this ways is that this definition gives an $e^{iS[w_2]}$ that is independent of Y for *any* choice of Y (regardless of whether the spin structure on X extends into Y), since

$$S[w_2(TY)] - S[w_2(TY')] = \pi \int_{Y \sqcup \bar{Y}'} \frac{F}{2\pi} \wedge \left(\frac{F}{2\pi} + w_2 \right) \cong 2\pi \int_{Y \sqcup \bar{Y}'} \frac{F}{2\pi} \wedge \frac{F}{2\pi} \in \overline{\mathbb{Z}}, \tag{468}$$

where the $\cong$ means working mod $\overline{\mathbb{Z}}$ (which is possible since $a \cup w_2 \cong a \cup a$ mod 2). The independence wrt the choice of Y means that the action actually only cares about $w_2|_{\partial Y}$, i.e., only cares about the spin structure on X.

Now consider changing the spin structure. The difference between any two spin structures on X is given by an element in $H^1(X; \mathbb{Z}_2)$. So, let $\beta \in H^1(X; \mathbb{Z}_2)$, and consider extending β into Y. The claim is that making this shift induces the corresponding shift

$$w_2 \mapsto w_2 + d\beta. \tag{469}$$

Of course if β is nontrivial we will not be able to keep β closed in Y if the Poincare dual cycle $\widehat{\beta}$ satisfies $\widehat{\beta} \sim 0$ in $H_{\dim Y - 1}(Y; \mathbb{Z}_2)$, so this will be a non-trivial change in w_2 in general.

But wait, even if $d\beta \neq 0$, why do we say that this shift is nontrivial? Don't we only care about the cohomology class of w_2? The answer to this is that the cohomology relevant to determining the spin structure (or lack thereof) of Y is in fact *relative* cohomology—w_2 is determined by a class in $H^2(Y, \partial Y; \mathbb{Z}_2)$, not $H^2(Y, \partial Y; \mathbb{Z}_2)$. So, while $d\beta$ is exact in $C^2(Y)$, it is *not* exact in $C^2(Y, \partial Y)$, since $\beta|_{\partial Y} \neq 0$, and thus shifting w_2 by $d\beta$ is nontrivial.

A more easily visualized way of describing how w_2 transforms is via Poincare duality (recall that $\widehat{w}_2$ is the submanifold on which the framing degenerates; fermions can be defined only in $Y \setminus \widehat{w}_2$). The shift of w_2 by $d\beta$ translates into

$$\widehat{w}_2 \mapsto \widehat{w}_2 + \partial_R \widehat{\beta}, \tag{470}$$

[44] Or rather, its lift to a cocycle on $H^2(X; \mathbb{Z})$ so that it may cup with the first Chern class c_1; we will not bother to make this distinction.

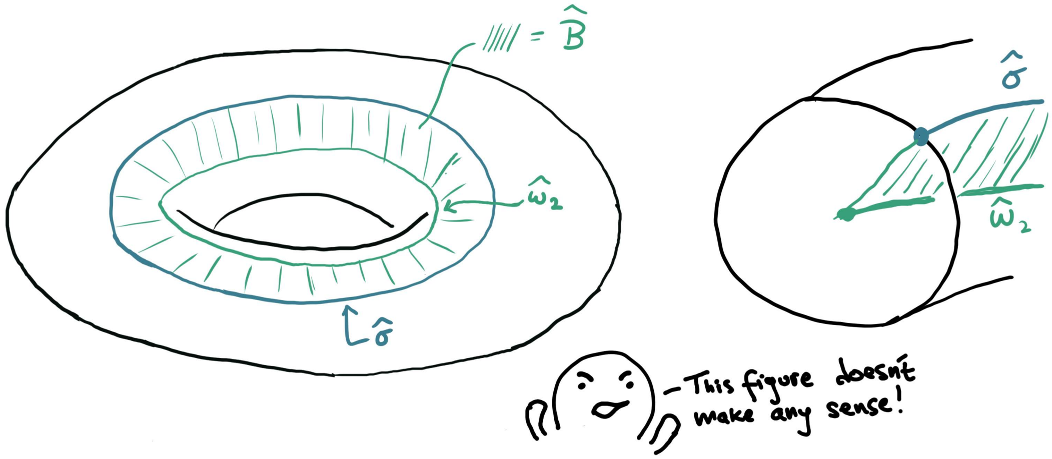


Figure 2: An example of the geometric picture of a spin torus that's been filled in "the wrong way". The left part is the full torus; the right is supposed to show a cut-out view; sorry for the rather poor artistry.

where β is still the 1-cochain extension into Y and ∂_R is the relative boundary operaton

$$\partial_R : C_n(Y, \partial Y) = C_n(Y)/C_n(\partial Y) \to C_{n-1}(Y, \partial Y), \tag{471}$$

which selects out the boundary of the submanifold it acts on, modulo the part of that submanifold which is contained in ∂Y. That is,

$$\partial_R \widehat{\beta} = \partial \beta + \alpha, \qquad \alpha \subset \partial Y. \tag{472}$$

Since Lefschetz duality gives the isomorphisms

$$H_p(Y, \partial Y; R) \cong H^{\dim Y - p}(Y; R), \qquad H_p(Y; R) \cong H^{\dim Y - p}(Y, \partial Y; R), \tag{473}$$

we see that if we identify w_2 as a class in $H^2(Y, \partial Y; \mathbb{Z}_2)$ then Lefschetz duality relates it to a class in $H_{\dim Y - 2}(Y; \mathbb{Z}_2)$: since $\partial_R \widehat{\beta}$ can be nontrivial in the later group, this shift is indeed a nontrivial shift of w_2.

All this is perhaps best illustrated by a picture of an example. Consider figure 2. The left part shows a solid torus. The boundary of this torus has R spin structure around the cycle transverse to the cycle $\widehat{\sigma}$ marked in blue (which is drawn on the surface of the solid torus) and N spin structure around the other cycle. This spin structure cannot be extended into the solid torus; the obstruction is measured by the class w_2. The dual 1-cycle $\widehat{w}_2$ is drawn in green, and is supposed to be inside the torus: it represents a "vortex" where the spin framing can not be defined. Note that

$$\partial \widehat{\mathcal{B}} = \widehat{w}_2 + \widehat{\sigma}, \qquad \partial_R \widehat{\mathcal{B}} = \widehat{w}_2, \tag{474}$$

where $\widehat{\mathcal{B}}$ is the "branch cut sheet" indicated in the figure. Now $\sigma \in H^1(T^2; \mathbb{Z}_2)$ can be thought of as a spin structure on T^2 (relative to the NN torus). When it is extended into the bulk, it can no longer be flat. In fact, when extended into the bulk, it becomes precisely $\mathcal{B}$. Accordingly, $d\mathcal{B} = w_2$,[45] and so even though w_2 is exact, Y is not spin. Also, we see that a transformation that changes the spin structure of the bounding T^2 to NN can be realized by sending $w_2 \mapsto w_2 + d\sigma = 0$ (working mod 2 and yes, using σ to denote the extension of the boundary σ into the bulk, even though earlier we called it $\mathcal{B}$—sue me). We see in this case that w_2 is exact, and as such is entirely determined by its value on the boundary—or, since it is still exact on the boundary, it is entirely determined by the spin structure (choice of trivialization of w_2) on the boundary. Since the CS action we have defined only cares about $w_2|_{\partial Y}$, the situation where $w_2 = d\beta$ is determined just from the boundary spin structure is generic for our problem.

Just to be totally clear, when we say that a spin structure is a choice of trivialization $w_2 = d\mathcal{S}$, we are working on a closed manifold. On an open manifold, a spin structure is a choice of trivialization $w_2 = d\mathcal{S}_R$, where $\mathcal{S}_R \in H^1(Y, \partial Y; \mathbb{Z}_2)$ is a *relative* 1-cochain. In the example we above, we indeed saw that just having w_2 be exact is not enough to get a spin structure—instead, w_2 must be relatively exact.

