Symmetries and anomalies diary

Ethan Lake

August 14, 2020

1 *Charge quantization and the weak mixing angle confusion* ✓

In various places (Ryder's QFT book, stack exchange, etc.) it is often stated that if the weak mixing angle is irrational, then electric charge is not quantized since the $U(1)$ of electromagnetism is embedded non-compactly in the full $SU(2) \times U(1)_Y$ electroweak gauge group. Today we will explain why this is incorrect (working as always under the assumption that the full gauge group is $SU(2) \times U(1)_Y$ and not $SU(2) \times \mathbb{R}$).

First we will show why the $U(1)$ of electromagnetism is compact, and afterwards explain why it is often stated otherwise in the literature. We choose the conventions where the covariant derivative and hypercharge gauge transformations look like

$$D_\mu\psi = (\partial_\mu - igW_\mu^a T^a - ig' B_\mu)\psi, \qquad \psi \mapsto e^{i\alpha(x)}\psi, \qquad B_\mu \mapsto B_\mu + \frac{1}{g'}\partial_\mu\alpha. \tag{1}$$

Note that the transformation of the field is not $e^{ig'\alpha(x)}$ — g' is just a coupling constant, it is not the label of a $U(1)$ representation.

Upon Higgsing, the Higgs field (charged in the spinor of the $SU(2)$) gets a vev which we choose to be in the $\downarrow$ direction. Thus the $U(1)_e$ symmetry which we will identify with electromagnetism corresponds to rotations

$$\phi \mapsto e^{i\alpha T^3} e^{i\alpha/2}\phi, \tag{2}$$

which leaves the vev of the Higgs invariant (the second factor is a hypercharge rotation). Note that in this convention, the Higgs has charge $1/2$ under $U(1)_e$. We choose it to have charge $1/2$ since the generators for the $SU(2)$ are $T^a = \sigma^a/2$. This $U(1)_e$ is obviously compact, since the rotations about the 3 axis are compact, the $U(1)_Y$ hypercharge was assumed to be compact, and in the above action we have $\alpha \sim \alpha + 4\pi$.

In this formulation, what are the volumes of the two $U(1)$ factors? In general, we have the representation q of $U(1)$, which acts as

$$q : \mathbb{R}/(2\pi/q)\mathbb{Z} \mapsto U(1)_{2\pi}, \quad x \mapsto e^{iqx}, \tag{3}$$

where by $U(1)_{2\pi}$ we just mean the complex numbers of norm 1. Thus if our minimally charged field carries a representation q under $U(1)$, we are identifying $2\pi/q$ with 0, and so the gauge group is a circle with "volume" $\mathrm{vol}(G) = 2\pi/q$. Since for $SU(2)$ rotations we identify $\alpha \sim \alpha + 4\pi$, the $U(1)_{T^3}$ rotations come from the representation

$$1/2 : \mathbb{R}/(4\pi\mathbb{Z}) \to U(1)_{2\pi}, \quad x \mapsto e^{-ix/2}, \tag{4}$$

and so $\mathrm{vol}(U(1)_{T^3}) = 4\pi$. Since we know we have quarks with charge 1/6 (the left handed ones), the field with minimal hypercharge has charge 1/6, and so $\mathrm{vol}(U(1)_Y) = 12\pi$. Note that we could rescale things so that the minimal hypercharge is 1, decreasing the volume of $U(1)_Y$ by a factor of 6, but then we'd also have to change the charge of the Higgs under T^3 rotations, which is less preferable since the normalization of these comes from their embedding in the $SU(2)$ factor. Anyway, if we did this, we see that the aspect ratio of the torus $U(1)_{T^3} \times U(1)_Y$ would be preserved by the rescaling. Since $U(1)_e$ is embedded compactly within this torus (one can think of it as a path wrapping the $U(1)_{T^3}$ cycle three times and the $U(1)_Y$ cycle once, since the ratio of the volumes of the two $U(1)$s is 3), it is embedded compactly no matter what our conventions regarding charge normalization are.

So, why do people say that $U(1)_e$ is non-compactly embedded inside $SU(2) \times U(1)_Y$? The argument is as follows: suppose we stick with the convention where the gauge couplings appear in the exponentials of the gauge transformations:

$$\phi \mapsto e^{ig\alpha T^3} e^{ig'\beta/2}\phi. \tag{5}$$

ϕ is left invariant if we take $\beta = \alpha g/g'$. We then form the torus $U(1) \times U(1)$, and embed the $\beta = \alpha g/g'$ curve inside. If g/g' is irrational (which it generically is) then the $\beta = \alpha g/g'$ curve is dense in the torus $U(1) \times U(1)$, and hence electromagnetism is actually $\mathbb{R}_e$, a real line embedded in the full gauge group. Thus charge is not necessarily quantized, Polyakov monopoles cannot exist, and so on.

The problem with this is that the lengths of the sides of the torus are not 2π! Indeed, in the gauge transformations above, we are not identifying $\alpha \sim \alpha + 2\pi$ or $\beta \sim \beta + 2\pi$: the volumes of the gauge groups are not 2π, they are $2\pi/g$ and $2\pi/g'$. Thus the aspect ratio of the torus should actually be g/g', and the curve $\beta = \alpha g/g'$ is compactly embedded within such a torus.

There is another way to see that the quantization (or lack thereof) cannot possibly depend on the ratio of the gauge couplings: we just re-define $gW \mapsto W$ and $g'B \mapsto B$. Then the gauge couplings only appear in the kinetic terms for the gauge fields and cannot possibly affect the periodicity with which the fields transform. The gauge couplings only tell us about the dynamics of the gauge fields and how they couple to the matter fields: they have nothing to do with how the matter fields transform under the symmetry, or what representation of the symmetry group they carry. This is obvious but somehow still not discussed correctly in some hep textbooks!

2 *Anomaly constraints in $SU(3) \times SU(2) \times U(1)$* ✓

Today we're working out what gauge anomaly constraints exist in the standard model, and how they end up being saturated. The problem statement is as follows:

Consider a gauge theory with gauge group $G = SU(3) \times SU(2) \times U(1)$, and consider k massless fermions coupled to G in various different representations. Assume that there is at least one fermion charged under each factor of the gauge group, and assume wolog that none of the fermions are completely neutral. Also note that we can take all the k fermions to be left handed wolog.[1] Assume that no two fermions exist which are conjugate, in the sense that one carries the representation (R, S, q) under $SU(3) \otimes SU(2) \otimes U(1)$ and the other carries $(\bar{R}, \bar{S}, -q)$. Also assume that there is at least one field with nontrivial charge under each factor (like the left-handed quark doublet in the SM).

First, we will find the minimal value of k such that all anomalies (including the gravitational anomaly) are cancelled. We will address how this change if you assume the fields are all in either the trivial or fundamental of $SU(3)$ and $SU(2)$? We will also look to see if there are any other combinations of 5 fields besides the standard model ones that are anomaly free (with these assumptions).

Let's start with the case $k = 2$. The only option we have is the pair

$$(R, S, q), \qquad (R, S, -q). \tag{6}$$

The $U(1)^3$, $SU(2) \times U(1)^2$, and $SU(3) \times U(1)^2$ anomalies are zero, and the $SU(2)^3$ anomaly is zero since $SU(2)$ has no complex representations (recall that the anomaly indicator which diagnoses perturbative gauge anomalies vanishes if the representation R in question is isomorphic to R^*). Thus for this to work, we need a trivial $SU(3)^3$ anomaly from the R's. Unfortunately the only non-anomalous representations of $SU(3)$ are the real ones, which are of the form (i, i) (where the two indices label the values of the different diagonal generators). Since all the representations of $SU(2)$ are either real or pseudo-real, we in fact have that the conjugate of (R, S, q) is isomorphic to $(R, S, -q)$, which is a contradiction.

What about $k = 3$? After playing around a bit, we can write down the following:

$$(\mathbf{6}, \mathbf{3}, 1), \quad (\mathbf{6}, \mathbf{3}, -1), \quad (\bar{\mathbf{15}}', \mathbf{3}, 0), \tag{7}$$

where $\mathbf{6}$ is the six-dimensional $SU(3)$ representation $(2, 0)$, $\bar{\mathbf{15}}'$ is the 15-dimensional $SU(3)$ irrep $(1, 2)$, and $\mathbf{3}$ is the vector representation (spin one) of $SU(2)$. The anomaly coefficients of these $SU(3)$ irreps are $A(\mathbf{6}) = 7$ and $A(\bar{\mathbf{15}}') = -14$ (see e.g. Cutler & Kephart 2000), and so the $SU(3)$ anomaly vanishes:

$$A_{SU(3)^3}((\mathbf{6}, \mathbf{3}, 1) \oplus (\mathbf{6}, \mathbf{3}, -1) \oplus (\bar{\mathbf{15}}', \mathbf{3}, 0)) = 7 \times 3 + 7 \times 3 - 14 \times 3 = 0. \tag{8}$$

The $U(1)^3$ anomaly, $U(1)$-graviton anomaly, and the mixed $U(1) \times SU(2)^2$, $U(1) \times SU(3)^2$ anomalies are all zero since we have a symmetric coupling between 1 and -1 hypercharges

[1] This works because the spinor representation is pseudoreal, with the pseudoreal structure being given by an antisymmetric matrix J. Then for any left-handed (say) fermion ψ_L appearing in the action and transforming in the representation R under the gauge group, we can change variables to work with $\psi_R = J\psi_L^\dagger$, which since $\bar{\gamma}J = -J\bar{\gamma}$ has the opposite chirality to ψ_L. The kinetic term for ψ_R then involves a coupling to the gauge field as $\bar{\psi}_R^\dagger A_\mu^a \gamma^\mu [T^a]^T \psi_R$, and by Hermiticity of the Lie algebra generators we see that ψ_R transforms in the R^* representation. So we can work e.g. change all right-handed fermions to left-handed ones as long as we remember to reverse their charges.

to the rest of the representations. Since $\bar{\mathbf{6}} \ncong \mathbf{6}$, there are no problems there. Finally, the global $SU(2)$ anomaly is also zero, since it requires the Dynkin index

$$t_2(S) = d(S)\frac{d^2(S)-1}{12} \tag{9}$$

to be integral, which it is: $t_2(\mathbf{3}\oplus\mathbf{3}\oplus\mathbf{3}) = 3\times 2$ (this amounts to requiring that $\mathrm{Tr}(T_3^2)$ be integral rather than half-integral, where we define $\mathrm{Tr}(T_3^2) = 1/2$ for the fundamental representation. Essentially, we need there to be an even number of $SU(2)$ doublet fermions). So, evidently these three fields do the job.

This however made use of some big $SU(3)$ representations — what if we restrict ourselves to the trivial and fundamental representations? Cancellation of the global $SU(2)$ anomaly means that we need an even number of $SU(2)$ doublets. Can we do it with three fields? This would mean two doublets, and cancellation of the $SU(2)^2\times U(1)$ anomaly would mean that they would have opposite hypercharges, with the third field having zero hypercharge. Then cancellation of the $SU(3)^2\times U(1)$ anomaly would mean that both doublets are in the fundamental of $SU(3)$, but then the $SU(3)^3$ anomaly is non-zero. So at least four fields are needed.

For four fields, we either need two $SU(2)$ doublets or four for cancellation of the global $SU(2)$ anomaly. If we have two then they must have opposite hypercharges, and then must both be in $\mathbf{3}$ under $SU(3)$ by our assumptions about not having conjugate fields. But then the remaining two fields cannot possibly cancel the $SU(3)^3$ anomaly, so this doesn't work. So, can we have four $SU(2)$ doublets? Yes we can; consider the quadruplet

$$(\mathbf{3},\mathbf{2},a),\quad (\bar{\mathbf{3}},\mathbf{2},-a),\quad (\mathbf{3},\mathbf{2},b),\quad (\bar{\mathbf{3}},\mathbf{2},-b), \tag{10}$$

where a, b are subject only to the constraint that $a \neq \pm b$. One sees that all the anomalies cancel.

Now we assume that there are five fields, as in the first generation of the Standard Model (of course in the SM, grouped according to their $SU(2)$ charges, these are $(e_L,\nu_L), (u_L,d_L), u_R, d_R, e_R$, with a possible sterile ν_R omitted). We can either have two or four $SU(2)$ doublets. Suppose first that there are four $SU(2)$ doublets. Then since we need at least one field charged under everything and we need the $SU(3)^3$ anomaly to vanish, we can either have two of these doublets in the fundamental of $SU(3)$ and two in the anti-fundamental, or one in the fundamental, one in the anti-fundamental, and two in the trivial. Consider first the former situation. Then the vanishing of the $SU(2)^2\times U(1)$ anomaly in this case means that

$$\sum_{\text{doublets}} q_i = 0, \tag{11}$$

where q_i is the hypercharge. But then the $U(1)$-graviton anomaly implies that $q_5 = 0$, where q_5 is the hypercharge of the last field. But then the last field is completely trivial, and we get the same answer as the $k=4$ case. So, now consider the case where one of the doublets is in the fundamental of $SU(3)$ and the other is in the anti-fundamental. Then the $SU(3)^2\times U(1)$ anomaly means that the hypercharges of the two fields coupled to $SU(3)$ are equal. Let their hypercharges be a, and let the charges of the other two $SU(2)$ doublets be b, c. Then we need $6a + c + d = 0$. Thus the $U(1)$-graviton anomaly reads $12a + 2c + 2d + q_5 = 0$, and so again we conclude that $q_5 = 0$ and the last field is completely decoupled.

So, we can assume wolog that there are only two $SU(2)$ doublets, as in the (first generation of the) SM. Suppose they are both charged under $SU(3)$. They cannot both be in the same $SU(3)$ irrep, since then there is no way to cancel the $SU(3)^3$ anomaly with the other $SU(2)$-neutral fields (the dimensions of the other fields are too small). So, one must be in $\mathbf{3}$ with the other in $\bar{\mathbf{3}}$. But then cancellation of the $SU(2)^2 \times U(1)$ anomaly means that the two doublets have opposite hypercharges, which gives a contradiction since we now have two fields that are conjugate to one another.

This means we can without loss of generality take the two doublets to be of the form

$$(\mathbf{3}, \mathbf{2}, a), \quad (\mathbf{0}, \mathbf{2}, b). \tag{12}$$

Cancellation of the $SU(3)^3$ anomaly means that the other three fields have to be of the form

$$(\bar{\mathbf{3}}, \mathbf{0}, c), \quad (\bar{\mathbf{3}}, \mathbf{0}, d), \quad (\mathbf{0}, \mathbf{0}, e), \tag{13}$$

which is starting to look increasingly like the Standard Model. Now one just needs to solve for the $U(1)$ hypercharges. We then have to solve the constraints

$$\begin{aligned} &6a + 2b + 3c + 3d + e = 0 \qquad 6a^3 + 2b^3 + 3c^3 + 3d^3 + e^3 = 0, \\ &2a - c - d = 0, \qquad 3a + b = 0. \end{aligned} \tag{14}$$

I'm not going to write out all the algebra, which is straightforward. One finds that there are two solutions. One has all of the hypercharges equal to zero except for $c = -d$. Again, this leaves us with a completely decoupled field, and we are back to the case of $k = 4$ (and it violates our assumptions about there being at least one field charged under everything). The other solution is the Standard Model. So, given these assumptions, the SM is unique.

What other sorts of possible matter content can we have? There are many options, since there are many semisimple Lie groups with real or pseudo-real representations. There are also a few groups with complex representations that always have zero anomaly coefficients, like $SO(4n + 2)$ with $n > 2$ (see Weinberg II). In particular, any gauge group that doesn't include factors of $SU(n)$ with $n \geq 3$ or factors of $U(1)$ will be automatically anomaly-free, regardless of the field content. If we have the group $SU(3) \times G$ where G is anomaly-free, then under the above restrictions we can have two fields $(R, g), (\bar{R}, g)$ where g is any complex irrep of G, or we can have $(R, g), (\bar{R}, h)$, where g, h are two distinct real representations of the same dimension. We can do a similar thing if the group is instead $U(1) \times G$. At this point, making concrete statements about what is allowed and what is not allowed has to be done on an increasingly case-by-case basis.

3 *Chiral anomaly via Pauli-Villars* ✓

This is the first logical half of P&S problem 19.4, whose problem statement is as follows:

Derive the chiral anomaly for QED in four dimensions using a PV regulator, i.e. by adding a regulator fermion Ψ with mass M to the theory and showing that $\langle \partial_\mu j^{5\mu} \rangle$ is non-zero in the $M \to \infty$ limit (of course the current is explicitly not conserved when $M > 0$, but in the absence of an anomaly the extent of the non-conservation will vanish as $M \to \infty$). As usual, treat the gauge field as a background field. The relevant integral you will need to do should be UV-finite, even before the $M \to \infty$ limit is taken. Show that the anomaly is given by the limit

$$\langle d^\dagger j_5 \rangle = \lim_{M\to\infty} \left(\langle p, k | 2iM\bar{\Psi}\gamma^5\Psi | 0 \rangle \right), \tag{15}$$

where the state $|p, k\rangle$ contains the two photons of momenta p, q which appear as two of the legs in the usual triangle diagram (the current insertion is the third leg).

We will do the regularization by adding the massive fermion to the Lagrangian with the kinetic term $-\bar{\Psi}(i\not{D}_A - M)\Psi$. Consider now calculating the anomaly by e.g. point-splitting the fermion operators occuring in the expression for the chiral current and connecting the two fermions with a Wilson line. This calculation will receive contributions from both the original ψ fermion and the regulator fermion Ψ. As we saw in the diary entry on the anomaly in non-Abelian gauge theories, we will end up doing an integral like

$$\int_{k,q} e^{ik\cdot\eta - iq\cdot x} \frac{\epsilon^{\mu\nu\lambda\sigma}(k+q)_\nu A_\lambda k_\sigma}{(k^2 - m^2)((k+q)^2 - m^2)}, \tag{16}$$

where η is an infinitesimal distance between two fermion operators which we are point-splitting and m is either 0 or M. The hope is that since we are sending $\eta \to 0$, we can get away with doing this before sending $M \to \infty$, and expand the integral at large k (viz. larger than M), getting a divergent contribution which is independent of M (and only the divergent contribution matters for the anomaly, since this integral is multiplied by an infinitesimal term proportional to η coming from the expansion of the operators in the OPE). So, this will result in the contributions from the two fermions cancelling, even though one is massive and the other isn't. The reason that we might expect this to work is basically that the anomaly is sensitive to UV physics (it comes from doing an OPE of two fermions defined at the same point), and so if we are allowed to take $M \to \infty$ at the end of the calculation in the spirit of PV regularization, the anomalous term should be independent of M.

Of course, we haven't actually gotten rid of the anomaly, since the Ψ fermion could still contribute to it directly. The chiral current for the massive fermion is

$$j^{5\mu} = \bar{\Psi}\gamma^\mu\gamma^5\Psi. \tag{17}$$

We then take the divergence and use the Dirac equation for Ψ, so that

$$\langle \partial_\mu j^{5\mu} \rangle = \langle 2iM\bar{\Psi}\gamma^5\Psi \rangle. \tag{18}$$

Since this vev can only comes from loops we'd naively expect it to be zero after we send $M \to \infty$, but as we will see this is not the case.

The first contributions to this expectation value come from triangle-diagram-like diagrams, with two outgoing photons radiating from a Ψ loop and with one insertion of $2iM\bar{\Psi}\gamma^5\Psi$ on the Ψ loop. These diagrams are derived in the usual way by adding $\int 2i\omega M\bar{\Psi}\gamma^5\Psi$ to the action, integrating out the fermions, expanding the ln det, and then differentiating with respect to ω by selecting out only those diagrams with a single insertion of $2iM\bar{\Psi}\gamma^5\Psi$. The relevant diagram (a fermion bubble with two photon legs and an insertion of $M\bar{\Psi}\gamma^5\Psi$) gives us

$$\begin{aligned}
\langle p,k|2iM\bar{\Psi}\gamma^5\Psi|0\rangle &= 2iMe^2i^3\int_q \mathrm{Tr}\left[\gamma^\nu\frac{1}{\not{q}-M}\gamma^\mu\frac{1}{\not{q}-\not{p}-M}\gamma^5\frac{1}{\not{q}+\not{k}-M}\right]\epsilon^*_\mu(p)\epsilon^*_\nu(k)\\
&= -8iM^2e^2\int_q \epsilon^{\nu\lambda\mu\sigma}\frac{q_\lambda(q-p)_\sigma - q_\lambda(q+k)_\sigma + (q-p)_\lambda(q+k)_\sigma}{(q^2-M^2)((q-p)^2-M^2)((q+k)^2-M^2)}\epsilon^*_\mu(p)\epsilon^*_\nu(k)\\
&= -8iM^2e^2\int_q\int_{x,y} 2\frac{q^2g_{\lambda\sigma}/4 - p_\lambda k_\sigma}{(q^2-M^2+\dots)^3}\epsilon^*_\mu(p)\epsilon^*_\nu(k),
\end{aligned} \tag{19}$$

where in the last step the 2 comes from passing to Feynman parameters and the $\dots$ are a standin for terms involving the Feynman parameters and the photon momenta. We do the shift to simplify the denominator in the usual way, but we don't have to really keep track of what happens since the q^2 term in the numerator dies by antisymmetry anyway. So we get

$$\langle p,k|2iM\bar{\Psi}\gamma^5\Psi|0\rangle = 16iM^2e^2\int_q\int_{x,y}\epsilon^{\mu\nu\lambda\sigma}p_\lambda k_\sigma\frac{\epsilon^*_\mu(p)\epsilon^*_\nu(k)}{(q^2-\Delta)^3}, \tag{20}$$

where $\Delta = M^2 + \dots$. This integral is perfectly UV finite, even without the $M\to\infty$ limit. We get

$$\int_q \frac{1}{(q^2-\Delta)^3} = -\frac{i}{2(4\pi)^2}\Delta^{-1}. \tag{21}$$

Taking now $M\to\infty$ the integral over the Feynman parameters just gives 1 since $\Delta\to M^2$, and so we get

$$\langle k,p|\partial_\mu j^{5\mu}|0\rangle = \frac{e^2}{2\pi^2}\epsilon^{\mu\nu\lambda\sigma}\epsilon^*_\mu(p)\epsilon^*_\nu(k)p_\lambda k_\sigma, \tag{22}$$

which is the expected result. Evidently the somewhat dubious arguments we made in the beginning about terms cancelling worked out this time.

4 *Trace anomaly via Pauli-Villars* ✓

This is the second logical half of the P&S problem that we started in the previous diary entry. The to-do list is:

Reproduce the trace anomaly for *QED* in four dimensions with a PV regularization scheme (you will also need to dimensionally-regularize one integral). In a similar spirit to the last problem, you should show that the anomaly is computed by taking the $M\to\infty$ limit of

$$\langle M\bar{\Psi}\Psi\rangle. \tag{23}$$

Recall that the generator of dilitations is $D^\mu = T^{\mu\nu}x_\nu$, where $T^{\mu\nu}$ is the "improved" (symmetric, gauge-invariant) stress tensor. Now $\partial_\mu D^\mu = T^\mu{}_\mu$, so we get a trace anomaly when $\langle T^\mu{}_\mu\rangle$ is "unexpectedly" nonzero (once we learn about RG this is of course not surprising at all, since when $e \neq 0$ we are not at a fixed point).

Let's first write down what we expect to get for $\langle T^\mu{}_\mu\rangle$. In four dimensions e is dimensionless at the free fixed point, but since it is marginally irrelevant a mass scale is still introduced into the theory when doing renormalization. Performing a dilitation $x \mapsto e^{-\lambda}x$ for λ infinitesimal, the gauge coupling changes as (apologies for the profusion of e's — should change to g, but too late)

$$e \mapsto e + \partial_M e \delta M = e + \partial_M e(e^\lambda M - M) \implies \delta e = \beta(e). \tag{24}$$

Performing this change of coordinates in the path integral, by the usual procedure we get

$$\langle \partial_\mu D^\mu\rangle = \langle T^\mu{}_\mu\rangle = \lambda\beta(e)\partial_e\mathcal{L} = \frac{\beta(e)}{2e^3}F_{\lambda\sigma}F^{\lambda\sigma} = \frac{1}{24\pi^2}F_{\lambda\sigma}F^{\lambda\sigma}, \tag{25}$$

where in the last step we've used the first-order result for the β function in QED.

The goal now is to see how we can reproduce this with an explicit calculation. To find $\langle T^\mu{}_\mu\rangle$, we can formally add $\int \eta T^\mu_\mu$ to the action, integrate out the fermions, and then differentiate with respect to η. To do this we need the energy-momentum tensor for QED. Now we know the Hamiltonian is

$$H = \frac{1}{2e^2}(E^2+B^2) - i\bar{\psi}(\gamma^i\partial_i - m)\psi, \tag{26}$$

since when going from the Lagrangian to the Hamiltonian the ∂_t part in the Dirac operator gets cancelled. So then we just need to write down a more covariant expression of μ,ν that reduces to the above when $\mu=\nu=0$. Such an expression is

$$T^{\mu\nu} = -\frac{1}{e^2}F^{\mu\sigma}F^\nu_\sigma + \frac{1}{4e^2}g^{\mu\nu}F^2 + i\bar{\psi}(\gamma^{\{\mu}D_A^{\nu\}})\psi - g^{\mu\nu}\bar{\psi}(i\not{D}_A - m)\psi. \tag{27}$$

When this gets stuck in diagrams, it gives us source counterterms for the gauge field propagator, the fermion propagator, and the electron / photon vertex. If we hadn't added the regulator fermion, we would calculate the anomaly by computing $\langle T^\mu{}_\mu\rangle$ perturbatively in dimensional regularization, as outlined in P&S chapter 19. Using $g^\mu_\mu = g^{\mu\nu}g_{\nu\mu} = d$ in dimensional regularization, we have, for a massless fermion,

$$\langle T^\mu{}_\mu\rangle = F^2\left(\frac{d}{4}-1\right) + (1-d)\bar{\psi}i\not{D}_A\psi. \tag{28}$$

The second term ends up not contributing to diagrams because of the equations of motion. However the first term does — it acts as a counterterm in photon propagators, and when these are stuck onto the legs of polarization bubbles with fermions running in the loop, the $(d/4-1)$ is rendered finite when it hits a divergent part of a $\Gamma(\epsilon/2)$ coming from the fermion integration. When we add in the regulator fermion, we get contributions to the polarization bubble from both the original fermion and the regulator. One can fairly quickly check (I won't write it out since it's similar to the calculation we'll do below) that the relevant one-loop contribution to the trace anomaly due to the regulator fermions comes from the integral

$$\frac{\epsilon}{4}\int_q \frac{f(k,p)}{(q^2-\Delta)^2}, \tag{29}$$

where p, k are external photon momenta and Δ goes to M^2 in the large M limit. In four dimensions, we get something like

$$\epsilon\left(\Gamma(\epsilon/2) - \ln\Delta + \dots\right), \tag{30}$$

and so when we take $\epsilon \to 0$, we get something which is independent of M! Thus, the regulator fermions exactly cancel the contribution to anomalous terms in the polarization bubbles that the massless fermions make! This is similar to the reasoning in the last problem — we're still arguing that we can get away with waiting until the very end of the calculation to take $M \to \infty$.

Of course, this doesn't eliminate the anomaly, since the regulator fermions contribute to the trace of T explicitly through their mass term. So, we have

$$\langle T^\mu{}_\mu \rangle = \langle M\bar{\Psi}\Psi \rangle. \tag{31}$$

Our task is to evaluate this in the limit $M \to \infty$ and see whether it is zero or not.

As usual, the first diagram that shows up is a triangle-type diagram with two photon legs and an insertion of the relevant operator (here $M\bar{\Psi}\Psi$) on the fermion loop. We get

$$\langle p, k|M\bar{\Psi}\Psi|0\rangle = iMe^2 \int_q \mathrm{Tr}\left[\gamma^\nu \frac{1}{\not{q} - M}\gamma^\mu \frac{1}{\not{q} - \not{p} - M}\frac{1}{\not{q} + \not{k} - M}\right] \epsilon^*_\mu(p)\epsilon^*_\nu(k) \tag{32}$$

Only terms with even numbers of γ matrices contribute. Their contributions can be found using standard rules for tracing out products of γ matrices. Note that we cannot at this stage ignore linear in $\not{q}$ terms in the numerators, since we haven't Feynman-ized the denominator yet to make it a function of just q^2.

We need two Feynman parameters to do the integral. After a bit of algebra, one sees that the appropriate shift in momentum needed to eliminate the linear-in-q terms in the denominator is

$$q \mapsto q + kx - py, \tag{33}$$

where x, y are the Feynman parameters. The numerator, before the shift, is

$$4g^{\mu\nu}M^3 + M\mathrm{Tr}\left[(\not{p} + \not{q})\gamma^\mu\not{q}\gamma^\nu + (\not{p} + \not{q})\gamma^\mu\gamma^\nu(\not{q} - \not{k}) + \gamma^\mu\not{q}\gamma^\nu(\not{q} - \not{k})\right], \tag{34}$$

where we have dropped things with an odd number of gamma matrices. The possible terms that we get after the shift which have an even number of qs go like

$$g^{\mu\nu}, \quad g^{\mu\nu}q^2, \quad q^\mu q^\nu, \quad k^2 g^{\mu\nu}, \quad p^2 g^{\mu\nu}, \quad k^\mu p^\nu, \quad k^\nu p^\mu, \quad p^\mu p^\nu, \quad k^\mu k^\nu. \tag{35}$$

The k^2 and p^2 terms are zero since we're dealing with photons. Anything with a p^μ or a k^ν is zero, since these momenta will be contracted with the polarizations and will thus be killed. To do the trace over the surviving terms, we use

$$\begin{aligned}\mathrm{Tr}[\not{r}\gamma^\mu\not{s}\gamma^\nu] &= 4(r^\mu s^\nu + r^\nu s^\mu - r\cdot s g^{\mu\nu}),\\ \mathrm{Tr}[\not{r}\gamma^\mu\gamma^\nu\not{s}] &= 4(r^\mu s^\nu - r^\nu s^\mu + r\cdot s g^{\mu\nu}).\end{aligned} \tag{36}$$

I won't write out all the algebra for expediency's sake, since it is straightforward. One can check that the surviving $p^\nu k^\mu$ term appears as

$$4Mp^\nu k^\mu(1-4xy), \tag{37}$$

and that the $p\cdot k$ appears with the same coefficient but with opposite sign (as dictated by gauge invariance).

What of the M^3 term in the numerator? This turns out to get killed by other things in the trace. We have

$$\begin{aligned}\mathrm{Tr}\left[\not{q}\gamma^\mu\not{q}\gamma^\nu+\not{q}\gamma^\mu\gamma^\nu\not{q}+\gamma^\mu\not{q}\gamma^\nu\not{q}\right] &= 16q^\mu q^\nu-4g^{\mu\nu}q^2\to g^{\mu\nu}\left(\frac{16}{4-\epsilon}-4\right)q^2,\\ &= g^{\mu\nu}\epsilon q^2\end{aligned} \tag{38}$$

where we've anticipated the use of dimensional regularization in the integral. This term is integrated with a denominator of $(q^2-\Delta)^3$, where as usual Δ is a function of x,y,p,k. This term then cancels the M^3 term:

$$\begin{aligned}\int_{q,x,y}\frac{4M^3+\epsilon Mq^2}{(q^2-\Delta)^3} &= \frac{-4iM^3}{32\pi^2\Delta}+\frac{Mi\epsilon}{16\pi^2}\left(\Gamma(\epsilon/2)+\text{finite}\right)\\ &= \frac{i}{8\pi^2}(-M^3/\Delta+M)\to 0,\end{aligned} \tag{39}$$

since $\Delta\to M^2$ in the $M\to\infty$ limit. So, we have

$$\langle p,k|M\bar{\Psi}\Psi|0\rangle=-2e^2Mi\int_{q,x,y}\frac{4(4xy-1)(p\cdot kg^{\mu\nu}-p^\nu k^\mu)}{(q^2-\Delta)^3}\epsilon^*_\mu(p)\epsilon^*_\nu(k). \tag{40}$$

The integral over the Feynman parameters is done over the range $\int_0^1\int_0^{1-x}dxdy$ since the integral is over the face of a 3-simplex in $\mathbb{R}^3$. Since we can set $\Delta\to M^2$ independent of x,y in the $M\to\infty$ limit, the integral over the Feynman parameters just yields 2/3. So we get

$$\begin{aligned}\langle T^\mu{}_\mu\rangle &= -\frac{16e^2Mi}{3}\int_q\frac{p\cdot kg^{\mu\nu}-p^\nu k^\mu}{(q^2-M^2)^3}\epsilon^*_\mu(p)\epsilon^*_\nu(k)\\ &= \frac{e^2}{6\pi^2}\left(p\cdot k\epsilon^*(p)\cdot\epsilon^*(k)-p\cdot\epsilon^*(k)k\cdot\epsilon^*(p)\right).\end{aligned} \tag{41}$$

Is this a sensible answer? Yes! We see that this is exactly equal to what we expected from the argument given earlier relating the trace anomaly to the beta function for the gauge coupling (note to self — did we drop a factor of two somewhere?)