Anyway, the whole point of this rather garrulous discussion is that it tells us how our CS action changes when we change the spin structure on X. If we change the 1-cocycle $\sigma \in H^1(X; \mathbb{Z}_2)$ parametrizing the spin structure[46] by $\sigma \mapsto \sigma + \beta$, then $w_2 \mapsto w_2 + d\beta$ means that the action for $U(1)_k$ changes as

$$(\delta S)[\beta] = k\pi \int_X c_1(E) \cup \beta, \tag{475}$$

where E is the gauge bundle. The physical meaning of this $\int c_1(E) \cup \beta$ term is very intuitive: we may write it as $\int_{\widehat{c}_1(E)} \beta$, where $\widehat{c}_1(E)$ is the 1-chain corresponding to the worldlines of any magnetic monopoles that happen to be present. Therefore, we can interpret the (change in the) spin structure as producing a coupling of the current for the magnetic symmetry (namely $\star F/2\pi$) to the spin structure, where the spin structure is thought of as a background $\mathbb{Z}_2$ gauge field.[47] The fact that the spin structure of the manifold acts as a background $\mathbb{Z}_2$ gauge field tells us that when monopoles travel around cycles where the spin framing rotates, they pick up minus signs if k is odd—that is, for odd k, the monopoles are fermions. Of course, physically we know that this is what happens by flux attachment arguments; this procedure makes it rigorous.

Now we generalize to non-Abelian CS theory. This is quite straightforward—there is only one possible generalization of the result for the $U(1)$ case, viz. that the action changes by[48]

$$(\delta S)[\beta] = k\pi \int_X w_2(E) \cup \beta, \tag{476}$$

[45] This is not in contradiction with $\partial\widehat{\mathcal{B}} = \widehat{w}_2 + \widehat{\sigma}$ since $\partial \leftrightarrow d$ under Lefschetz duality only when ignoring contributions from boundary chains.

[46] Spin structures are really an $H^1(X; \mathbb{Z}_2)$ torsor; I will ignore this from now on.

[47] After all, to change the spin structure we may tensor either the spinor bundle or the gauge bundle with a real line bundle whose first SW class is β, so that spin structures indeed behave as background $\mathbb{Z}_2$ gauge fields.

[48] Again, we are being a bit fast and loose with the precise product operations between the various cochains appearing here—sorry.

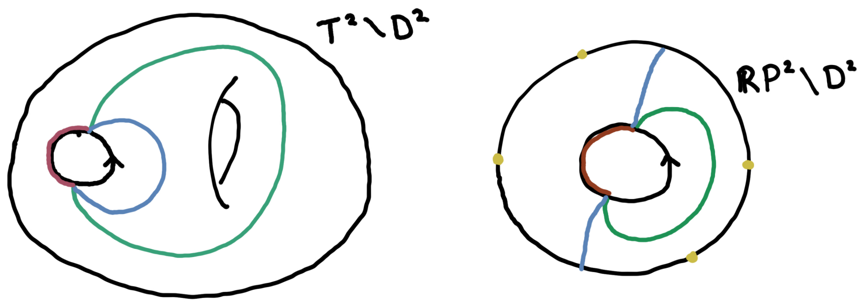


Figure 3

where $w_2(E)$ is the second SW class of the gauge bundle (for complex vector bundles this is the mod-2 reduction of the first Chern class, and so reduces to the above result in the case when E is a complex line bundle). This may be derived by defining the spin CS action as

$$S[w_2] = kS_{CS} + \pi k \int_Y w_2(E) \cup w_2(TY). \tag{477}$$

This definition of the action works because as before, the extra term allows the action to be independent of the choice of Y. For example, suppose that the gauge bundle E is real (e.g. if the gauge group is $SO(N)$). Then $S_{CS} = \pi p_1(E)$ where $p_1 = \frac{1}{2}\mathrm{Tr}(F/2\pi \wedge F/2\pi)$ is the Pointryagin density, and the independence of $S[w_2]$ on Y is demonstrated by (letting $Z = Y \sqcup \bar{Y}'$ as usual)

$$S[w_2(TY)] - S[w_2(TY')] = \pi k p_1(E_Z) + \pi k \int_Z w_2(E_Z) \cup w_2(TZ) \in \overline{\mathbb{Z}}, \tag{478}$$

where in the last step we used

$$w_2(E_Z) \cup w_2(TZ) \cong w_2(E_Z)^2 \tag{479}$$

and

$$p_1(E_Z) \cong w_2(E_Z)^2, \tag{480}$$

where again $\cong$ means mod 2, and where we are not bothering to distinguish the Pointryagin square from the cup square for notational simplicity.

Note that as a byproduct we may conclude that all CS theories based on simply connected Lie groups (so that $w_2(E) = 0$—these theories do not have monopoles) are non-spin for any value of the level. *The spin structure dependence of CS theories thus comes entirely from needing to give a framing to monopole worldlines*, and so gauge groups without monopoles will never have spin CS actions. For example, $SU(N)_k$ is never spin. On the other hand, if $\pi_1(G) \neq 0$, then whether or not the theory is spin will depend on the level (in general, spin if the level is odd, non-spin if the level is even).

23 *θ angles and deconfinement in two dimensions from the strong coupling expansion on the lattice ✓*

Today we're going to elaborate on some of the details implicitly contained in Seiberg's old paper on θ angles and lattice gauge theory [**?**]. The goal will be to consider Euclidean lattice $U(1)$ gauge theory with a θ term in two dimensions, and to find the free energy and Wilson loop vevs in the strong coupling limit. Results for the (weak-coupling) continuum limit can then be argued for schematically.

First we need to write down an appropriate Euclidean lattice action. We write it as

$$S = \beta S_{\text{matter}} + \frac{i\theta}{2\pi}\sum_{\square}\Phi_\square, \tag{481}$$

where the flux $\Phi_\square$ is, for a plaquette $\square$ with bottom-left corner at the lattice site i,

$$\Phi_\square = -i\ln e^{i\oint_{\partial\square}A} = [A_x(i_x,i_y)+A_y(i_x+1,i_y)-A_x(i_x+1,i_y+1)-A_y(i_x,i_y+1)]_{[-\pi,\pi)}, \tag{482}$$

where the subscript on the brackets indicates that we choose a branch of the logarithm such that $\Phi_\square$ is always valued in $[-\pi,\pi)$. In keeping with this branch, we also choose our lattice gauge fields $A_\mu(i)$ to be valued in $[-\pi,\pi)$;[49] in this notation the A_μ fields are just the logarithms of the $U(1)$ variables on each link for the aforementioned branch of the logarithm, and in particular they do not live in $\mathbb{R}$. Now with this convention $\sum_\square \Phi_\square \in 2\pi\mathbb{Z}$ when summed over the whole lattice. This means $\theta\sim\theta+2\pi$, and in what follows we will always take $|\theta|\le\pi$.

Let's first calculate the free energy $\mathcal{F}[\theta]$ in the infinite gauge coupling limit $\beta=0$, where the action is purely the topological term. The free energy won't depend on the boundary conditions for the lattice in the thermodyamic limit, and so to be consist with the Wilson line calculations we'll do later, we take the lattice to be a cylinder of length L_x in the x direction, and circumference L_y in the y direction (the compact direction). We can then fix a gauge such that $A_x=0$ (we can't choose $A_y=0$ since $H^1(\text{Cyl};\mathbb{R})\neq 0$; the holonomy $e^{i\oint dy\,A_y}$ is gauge invariant). The partition function is

$$Z[\theta] = \int\prod_{x,y=0}^{L_x,L_y} d\gamma_x^y \, \exp\left(\frac{i\theta}{2\pi}\sum_\square \Phi_\square\right), \tag{483}$$

where we have written $\gamma_x^y = A_y(x,y)$ to save on the notation.

Note that the γ_x^y for different y are completely decoupled in this gauge: thus we can write

$$Z[\theta] = (Z_1[\theta])^L_y = \left(\int\prod_{x=0}^{L_x} d\gamma_x\, e^{i\bar\theta\sum_\square\Phi_\square}\right)^{L_y}, \qquad \bar\theta\equiv\theta/2\pi. \tag{484}$$

[49]The reason that we choose the $[-\pi,\pi)$ branch instead of $[0,2\pi)$ is because it will make it easier to work with the θ angle, on which the free energy will depend on in a way that's symmetric about $\theta=0$, not $\theta=\pi$.

We start with the integral over γ_0, since γ_0 only appears in one place. Thus

$$Z_1[\theta] = \int \prod_{x=0}^{L_x} d\gamma_x \, \exp\left(i\bar{\theta}[\gamma_1 - \gamma_0]_{[-\pi,\pi)}\right)(\dots), \tag{485}$$

where $\dots$ involves things that don't contain γ_0.

The integral is easy to get confused about, so we will be pedantic. Suppose first that $\gamma_1 > 0$. Then the []s will come into affect when $\gamma_1 - \gamma_0 = \pi$, i.e. when $\gamma_0 = \gamma_1 - \pi$. Thus

$$\begin{aligned}
\int_{-\pi}^{\pi} d\gamma_0 \, e^{i\bar{\theta}[\gamma_1-\gamma_0]_{[-\pi,\pi)}} &= \int_{-\pi}^{\gamma_1-\pi} d\gamma_0 \, e^{i\bar{\theta}(-2\pi+\gamma_1-\gamma_0)} + \int_{\gamma_1-\pi}^{\pi} d\gamma_0 \, e^{i\bar{\theta}(\gamma_1-\gamma_0)} \\
&= \frac{i}{\bar{\theta}}\left(e^{i(-\theta+\gamma_1\bar{\theta})}(e^{i\bar{\theta}(\pi-\gamma_1)} - e^{i\theta/2}) + e^{i\bar{\theta}\gamma_1}(e^{-i\theta/2} - e^{i\bar{\theta}(\pi-\gamma_1)})\right) \\
&= \frac{i}{\bar{\theta}}\left(e^{-i\theta/2} - e^{-i\theta/2+i\gamma_1\bar{\theta}} + e^{i\bar{\theta}\gamma_1-i\theta/2} - e^{i\theta/2}\right) \\
&= \frac{2}{\bar{\theta}}\sin(\theta/2).
\end{aligned} \tag{486}$$

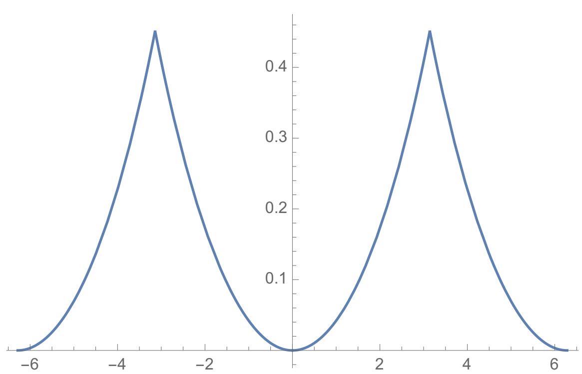
Since this is independent of γ_1, we of course get the same result if we take $\gamma_1 < 0$. The important thing here is that after γ_0 is integrated out, the resulting partition function looks *exactly* like the one for the partition function of a system with $L_x \mapsto L_x - 1$, multiplied by a factor of $2\sin(\theta/2)/\bar{\theta}$. Thus we can simply write[50]

$$Z[\theta] = \left(\frac{\sin(\theta/2)}{\theta/2}\right)^{L_x L_y}, \tag{487}$$

so that the free energy per unit area is

$$\mathcal{F}[\theta] = -\ln\left(\frac{\sin(\theta/2)}{\theta/2}\right). \tag{488}$$

Note that $\mathcal{F}[\theta] = \mathcal{F}[-\theta]$ as required. However, we also know that $\mathcal{F}[\theta] = \mathcal{F}[\theta \pm 2\pi]$: imposing this condition leads to a non-analyticity of $\mathcal{F}[\theta]$ at $\theta = \pm\pi$, which comes from the twofold GSD at the points where $\theta \in \pi(2\mathbb{Z}+1)$. $\mathcal{F}[\theta]$ looks like



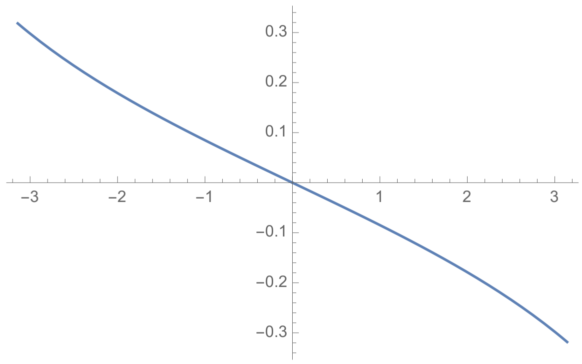
(489)

[50]we are replacing $\bar{\theta} \to \theta$ here, since in retrospect it is nicer to have a normalization factor of $1/2\pi$ multiplying each $d\gamma_x^y$ in the partition function.

Now $\langle \Phi \rangle$ is odd under P and C (but unlike in 3+1D it is even under T if $A \mapsto -A$ as a differential form under T). The vev of the topological charge density is

$$\langle Q_{top} \rangle = -i\partial_\theta \mathcal{F}[\theta] = i\frac{\theta}{\sin(\theta/2)}\left(\frac{\cos(\theta/2)}{2\theta} - \frac{\sin(\theta/2)}{\theta^2}\right) = i\left(\frac{\cot(\theta/2)}{2} - \frac{1}{\theta}\right), \tag{490}$$

which is non-vanishing (equal to $\mp i/\pi$) at the P symmetric points $\theta = \pm\pi$ (if we approach them from within $|\theta| \leq \pi$): thus, we have SSB for P (as well as for C). $\langle Q_{top} \rangle \neq 0$ just indicates the presence of a nonzero background electric field. As expected, the points $\pm\pi$ are the points of largest $|\langle Q_{top} \rangle|$: the plot of $\langle Q_{top} \rangle$ looks like



(491)

Now let's diagnose confinement by looking at $\langle W_q(C_x)W_{-q}(C_{x+L})\rangle$, where $q \in \mathbb{Z}$ is a charge, C_x is a curve wrapping the y direction at x-coordinate x, and L is the lattice distance between the two Wilson lines. Because the partition function in the absence of Wilson lines factorizes as a product of partition functions on each x coordinate (due to the $\beta \to 0$ limit we've taken), when we calculate the expectation value we can, without loss of generality, take $x = 0$ and $L_x = L$. Therefore

$$\langle W_q(C_x)W_{-q}(C_{x+L})\rangle = \frac{1}{Z[\theta]}\int \prod_{x,y=0}^{L,L_y} \frac{d\gamma_x^y}{2\pi} \exp\left(iq\sum_{y=0}^{L_y}\gamma_0^y - iq\sum_{y=0}^{L_y}\gamma_L^y + i\bar{\theta}\sum_{x,y=0}^{L-1,L_y}[\gamma_{x+1}^y - \gamma_x^y]_{[-\pi,\pi)}\right). \tag{492}$$

As before, the γ_x^y variables for different y are completely independent—the only variables that are linked together are the ones in the brackets. So then

$$\langle W_q(C_x)W_{-q}(C_{x+L})\rangle = \frac{1}{Z[\theta]}\left[\int \prod_{x=0}^{L} \frac{d\gamma_x}{2\pi} \exp\left(iq(\gamma_0 - \gamma_L) + i\sum_{x=0}^{L-1}\bar{\theta}[\gamma_{x+1}^y - \gamma_x^y]_{[-\pi,\pi)}\right)\right]^{L_y}. \tag{493}$$

Let's look at just the part involving γ_0. Assume for simplicity that $\gamma_1 > 0$. Then similarly to before, the relevant integral is

$$\int_{-\pi}^{\gamma_1-\pi} d\gamma_0\, e^{iq\gamma_0+i\bar\theta(-2\pi+\gamma_1-\gamma_0)} + \int_{\gamma_1-\pi}^{\pi} d\gamma_0\, e^{iq\gamma_0+i\bar\theta(\gamma_1-\gamma_0)} = \frac{e^{i(-\theta+\bar\theta\gamma_1)}}{i(q-\bar\theta)}\left(e^{i(q-\bar\theta)(\gamma_1-\pi)} - e^{-i\pi(q-\bar\theta)}\right) + \frac{e^{i\bar\theta\gamma_1}}{i(q-\bar\theta)}\left(e^{i\pi(q-\bar\theta)} - e^{i(q-\bar\theta)(\gamma_1-\pi)}\right) = -(-1)^q\frac{2e^{iq\gamma_1}\sin(\theta/2)}{q-\bar\theta}. \tag{494}$$

Therefore after integrating out γ_0, we get (assuming for simplicity that $L_y \in 2\mathbb{Z}$ to get rid of the $(-1)^q$)

$$\langle W_q(C_x)W_{-q}(C_{x+L})\rangle = \frac{1}{Z[\theta]}\left[\frac{2\sin(\theta/2)}{q-\bar\theta}\int\prod_{x=1}^{L}\frac{d\gamma_x}{2\pi}\,\exp\left(iq(\gamma_1-\gamma_L)+i\sum_{x=0}^{L-1}\bar\theta[\gamma^y_{x+1}-\gamma^y_x]_{[-\pi,\pi)}\right)\right]^{L_y}. \tag{495}$$

Note that this looks, up to the multiplicative factor, exactly the same as what we had before, just with the left Wilson line moved one lattice spacing closer to the right one. Therefore we can easily iterate and integrate out the rest of the γ_xs:

$$\langle W_q(C_x)W_{-q}(C_{x+L})\rangle = \frac{1}{Z[\theta]}\left[\frac{2\sin(\theta/2)}{2\pi q-\theta}\right]^{LL_y}. \tag{496}$$