5 *Mixed t Hooft anomaly for the compact boson / Luttinger liquid* ✓

Consider the compact boson in two dimensions at radius R:

$$S=\frac{R^2}{4\pi}\int\partial_\mu\phi\partial^\mu\phi. \tag{42}$$

Today's problem statement is as follows:

Working on the spacetime $S^1 \times \mathbb{R}$, write down the charge operators for the momentum and winding number global symmetries. Carry out the analysis using Abelian duality to define a field σ such that $d\phi \propto \star d\sigma$, and find the commutation relations between ϕ and σ.

Next, consider a field with non-zero charge under both momentum and winding number symmetries, and find the algebra of the symmetry charges. By considering subregion charge operators which act on submanifolds of the spatial circle, show that the symmetry is realized projectively. Demonstrate the nontrivial third $U(1)$ cohomology class which parametrizes the mixed t Hooft anomaly.

I first learned about this approach to anomalies from Alvarez's old paper [**?**]. A great reference for this subregion charge operator approach to anomalies is in [**?**]; see also the earlier work [**?**] for a good discussion of this way of thinking about anomalies as well as an example similar to today's where the symmetries in question are reduced to $\mathbb{Z}_2$ subgroups.

First let us write down the dual theory. This is done in the standard way by promoting $d\phi \mapsto D_B\phi$ for B a dynamical one-form, adding the term $\frac{i}{2\pi}\int B \wedge d\sigma$ for σ a zero-form, choosing unitary gauge to set $\phi \to 0$, and finally doing the Gaussian integral over B. This produces the dual action

$$S = \frac{1}{4\pi R^2}\int \partial_\mu\sigma\partial^\mu\sigma. \tag{43}$$

We can find the relation between the field strengths by looking at where $d\phi$ goes under the mapping. We take $d\phi \to D_B\phi$ and then kill ϕ to get B. After we do the shift on B to eliminate the $B \wedge d\sigma$ coupling and then do the integral over B, we are left with $R^{-2} \star d\sigma$. Thus under the duality, we have

$$d\phi \leftrightarrow \frac{1}{R^2} \star d\sigma. \tag{44}$$

From the original action, we see that the canonical momentum for ϕ is[2]

$$\pi_\phi = \frac{R^2}{2\pi}\partial_t\phi = \frac{R^2}{2\pi}\frac{1}{R^2}\epsilon^{tx}\partial_x\sigma = \frac{1}{2\pi}\partial_x\sigma, \tag{45}$$

where x is the direction along the spatial S^1. Likewise, we could also do canonical quantization while choosing σ as the coordinate; in this case

$$\pi_\sigma = -\frac{1}{2\pi}\partial_x\phi. \tag{46}$$

The symmetry here is part of the reason why we chose the factors of 2π in the action as we did.

From the commutation relation

$$[\phi(x), \partial_y\sigma(y)] = 2\pi i\delta(x-y), \tag{47}$$

we get

$$[\phi(x), \sigma(y)] = -2\pi i\Theta(x-y). \tag{48}$$

[2]Here we write an $=$ when replacing something with its image under T-duality. This is legit since we are considering a situation in which both fields appear in S only quadratically, and hence can be integrated out exactly, so that duality can be used as an equality inside partition functions.

We could also have chosen the RHS to be $-\pi i \mathrm{sgn}(x-y)$, but the Θ function will be more convenient later on. This commutator shows that ϕ, σ are mutually very non-local, which makes sense due to them (or rather their derivatives) essentially being Fourier transforms of one another. Also note that we said we were working on a spatial S^1, which means that strictly speaking the Θ function above requires a basepoint. Technicalities related to this are no fun to keep track of and don't change the picture of the results we'll be getting, and so we will ignore these issues in what follows.

Using our expressions for the momenta of the fields, we see that the charges for the winding number and momentum symmetries are represented on charge w, n fields as

$$Q^w = \exp\left(\frac{iw}{2\pi}\int \partial_x \sigma\right), \qquad Q^n = \exp\left(-\frac{in}{2\pi}\int \partial_x \phi\right), \tag{49}$$

where the integrals are over the whole S^1. Note that the charge operators involve exponentials of operators which don't commute, which will be important later.

To see if there are any anomalies, we should ask whether or not the symmetry is "splittable", i.e. whether or not the charge operators can be well-defined as operators acting on submanifolds of the spatial S^1 rather than on all of space. If the symmetry is splittable then we have a well-defined local current (obtained by making the submanifolds infinitesimally small); this can then be coupled to a gauge field and the symmetry can be gauged, implying the theory is anomaly free. If on the other hand the symmetry is not splittable then there must not be a well-defined local current, which prevents the symmetry from being gauged.[3]

To this end, consider the symmetry generator for an element $(\alpha, \beta) \in U(1)_{\text{momentum}} \times U(1)_{\text{winding}}$ on a given interval $I = [a, b]$:

$$U_I(\alpha, \beta) = \exp\left(-\frac{in[\alpha]}{2\pi}\int_a^b \partial_x \phi\right) \exp\left(\frac{iw[\beta]}{2\pi}\int_a^b \partial_x \sigma\right), \tag{50}$$

where we have used the notation

$$[x] \equiv x \bmod 2\pi. \tag{51}$$

We now want to know whether or not these subsystem symmetry generators realize a linear representation of the $U(1) \times U(1)$ symmetry group. Noting that the two exponentials in the definition of U_I commute,[4] one sees that

$$U_I(\alpha, \beta) U_I(\alpha', \beta') = \Omega_1^2((\alpha, \beta), (\alpha', \beta'); I) U_I(\alpha + \beta, \alpha' + \beta'), \tag{53}$$

[3]Of course while we have used the word "current" these statements about splittability continue to hold when the symmetry group is discrete — indeed the failure of splittability (i.e. onsite-ness of the symmetry action) is usually how cond-mat people define the anomaly.

[4]This is because exponentials can be re-arranged at the cost of a factor

$$\exp\left(i\frac{nw[\alpha][\beta]}{2\pi}\int_a^b dx\, dy\, \partial_x \delta(x-y)\right) = 1, \tag{52}$$

which only works because the two integrals over x and y are over the exact same domain; if one wiggles the support of one of the integrals this property disappears. This non-commutativity is another way to see the mixed anomaly and is elaborated on elsewhere. Different regularization conventions produce extra phase factors that don't affect the cohomology classes of the obstructions derived below, and so we won't bother keeping track of them.

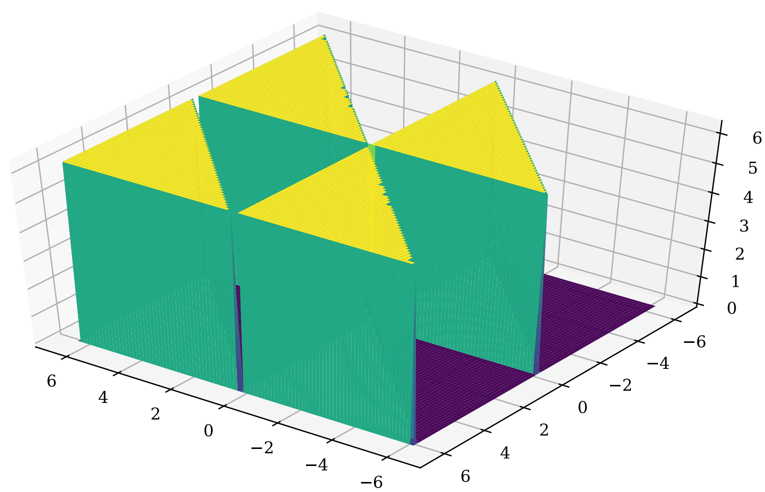
where the operator Ω_1^2 is a function of two group elements integrated over a 1-submanifold:

$$\Omega_1^2((\alpha,\beta),(\alpha',\beta');I) = \exp\left(-\frac{in\Delta(\alpha,\alpha')}{2\pi}\int_b^a \partial_x\phi\right)\exp\left(\frac{iw\Delta(\beta,\beta')}{2\pi}\int_b^a \partial_x\sigma\right), \tag{54}$$

where we have defined the coboundary operator $\Delta : \mathbb{R}^2 \to 2\pi\mathbb{Z}$ by

$$\Delta(x,y) = [x]+[y]-[x+y]. \tag{55}$$

$\Delta(x,y)$ is nontrivial for certain triangular regions; a plot looks like this:



(56)

Since Δ maps into $\overline{\mathbb{Z}}$, Ω_1^2 is trivial when we take the interval I to extend over all of the spatial manifold Σ, since then the integrals are then $\int_\Sigma d\phi, \int_\Sigma d\sigma \in \overline{\mathbb{Z}}$ (assuming periodic boundary conditions on Σ).

Anyway, nontriviality of Ω_1^2 indicates that "symmetry fractionalization" occurs, at least in the extent that the split charge operators only act projectively on the Hilbert space of the theory (although here, the extent to which they are projective is caputred by a nontrivial operator rather than a simple phase factor). Note that the true global symmetry doesn't really get fractionalized, since the full charge operators still act in a linear representation ($\Omega_1^2 = 1$ when $I = \Sigma$). This is the same as saying that the symmetry can fractionalize on individual excitations, but only if those excitations obey a selection rule where they only come in configurations that transform in a linear representation.[5] This type of symmetry fractionalization is not sufficient to have an anomaly, although it is necessary, as we will see shortly.

Note that $\Omega_1^2((\alpha,\beta),(\alpha',\beta');I)$ only actually operates on ∂I (consistent with it being trivial when $I = \Sigma$), since the fact that $\phi(a)$ commutes with $\sigma(b)$ for $a < b$ means that we

[5]SPTs are thus basically the concept of fractionalization applied to group representations: we break up something transforming in a linear representation to several things transforming in projective ones, with an overall charge neutrality constraint. More on this in a separate diary entry.

can write

$$\Omega_1^2((\alpha,\beta),(\alpha',\beta');I) = \Omega_0^2((\alpha,\beta),(\alpha',\beta');b)[\Omega_0^2((\alpha,\beta),(\alpha',\beta');a)]^*, \tag{57}$$

where

$$\Omega_0^2((\alpha,\beta),(\alpha',\beta');x) = \exp\left(-\frac{in\Delta(\alpha,\alpha')}{2\pi}\phi(x)\right)\exp\left(\frac{iw\Delta(\beta,\beta')}{2\pi}\sigma(x)\right). \tag{58}$$

For example, if we choose $(\alpha,\alpha') = (\beta,\beta') = (\pi,\pi)$, then

$$\Omega_0^2((\pi,\pi),(\pi,\pi)) = e^{-in\phi(x)}e^{iw\sigma(x)}. \tag{59}$$

The fact that the commutation relations for the operators appearing in Ω_0^2 are very non-local implies that the symmetry is not realized in an "onsite" way on the boundary.

Also, as we will see this form for Ω_0^2 is in keeping with the descent equations for anomalies, which we will elaborate on later. Also, note that Ω_0^2 is only defined up to a coboundary of a group 1-cocycle, since we are allowed to re-define the U_I operators by phases via $U_I((\alpha,\beta)) \mapsto e^{ig((\alpha,\beta))}U_I((\alpha,\beta))$.

Now we need to figure out what the anomaly is, by going one level up in group cohomology (Ω_2^1 is a group 2-cochain, and the anomaly will be captured by a 3-cochain). To do this, we examine associativity of the symmetry operators, since the failure of associativity is captured by 3-cochains. The associativity of the U_I operators requires the cocycle condition on Ω_1^2, namely

$$(\delta\Omega_1^2)((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta'')) = 0, \tag{60}$$

where δ is the coboundary operator. One can check that this will hold provided that

$$\Delta(\alpha,\alpha') + \Delta(\alpha+\alpha',\alpha'') - \Delta(\alpha',\alpha'') - \Delta(\alpha,\alpha'+\alpha'') = 0, \tag{61}$$

which is indeed true. Note that we should really be writing e.g. $\Delta([\alpha+\alpha'],\alpha'')$, but we aren't since $\Delta([\alpha+\alpha'],\alpha'') = \Delta(\alpha+\alpha',\alpha'')$. We stress that if this associativity condition did not hold our theory would not make sense — it is merely a consistency check, not an anomaly test.

Now we ask if the action is realized associatively on individual points, i.e. whether or not the Ω_0^2 operators are annihilated by δ. This may or may not be true, and if not, it signals an anomaly. We define the cochain Ω^3 (which is just a phase factor — no ϕ or σ operators included) to measure the lack of associativity through

$$\Omega^3((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta'')) \equiv (\delta\Omega_0^2)((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta'')). \tag{62}$$

This formula requires some explanation. Writing group elements momentarily as g,h,k, associativity tells us that

$$\begin{aligned}
[U_I(g)U_I(h)]\,U_I(k) &= U_I(g)[U_I(h)U_I(k)] \\
\Omega_1^2(g,h)U_I(gh)U_I(k) &= U_I(g)\Omega_1^2(h,k)U_I(hk) \\
\Omega_1^2(g,h)\Omega_1^2(gh,k)U_I(ghk) &=^{U_I(g)} \Omega_1^2(h,k)\Omega_1^2(g,hk)U_I(ghk),
\end{aligned} \tag{63}$$

where $U_I(g)$ acts on Ω_1^2 in the way needed to allow us to pull it through to meet $U_I(hk)$. Thus as checked above, we require that $\delta\Omega_1^2 = 0$, where the coboundary operator includes the action of U_I (we didn't mention the action earlier since Ω_1^2 commutes with the U_I's). However, the individual operators Ω_0^2 which are localized at ∂_I do not need to be co-closed, and so in general we have

$$\Omega_0^2(g,h)\Omega_0^2(gh,k) = \Omega^3(g,h,k)\left({}^{U_I(g)}\Omega_0^2(h,k)\Omega_0^2(g,hk)\right). \tag{64}$$

We can then use the commutation rules to derive the action

$${}^{U_I(\alpha,\beta)}\Omega_0^2((\alpha',\beta'),(\alpha'',\beta'')) = \exp\left(-\frac{inw}{4\pi}\left[\beta\Delta(\alpha',\alpha'') + \alpha\Delta(\beta',\beta'')\right]\right)\Omega_0^2((\alpha',\beta'),(\alpha'',\beta'')). \tag{65}$$

Then using $\Theta(0) = 1/2$ in the commutation relations for ϕ and σ operators at the same point, some algebra gives

$$\Omega^3((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta'')) = \exp\left(\frac{inw}{4\pi}f((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta''))\right), \tag{66}$$

where

$$\begin{aligned} f((\alpha,\beta),(\alpha',\beta'),(\alpha'',\beta'')) = \Delta(\beta,\beta')\Delta(\alpha+\alpha',\alpha'') - \Delta(\beta',\beta'')\Delta(\alpha,\alpha'+\alpha'') \\ + \beta\Delta(\alpha',\alpha'') + \alpha\Delta(\beta',\beta''). \end{aligned} \tag{67}$$

As with Ω_1^2, Ω^3 is only defined up to a 3-coboundary, since we can re-define the Ω_0^2 operators by group 2-cocycles without changing the value of Ω_1^2.

As an example, consider the case where we consider only the $\mathbb{Z}_2 \subset U(1)$ symmetry of translations of both ϕ and σ by π. Setting all of the group elements in the above to π, we find

$$\Omega^3((\pi,\pi),(\pi,\pi),(\pi,\pi)) = (-1)^{nw}. \tag{68}$$

One can check that we would have obtained the form for Ω^3 had we used different conventions for the commutation relations and symmetry generators, e.g. if we had taken the ϕ,σ commutator to go like $-\pi i \mathrm{sgn}(x-y)$ and taken the $\partial_x\phi$ integral in the symmetry generator to be from $a-\epsilon$ to $b+\epsilon$, where $\epsilon \to 0$ is used to ensure that the terms in U_I commute with one another.

Ω^3 represents the anomaly, and lives in $H^3(U(1)\times U(1);U(1))$ as expected. First, note that $\Omega^3 = 0$ if either $n = 0$ or $w = 0$. This signals the fact that the anomaly is a mixed anomaly between the two symmetries: either one by itself is non-anomalous, but together they become anomalous (essentially because their respective subsystem symmetry generators do not commute). Also, note that the way it was defined makes Ω^3 look like a coboundary and hence a trivial cohomology class, but this is not so since although $\Omega^3 = \delta\Omega_0^2$, Ω_0^2 is not an element of $C^2(U(1)\times U(1);U(1))$ (group 2-cochains), since it contains the nonlocal ϕ,σ operators.

As a sanity check, we should make sure that $H^3(U(1)\times U(1);U(1))$ is non-zero using other methods. The Kunneth formula for group cohomology with $U(1)$ coefficients is a little dicey, so we instead calculate $H^4(U(1)^2;\mathbb{Z})$. We use that $H^p(U(1);\mathbb{Z}) = \mathbb{Z}$ if p is even and $H^p(U(1);\mathbb{Z}) = 0$ else, which is true because the group cohomology of $U(1)$ (with $\mathbb{Z}$ coefficients) is the simplicial cohomology of the classifying space of $U(1)$, namely $\mathbb{CP}^\infty$, which has a single generator in each even degree. Thus we get

$$H^4(U(1)^2;\mathbb{Z}) \cong \bigoplus_{i=0}^{4} H^i(U(1);\mathbb{Z}) \otimes_{\mathbb{Z}} H^{4-i}(U(1);\mathbb{Z}) \cong \mathbb{Z}^3. \tag{69}$$

In our example, we are only accessing the $\mathbb{Z}$ factor coming from $H^2 \otimes H^2$ (note to self: come back and pay more attention to group cohomology with a topological group — should probably be using the discrete topology.) Since group cohomology with topological groups is kinda weird, suppose as a consistency check that our symmetry group was instead $\mathbb{Z}_n \times \mathbb{Z}_n$. Then the Kunneth formula gives (recall that $H^p(\mathbb{Z}_n;\mathbb{Z})$ is $\mathbb{Z}$ if $p = 0$, $\mathbb{Z}_n$ if $p > 0$ is even, and 0 if p is odd)

$$H^3(\mathbb{Z}_n^2; U(1)) \cong \mathbb{Z}_n^3, \tag{70}$$

which agrees with our $U(1)$ result in the limit $n \to \infty$.

Finally, we point out that even if we restrict the symmetry group to $\mathbb{Z}_2^2 \subset U(1)^2$ with $\phi \mapsto \phi + \pi, \sigma \mapsto \sigma + \pi$, we still have an anomaly, as Ω^3 becomes a nontrivial class in

$$H^3(\mathbb{Z}_2 \times \mathbb{Z}_2; U(1)) \cong \mathbb{Z}_2^3. \tag{71}$$

On a related note, even if only the $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry is preserved, the theory can still not be gapped without breaking the symmetry: for example, mass terms like $\cos\phi$ are not allowed since they are not symmetric, and while $\cos 2\phi$ is allowed, it is either irrelevant or leads to SSB since it has two inequivalent minima. This is in keeping with the rule that anomalous theories cannot be trivial in the IR unless the symmetry is explicitly broken.

6 *Mixed t Hooft anomalies for free fields* ✓

In a previous diary entry, we saw that the $U(1) \times U(1)$ symmetry of the compact boson in 1+1D was realized projectively, signaling the existence of a mixed t Hooft anomaly. We did this by looking at the algebra of various split symmetry operators. Today we will understand this in a different light by explicitly showing that the $U(1)\times U(1)$ symmetry cannot be gauged, and will write down the anomaly polynomial (which in the 1+1D case is a 4-form). More generally, we will show this in pedantic detail for the $U(1) \times U(1)$ $(D/2-1)$-form symmetry possessed by $(D/2-1)$-form $U(1)$ gauge theory in D dimensions, where D is even.

We will write the action in D dimensions as

$$S = \frac{1}{2g^2} \int_M F \wedge \star F, \tag{72}$$

where F is some $D/2$-form field which is locally $F = dA$ (depending on conventions there should potentially be some combinatorial prefactors dependent on D up front, but we will ignore them). Since the field strength F is a $D/2$ form, this theory is self-dual, as $\star F$ is also a $D/2$ form. The dual action is

$$S = \frac{g^2}{2(2\pi)^2} \int \mathcal{F} \wedge \star\mathcal{F}, \tag{73}$$

where the dual field $\mathcal{A}$ (locally $\mathcal{F} = d\mathcal{A}$) is mapped to the original through $\mathcal{F} \leftrightarrow g^2 \star F/2\pi$ (with the meaning here that the correlation functions of $g^2 \star F/2\pi$ in the $F \wedge \star F$ action are the same as the correlation functions of $\mathcal{F}$ in the $\mathcal{F} \wedge \star\mathcal{F}$ action). In writing both of these actions, we are *not* assuming that A and $\mathcal{A}$ are globally well-defined. In fact, in the partition function we will need to sum over all possible line bundles for A and $\mathcal{A}$.

Suppose now that we try to gauge the two symmetries in the theory (the symmetries of shifting A and its dual $\mathcal{A}$ by flat $(D/2-1)$-forms). In order to gauge the shift symmetries, we introduce a gauge field B which lets us locally transform $A \mapsto A + \alpha$, where α is not necessarily in the kernel of d. There are many ways to see that doing so implies that the $\mathcal{A}$ shift symmetry cannot also be simultaneously gauged, signaling an anomaly. This is essentially because A and $\mathcal{A}$ are related to eachother by a "Fourier transform" and so the way in which a gauge field acts on them cannot be "simultaneously diagonalized" on both of the fields.

In order for the shift symmetry of A to be gauged, we must minimally couple to a background field B as[6]

$$S[B] = \frac{1}{2g^2}\int (F-B) \wedge \star(F-B). \tag{74}$$

For example, if $D = 4$, then we have regular Maxwell theory. Under gauge transformations, both F and B shift by F', where F' is allowed to be the curvature of any connection on any line bundle (not necessarily trivial, since we are summing over all line bundles for the original gauge field A), and B is a 2-connection on a trivial 2-bundle ($U(1)$ gerbe).

Schematically, we can argue that this leads to a breaking of the shift symmetry for $\mathcal{A}$ as follows. The upgraded gauge-invariant current for the shift symmetry of A is $(B-F)/g^2$, which we get by computing $\delta S[B]/\delta B$. This is conserved provided that the sources of the A gauge field match up with $d^\dagger B$, i.e. provided that $d^\dagger F = d^\dagger B$. This holds as the equation of motion for A. We expect that the current for the dual shift symmetry on $\mathcal{A}$ is then

$$\mathcal{J} \sim \star(F-B). \tag{75}$$

However, the conservation of $\mathcal{J}$ is broken by the curvature of B, since

$$d^\dagger \mathcal{J} \sim d^\dagger \star (F-B) \propto \star dB \neq 0. \tag{76}$$

An analogous problem would have occurred had we started with the dual action, and gauged the shift symmetry on $\mathcal{A}$ by taking

$$\mathcal{F} \mapsto (\mathcal{F} - \mathcal{B}). \tag{77}$$

Doing this would lead to a violation of the conservation of the J current of the form

$$d^\dagger J \propto \star d\mathcal{B}. \tag{78}$$

[6] In writing this expression, we are tacitly assuming that B is a connection on a trivial $(D/2)$-bundle, so that B is globally well-defined. Relaxing this assumption can be done but leads to keeping track of more details which aren't super illuminating.

Essentially, the point is that no matter which formulation we choose, one of the currents must not be conserved after we introduce a gauge field.[7]

Now for a more careful argument. To figure out what happens in the dual formulation, we just run Abelian duality on this in the standard way. If the coupling to the background gauge fields is not anomalous, we should get an action like $S[\mathcal{B}] \sim g^2 \int (\mathcal{F} - \mathcal{B})^2$. The usual duality recipe tells us to first add a gauge field, and then integrate out the original field variables. So, we add a $D/2$-form gauge field a to the action, and then add *another* $(D/2-1)$-form field $\mathcal{A}$ with curvature $\mathcal{F}$ whose job is to kill off a to reproduce the original non-gauged theory. Summing over all line bundles L for A and $\mathcal{L}$ for $\mathcal{A}$, we have the partition function

$$Z = \sum_{L,\mathcal{L}} \int \mathcal{D}A\mathcal{D}a\mathcal{D}\mathcal{A} \exp\left(\frac{1}{2g^2}\int (F-a-B)\wedge\star(F-a-B) + \frac{i}{2\pi}\int \mathcal{F}\wedge a\right). \tag{83}$$

Here we are being slightly sloppy and not writing factors of $1/\mathrm{vol}\, G$ for the various groups of gauge transformations.

In the present partition function A and $\mathcal{A}$ are independent fields. Thus the coupling $\mathcal{F}\wedge a$ is gauge-invariant since $\mathcal{F}$ does not transform under shifts in A (this can be corroborated by checking that under duality, both F and $F + d\alpha$ map to $g^2\mathcal{F}/2\pi$, instead of e.g. $F + d\alpha$ mapping to $\mathcal{F} + g^2 \star d\alpha/2\pi$). Also note that the $\mathcal{F}\wedge a$ coupling is gauge invariant under the $\mathcal{A} \mapsto \mathcal{A} + \alpha$ shift symmetry: since the integral over $\mathcal{A}$ sets a to be exact, after integrating out $\mathcal{A}$ we are left with $\int_M d\alpha \wedge db$ for some b with $a = db$, and so since $\alpha|_{\partial M} = 0$ (as α is a gauge transformation), then $\int_M d\alpha\wedge db = 0$.

To see that we haven't done anything by writing the partition function in this way, we see that the integral over the globally well-defined part of $\mathcal{A}$ sets $da = 0$, and the sum over

[7] We have been setting $\theta = 0$ so far for simplicity. With a θ term, it's easiest to work in terms of the self-dual and anti-self-dual field strengths. We write the action as

$$\frac{i}{4\pi}\int (\tau F_+\wedge F_+ + \bar{\tau}F_-\wedge F_-) = \frac{i}{4\pi}\left(\tau\langle F_+, F_+\rangle - \bar{\tau}\langle F_-, F_-\rangle\right), \tag{79}$$

where the modular parameter is $\tau = \frac{\theta}{2\pi} - \frac{2\pi i}{g^2}$. Running Abelian duality tells us that the appropriate dual fields should be identified as

$$\mathcal{F}_+ = i\tau F_+, \qquad \mathcal{F}_- = i\bar{\tau}F_-. \tag{80}$$

So, without the θ term the duality between F and $\mathcal{F}$ is just Hodge duality plus an inverting of the coupling constant, while the θ term mixes F and its Hodge dual together.

Now we insert a background field for F, making the replacements $F_\pm \mapsto F_\pm - B_\pm$ in the action. We then run duality on this, which won't be written out explicitly for the sake of expediency. The appropriate dualized background field is

$$\mathcal{B} = \mathrm{Re}(\tau)B + \mathrm{Im}(\tau)\star B, \tag{81}$$

and we find that running the duality produces the action we'd expect, up to a contact term for the dual background fields:

$$\begin{aligned} S_{dual} &= -\frac{i}{4\pi}\int\left(\frac{1}{\tau}(\mathcal{F}_+ - \mathcal{B}_+)\wedge\star(\mathcal{F}_+ - \mathcal{B}_+) - \frac{1}{\bar{\tau}}(\mathcal{F}_- - \mathcal{B}_-)\wedge\star(\mathcal{F}_- - \mathcal{B}_-)\right) - S_{\mathcal{B}}, \\ S_{\mathcal{B}} &= -\frac{i}{4\pi}\int\left(\frac{1}{\tau}\mathcal{B}_+\wedge\star\mathcal{B}_+ - \frac{1}{\bar{\tau}}\mathcal{B}_-\wedge\star\mathcal{B}_-\right). \end{aligned} \tag{82}$$

As explained in the main text, this extra contact term signals the mixed anomaly.

$\mathcal{L}$ gives a delta function setting

$$\int_N a \in 2\pi\mathbb{Z} \to [a] \in H^{D/2}_{2\pi\mathbb{Z}}(M;\mathbb{R}), \tag{84}$$

where $N \subset M$ is any closed $D/2$ manifold, and the subscript on the cohomology group indicates that a must integrate over any closed $D/2$ manifold to something in $2\pi\mathbb{Z}$. This quantization condition on a means that it is just the same as F: a closed $D/2$ form which is subject to Dirac quantization. Thus since we are summing over all line bundles for A, we can simply do a change of integration variables for A to absorb a into F, recovering the original action without a or $\mathcal{A}$.

Now we proceed by using the gauge freedom of a to kill off F, which is allowed since F is exact (this amounts to choosing "unitary gauge"). Then

$$Z = \sum_{\mathcal{L}} \int \mathcal{D}a\mathcal{D}\mathcal{A} \exp\left(\frac{1}{2g^2}\int (a+B)\wedge\star(a+B) + \frac{i}{2\pi}\int \mathcal{F}\wedge a\right). \tag{85}$$

Now we kill off the $\mathcal{F}\wedge a$ term by shifting the integration as (the unsavory i here is just because we're working in $i\mathbb{R}$ time)

$$a \mapsto a - B - \frac{g^2 i}{2\pi}\star\mathcal{F}. \tag{86}$$

Quietly absorbing the Gaussian integral into the integration measure, we get the dual action

$$S_{dual} = \frac{1}{2g^2_{dual}}\int (\mathcal{F}\wedge\star\mathcal{F} + \frac{i}{g}(B\wedge\mathcal{F} + \star\mathcal{F}\wedge\star B)), \tag{87}$$

where the dual coupling constant is $g_{dual} = 2\pi/g$. To be a but more suggestive, we define

$$\mathcal{B} = \frac{2\pi i}{g^2}\star B. \tag{88}$$

With this we get

$$S_{dual} = \frac{1}{2g^2_{dual}}\int ((\mathcal{F}-\mathcal{B})\wedge\star(\mathcal{F}-\mathcal{B}) - \mathcal{B}\wedge\star\mathcal{B}) \tag{89}$$

Note that this action is what we would expect to get if we had started by gauging the shift symmetry on $\mathcal{A}$, except for the anomalous $\mathcal{B}\wedge\star\mathcal{B}$ term, which looks like a mass for the background field. We see that this presentation is not gauge invariant no matter how we choose B to transform under the $\mathcal{A}$ shift symmetry, since the action transforms under gauge transformations as

$$\delta S \sim \int_M \alpha\wedge d\star\mathcal{B} \sim \int_M \alpha\wedge dB, \tag{90}$$

and so the gauge invariance of the dual action is broken by the curvature of the background field.