This means that

$$\langle W_q(C_x)W_{-q}(C_{x+L})\rangle = e^{-L_yT}, \tag{497}$$

where the line tension is

$$T = L\ln\left|\frac{2\pi q-\theta}{\theta}\right|, \tag{498}$$

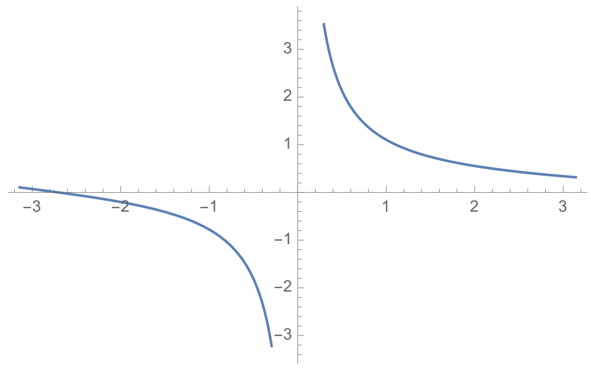
where as before $|\theta| \le \pi$. Thus if $\theta \neq \pm\pi$, we will for sure have area-law confinement. However, suppose $\theta = \pi$: we get a completely tensionless and deconfined Wilson line, provided that $q = 1$. If we choose $q = -1$, we get a strong line tension, and an area law. The opposite is true if we take $\theta = -\pi$, with $q = -1$ giving a tensionless Wilson line: these two choices are related by C, which sends both $\theta \to -\theta$ and $q \to -q$. What's going on here is the following: the θ term sets up a background electric field of strength $\bar\theta$, pointing along the x axis. This electric field causes charges to be linearly confined, except when $\theta = \pm\pi$: in this case, the ± 1 strength electric flux created by the Wilson line insertions can screen the background electric field, leading to an electric field which is of uniform magnitude everywhere, but which has a reversal of sign in the domain enclosed by the two Wilson lines (the Wilson lines are domain walls for the spontaneously broken C symmetry, which is broken by the orientation of the electric field). The reason why one of the $q \pm 1$ Wilson line charges is confined while one is tensionless at $\theta = \pm\pi$ is just because only one choice of q allows for an electric field that is everywhere uniform in strength.

To back up this conclusion, we can calculate the vev of the topological charge density as a function of x. For x to the left of C_x or to the right of C_{x+L}, $\langle Q_{top}(x)\rangle = i(\cot(\theta/2)/2 - 1/\theta)$

as before. But for x in between the two loops, we find

$$\langle Q_{top}(x)\rangle = i\partial_\theta \ln\left[\frac{2\sin(\theta/2)}{2\pi q-\theta}\right] = i\frac{2+(2\pi q-\theta)\cot(\theta/2)}{4\pi q-2\theta}. \tag{499}$$

The energetically favorable choices with P symmetry are $q=\pm1, \theta=\pm\pi$. For these choices we get $\langle Q_{top}(x)\rangle \to \pm i/\pi$. Note that this is *opposite in sign* to $\langle Q_{top}(x)\rangle$ for x not between the two Wilson lines: this is why the two Wilson lines are domain walls, across which the sign of $\langle Q_{top}(x)\rangle$ flips. At general θ, $\langle Q_{top}(x)\rangle$ for x in between the two Wilson lines (with $q=1$) looks like, as a function of θ,



(500)

The divergence at $\theta=0$ comes from the fact that when $\theta=0$, $\langle W_1(C_x)W_{-1}(C_{x+L})\rangle = 0$.

24 $SU(2)\times SU(2)$ *chiral symmetry breaking* ✓

Today we're going through the details of the setup of the chiral lagrangian for the breaking of $SU(2)\times SU(2)$ in QCD (with the first generation of quarks only). This is standard stuff; I just wanted to have the details that were skipped over in the book I was reading (Zinn-Justin) spelled out somewhere.

In what follows we will be adding things to the free action

$$S_\psi = \int \bar\psi i\partial\!\!\!/\psi, \qquad \psi = (u,\, d)^T, \tag{501}$$

where u, d are the up and down quarks. This possesses a $U(2)^2$ flavor symmetry, where the two factors act separately on the left- and right-handed parts of the two quarks.

To discuss spontaneous chiral symmetry breaking phenomenologically, we will start by adding to the action the term

$$S_{\psi M} = -g\int \bar{\psi}(\Pi_+ M + \Pi_- M^\dagger)\psi, \qquad \Pi_\pm = \frac{1}{2}(\mathbf{1} \pm \bar{\gamma}), \tag{502}$$

where M is some 2×2 matrix field that will serve as the order parameter for the spontaneously broken symmetry. We need to know how the discrete symmetries P and C are implemented on M (we are in Euclidean time, so we only care about C and P). We will take P to act as reflection of a single spacetime coordinate for simplicity—following the discussion in the diary entry on fermions, we recall that P is implemented by $P : \psi \to \bar{\gamma}\gamma_i\psi, \bar{\psi} \mapsto \bar{\psi}\gamma_i$ for some i ($\bar{\psi}$ is a field independent from ψ, transforming in the inverse way under $\mathrm{Spin}(d)$ as ψ). Requiring that $S_{\psi M}$ is P-invariant means that

$$\Pi_+ M + \Pi_- M^\dagger = \gamma_i(\Pi_+ P(M) + \Pi_- P(M^\dagger))\gamma_i = \Pi_- P(M) + \Pi_+ P(M^\dagger) \tag{503}$$

and so evidently P acts as

$$P(M) = M^\dagger. \tag{504}$$

Charge conjugation symmetry gives us the constraint

$$[C\Pi_+ C^\dagger \otimes C(M) + C\Pi_- C^\dagger \otimes C(M^\dagger)]^T = \Pi_+ M + \Pi_- M^\dagger. \tag{505}$$

Hopefully the notation here isn't too confusing: $C(M)$ is the charge-conjugated image of M, while the $\otimes$s are used since M and $\Pi_\pm$ act on different factors of the Hilbert space.[51] Now $C\gamma^\mu C^\dagger = -\gamma_\mu^T$, and thus $[C\bar{\gamma}C^\dagger]^T = \bar{\gamma}$ in $d = 4$ dimensions, while $[C\bar{\gamma}C^\dagger]^T = -\bar{\gamma}$ in $d = 2$. Thus $C\Pi_\pm C^\dagger = \Pi_\pm$ if $d = 4$, and $C\Pi_\pm C^\dagger = \Pi_\mp$ in $d = 2$. Therefore C symmetry tells us that

$$C(M) = M^T \quad (d = 4), \qquad\qquad C(M) = M^* \quad (d = 2). \tag{506}$$

With these transformations, $S_{\psi M}$ is symmetric.

The flavor symmetry of the free term is $U(2)_+ \times U(2)_-$, which acts as

$$U(2)_+ \times U(2)_- : \psi \mapsto U_+^{\Pi_+} U_-^{\Pi_-}\psi, \qquad \bar{\psi} \mapsto \bar{\psi}P[(U_+^{\Pi_+}U_-^{\Pi_-})^\dagger] = \bar{\psi}(U_+^{\Pi_-}U_-^{\Pi_+})^\dagger, \qquad U_\pm = e^{i\theta_a^\pm t^a}, \tag{507}$$

where t^a are the (Hermitian) generators for the $\mathfrak{u}(2)$ Lie algebra and the notation is $U_\alpha^{\Pi_\beta} = e^{i\theta_a^\alpha t^a \Pi_\beta}$, so that $U_\alpha^{\Pi_\beta}$ acts as $\mathbf{1}$ on spinors of chirality opposite to that of the index β. By looking at what happens when we decompose ψ as a sum of chiral spinors, we see that we can also write this as

$$U(2)_+ \times U(2)_- : \psi \mapsto (\Pi_+ U_+ + \Pi_- U_-)\psi, \qquad \bar{\psi} \mapsto \bar{\psi}(\Pi_+ U_-^\dagger + \Pi_- U_+^\dagger). \tag{508}$$

Since $e^{i\Pi_\pm}\partial\!\!\!/ = \partial\!\!\!/ e^{-i\Pi_\pm}$, the free part of the action is invariant.