Before we move on, we should point out that we have been using one higher $U(1)$ gauge field B to attempt to gauge both symmetries. In general, one might have thought that we should be allowed to use two higher gauge fields, since the full symmetry is a $U(1) \times U(1)$ 1-form symmetry. This turns out to not work, essentially because the two putative gauge fields have to be related to one another in an inconsistent way by duality. Let the two fields be denoted as B_e and $\star B_m$, with B_m neutral under the shift symmetry of A and B_e neutral under that of $\mathcal{A}$. In order to get a dual action which is electrically gauge-invariant, the B_m fields need to be included in the original action, and so we can write

$$S = \frac{1}{2g^2} \int (dA - B_e - \star B_m) \wedge \star (dA - B_e - \star B_m). \tag{91}$$

In order for this to be magnetically gauge invariant, we need the terms with B_m to integrate to zero. But this is possible only if

$$\int dA \wedge B_m = 0, \tag{92}$$

i.e. only if $dB_m = 0$, and as such we cannot gauge the $\mathcal{A}$ symmetry with a genuine background field (one which is allowed to be non-flat). Any way we look at it, there's a mixed t Hooft anomaly.

7 *Simple example of a discrete 't Hooft anomaly*

This was inspired by wanting to work through and elaborate on appendix D of "Theta, TR, and Temperature" [?]. I'm sure all of what follows exists in the literature somewhere; in any case it's just a slight elaboration on the TTrT paper. Our problem statement is as follows:

Consider the quantum mechanics of a free fermion with periodic boundary conditions in time, with Lagrangian $\mathcal{L} = i\psi^\dagger \partial_t \psi$. Show that this system has a mixed t Hooft anomaly between charge conjugation and $U(1)$. What is the corresponding bulk term that needs to be added to allow the system to be symmetric? Finally, briefly explain why this is equivalent to the bosonic particle-on-a-ring-with-flux model considered in Theta TR and Temperature, which is in turn equivalent to QED_2 at $\theta = \pi$. Consider this model with a cosine potential, and find out whether or not the GSD at $\theta = \pi$ is lifted.

Projective symmetry action

Formally, from the cohomology classification of anomalies (that can be captured this way), we know that in quantum mechanics, anomalies in a symmetry group will be classified by the group cohomology $H^2(G; U(1))$, which classifies central extensions of G by $U(1)$ (aka projective representations).

Let's first see that in quantum mechanics, the action of the classical symmetry group (in our case $O(2)$) must either be enlarged to a linear action of a central extension of the classical symmetry group, or be represented projectively on the Hilbert space of the theory (depending on the way we choose to look at things).

Let us write the group elements of $U(1)$ as α, the generator of $\mathbb{Z}_2^C$ as C, and their representations on the Hilbert space as R_α, R_C, which we will first assume to be unitary linear representations. They act on the ψ operators as

$$\alpha\psi\alpha^{-1} = e^{i\alpha}\psi, \qquad C\psi C^{-1} = \psi^\dagger. \tag{93}$$

Thus from $R_\alpha \psi^\dagger R_\alpha^{-1} = e^{-i\alpha}\psi^\dagger$ (since the representation is assumed to be unitary), we have

$$e^{-i\alpha}\psi^\dagger = R_\alpha R_C \psi R_C^{-1} R_\alpha^{-1}, \tag{94}$$

and so evidently

$$R_\alpha R_C = R_C R_\alpha^{-1}, \tag{95}$$

and so the symmetry group we expect to get is $U(1) \rtimes \mathbb{Z}_2^C = O(2)$.

Is this symmetry realized linearly on the Hilbert space? The Hilbert space is $\mathcal{H} = \{|0\rangle, |1\rangle = \psi^\dagger|0\rangle\}$. Then

$$R_C|1\rangle = R_C\psi^\dagger R_C^{-1} R_C|0\rangle = \psi R_C|0\rangle. \tag{96}$$

If $R_C|0\rangle \propto |0\rangle$ then $R_C|1\rangle = 0$, and so R_C is not a unitary representation. Thus we can take $R_C|1\rangle = |0\rangle$ (inserting a phase factor in this definition will not change the cohomological classification of the anomaly discussed below), and consequently $R_C|0\rangle = |1\rangle$.

Suppose $|0\rangle$ has charge q_0 under $U(1)$. Then

$$R_\alpha|1\rangle = R_\alpha\psi^\dagger R_\alpha^{-1} R_\alpha|0\rangle = e^{-i\alpha}e^{iq_0\alpha}|1\rangle. \tag{97}$$

Thus, if q_1 is the charge of $|1\rangle$, we have

$$q_1 = q_0 - 1. \tag{98}$$

Now consider evaluating this using C:

$$R_\alpha|1\rangle = R_\alpha R_C|0\rangle = R_\alpha R_C R_\alpha^{-1} R_\alpha|0\rangle = R_C R_\alpha^{-1}|0\rangle = e^{-iq_0\alpha}|1\rangle, \tag{99}$$

and hence we also require

$$q_1 = -q_0. \tag{100}$$

Evidently we must have $q_0 = 1/2, q_1 = -1/2$. But this is a contradiction to our assumption about how the symmetry is realized on the Hilbert space, since now

$$R_{\alpha=2\pi}|0\rangle = -|0\rangle, \tag{101}$$

and so $R_{2\pi}$ is not represented trivially on the Hilbert space! In fact, it is represented as $R_{2\pi} = -1$. Thus the symmetry group actually acts projectively, with relations between representations of generators holding only modulo elements of $\mathbb{Z}_2$[8].

[8]Re-defining $C|0\rangle = e^{i\beta}|1\rangle$ for some phase $e^{i\beta}$ would not have changed the fact that $R_{2\pi} = -1$.

Group extensions:

Projective representations are classified by central extensions of the symmetry group. In quantum mechanics, we are allowed to extend the symmetry group by $U(1)$ — that is, we allow the representation of the symmetry to not be linear, as long as the non-linearity manifests itself purely as $U(1)$ phases. Thus in our case the full quantum symmetry group should fit into the exact sequence

$$1 \to U(1) \to G \to O(2) \to 1. \tag{102}$$

Central extensions of the above form are classified by the cohomology group

$$H^2(O(2); U(1)) = \mathbb{Z}_2, \tag{103}$$

which can be calculated with spectral sequences starting from the exact sequence $1 \to U(1) \to O(2) \to \mathbb{Z}_2 \to 1$. The nontrivial element of $H^2(O(2); U(1))$ is the central extension realized by our fermions, where rotations by 2π act as -1. We can write the extension realized by the fermions as

$$1 \to U(1) \to pin_+(2) \to O(2) \to 1, \tag{104}$$

where $pin_+(2)$ is the double-cover of $O(2)$ for which $C^2 = 1$. The extension is central since the image of $\mathbb{Z}_2$ in $pin_+(2)$, namely the identity and the rotation by 2π, is central: $R_C R_{2\pi} = R_{-2\pi} R_C = R_{2\pi} R_C$. Note that since

$$H^2(U(1); U(1)) = H^2(\mathbb{Z}_2^C; U(1)) = 0, \tag{105}$$

we need both the $U(1)$ and the $\mathbb{Z}_2^C$ symmetries to get the anomaly: this is why it is referred to as a mixed t Hooft anomaly.

Let us briefly digress by talking about the difference between $U(1)$ coefficients and $\mathbb{Z}_2$ coefficients. If we instead performed an extension by $\mathbb{Z}_2$, then our allowed quantum symmetry groups would be given by $H^2(O(2); \mathbb{Z}_2)$. We can calculate this given the facts that

$$H^2(O(2); U(1)) \cong \mathbb{Z}_2, \qquad H^3(O(2); U(1)) \cong \mathbb{Z} \times \mathbb{Z}_2. \tag{106}$$

We do this by applying the Kunneth formula for group cohomology to the cohomology $H^2(1 \times O(2); \mathbb{Z}_2 \otimes_{\mathbb{Z}} U(1))$. Since $\mathbb{Z}_2 \otimes_{\mathbb{Z}} U(1) = \mathbb{Z}_2$ as $0 = (1+1) \otimes \alpha = 1 \otimes (2\alpha)$ implies that taking the $\otimes$ with $\mathbb{Z}_2$ kills all of the elements in $U(1)$, this cohomology group is actually $H^2(1 \times O(2); \mathbb{Z}_2 \otimes_{\mathbb{Z}} U(1))$. Since $H^k(1; \mathbb{Z}_2)$ is $\mathbb{Z}_2$ if $k = 0$ and is 0 else, we get the following exact sequence:

$$1 \to \mathbb{Z}_2 \otimes H^2(O(2); U(1)) \to H^2(O(2); \mathbb{Z}_2) \to \mathrm{Tor}\left[\mathbb{Z}_2, H^3(O(2); U(1))\right] \to 1. \tag{107}$$

Plugging in the cohomology groups and using $\mathrm{Tor}[\mathbb{Z}_2, \mathbb{Z} \oplus \mathbb{Z}_2] = 0 \oplus \mathbb{Z}_2$, we see that

$$H^2(O(2); \mathbb{Z}_2) \cong \mathbb{Z}_2^2. \tag{108}$$

Thus, restricting the coefficients to $\mathbb{Z}_2$ leads to a doubling of the number of extensions. This is because with $\mathbb{Z}_2$ coefficients we can fractionalize the relation $C^2 = 1$ to $C^2 = -1$, as well as fractionalizing the 2π rotation. Thus we get $pin_\pm(2)$, $O(2) \times \mathbb{Z}_2$, and an extension where a 2π acts as the identity but $C^2 = -1$. When we use $U(1)$ coefficients instead this classification collapses to just two extensions, since the relation $C^2 = -1$ can be turned into $C^2 = +1$ by redefining $C \mapsto iC$. Thus since in quantum mechanics we need to choose $U(1)$ for the coefficient group, we can take $C^2 = +1$ and set the symmetry group realized by our fermions to be $pin_+(2)$ without loss of generality.

Diagnosing the anomaly by gauging: Now we will diagnose the anomaly by attempting to gauge the $U(1)$ symmetry which sends ψ to $e^{i\alpha}\psi$. Adding a background field A (since A only has a time component, we could also rename it as μ), the new Lagrangian is

$$\mathcal{L} = i\psi^\dagger(\partial_t - iqA)\psi. \tag{109}$$

Charge conjugation sends $\psi \mapsto \psi^\dagger$, and $A \mapsto -A$.[9] The Lagrangian is invariant under charge conjugation because the integration by parts and fermion anticommutation give cancelling minus signs. It is also invariant under local $U(1)$ transformations by construction.

The Lagrangian is gauge invariant, but is the theory invariant under the full $O(2) = U(1) \rtimes \mathbb{Z}_2^C$ symmetry when quantized? To answer this, we look at the partition function. The Hamiltonian with $A = 0$ is zero, while the A term adds a chemical potential to the Hamiltonian so that $H = -q\psi^\dagger A_0 \psi$. The partition function is thus ethan: *what partition function are we computing? Whey do we need to sum over boundary conditions to see the anomaly?*

$$\begin{aligned} Z[A] &= \mathrm{Tr}_{\mathcal{H}} e^{-iH} = \langle 0| e^{iq\int \psi^\dagger A_0 \psi} |0\rangle + \langle 1| e^{iq\int \psi^\dagger A_0 \psi}|1\rangle \\ &= 1 + e^{iq\int A_0}. \end{aligned} \tag{110}$$

Under charge conjugation, this maps to

$$C : Z[A] \mapsto 1 + e^{-iq\int A_0} = e^{-iq\int A_0} Z[A], \tag{111}$$

which is not equal to $Z[A]$ since in general the holonomy of A will not be in $(2\pi/q)\mathbb{Z}$. Thus, we conclude that when we try to couple the theory to a background $U(1)$ field, $\mathbb{Z}_2^C$ is broken—this indicates the presence of a t Hooft anomaly.

How might we cancel this anomaly? Observe that if we added the term $-\frac{q}{2}\int A$ to the Lagrangian, then the partition function would be equal to

$$\begin{aligned} \widetilde{Z}[A] &= \mathrm{Tr}_{\mathcal{H}} \exp\left(iq\int \psi^\dagger A \psi - \frac{iq}{2}\int A\right) \\ &= (1 + e^{iq\int A_0})e^{-iq/2\int A_0}. \end{aligned} \tag{112}$$

Under charge conjugation,

$$C : \widetilde{Z}[A] = 2\cos\left(\frac{q}{2}\int A_0\right) \mapsto 2\cos\left(-\frac{q}{2}\int A_0\right) = \widetilde{Z}[A], \tag{113}$$

and so the partition function is invariant under $\mathbb{Z}_2^C$. The problem of course is that we have added an incorrectly quantized Chern-Simons term, which breaks the invariance of

[9]The usual discussion of A being a background field and hence not literally transforming under the symmetry action of course applies.

the Lagrangian under large gauge transformations which change the holonomy of A by $(2\pi/q)\mathbb{Z}$ (and is also not well-defined because of Cech-y reasons; see a later diary entry on the CS term). So this solution doesn't let us have a symmetric theory either, provided our background gauge field is a $U(1)$ gauge field and not an $\mathbb{R}$ gauge field.

Anomaly inflow then happens by simply allowing the time circle to bound a disk. This allows us to write the $-\frac{q}{2}\int A$ term in a gauge-invariant way as $-\frac{q}{2}\int_{D^2} F$. Simply letting the circle bound the disk dispenses with the large gauge transformation issue and lets the full theory, namely

$$S = \int_{\partial D^2} i\psi^\dagger(\partial_t - iqA)\psi - \frac{q}{2}\int_{D^2} F, \tag{114}$$

which is now invariant under the full $U(1) \rtimes \mathbb{Z}_2^C$ symmetry.

8 *Fermion nonconservation and the ABJ anomaly in 3+1D from Hamiltonian spectral flow* ✓

Just realized that there was still a problem in P&S that I wanted to do—it's from chapter 19. Here's the problem statement:

Examine the ABJ relation

$$\Delta N_L - \Delta N_R = -\frac{e^2}{16\pi^2}\int \epsilon^{\mu\nu\lambda\sigma}F_{\mu\nu}F_{\lambda\sigma} \tag{115}$$

in four dimensions, where $N_{L/R}$ are the number of left- and right-handed fermions $\psi_{L/R}$ and the Δ measures differences between $t=\infty$ and $t=-\infty$. Take the background field to be

$$A^\mu = (0,\ 0,\ Bx,\ A), \tag{116}$$

with B constant and A constant in space but possibly varying adiabatically in time. First, find the Hamiltonian. Then, solve the Schrodinger equation for the two fields $\psi_{L/R}(x)$. You should encounter a harmonic oscillator at some point during the calculation, just like when doing the analogous problem in two dimensions.

Consider putting the fermions in a box with sides of length L, with periodic boundary conditions. Find the degeneracy of the energy levels. Then change A adiabatically. What happens to the number of left- and right-handed fermions? Show that this checks out with the ABJ equation.

Now $\epsilon^{\mu\nu\lambda\sigma}F_{\mu\nu}F_{\lambda\sigma} = 4\epsilon^{0ijk}F_{0i}F_{jk} = 8E_iB^i$, so that the ABJ formula is

$$N_L - N_R = -\frac{e^2}{2\pi^2}\int E_iB^i. \tag{117}$$

The goal of this problem is to check this relation in a rather direct way.

First let's get the Hamiltonian. We have

$$\mathcal{H} = \pi_\psi D_0 \psi - \mathcal{L}. \tag{118}$$

Here π_ψ is the regular canonical momentum for ψ, namely $\pi_\psi = \bar{\psi} i\gamma^0 = i\psi^\dagger$. Note that the $p\partial_t q$ term is modified by replacing ∂_t with the covariant version D_0 (we are treating A^μ as a background field, not a dynamical one), which completely cancels the D_0 part of the Lagrangian.

We will work in mostly negative signature, so that the gamma matrices are $\gamma^i = iY \otimes \sigma^i$. When combined with the $\gamma^0 = X \otimes \mathbf{1}$ from $\bar{\psi}$, we get the matrix $-Z \otimes \sigma^i$. Thus in the basis $\psi = (\psi_L, \psi_R)^T$, we have

$$\mathcal{H} = -i\psi_R^\dagger \sigma^i (\partial_i - ieA_i)\psi_R + i\psi_L^\dagger \sigma^i (\partial_i - ieA_i)\psi_L. \tag{119}$$

Now the only spatial coordinate that appears in the vector potential is x, so that we have translation symmetry for y and z. Thus if we have $\int \mathcal{H} = E$ for e.g. the right-handed fermions then we are led to consider the Eigenvalue problem

$$-i\sigma^i(\partial_i - ieA_i)\psi_R = E\psi_R, \qquad \psi_R = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix} e^{ik_y y + ik_z z}. \tag{120}$$

Writing out the eigenvalue equation, we get the coupled equations

$$\begin{aligned} (E - k_z + eA)\phi_1 + (i\partial_x + ik_y - ieBx)\phi_2 = 0 \\ (E + k_z - eA)\phi_2 + (i\partial_x - ik_y + ieBx)\phi_1 = 0. \end{aligned} \tag{121}$$

So then after some algebra,

$$(\partial_x^2 + E^2 + B)\phi_1 = [(k_z - A)^2 + (k_y - Bx)^2]\phi_1, \tag{122}$$

where we are temporarily letting $e = 1$. More suggestively,

$$(-\partial_x^2 + (k_y - Bx)^2 + (k_z - A)^2 - B)\phi_1 = E^2\phi_1, \tag{123}$$

which is the Harmonic oscillator we were told we were going to find. Notice that k_y just sets the location of the center of the oscillator potential, but does not actually appear in the expression for the energy levels E (which we can get explicitly if we want, but we won't need the exact expressions).

What would have happened if we had done this with ψ_L instead? The only difference for ψ_L is that the eigenvalue equation is

$$+i\sigma^i(\partial_i - ieA_i)\psi_L = E\psi_L, \qquad \psi_L = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \end{pmatrix} e^{ik_y y + ik_z z}, \tag{124}$$

and so the only change is to replace E with $-E$, which won't affect conclusions about the degeneracy.

Now we put the Fermions in a box with side length L. The momenta are then $k_i \in (2\pi)/L\mathbb{Z}$. From the oscillator equation we saw that k_y sets the center of the oscillator potential through $x_c = k_y/B$ and doesn't affect the energy, while k_z sets the energy. So for a given energy (determined by k_z), different choices of k_y will lead to degenerate levels. However, we only have access to a certain number of k_y due to the finiteness of the box: taking the harmonic oscillator wavefunctions to be localized on a scale in the y direction much smaller than L, we require that $x_c < L$, or

$$k_y < BL \tag{125}$$

since of course we need the wavefunction to be within the box. Since each value of k_y is spaced $\delta k_y = 2\pi/L$ apart, the degeneracy of each k_z level is

$$n = \frac{BL^2}{2\pi}, \tag{126}$$

independent of k_z.

Now we consider an adiabatic change in A. Since this gives us a nonzero $E_z = \partial_t A$ and since the magnetic field is $B\hat{z}$, we have a nonzero E_iB^i during this process and expect $\Delta N_L - \Delta N_R \neq 0$ from the ABJ formula. We will consider a change in A so that the initial and final configurations are related by a gauge transformation with winding 2π around the x direction. So, we can set the initial A to be $A_i = 0$ and the finial A to be $A_f = 2\pi/L$, which is a large gauge transformation since the holonomy of the gauge field around the x direction in the final state is in $2\pi\mathbb{Z}$. This gives the minimal value we can have for $\int F \wedge F$, since in this context $\int F \wedge F$ measures the number of large gauge transformations (each a winding by 2π) that occur.

Since the initial and final states are related by a large gauge transformation, the solving of the eigenvalue problem proceeds in the same way at t_i and t_f, and we get the same harmonic oscillator spectrum. However, let's look at where the levels go during the evolution: solving for E in the harmonic oscillator Hamiltonian for ϕ_1, we see that E depends on A through $E \sim k_z - A$. Thus increasing A by $A \mapsto A + 2\pi/L$ is tantamount to *decreasing* k_z by $2\pi/L$, and so during the adiabatic evolution one level for the ψ_R fermions sinks below the Fermi level (wherever that may be). Conversely, we saw that when solving for the left-handed analogue φ_1, we just had to replace $E \mapsto -E$, and so for the left-handed fermions we have $E \sim A - k_z$, and so the change is equivalent to *increasing* k_z by $2\pi/L$, which pulls up a level's worth of ψ_L fermions out of the Fermi sea. Thus during the change in A, the vacuum looses R fermions but gains L fermions.

Since there are $BL^2/(2\pi)$ states in each level and since doing a large gauge transformation moves the ψ_R fermions down a level and the ψ_L fermions up a level, we have (now restoring e)

$$\Delta N_R - \Delta N_L = -\frac{BL^2}{2\pi} - \frac{BL^2}{2\pi} = -\frac{Be^2L^2}{\pi}. \tag{127}$$

Does this agree with the anomaly formula? Let's check: during the time when A is changing (which is the only time during which $E \neq 0$), the non-zero part of E is $E_z = \frac{2\pi}{LT}$, where $T = t_f - t_i$ is the time of the adiabatic evolution. Thus according to the ABJ formula,

$$\Delta N_R - \Delta N_L = -\frac{e^2}{2\pi^2}\int d^4x E_iB^i = -\frac{e^2}{2\pi^2}\int d^3x B\frac{2\pi}{L} = -\frac{Be^2L^2}{\pi}, \tag{128}$$

which is exactly what we got with the more direct approach. Yay!

z

9 *Orbifolding orbifolds and gauging higher form discrete symmetries* ✓

Today we are revisiting an old diary entry which explained why orbifolding is an involution. Here we look at the same sort of result from a different perspective, as well as go into rather garrulous detail about the relation between orbifolding and gauging. A good reference to get inspired for this diary entry is the introduction of [**?**], specifically the part that talks about gauging 1-form $\mathbb{Z}_2$ symmetries.

We will work throughout mostly with a theory on a two dimensional spacetime with a discrete global 0-form symmetry G, which we will take to be $\mathbb{Z}_N$ for simplicity. Nothing prevents us from going to more general Abelian G and to higher form symmetries, other than a desire to make the notation simple. First we will discuss rather pedantically how this symmetry gets gauged in the continuum. Next we will explain orbifolding and its precise relation to gauging, and finally we will talk about what happens when these operations are done twice.

Gauging

Let us couple this theory to a flat background 1-form $\mathbb{Z}_N$ gauge field A, which we assume is done through some sort of minimal-coupling prescription in an action. The only gauge-invaraint data in the gauge field comes from its cohomology class. Letting the partition function on a D-manifold X in the presence of the background field A be $Z[X; A]$, the gauged partition function is obtained by path-integrating over A, i.e. by summing over all cohomology classes in $H^1(X; \mathbb{Z}_N)$ (coboundaries are pure gauge). The gauge partiton function $Z[X]$ is[10]

$$Z[X] = \frac{1}{\dim H^0(X; \mathbb{Z}_N)} \sum_{A \in H^1(X;\mathbb{Z}_N)} Z[X; A]. \tag{129}$$

Here, the normalization out front is the volume of the group of global symmetries, viz. the $\mathbb{Z}_N$-valued functions that are constant on each path component of X. This is the "global part of the group of gauge transformations". Depending on the situation, we may or may not regard this as being gauged. Usually if we are in the setting where $X = \Sigma \times \mathbb{R}$ with $\partial\Sigma \neq 0$, these transformations are not gauged, since the boundary conditions on the fields are fixed and physical, with gauged gauge transformations required to vanish on $\partial\Sigma$ (think of CS theories and the physicality of the phase differences in Josephson junctions). We will elaborate more on this in the next subsection, but for now we will assume that X is closed (and we will continue to denote the slice on which we quantize by Σ). From here on we

[10]Since we are choosing a specific representative element for each cohomology class, this is a gauge-fixed partition function.

will assume X has only a single connected component, saving us the trouble of writing the normalization factor.

A Hamiltonian / wavefunction approach to the gauging procedure (i.e. operators and commutation relations on Σ rather than classical fields on X), which doesn't require thinking about an action as much, is as follows. First, we equip Σ with a cell decomposition (dual to a triangulation).[11] We then define the split charge operators $U_M(g) = e^{iQ_M(g)}$, where $M \subset \Sigma$ is a d-cell ($d = \dim \Sigma$) and $Q_M(g)$ is the charge operator $Q(g)$ for the element g restricted to M. $e^{iQ_M(g)}$ is defined to create an oriented domain wall on ∂M across which charged operators are acted on by R_g (in our case, a representation of $\mathbb{Z}_N$)[12]. Just for notation's sake, we might want to write this as $Q_M(g) = g \int_M \star j$ for j some 1-form conserved ($d^\dagger j = 0$) current. Here the notation is schematic; it is only meant to convey that $Q_M(g)$ is given by an integral (or sum on the lattice) of something local (this may not be particularly nice, e.g. in the Ising model it would be $g \sum_{i \in M} \ln X_i$).

Now by assumption the Hamiltonian H is symmetric and local, and so $U_M(g)^\dagger H U_M(g)$ fails to be H by something supported on ∂M. To change this, we of course add a gauge field A to H which modifies the energies of G domain walls — one can think of this as promoting the couplings of the interaction / gradient terms to dynamical fields (best thought of on the lattice), but really what's going on is that the gauge field is just shorthand notation for the transition functions of the $\mathbb{Z}_N$ bundle the fields of the theory are defined on. Anyway, H then commutes with the operators

$$\mathcal{G}_M(g) = U_M(g) T_{\partial M}(g), \tag{130}$$

where $T_{\partial M}(g)$ is an operator supported on ∂M which implements the change in the gauge field needed so that $[\mathcal{G}_M(g), H] = 0$ (this is just saying that a change in the sections plus a change in the transition functions is a trivial automorphism of the bundle). In the $\mathbb{Z}_N$ case, which we will be specializing to in the following for simplicity of notation, we have $T_{\partial M}(g) = e^{i \int_{\partial M} \star E}$, where E is the momentum of the gauge field (written as a 1-form).

To gauge the symmetry, i.e. to make the symmetry local, we need to work in a space where all of the $\mathcal{G}_M(g)$ with $M \subsetneq \Sigma$ act as the identity. We do this by cutting up Σ into a bunch of pieces (the cells in its cell decomposition) and sowing them back together with added g twists. That is, to gauge the symmetry we act with all possible split charge operators, by applying the operator

$$\Pi = \prod_{\text{d-cells } M} \sum_g \mathcal{G}_M(g) = \prod_{\text{d-cells } M} \Pi_{\mathbf{1}}(M) \tag{131}$$

on some reference vacuum state $|0\rangle$, where $\Pi_{\mathbf{1}}(M)$ is the projector which projects onto states on which $\mathcal{G}_g(M)$ acts as $\mathbf{1}$. A particular term in the resulting sum for Π looks like a "g soap bubble foam operator". Note that any operator which is gauge variant has zero expectation

[11] If we are working in the continuum, this should be thought of as a good cover of Σ. If we are working on a lattice, we should think of the cell decomposition as defining the lattice. We will largely use continuum notation in what follows, but this is only for concreteness.

[12] If our symmetry was anomalous, this step would be problematic — the charge operators would not be splittable in this way.

value in the state $\Pi|0\rangle$, because of the phase interference accumulated during the sum over all possible soap bubble films.

The holonomy of the gauge field is determined by the net group element accumulated as one passes through a closed loop in the foam. Now products of $\mathcal{G}_M(g)$ cannot change the holonomy of the gauge field since all the M's are homologically trivial (they are patches in a good cover of Σ).[13] In order to change the holonomies, we act with the operators $T_N(g)$, where $N \in H_{d-1}(\Sigma, \mathbb{Z})$.

Now, do we want to sum over all holonomies, or not? If by "gauge G" we mean "make G local", i.e. "add a gauge field so that charge operators for G acting on submanifolds of space act as $\mathbf{1}$", then there is no a priori reason why the holonomies of the gauge field need to be summed over, since the holonomies are not shifted by any of the operators that we are projecting to $\mathbf{1}$. However, the natural thing to do is to sum over them: in the case of e.g. a $U(1)$ gauge field the holonomies are automatically summed over anyway since they are all continuously connected to one another, and for the discrete case the only gauge-invariant data in the field is its holonomy, and so if we don't sum over holonomies then after choosing a gauge-fixing condition we see that we have in fact done nothing at all.

Now we have been working in the Hamiltonian picture, and seen that we need to sum over all spatial holonomies of the gauge fields. When we switch to a path integral picture, where do the temporal holonomies of the gauge field come in? They come from a projection onto the trivial representation of the global symmetry. We have been assuming that the operators $\mathcal{G}_M(g)$ for $M \subsetneq \Sigma$ act as $\mathbf{1}$ on the physical Hilbert space but also that $\partial\Sigma = 0$, and so we need $\mathcal{G}_\Sigma(g)$ to act as $\mathbf{1}$. Inserting the appropriate projector turns out to implement a sum over temporal gauge field holonomies, as explained in the next subsection. Note however that when $\partial\Sigma \neq 0$ we again need to deal with boundary condition issues, and whether or not we insert this projector and sum over temporal holonomies is a question of what boundary conditions we choose.

Why is orbifolding the same as gauging?

We will now see why orbifolds and gauging are essentially the same thing. For notational simplicity, we will focus on the concrete example of orbifolding a $\mathbb{Z}_N$ subgroup of a $U(1)$ symmetry. We will also work in two spacetime dimensions, and will take the closed spatial manifold on which we quantize to be Σ (which in general is a disjoint union of S^1s). The length of the temporal circle will be T. All of this is really only for notational reasons; the physics ideas here are general.

Anyway, to do an orbifold we insert symmetry twist operators along the cycles of spacetime, and sum over all such possible insertions. Now when we choose a spatial slice and quantize, the symmetry twist operators along the spatial cycle are clear: they are the charge operators for the symmetry, since moving charge operators through them (by time-evolving) results in the appropriate $\mathbb{Z}_N$ action on the operator according to its charge, by virtue of the commutation relations. Therefore along the spatial cycle(s), we will want to insert $e^{in\frac{2\pi}{N}\int_\Sigma \star j}$ operators, for $n \in \mathbb{Z}_N$. We've written the $\star j$ here since again we are orbifolding a subgroup

[13] Inserting these operators at a given time slice also doesn't change the holonomy of any temporal Wilson lines whose support transversely intersects M, since the time component of the gauge field has no canonical momentum.

of $U(1)$, but in general this exponential would just be replaced by the operator $U_g(C)$ implementing the group action. The appropriate operators to insert along other cycles are then determined by modular invariance / freedom to choose the slice on which we quantize to be the same expression with $\star j$ integrated over the appropriate cycle. Then the orbifolded partition function is

$$Z_O[X] = \left\langle \sum_{C \in H_1(X;\mathbb{Z}_N)} e^{i\frac{2\pi}{N}\int \star j \wedge \widehat{C}} \right\rangle \tag{132}$$

Now from this way of writing it, we can basically see why this is gauging: each operator insertion is the appropriate $\int j \wedge \star A$ term in the gauged action (where $A = \widehat{C}$), and the gauge field is integrated over by the sum over cohomology classes of $\widehat{C}$.[14]

Note that since we are only summing over a single representative element of each cohomology class, the above partition function is written in a gauge-fixed form. Indeed, form of the partition function changes if we do a gauge transformation by inserting e.g. $e^{i\frac{2\pi}{N}\int_I \star j}$ where $\partial I \neq 0$. Therefore there is no local symmetry in the above partition function, in that split charge operators don't act as the identity. To get a gauge-invariant formulation, we would want to actually quantize the gauge field so that we could introduce a canonical momentum for it, which would allow us to construct the gauge transformation generators from the split charge operators. Then a gauge-invariant partition function would be the same as the above, just with the sum over C running over all possible closed 1-dimensional submanifolds of spacetime.[15]

Now that we know why the above $Z_O[X]$ is the gauged partition function in a certain gauge-fixing choice, we will go into a little pedantic digression about why exactly we referred to it above as representing a partition function on a manifold with symmetry-twisted boundary conditions.

`<digression>`

Let's look at what the various operator insertions in $Z_O[X]$ do, from a Hamiltonian perspective. It's pretty clear what the spacelike $C \subset \Sigma$ insertions do: these operators appear in the form $\sum_{n\in\mathbb{Z}_N} e^{inQ}$, which obviously just implements the projector onto the trivial representation of the symmetry (since we have assumed $\partial\Sigma = 0$, no complications with boundary conditions here!). As we said, this can be thought of as twisting the boundary

[14] This holds true for the discrete case, but in case you don't like the notation with $\int$s and $\wedge$s, you can write it in terms of $U_g(C)$ and realize that the $U_g(C)$s are just the transition functions for a $\mathbb{Z}_N$ bundle where the patch overlaps are along the curves C — this collection of transition functions is then the same thing as a $\mathbb{Z}_N$ gauge field.