Under $U(2)_+ \times U(2)_-$, $S_{\psi M}$ transforms as

$$U(2)_+ \times U(2)_- : S_{\psi M} \mapsto -g\int \bar{\psi}\left(\Pi_+ U_-^\dagger M' U_+ + \Pi_- U_+^\dagger (M')^\dagger U_-\right)\psi, \tag{509}$$

[51] We will of course usually omit the $\otimes$s, but here we've written them since e.g. $C\Pi_\pm C^\dagger C(M)$ is likely to cause confusion.

where M' is the image of M under $U(2)_+ \times U(2)_-$. Therefore this interaction will be symmetric provided that M transforms as

$$U(2)_+ \times U(2)_- : M \mapsto U_- M U_+^\dagger, \tag{510}$$

which is the expected transformation law for a Goldstone field.

The next terms we need to add to the action are a kinetic term for M, and a potential to give M a vev:

$$S_M = \frac{1}{2\alpha} \int \left[\mathrm{Tr}[\partial_\mu M \partial^\mu M^\dagger] - V(MM^\dagger) \right], \tag{511}$$

which preserves C and P. This is done at a purely phenomenological level, but microscopically we might imagine this as coming from the result of adding M to the action to decouple some type of fermion interaction, with expectation values of M being equal to expectation values of the corresponding chiral fermion bilinears that are condensed in the SSB regime. The above S_M terms are then assumed to be an EFT way of capturing the effective interactions for the order parameter M induced by the fermion dynamics.

We will also include a term that explicitly breaks the $U(2)_+ \times U(2)_-$ symmetry while preserving C and P. This is done with

$$S_B = -\frac{1}{\sqrt{2}} \int \mathrm{Tr}[B(M + M^\dagger)], \tag{512}$$

where B is some fixed matrix, that we think of as a classical source / a "magnetic field" used to generate correlation functions of the order parameter. S_B preserves P for any choice of B since $M + M^\dagger$ is a scalar, while it preserves C if $B^T = B$. Note that $M - M^\dagger$ is a pseudoscalar though, so such a term would break P explicitly if added.

Let's now look at how all the terms we've introduced affect the axial and vector currents. The vector current is found by taking $U_+ = U_-$ and performing the variation $\psi \mapsto U\psi$, $M \mapsto UMU^\dagger$. Taking $U = e^{i\theta_a t^a}$ for θ_a small, we use $\partial_\mu M \mapsto i\partial_\mu \theta^a [t^a, M]$ to get (I'm not being super careful about signs and is)

$$\delta_V S = \int \left[\partial_\mu \theta^a \left(\bar\psi \gamma^\mu t^a \psi + i\mathrm{Tr}[t^a([M, \partial_\mu M^\dagger] + [M^\dagger, \partial_\mu M])] \right) + \frac{i}{\sqrt{2}} \theta_a \mathrm{Tr}[[t^a, B](M + M^\dagger)] \right], \tag{513}$$

where we used the cyclicity of the trace and that $M^\dagger \mapsto UM^\dagger U^\dagger$. The part contracted with $\partial_\mu \theta^a$ is the vector current, and so the Ward identity tells us

$$\partial_\mu j_V^{\mu a} = \partial_\mu \left(\bar\psi \gamma^\mu t^a \psi + i\mathrm{Tr}[t^a([M, \partial_\mu M^\dagger] + [M^\dagger, \partial_\mu M])] \right) = -\frac{i}{\sqrt{2}} \mathrm{Tr}[[t^a, B](M + M^\dagger)]. \tag{514}$$

The axial current is found by taking $U_- = U_+^\dagger$, so that $\psi \mapsto (\Pi_+ U + \Pi_- U^\dagger)\psi = e^{-i(\Pi_+ - \Pi_-)\theta_a t^a} \psi$ and $M \mapsto U^\dagger M U^\dagger$. This leads to a similar situation as the vector current, except with commutators replaced by anti-commutators:

$$\delta_A S = \int \left[\partial_\mu \theta^a \left(\bar\psi \gamma^\mu \bar\gamma t^a \psi + \mathrm{Tr}[t^a(\{M, \partial_\mu M^\dagger\} + \{M^\dagger, \partial_\mu M\})] \right) - \frac{i}{\sqrt{2}} \theta_a \mathrm{Tr}[\{t^a, B\}(M - M^\dagger)] \right], \tag{515}$$

so that

$$\partial_\mu j_A^{\mu a} = \frac{i}{\sqrt{2}}\mathrm{Tr}[\{t^a, B\}(M - M^\dagger)]. \tag{516}$$

Therefore the vector current is not conserved unless $B \propto \mathbf{1}$, while the axial current is not conserved for all $B \neq 0$.[52] Therefore B acts as a classical source for the axial current, and can also act as a source for the vector current if $B \not\propto \mathbf{1}$. Also note that as required, the divergence of j_V is a scalar, while the divergence of j_A is a pseudoscalar.

To discuss SSB, we will pick an explicit form for B and B — this is just the usual procedure of discussing the SSBroken state by adding a small background symmetry-breaking field to select out a particular ground state. For simplicity we will specialize to the case where $B = b\mathbf{1}$, which conserves j_V but explicitly breaks j_A. After M gets a vev from the potential, this choice of B will give (equal) masses to the u and d quarks. For the potential, we take the usual ($m^2 < 0$)

$$V(x) = \frac{1}{2}m^2 x + \frac{1}{24}\lambda x^2. \tag{517}$$

Therefore we will parametrize M as

$$M = \frac{1}{\sqrt{2}}(\sigma\mathbf{1} + i\pi_a\sigma^a). \tag{518}$$

We can now solve for $p = \langle\pi\rangle$ and $s = \langle\sigma\rangle$. If $b \neq 0$ then $p = 0$, which then after some straightforward algebra gives

$$s(m^2 + \lambda s^2/6) = b. \tag{519}$$

Solving this to first order in b, we get

$$s = \sqrt{-6m^2/\lambda} + c\sqrt{-3/(2\lambda m^2)}. \tag{520}$$

We can then plug this back into the potential by taking $\sigma \mapsto \sigma - s$ to find the masses of the σ and the π—we get, to first order in b,

$$m_\pi^2 = m^2 + s^2\lambda/6 = \frac{b}{4|m^2|}, \qquad m_\sigma^2 = m^2 + s^2\lambda/2 = 2|m^2| + \frac{3b}{4|m^2|}. \tag{521}$$

As required, the π mass is zero when $b = 0$, since when $b = 0$ (i.e. when the external symmetry-breaking field is turned off) the π is a Goldstone boson. This vev for M also gives rise to a mass term for the fermions via the Yukawa coupling in $S_{\psi M}$; in this simple case where $B \propto \mathbf{1}$, at $m_\sigma \to \infty$ both quarks have the same mass, $m_u = m_d \propto gs$ (we don't really think of $\langle\pi\rangle$ as contributing to the mass since it only really does so when b is large, but this is outside of our approximation scheme).

A last comment is that this approach lets us easily deal with a possible θ term. If the quarks are coupled to an $SU(3)$ gauge field with $\theta\int c_2[A_{SU(3)}]$ in the action, then we can eliminate the θ term with a chiral rotation. This is the equivalent to performing the shift

[52] This is because in order for $\partial_\mu j_A^{\mu a} = 0$ for all a with $B \neq 0$, we need B to anti-commute with all the t^a. This is impossible since this implies $B[t^a, t_b] = [t^a, t^b]B$, which cannot be true since $[t^a, t^b] = if^{abc}t^c$ is linear in the t^as, and hence B must anti-commute with it—a contradiction.

$M \mapsto e^{-2i\theta} M$, which then shows up equivalently as a shift $B \mapsto e^{i\theta Z} B$. This has the effect of doing $b \mapsto b\cos\theta$ for the purposes of computing $\langle \sigma \rangle$ and $\langle \pi \rangle$.

25 *Gluon screening of Wilson lines in non-Abelian gauge theory and some useful representation theory computations* ✓

Today we're going over something that took me forever to finally understand: how exactly gluons can screen Wilson lines in certain representations to turn area law behavior into perimeter law behavior.