[15] For posterity's sake, something I was getting confused about: how does inserting the gauge transformation operators, given that they only act within a single spatial slice, implement a transformation which changes A by something which is exact on X, not just on Σ (of course restrictions to Σ of forms which are exact on X are exact while the converse is not true). Basically, the insertion of the gauge transformation operator $\mathcal{G}_I = e^{i\frac{2\pi}{N}\int_I(\star j - d\pi_A)}$ for $I \subset \Sigma_t$ (Σ_t a given time slice) projects the sum over A onto configurations where $A|_{\Sigma_t} = d\widehat{I}$. This happens because $\pi_A(t)$ appears in the path integral in the form $e^{i\pi_A(t)(d\widehat{I}-A(t))}$, which when integrated out sets $A(t) = d\widehat{I}$. Therefore when $\mathcal{G}_I$ is inserted into the path integral, in the case we're currently focusing on where the gauge field is flat, the sum over A gets projected onto configurations whose Poincare duals pass through the points $\partial I \subset \Sigma_t$. For the above orbifolded partition function the insertion of $\mathcal{G}_I$s sets the partition function to zero; hence it is not gauge invariant (but rather gauge-fixed).

conditions in time, since time ordering means that we can relate $\mathcal{O}(t)$ with $\mathcal{O}(t+T)$ by commuting it through the charge operator, picking up a $\mathbb{Z}_N$ action on $\mathcal{O}$.

What do the other (timelike) C insertions do? From what we know about orbifolding, these are supposed to twist the boundary conditions on the fields in some way. But how do they do this? They are built from the spatial component of the current, which commutes with the fields, and so they are just numbers — they can't act on fields as nontrivial operators to implement any kind of twisting action.

The key here is to realize that these insertions modify the momentum operator (the real physical momentum operator, i.e. the 01 component of the stress tensor). Indeed, we see that they can be incorporated into the action by adding the term

$$\delta S = \int \frac{2\pi}{N} \star j \wedge \widehat{C}_t, \tag{133}$$

where in this case $\widehat{C}$ is the 1-form Poincare dual to a 1-cycle homologous to the temporal circle. Varying this with respect to the metric (which appears in δS only implicitly via the Hodge star $\star$—explicitly, $\star j \wedge \widehat{C}_t = g^{\mu\nu} j_\mu (\widehat{C}_t)_\nu \, dt \wedge dx$), we see that this term modifies the stress tensor by the term

$$T_{\mu\nu} \supset \frac{2\pi}{N} j_{(\mu} (\widehat{C}_t)_{\nu)}. \tag{134}$$

Now for a choice of C_t which is constant in time, we can write $\widetilde{C}_t = B(x)dx$ for some function B. Therefore the physical momentum is modified by a term[16]

$$\delta P(t) = \frac{2\pi}{N} \int_\Sigma dx \, j_0(t) B(x). \tag{135}$$

Now if C_t is a thin line wrapping the temporal cycle then $B(x)$ is a bump function, and we see that the momentum receives a contribution from the charge density at that location. On the other hand, if we take C_t to be completely smeared out[17] so that $B(x)$ is constant, the contribution to the momentum is

$$\delta P = \frac{2\pi}{N} Q \, \mathrm{hol}(B) \in \frac{2\pi}{N}\mathbb{Z}, \tag{136}$$

where Q is the global charge operator. This is why we say that the operator insertions in the partition function for timelike C twist the boundary conditions. While they don't actually do this, since our fields are always single-valued on spacetime, we do see that these operator insertions modify the momentum operator to give it the value it would have if the fields were appropriately multivalued, with twisted boundary conditions.

`</digression>`

[16]Note that this change involves j_0, even though δS in this case involves only j_x! This is because we obtained P by considering the effect of adding in an infinitesimal off-diagonal component of the metric, which makes j_0 show up.

[17]So that technically $C \in H_1(X; \mathbb{R})_\mathbb{Z}$, where the subscript means that the elements have $\mathbb{Z}$ periods. The distinction between the continuum and the lattice here is really just a matter of notation, and I won't be too pedantic about talking explicitly about both cases.

Before moving on, we note another nice way of writing down the above orbifolded (aka gauge-fixed) partition function. We do this in a Hamiltonian perspective by writing

$$Z_O[X] = \left\langle \sum_{C\in H_1(X;\mathbb{Z}_N)} e^{i\frac{2\pi}{N}\int_C \star\pi_A} \right\rangle = \sum_{C\in H_1(X;\mathbb{Z}_N)} \langle T_C \rangle \,, \tag{137}$$

where the T_Cs are the 't Hooft lines (see caveat below for what this notation actually means when C is spacelike). Here the expectation value is taken as $\langle\cdot\rangle = \sum_\phi \langle \phi, A = 0|T(\cdot)e^{-iHT}|\phi, A = 0\rangle$, where ϕ represents the matter fields. That is, we fix a definite configuration for the gauge fields, but sum over all the matter fields. Additionally, the Hamiltonian implicit in the expectation value contains the $\int_\Sigma \star j \wedge A$ term. The needed sum over A is implemented by the 't Hooft operators. For C timelike this is clear — then the insertion of T_C just modifies the Hamiltonian by $e^{i(2\pi n/N)\pi_{A_x}(y)}$ for the point $y = C \cap \Sigma$; when acting on the $|A = 0\rangle$ this translates the state to something which has a gauge field with $2\pi n/N$ holonomy around Σ, and hence summing over cohomology classes of timelike C reproduces the sum over all possible spatial holonomies for A.

The spacelike T_C insertions are slightly different — indeed, since A_0 has vanishing canonical momentum, the notation doesn't even make sense! Instead, when C is spacelike, we define $T_C = e^{i\frac{2\pi}{N}\int_C \star j}$ to be the global charge operator. Why does this operator create holonomy for A_0? Recall that A_0 appears in the path integral via $e^{i\int_\Sigma A_0(\star j - d\pi_A)}$ as a way of enforcing the gauge constraint at each time step — the momentum π_A is then integrated out, giving an action in terms of A_0. Therefore with T_C inserted at time t_0, and for C the n-fold multiple of Σ, the partition function will contain the term $e^{i\int_\Sigma(\star j - d\pi_A)(A_0 + 2\pi n/N)}$, since $\int d\pi_A 2\pi n = 0$. Therefore the insertion of T_C is exactly equivalent to shifting A by $\widehat{C} = (2\pi n/N)\delta(t - t_0)dt$, which indeed changes the holonomy of A around the temporal cycle by $2\pi n/N$. Therefore the insertions of the (properly defined) T_Cs as above indeed is equivalent to the orbifolded partition function.

Orbifolding the orbifold; gauging the "dual" symmetry

Now we will do another orbifold. We will see that this is equivalent to ungauging. Indeed, suppose we want to undo the gauging procedure. Then in the gauge-fixed presentation where we only sum over cohomology classes, the gauge field A can be frozen out if we project the holomony of A around every cycle to be trivial, i.e. if we project onto states with no "magnetic flux" threading the cycles of X (killing the holonomy of A is tantamount to setting $A = 0$, since the gauge group is discrete). To this end, let

$$W_C \equiv e^{\frac{2\pi i}{N}\int_C A} \tag{138}$$

be the Wilson loop around some cycle C. Then we can project onto the trivial gauge field by inserting a sum of W_C's raised to all possible powers in the parition function to act as the appropriate projector:

$$Z_{O^2}[X] = Z[X] = \langle \sum_{C\in H_1(X;\mathbb{Z}_N)} W_C \rangle_A = \sum_{A\in H^1(X;\mathbb{Z}_N)} \sum_{C\in H_1(X)} W_C Z[X; A]. \tag{139}$$

Again note that we haven't had to write a sum over $n \in \mathbb{Z}_N$ of powers W_C^n, since this is automatically included in the sum over homology classes.

Why is this equivalent to orbifolding the orbifold? Recall that to do the first orbifold, we performed a weighted sum of partition functions, where each term was weighted by a phase determined by the charge along each cycle C of X, as measured by $\int \star j \wedge \widehat{C}$. We saw that this was equivalent to gauging the theory.

The claim is that inserting the Wilson lines as above does exactly the same thing, i.e. that the Wilson lines are the charge operators for a certain dual symmetry (viz. $W_C = \exp(\frac{2\pi}{N} i \int \star j^\vee \wedge \widehat{C})$), and that when this symmetry is gauged, we reproduce the original partition function. The duality here is nicely seen by looking at (137), which has the exact same form as (139).

We see from the formulae for the orbifold that it is basically doing a Fourier transform, while the second orbifold is basically doing the inverse Fourier transform. Thus the symmetry and its dual referred to above are essentially Fourier transforms of one another. To connect this with the notation that one usually sees in CFT books, recall that in the CFT example of a free boson with a $\mathbb{Z}_N$ orbifold on $X = T^2$, we have

$$Z \xrightarrow{orb} \frac{1}{N} \sum_{g,h \in \mathbb{Z}_N} Z_{g,h} \xrightarrow{orb} \frac{1}{N^2} \sum_{f,f' \in \mathrm{Rep}(\mathbb{Z}_N)} \sum_{g,h \in \mathbb{Z}_N} f(g) f'(h) Z_{g,h} = Z, \tag{140}$$

where $Z_{g,h}$ is the partition function on the torus with boundary conditions twisted or untwisted according to g, h. The above formulae are just a more general way of writing this.

The way to think about this dual symmetry, the fact that it always exists, the fact that its charge operators are the Wilson lines, and the reason why gauging it gets one back to the original theory are all explained in a separate diary entry on gauging higher symmetries, so we won't go into any further depth here.

10 *Cooler proof of Goldstone's theorem for p-form symmetries* ✓

Today we looking at a proof of the generalized Goldstone's theorem which was presented in [?] and is complimentary to the one in my paper. Today we are just going to look through their arguments and fill in some details.

The proof of Goldstone's theorem will start from the Ward identity, and basically just uses dimensional analysis. In what follows, we will ignore factors of i and q (charge), as well as signs (these are all irrelevant for our purposes).

Let $\mathcal{O}_C$ be an operator charged under a p-form symmetry, with $\mathrm{Supp}(\mathcal{O}_C) = C$ a p-dimensional submanifold of spacetime X (or an appropriately smeared bump-function version of a p-dimensional submanifold if we are being pedantic). We will assume that SSB occurs. If $p \geq 1$, the vacuum expectation value of the un-renormalized operator $\mathcal{O}$ is then allowed to vanish up to as fast as a "perimeter law", meaning that it may vanish as fast as $e^{-g^2 L^p/a}$, where g is a coupling constant and a is a UV cutoff. If this is the case, we will always renormalize $\mathcal{O}$ by subtracting off the UV divergence; this can be done with a simple multiplicative renormalization and ensures that the renormalized $\mathcal{O}$ has a finite, cutoff-independent vev. We will assume that such renormalization has been done in what follows.

Under an infinitesimal transformation, we let $\mathcal{O}$ transform as

$$\mathcal{O} \mapsto \mathcal{O} + \mathcal{O} \int_C \lambda, \tag{141}$$

for some p-form λ. Since we have a p-form symmetry when $d\lambda = 0$, the action must vary as

$$e^{-S} \mapsto e^{-S} \left(1 + \int \star J \wedge d\lambda\right), \tag{142}$$

where J is a $(p+1)$-form current which is classically conserved, $d^\dagger J = 0$. This conservation law ensures that the charge operator $Q(M_{D-p-1}) \sim \int_{M_{D-p-1}} \star J$ is topological. So then after integrating by parts (we assume that λ is compactly supported) the Ward identity reads

$$\langle \mathcal{O}_C \int_X \lambda \wedge \widehat{C} \rangle = \langle \mathcal{O}_C \int_X d \star J \wedge \lambda \rangle. \tag{143}$$

We then conclude that

$$\langle \mathcal{O}_C \rangle \widehat{C}(x) = \langle \mathcal{O}_C (d \star J)(x) \rangle. \tag{144}$$

Note that because we have assumed SSB, we can choose an $\mathcal{O}_C$ which is both charged under the symmetry and is such that the LHS is nonzero.

Now we pick an open $D-p$ manifold M_{D-p} which intersects C transversely at a point, and then integrate the ward identity over this manifold. We get

$$\langle \mathcal{O}_C \rangle \int_X \widehat{M}_{D-p} \wedge \widehat{C} = \int_{\partial M_{D-p}} \langle \mathcal{O}_C \star J \rangle. \tag{145}$$

Now by our choice of M_{D-p}, this simplifies to

$$\langle \mathcal{O}_C \rangle = \int_{\partial M_{D-p}} \langle \mathcal{O}_C \star J \rangle. \tag{146}$$

Note that the LHS is *independent* of the choice of M_{D-p}! In fact, it is just a constant. Thus we can make ∂M_{D-p} have support arbitrarily far away from the support of $\mathcal{O}_C$, and the RHS must remain a constant. This implies that we have the correlator

$$\langle \mathcal{O}_C \star J(r) \rangle \sim \frac{1}{r^{D-p-1}}, \tag{147}$$

where r is some typical distance away from C. For example, for $p = 0$ C is just a point, and we can take M_D to be a D-ball of radius r centered on C. For e.g. $p = 1$ and $D = 3$, we might take C to be the z axis and M_2 to be a solid disk in the xy plane centered at the origin and with radius r.

Anyway, the point is that this power law correlation function implies that we must have massless particles in the spectrum: if we had no massless particles, such a long-ranged correlation function would not be possible. Now we usually expect that for SSB the current will be realized as $J = dA$ for some p-form A[18]. The action for the Goldstones is then the usual

$$S = \frac{1}{2g^2}\int dA \wedge \star dA, \tag{148}$$

where g^{-2} is the "superfluid stiffness".

Is this compatible with the Ward identity when SSB is assumed? Let us test it for the case where $\mathcal{O}_C = \exp\left(i\int_C A\right)$. The AA correlator, going as $1/k^2$, goes as $\langle A(r)A(0)\rangle \sim \int d^Dk k^{-2}e^{ikr} \sim r^{2-D}$, so that

$$\langle e^{i\int_C A} \star J(r)\rangle \sim \partial_r\langle\int_C A(0)A(r)\rangle \sim \partial_r r^{2-D+p} \sim \frac{1}{r^{D-p-1}}, \tag{149}$$

which is indeed what the Ward identity requires.

11 *Anomalies and current OPEs* ✓

Today we try to get a better understanding of why central-extension-y terms which appear in current OPEs encode information about anomalies (in two dimensions throughout so that we can use CFT language).

Let us consider a theory with holomorphic and antiholomorphic currents, with OPEs

$$J(z)J(w) \sim \frac{k}{(z-w)^2} + \dots, \qquad \bar{J}(\bar{z})\bar{J}(\bar{w}) \sim \frac{\bar{k}}{(\bar{z}-\bar{w})^2} + \dots, \qquad J\bar{J} \sim 0, \tag{150}$$

where the ... could include further terms linear in the current, like e.g. for WZW models / current algebras and stuff (the ... does *not* stand for nonsingular stuff). We assume the classical eom for the currents are $\bar{\partial}J = \partial\bar{J} = 0$.[19] By coupling the theory to background gauge fields, we will find information about the various anomalies in terms of the parameters $k, \bar{k}$ appearing in the OPEs. Finally we will take a slightly more general approach and deduce a more general form of the current-current OPEs. By keeping track of a non-singular contact term for the $J\bar{J}$ OPE, we will demonstrate the mixed anomaly between the current and the appropriately-defined axial current.