Since we are looking at pure gauge theory and are discussing screening, only non-Abelian gauge groups will be relevant, since only for non-Abelian groups are the gauge bosons charged. We will work on the lattice, since it will make the screening phenomenon most clear. We will show that for certain choices of representation R, the Wilson line $W_R(C)$ will have perimeter-law behavior in the strong-coupling expansion. Now in the weak coupling limit[53] the Wilson line always has perimeter law: we expand it to order g^2 and get $\langle W_R(C) \rangle \approx 1 + Ng^2 \oint \oint \langle A^a_\mu(x) A^a_\nu(y) dx^\mu \wedge dy^\nu \rangle$, where $N = \dim(R) C_2(R)/2$ and a is an arbitrary group index (no sum). We can replace the correlator to $O(g^2)$ with the Fourier transform of Π_T/p^2, which gives a perimeter law (in 3+1 D). This means that there are representations for which $W_R(C)$ has a perimeter law both at strong coupling and at weak coupling, and so the Wilson lines are tensionless throughout the phase diagram. From the conceptual point of view, this is rather obvious: we know that Wilson lines in the fundamental are perimeter-law when dynamical fundamental matter is included, due to screening effects. Since the gauge fields themselves are charged in the adjoint, they can screen Wilson lines in representations obtained from tensor products of the adjoint, and hence we expect that such Wilson lines should be perimeter-law even in the pure gauge theory. Actually showing results like this in detail is a bit technical, however.

[53] The action is $\beta \sum_\square (\dots)$. For large β we can fix a gauge in which the product of Us around a plaquette goes to $e^{-a^2 F_{\mu\nu} + \cdots}$, where $\mu\nu$ label the plane the plaquette is in and a is the lattice spacing (and $U_l = e^{i\int_l A} \approx \mathbf{1}$). This gives $S \approx \beta a^4 \sum_\square \mathrm{Tr}[F_{\mu\nu} F^{\mu\nu}]$, and so (in four dimensions)

$$\frac{1}{g^2} \leftrightarrow \beta a^4. \tag{522}$$

The continuum limit is thus the weak coupling ($\beta \to \infty$) limit. Note that we can't get access to the strong-coupling side of the gauge theory; β has to remain large for the continuum formulation to work.

Now for a given representation R, we want to compute

$$\langle W_R(C)\rangle = \left\langle \mathrm{Tr}_R \left[\prod_{l\in C} U_R(g_l)\right]\right\rangle = \int \prod_l \mathcal{D}g_l \,\mathrm{Tr}_R\left[\prod_{l\in C} U_l\right] \exp\left(-\beta\sum_{\Box}\mathrm{Tr}_f\left[\prod_{l\in\partial\Box} U_l\right]\right), \tag{523}$$

where f is the fundamental representation of the gauge group, and the sum over plaquettes includes a sum over plaquette orientations.[54] Here the notation is such that U_l is the representation of g_l in the fundamental or antifundamental, depending on the orientation of the link. A more verbose notation is $U_f(g_l)$, which we use for the representation matrices appearing in the Wilson loop. If a representation subscript on a $U(g)$ is omitted, it is implied that the representation is taken to be the fundamental. As usual, each plaquette term has two fundamental matrices and two antifundamental matrices: if the group elements on the links of a plaquette are $g_1,\ldots,g_4$ labeled counterclockwise from the "bottom", then $U_\Box \equiv \prod_{l\in\partial\Box} U_l = U(g_1)U(g_2)U^*(g_3)U^*(g_4)$. Flipping the orientation of a plaquette is therefore equivalent to taking the trace in the conjugate representation f^*.

Since the weak-coupling limit gives a perimeter law, to address confinement we just need to look at strong coupling. For strong coupling, we expand the exponential in powers of β. Now $\int \mathcal{D}g_l\, U_l = 0$ by the properties of the Harr measure. If we were doing Abelian gauge theory, we would then derive an area law by expanding the exponential to Ath order (here A is the area enclosed by C), which would give us enough U_ls from the exponential to cancel those from the Wilson loop.

However, with non-Abelian gauge theories, we can do something different. Physically, this is because gluons are charged, so that in non-Abelian gauge theories, the gauge field itself can screen sources. This allows us to form tubes of glue around the Wilson loop, which can potentially screen it, depending on its charge.

For concreteness, consider a 3D (Euclidean) theory, with a Wilson loop inserted as above in a representation R. Now form a tube out of cubes, with the Wilson line located along one of the sharp edges of the tube. A section of this tube is shown in Figure 4. This tube will appear at order β^{4P} in perturbation theory, where P is the perimeter of the Wilson line (each corner contributes an extra β^4).

This tube will screen the Wilson line if

$$R^* \in f\otimes f^*. \tag{525}$$

If the Wilson line weren't there, the tube would give a non-zero contribution to the partition function since each U_l on an edge appears with a corresponding U_l^* from a neighboring plaquette, allowing the integral over g_l to be nonzero. This is because $1 \in f\otimes f^*$. If

[54]This is needed so that the action is real: if $\sum'_\Box$ is a sum over all plaquettes without counting the orientation separately, then

$$\sum_\Box \mathrm{Tr}_f \prod_{l\in\partial\Box} U_l = \sum_\Box{}' \mathrm{Tr}_f\left(\prod_{l\in\partial\Box} U_l + \prod_{l\in\partial\Box} U_l^*\right), \tag{524}$$

which is real (if the gauge group is real, like $SO(n)$, then a term like $\mathrm{Tr}_f(U_1U_2U_3U_4)$ has a partner $\mathrm{Tr}_f(U_4^{-1}U_3^{-1}U_2^{-1}U_1^{-1}) = \mathrm{Tr}_f[(U_1U_2U_3U_4)^T]$, and so the sum over orientations is redundant.

$R^* \in f \otimes f^*$ then $1 \in R \otimes f^* \otimes f$, and so with a Wilson line with such an R can be added in the position shown without making the result vanish under integration.

To argue this precisely, we need to recall the orthogonality relation[55]

$$\int \mathcal{D}g\,[U_R(g)]^i_j [U_{R'}(g)]^k_l = \frac{1}{\dim R}\delta_{R^*,R'}\delta^i_l\delta^k_j. \tag{529}$$

The LHS is basically a group average over all similarity transforms of the matrix $[E^k_l]^i_j = \delta^i_l\delta^k_j$, and the RHS tells us that this average is zero if E^k_l has the 1 off of the diagonal, while E^k_l averages out to ($1/\dim R$ times) the identity if the 1 is on the diagonal. In particular, $\int \mathcal{D}g\,[U_R(g)]^i_j = 0$ unless $R = 1$.

Now consider expanding the expectation value to the order of β at which the tube geometry appears. We have

$$\langle W_R(C)\rangle = \int \prod_l \mathcal{D}g_l\,\mathrm{Tr}_R\left[\prod_{l\in C} U_R(g_l)\right]\prod_{\square\in T}\mathrm{Tr}[U_\square], \tag{530}$$

where T is the tube. Combining these into one trace,

$$\langle W_R(C)\rangle = \int \prod_l \mathcal{D}g_l\,\mathrm{Tr}\left[\prod_{l\in C} U_R(g_l)\otimes\bigotimes_{\square\in T} U_\square\right]. \tag{531}$$

Let us isolate just the terms that depend on g_1, where g_1 is the first link of the Wilson line:

$$\begin{aligned}\langle W_R(C)\rangle = \int \prod_l \mathcal{D}g_l\,\mathrm{Tr}\Bigg[&(U_R(g_1)\otimes U(g_1)\otimes U^*(g_1)\otimes \mathbf{1})\cdot\Bigg(\prod_{l>1} U_R(g_l)\otimes\prod_{l\in\square_1\backslash l_1} U_l\\ &\otimes\prod_{l\in\square_1'\backslash l_1} U_l\otimes\bigotimes_{T\ni\square\neq\square_1,\square_1'} U_\square\Bigg)\Bigg].\end{aligned} \tag{532}$$

Oh god, the notation. Sorry. Here, $\square_1, \square_1'$ are the two plaquettes in T that have the link l_1 as a side, and $\mathbf{1}$ is the identity acting on the tensor factors of every plaquette but these two.

[55]The proof goes as follows: using the invariance of the Harr measure under shifts, one shows that

$$U_R(h)\int \mathcal{D}g\,U_R(g)E^k_l U_{(R')^*}(g^{-1}) = \left(\int \mathcal{D}G\,U_R(g)E^k_l U_{(R')^*}(g^{-1})\right)U_{(R')^*}(h), \tag{526}$$

where $[E^k_l]^i_j = \delta^i_l\delta^k_j$. Therefore using Schur's lemma, since both R, R' were assumed to be irreducible,

$$\int \mathcal{D}g\,[U_R(g)E^k_l U_{R'}(g)]^i_j = \delta_{R^*,R'}\delta^i_j C(E^k_l), \tag{527}$$

where $C(E^k_l)$ is a constant. If we set $R^* = R'$, take the trace of both sides, and use the cyclicity of the trace, we get

$$\mathrm{Tr}[E^k_l] = \dim(R)C(E^k_l) \implies C(E^k_l) = \frac{1}{\dim R}\delta^k_l. \tag{528}$$

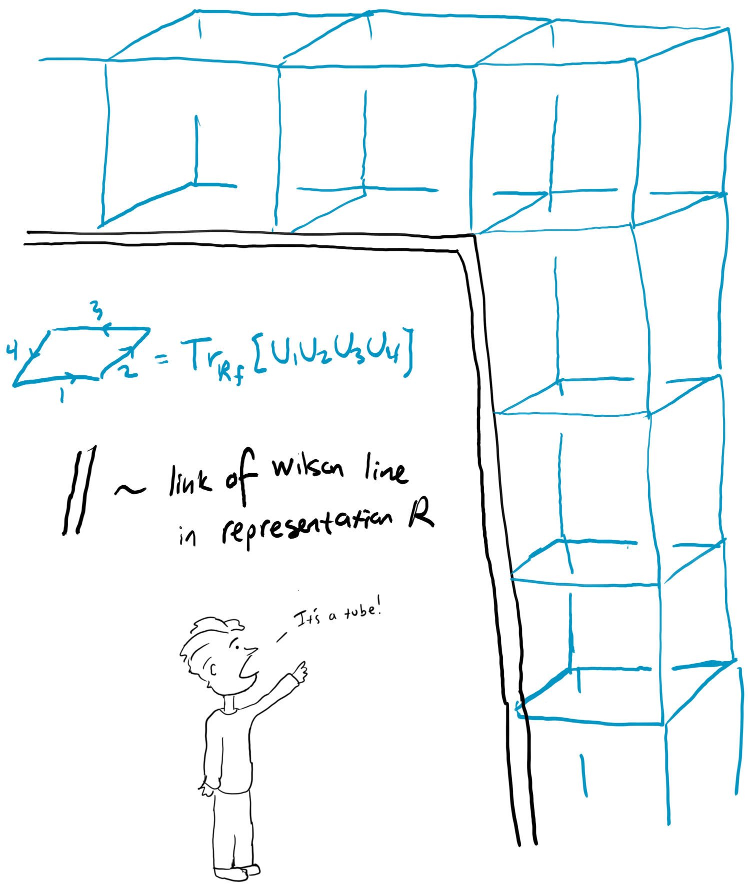


Figure 4: Geometry for how a tube of glue can screen a Wilson line.

Now let S be the matrix such that

$$S(U_R(g_1)\otimes U(g_1)\otimes U^*(g_1))S^{-1} = \bigoplus_{R_i\in R\otimes f\otimes f^*} U_{R_i}(g_1). \tag{533}$$

For example, for $1/2\otimes 1/2$ in $SU(2)$, we have (see this footnote for a reminder about how to decompose $\otimes$s of irreps[56])

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}, \tag{536}$$

which takes the basis $\chi^i\otimes\psi^j$ to the basis $(\chi^0\psi^0,\ [\chi^0\psi^1+\chi^1\psi^0]/\sqrt{2},\ \chi^1\psi^1,\ [\chi^0\psi^1-\chi^1\psi^0]/\sqrt{2})^T$. Actually, since we will want to calculate decompositions involving $f\otimes f^*$ and not $f\otimes f$ (in $SU(2)$ they happen to be isomorphic), we will want to work e.g. in the basis $\chi^i\otimes\psi_j$ instead. The matrix which takes the $\chi^i\otimes\psi_j$ basis to the $(\chi^0\psi_1,\ [\chi^0\psi_0-\chi^1\psi_1]/\sqrt{2},\ -\chi^1\psi_0,\ [\chi^0\psi_0+\chi^1\psi_1]/\sqrt{2})^T$ basis is

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 0 & -1 & 0 \\ 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \end{pmatrix} \tag{537}$$

Note that the trivial representation appears symmetrically in this basis instead of representation, since it is just the trace.

Now consider the integration over g_1. Since all matrix elements of any representation other than the trivial one have zero average over the group, we will only get something nonzero if the trivial representation appears in the above $\oplus$. We can then (not worrying about constant factors) do the g_1 integral and write

$$\langle W_R(C)\rangle = \int \prod_{l>1}\mathcal{D}g_l \operatorname{Tr}\left[\left(1\oplus\bigoplus_{R_i\in R\otimes f\otimes f^*} 0_{d_{R_i}\times d_{R_i}}\right) S(\cdots)S^{-1}\right], \tag{538}$$

[56] The general strategy is to take a tensor product of tensors transforming in two given representations, and then look for invariant subspaces among the collection of tensors in the tensor product, with each invariant space constituting a term in the direct sum decomposition. For example, we can work out $3\otimes 8$ in $SU(3)$, with 3 the fundamental and 8 the adjoint. Consider then $\chi^i A^j_k$. First, we can contract the i with the k, obtaining a single fundamental index and implying a 3 in the $\oplus$ decomposition. Next, after taking out the contracted piece, form $S^{ij}_k = \chi^{(i}A^{j)}_k$ and $A^{ij}_k = \chi^{[i}A^{j]}_k$. We can now contract both of these with ϵ_{lmn}. The former dies, and hence gives us an irrep. With two symmetrized upper indices and one lower index, we naively have an 18-dimensional irrep. But we have to remember that we have taken out the triplet that came from the contraction, and so $S^{kj}_k = 0\,\forall j$. This means that the S tensors define a $18-3=15$ dimensional irrep. When A^{ij}_k is contracted with the epsilon symbol, we get the tensor $\widetilde{A}_{kn} = \epsilon_{nij}A^{ij}_k$. Contracting this again with ϵ^{nlm}, we get

$$\epsilon^{nlm}\epsilon_{nij}A^{ij}_k = (\delta^l_i\delta^j_m - \delta^m_i\delta^l_j)A^{ij}_k = 0, \tag{534}$$

since the trace part of A^{ij}_k has already been subtracted out. Thus $\widetilde{A}_{kn}$ is symmetric in its indices, and hence transforms as the 6^* irrep. Summarizing,

$$3\otimes 8 = 3\oplus 6^*\oplus 15. \tag{535}$$

where $\cdots$ are the terms not containing g_1. We then write

$$S^{-1}\left(1 \oplus \bigoplus_{R_i \in R\otimes f\otimes f^*} 0_{d_{R_i}\times d_{R_i}}\right) S = \Pi_{\mathbf{1}}^{R\otimes f\otimes f^*}, \tag{539}$$

where $\Pi_{\mathbf{1}}^{R\otimes f\otimes f^*}$ is the projector onto the trivial representation in $R\otimes f\otimes f^*$, expressed in the $\otimes$ basis. For example, in the $SU(2)$ example,

$$\Pi_{\mathbf{1}}^{1/2\otimes 1/2^*} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{540}$$

Therefore we have

$$\langle W_R(C)\rangle = \int \prod_{l>1} \mathcal{D}g_l \operatorname{Tr}\left[(\Pi_{\mathbf{1}} \otimes \mathbf{1}) \cdot (\cdots)\right], \tag{541}$$

and we have successfully done the g_1 integral, getting a nonzero result.

Now we repeat this for all $g_l, l \in C$. Each integration gives us a factor of $\Pi_{\mathbf{1}}^{R\otimes f\otimes f^*}$ in the trace. We then do the integrals over the g_l for l a link in the tube not containing the Wilson line. These integrations make $\Pi_{\mathbf{1}}^{f\otimes f^*}$ matrices appear in the trace. When all is said and done, we get a trace over a bunch of products of things like $\mathbf{1}\otimes\ldots\otimes\mathbf{1}\otimes\Pi_{\mathbf{1}}^{R\otimes f\otimes f^*}\otimes\mathbf{1}\otimes\ldots\otimes\mathbf{1}$ and $\mathbf{1}\otimes\ldots\otimes\mathbf{1}\otimes\Pi_{\mathbf{1}}^{f\otimes f^*}\otimes\mathbf{1}\otimes\ldots\otimes\mathbf{1}$. These are giant matrices, and because of the projectors they are mostly made out of zeros. However, the trace does not vanish. Indeed, each matrix in the product always has a 1 in the upper-right hand corner: in the basis of the tensor product, the first basis vector in the Hilbert space that the matrix in the trace acts on is $T_{0\ldots 0}^{0\ldots 0}$, where T is the tensor product of the basis vectors for every single $\otimes$ factor. Since this entry always appears in the trace (the trivial representation), every matrix in the product appearing in the trace must not annihilate this vector; hence their product does not either. This gives us a positive, nonzero contribution to $\langle W_R(C)\rangle$. Since in this basis the matrix elements of the $\Pi_{\mathbf{1}}$ projectors are only either 0 or 1, this contribution cannot be canceled by any other parts of the trace, and so

$$\langle W_R(C)\rangle \sim \mathcal{N}\beta^{4P}, \tag{542}$$

where $\mathcal{N}$ is some number depending on R, the gauge group, etc. Therefore the Wilson line has a perimeter law, and is screened.

So, the Wilson line will get screened if $R^* \in f\otimes f^*$, or equivalently if $R \in f\otimes f^*$, since $f\otimes f^*$ is invariant under conjugation. In particular, adjoint Wilson lines are always screened, since $f\otimes f^*\otimes A \ni 1$, which can be proved either by noting that $A \in f\otimes f^*$ and recalling that A is self-dual[57], or by noting that the generator matrices $[T^a]^i_j$ constitute invariant symbols with one adjoint index, one fundamental index, and one anti-fundamental

[57]The adjoint representation is always self-dual. Indeed, from two tensors $\chi^i_j\eta^l_m$, we can always contract both pairs of indices to form a singlet, and so $A\otimes A \ni 1$. Alternatively, one could note that the Killing form on the Lie algebra provides an isomorphism between A and its dual.

index.[58] Physically this is intuitive since sources that come from tensor products of the adjoints should be able to be screened by a sufficient number of gluons, which are in the adjoint.

However, something more general is true: if R is any representation appearing in $(f \otimes f^*)^{\otimes N}$, for any N, then R is screened. Mathematically, this follows from simply repeating the above procedure, but going to β^{4NP} in perturbation theory, where an N-sheeted tube along the Wilson line appears. This N-sheeted tube will then be enough to screen the Wilson line.

The irreps that appear in the $\oplus$ decomposition of $(f \otimes f^*)^N$ are precisely those which transform trivially under the center of the gauge group. To put it another way, let G be the gauge group (and f its fundamental representation), $\widetilde{G}$ be its universal cover, and let $G' = \widetilde{G}/Z(\widetilde{G})$. Then the claim is that irreps appearing in the $\oplus$ decomposition of

$$\langle f \otimes f^* \rangle = \{(f \otimes f^*)^{\otimes N} \mid N \in \mathbb{N}\} \tag{545}$$

are precisely the irreps of G'.

To show this pedantically, note that any irrep in $\langle f \otimes f^* \rangle$ transforms trivially under $Z(\widetilde{G})$, and hence is an irrep of G', and so $\langle f \otimes f^* \rangle \subset \mathrm{Rep}(G')$. Conversely, any irrep of G' will appear in $\langle f \otimes f^* \rangle$. To show this, first note that the irreps of G' are a subset of the irreps of G, the later of which is generated by taking tensor powers of f. Thus all the irreps of G' will appear in the decomposition of some $f^{\otimes N}$ for some N. Let $m = |Z(G)|$. Then since irreps of G' are invariant under $Z(G)$, every irrep of G' appears in the decomposition of $f^{\otimes mk}$, for some $k \in \mathbb{N}$. Furthermore, $1 \in f^{\otimes m}$, and so likewise $1 \in (f^*)^{\otimes m}$. Thus $(f \otimes f^*)^{mk} \supset f^{\otimes mk} \otimes \mathbf{1}^{\otimes k}$, and so $\mathrm{Rep}(G') \subset \langle f \otimes f^* \rangle$. Therefore we actually have that $\mathrm{Rep}(G') = \langle f \otimes f^* \rangle$.

Summing up, we can say that a Wilson line in a representation R will be screened iff $R \in \mathrm{Rep}(G')$. Again, physically this follows from the fact that the gluons are in the adjoint, and so the tube of glue can only screen things that transform trivially under $Z(G)$.

For example, if $G = SU(2)$, then integer-spin Wilson lines are screened, while half-odd-integer spin lines are not. In general, for $G = SU(N)$, $N \otimes N^* = \mathbf{1} \oplus A$, and so any irrep appearing in $\langle A \rangle$ will be screened. These are precisely the irreps coming from tensors with an equal number of upper and lower indices.

For example, in $SU(3)$, the $3, 3^*, 6, 6^*, 15, 15^*, \dots$ representations are un-screened, while e.g. the $A = 8, 10, 10^*, 27, \dots$ representations are screened. The adjoint representation $A = 8$ can be screened by a single tube of glue, but e.g. if $R = 10$ then a 2-sheeted tube will do the job: the two sheets provide[59]

[58] Actually something more general is true, namely $R \otimes R^* \otimes A \ni 1$ for any irrep R. Indeed, consider an infinitesimal transformation by $U = \mathbf{1} + i\theta^b T^b_R$. Then thinking of a in T^a as an adjoint index, to $O(\theta)$ we have

$$\begin{aligned}
[T^a]^i_j \mapsto &(\delta^i_k + i\theta^b [T^b_R]^i_k)(\delta^{ad} + i\theta^b [T^b_A]^{ad})[T^d_R]^k_l(\delta^l_j - i\theta^b [T^b_R]^l_j) \\
&= [T^a]^i_j + i\theta^b \left([T^b_R, T^a_R]^i_j + (-if^{bad})[T^d_R]^i_j\right) \\
&= [T^a]^i_j,
\end{aligned} \tag{543}$$

since $[T^a_A]^{bc} = -if^{abc}$. In particular, taking $R = A$ and using the reality of A tells us that

$$A \otimes A = \mathbf{1} \oplus A \oplus \dots . \tag{544}$$

[59] To derive this, we just need to calculate $8 \otimes 8$. From two adjoint tensors $A^i_j B^k_l$ we can take traces in

$$(3 \otimes 3^*)^{\otimes 2} = (1 \oplus 8)^{\otimes 2} = 1^{\oplus 2} \oplus 8^{\oplus 4} \oplus 10 \oplus 10^* \oplus 27, \tag{547}$$

and the 10^* on the RHS gives us the representation needed to screen the Wilson line. Of course, for $G = PSU(N)$, all Wilson lines are screened.

So, we see that confinement (at least to the extent to which confinement is captured by line tensions) is very much dependent on the topological properties of the gauge group! This is perhaps not too surprising, since some sort of monopole condensation is usually thought to be the mechanism whereby confinement occurs, and the different types of monopoles which can exist depend on what the topology of the gauge group is.

(note to self: is the tube geometry really needed? Here's an alternative geometry: first take any "wall" of plaquettes that have the Wilson line on one edge, and the overlay the wall with its dual, forming a double-layer wall with a factor of $\mathrm{Tr}_{f\otimes f^*}[U_\square \otimes U_\square^*]$ for each plaquette. If $R^* \in (f \otimes f^*)^N$, then we take N copies of this wall, which provides enough glue to screen the Wilson line. This lets us do screening with slightly fewer plaquettes, and also gives us an approach which works in two dimensions)

two ways to create adjoints, and in one way to create a singlet, giving us $1 \oplus 8 \oplus 8$. Then we can take off the trace pieces and symmetrize / antisymmetrize the top two indices. Let $\widetilde{A}^{ij}_{kl}$ be antisymmetric in the upper two indices. Then form $\epsilon_{ijm}\widetilde{A}^{ij}_{kl}$, and contract with ϵ^{kln}; this gives zero since $\epsilon_{ijm}\epsilon^{kln}$ can be expanded as δ functions between the first and second triplet of indices, each term of which then vanishes by the tracelessness of $\widetilde{A}$. The same vanishing act happens if we instead contract with ϵ^{mkn} or ϵ^{mln}, and so $\epsilon_{ijm}\widetilde{A}^{ij}_{kl}$ is totally symmetric in mkl, and hence $\widetilde{A}^{ij}_{kl}$ gives us an irrep. A totally symmetric mkl gives us 10 independent tensors, and so this irrep is 10 dimensional. Antisymmetrizing the bottom two indices gives us another irrep, with three symmetric upper indices; this is also 10-dimensional. These two irreps are 10^* and 10, respectively. Note that 10 and 10^* are distinct, despite being invariant under the $\mathbb{Z}_3$ center of the gauge group (this is weird?); this is because the contraction $\widetilde{A}^{ij}_{kl}\widetilde{A}^{kl}_{ij} = 0$ due to the mixed symmetry / antisymmetry of the upper and lower sets of indices. Finally, we can symmetrize both the upper and lower indices: this gives 36 index choices, but there are 3×3 constraints on them coming from tracelessness, hence this irrep is 27-dimensional (and self-dual). Putting all of these together, we get

$$8 \otimes 8 = 1 \oplus 8 \oplus 8 \oplus 10 \oplus 10^* \oplus 27. \tag{546}$$