[18]What about in electromagnetism with electric charges; $d^\dagger F = \rho$, $dF = 0$? We expect that in the Coulomb phase the $U(1)_m$ is still spontaneously broken, otherwise given the explicit breaking of $U(1)_e$, the massless photon would not generically be around at low energies. On the other hand, the $U(1)_m$ current is $\star F$, and if this were realized as $\star F = d\widetilde{A}$, then we would have $d^\dagger F = \star d^2\widetilde{A} = 0$, a contradiction. Alternatively, we could write $\star F = d\widetilde{A}$, but $\widetilde{A}$ would be a singular field, for which $d^2\widetilde{A} \neq 0$. This is just the electromagnetic dual of the statement that the vector potential is singular when magnetic monopoles are around. This isn't just a global issue since we are treating the matter as dynamical (i.e. we are not defining the electric charges by excising little bits that change the topology of spacetime). Thus, we have found a counterexample to the claim that the current is realized as $J = dA$ whenever SSB occurs. Note that in the generalized global symmetries + holography paper, the authors say that $\star F \neq d\widetilde{A}$ means that $U(1)_m$ is unbroken, but again this can't be the case since we know the 't Hooft operators have a perimeter law in the Coulomb phase.

[19]In situations where we have a non-chiral current J with $d^\dagger J = \bar{\partial}J + \partial\bar{J} = 0$ we can form linear combinations from J and $\star J$ to create such (anti-)holomorphic currents, so working with these eoms is wolog in this setting.

A good reference for background material to get inspired about this kind of stuff is Zohar Komargodski's notes on RG flows. This is standard stuff, but I just hadn't seen it written up anywhere.

First let us remember what contact terms in the current-current correlators mean. We will be fast and schematic. Consider a two-point function with a contact term

$$\langle J(x)J(y)\rangle = f(x-y) + g\delta(x-y) + \dots . \tag{151}$$

Since we get this by taking $\delta_{A(x)}\delta_{A(y)}F[A]$ for a background field which couples to J, the contact term must come from a counterterm like $\int A^2$ which has been added to the action. Similarly we could have a counterterm like $\int A\partial A$ which would give us a contact term like $\partial\delta(x-y)$, and so on; higher derivative counterterms give more singular contact terms, and counterterms higher order in A give contact terms to higher point functions of J. Often these contact terms are non-universal and can be modified at will without affecting the physics. Sometimes this is not the case though, e.g. when the contact terms are determined by an OPE in a CFT (since because of the limiting process involved in computing the OPE these terms are actually determined by correlation functions at *separated* points), or when they are required by gauge invariance (like the $A_\mu A^\mu$ contact term required by gauge invariance in scalar QED). Since the contact terms tell us about counterterms involving gauge fields, if we know the contact term structure we can learn about whether or not the theory has a gauge-invariant action when background gauge fields are added.

Actually before going into detail and discussing what happens when gauge fields are added, we can first argue from a Hamiltonian perspective why k and $\bar{k}$ determine the anomaly structure. Indeed, consider the chiral currents $J, \bar{J}$, which classically are conserved. The $k, \bar{k}$ terms in the OPE mean that the charge generators $Q = \oint \Sigma dz\, J$, $\bar{Q} = \oint_\Sigma d\bar{z}\, \bar{J}$ (here Σ is the spatial manifold — we will take it to be a unit S^1 for concreteness) do not commute with themselves (seen by doing the usual evaluation of the OPE inside the commutator), with $[Q,Q] = ik, [\bar{Q},\bar{Q}] = -i\bar{k}$ (with the commutator taken as usual in the radial quantization way). Now consider gauging e.g. the holomorphic symmetry by adding a gauge field A. The operators generating gauge transformations are then generated by the operators

$$\mathcal{G}(I) = \int_I (\partial_x \pi_A - \star J), \tag{152}$$

with π_A the electric field operator and $I \subset \Sigma$ some interval in space. These operators don't commute with themselves, but this by itself isn't a problem — we are still allowed to work in a subspace where they act as $\mathbf{1}$ provided that either their commutator is also a gauge transformation (i.e. $[\mathcal{G}(I), \mathcal{G}(I')] \propto \mathcal{G}(I'')$, a structure that comes from the degree-1 simple poles in the current-current OPEs — this is what happens for non-Abelian symmetries) or unless their commutator is a c-number that can be done away with by a multiplicative renormalization. To make the calculation simple, examine the commutator $[e^{i\alpha\mathcal{G}(I)}, e^{i\beta Q}]$ for infinitesimal α. This produces (use the same trick for computing the commutator, viz. surrounding I by $\Sigma \cup \bar{\Sigma}$)

$$[e^{i\alpha\mathcal{G}(I)}, e^{i\beta \oint_\Sigma dz\, J}] \sim i\alpha\beta k|I|. \tag{153}$$

Since the RHS isn't an $\mathcal{G}(I)$ when $k \neq 0$, we cannot work in a subspace where all the $\mathcal{G}(I)$

act trivially, and hence neither of the chiral symmetries can be gauged.[20] Note however that if $k = \bar{k}$, then the vector symmetry with charge $Q + \bar{Q}$ can be gauged, while if $k = -\bar{k}$ the axial symmetry with charge $Q - \bar{Q}$ can be gauged. If $\bar{k}$ isn't k or $-k$ though, nothing can be gauged.

Now we take a look at this from the perspective of explicitly adding in a background field. For concreteness we will be adding a gauge field for the vector current (and so we know from above that we will end up needing $k = \bar{k}$). To couple to the $U(1)_V$ gauge fields in a gauge-invariant way, we need to couple $\bar{A}$ with J and A with $\bar{J}$ (the gauge field A is not necessarily holomorphic, the notation just means that it is the z-component of the 1-form $A_\mu dx^\mu$). Now we expand the partition function in the background fields to quadratic order:

$$Z[A, \bar{A}] \approx \left\langle 1 - \int (\bar{J}A + J\bar{A}) + \frac{1}{2}\int_{z,w} (\bar{J}A + J\bar{A})(z) \cdot (\bar{J}A + J\bar{A})(w) \right\rangle. \tag{154}$$

Let us now look at the gauge (in)variance of this expression. The linear term is gauge invariant since the equation of motion (viz. current conservation) holds on J with no other operators inserted. Any anomalous variation will thus come from the quadratic term. Using the OPEs (allowed since there are no other operator insertions; the $A, \bar{A}$s of course don't count since they are not dynamical), the variation of the partition function is (higher order terms in the OPE will vanish for $\oint dz\, z^n = 2\pi i \delta_{n,-1}$ reasons)

$$\delta Z[A, \bar{A}] = \delta \frac{1}{2}\int_{z,w} \left(\frac{k}{(z-w)^2} \bar{A}(z,\bar{z})\bar{A}(w,\bar{w}) + \frac{\bar{k}}{(\bar{z}-\bar{w})^2} A(z,\bar{z})A(w,\bar{w}) \right). \tag{155}$$

Let's just look at the first term. We have, for $A \mapsto A + d\gamma$, (and not being too careful about factors of π and 2, and using the same shitty notation where A and its z component are the same so that $A \mapsto A + d\gamma$ is in complex notation $A \mapsto A + \partial\gamma, \bar{A} \mapsto \bar{A} + \bar{\partial}\gamma$)

$$\begin{aligned} \delta \frac{1}{2}\int_{z,w} \partial_w \frac{k}{z-w} \bar{A}(w,\bar{w})\bar{A}(z,\bar{z}) &= \int \partial_w \frac{k}{z-w} \bar{A}(w,\bar{w}) \partial_{\bar{z}}\gamma(z,\bar{z}) \\ &= \int_{z,w} \partial_{\bar{z}} \frac{k}{z-w} \partial_w \bar{A}(w,\bar{w})\gamma(z,\bar{z}) \\ &= \int_{z,w} \delta^2(z-w, \bar{z}-\bar{w}) k \partial_w \bar{A}(w,\bar{w})\gamma(z,\bar{z}) \\ &= k\int_z \gamma \partial \bar{A}. \end{aligned} \tag{156}$$

The second term is essentially identical and hence the variation of the partition function is

$$\delta Z[A, \bar{A}] = \int \gamma \left(k\partial\bar{A} + \bar{k}\bar{\partial}A \right). \tag{157}$$

Now in the special case that $k = \bar{k}$, this can be canceled by adding the local counter-term

$$S_{ct} = k\int_z A\bar{A}, \tag{158}$$

[20]Note to self: can we show that the divergence of the current is zero if the curvature of the background field vanishes? There should be a Hamiltonian-centric way of seeing this based on operator commutators.

or since the variation in this case is the integral of $k\gamma F$, it can also be canceled by the variation of a Chern-Simons theory in three dimensions (note that I have not been keeping track of factors of 2π and stuff so the CS term won't have the proper normalization if we use the literal expression above—the needed numerical factors will enter e.g. from the step where we replaced $\partial_{\bar z}\frac{1}{z-w}$ with the delta function). Anyway, note that if $k \neq \bar{k}$, then no such local counterterm will do, and no anomaly cancellation is possible. Note that we would also derive an un-cancellable anomaly if one of the couplings $J\bar{A}$ or $\bar{J}A$ wasn't present in the gauged partition function. This further illustrates the general phenomenon of anomalies being tied to chirality: if the currents are intrinsically chiral, or if the gauging is done in a chiral way, there is an anomaly. This fits with our experience of gauge anomalies in hep-ph scenarios coming from fermions which are chirally coupled to gauge fields.

A rather simple example of a scenario in which the $\int A\bar{A}$ term needs to be employed is in scalar electrodynamics. For example, consider a compact scalar, with just a free action. The current associated with the shift symmetry is $J = \partial\phi, \bar{J} = \bar{\partial}\phi$. Suppose we want to gauge this symmetry via minimal coupling of a gauge field $(A, \bar{A})$ to the current. The OPEs are of course $JJ \supset 1/(z-w)^2, \bar{J}\bar{J} \supset 1/(\bar{z}-\bar{w})^2$, so that $k = \bar{k}$ and the anomaly can be canceled by a term $A\bar{A} = A_\mu A^\mu$. Of course such a term is present, since we know that the full gauged action has the kinetic term $(d\phi - A)^2$, which contains the A^2 term.

Now we back up a little bit and consider the current OPEs from a bit more general point of view. In momentum space, we have[21]

$$JJ = \frac{q_+^2}{q^2}k, \qquad J\bar{J} = -K, \qquad \bar{J}\bar{J} = \frac{q_-^2}{q^2}\bar{k}, \tag{159}$$

where K is some arbitrary constant that we can choose by hand (as part of how we define the regularization procedure — it comes from $\delta_{A(z)}\delta_{\bar{A}(z)}$, which picks up the $\int A\bar{A}$ counterterm. As we said, these counterterms only affect correlation functions at coincident points, hence why they are part of our regularization procedure). Since the K term is momentum-independent, it only contributes a δ function contact term in real space and so doesn't affect any correlation functions at separated points—this is why we needen't worry about its arbitraryness. Here q_+ Fourier transforms to $\partial = (\partial_0 - i\partial_1)/2$ and q_- goes to $\bar{\partial} = (\partial_0 + i\partial_1)/2$. The $\pm$ sign indicates their chirality, so that ∂ kills right-movers (negative chirality) and $\bar{\partial}$ kills left-movers (positive chirality). In real space this works since schematically we have e.g. $q_+^2/q^2 = q_+/q_- \to \partial\frac{1}{\bar{z}} = \frac{1}{z^2}$.

Anyway, now let's look at current conservation. In our notation $\partial_\mu J^\mu$ is written as $\partial\bar{J} + \bar{\partial}J$. Using the OPEs, we find

$$\langle(\partial\bar{J} + \bar{\partial}J)\bar{J}\rangle = q_-(\bar{k} - K), \qquad \langle(\partial\bar{J} + \bar{\partial}J)J\rangle = q_+(k - K). \tag{160}$$

Thus we can preserve current conservation for the $U(1)_V$ current only if $k = \bar{k}$ (as we saw before, we get a gauge anomaly if $k \neq \bar{k}$, which is compatible with this condition on current conservation), and provided that we also choose $K = k$.

[21] Here we are focusing just on the anomalous parts. More generally we can have terms like $JJ = \frac{q_+^2}{q^2}f(q^2)$ for $f(q^2)$ some dimensionless function. This won't appear in the examples we're interested in, and in any case we can absorb $f(\infty)$ into the k's so that in the UV the form for the OPEs below suffices.

Now let us define an axial current, whose holomorphic and anti-holomorphic components we will write as $\mathcal{J}, \bar{\mathcal{J}}$. The currents are related as[22]

$$\mathcal{J} = J, \qquad \bar{\mathcal{J}} = -\bar{J}. \tag{161}$$

Thus current conservation for the axial current is

$$d^\dagger \mathcal{J} = \partial \bar{\mathcal{J}} + \bar{\partial} \mathcal{J} = -\partial \bar{J} + \bar{\partial} J = -i dJ, \tag{162}$$

which is basically the usual Hodge duality formula for relating the currents of symmetries with mixed anomalies that arise in many free field theory contexts. Thus if both currents are conserved, then the regular (vector) current is both closed and co-closed. This in turn means that if both currents are conserved, we have

$$\partial \bar{J} = \partial \bar{\mathcal{J}} = 0, \qquad \bar{\partial} J = \bar{\partial} \mathcal{J} = 0, \tag{163}$$

so that if both currents are conserved the un-barred currents really are holomorphic and the barred currents really are antiholomorphic.

Using the OPE, we can check that

$$\langle d^\dagger \mathcal{J} J \rangle = q_+(k + K), \qquad \langle d^\dagger \mathcal{J} \bar{J} \rangle = -q_-(\bar{k} + K). \tag{164}$$

Thus the axial current is only conserved if we have both $k = \bar{k}$ (as usual), and if we choose $k = -K$. Now we see the mixed anomaly between J and $\mathcal{J}$ — for $k \neq 0$ it's impossible to choose our regularization conventions (alias K) in such a way that both $d^\dagger J = 0$ and $d^\dagger \mathcal{J} = 0$. Now the anomaly means that if we gauge the symmetry generated by J (as we usually do), $\mathcal{J}$ conservation will get broken, which we confirm by setting $K = -k = -\bar{k}$ and computing

$$\langle d^\dagger \mathcal{J} (\bar{J} A + J \bar{A}) \rangle = k(q_+ \bar{A} - q_- A) = k \star F. \tag{165}$$

This tells us that $\langle d^\dagger \mathcal{J} \rangle = k \star F$, at least to lowest order. This comes from putting $d^\dagger \mathcal{J}$ in the path integral and expanding the $e^{-\int J_\mu A^\mu}$ term to first order, to create the usual bubble

[22] One sees that the word axial is appropriate by e.g. by looking at the case of Dirac fermions: there we have $J = j_0 - ij_1 = 2L^\dagger L, \bar{J} = j_0 + ij_1 = 2R^\dagger R$, for $j^\mu = \bar{\Psi} \not{D}_A \Psi$ and $\Psi = (L, R)^T$. The chiral current is $\mathcal{J}_0 = L^\dagger L - R^\dagger R, \mathcal{J}_1 = i(L^\dagger L + R^\dagger R)$ for gamma matrices equal to the Pauli matrices, and so indeed $\mathcal{J} = 2L^\dagger L, \bar{\mathcal{J}} = -2R^\dagger R$.

diagram with one photon leg. We see this in a slightly different way by writing

$$
\begin{aligned}
\partial_\mu \langle \mathcal{J}^\mu(z)\rangle_{A,\bar{A}} &= -\partial \frac{\delta Z[A,\bar{A}]}{\delta A(z)} + \bar{\partial}\frac{\delta Z[A,\bar{A}]}{\delta \bar{A}(z)} \\
&= \left(-\partial \frac{\delta}{\delta A(z)} + \bar{\partial}\frac{\delta}{\delta \bar{A}(z)}\right)\Big\langle 1 - \int_u (J\bar{A} + \bar{J}A)(u) \\
&\quad + \frac{1}{2}\int_{u,v} (J\bar{A} + \bar{J}A)(u)(J\bar{A} + \bar{J}A)(v)\Big\rangle_{0,0} \\
&= \frac{1}{2}\left(-\partial \frac{\delta}{\delta A(z)} + \bar{\partial}\frac{\delta}{\delta \bar{A}(z)}\right)\int_{u,v}\Big(\frac{k}{(u-v)^2}\bar{A}(u)\bar{A}(v) \\
&\quad + \frac{\bar{k}}{(\bar{u}-\bar{v})^2}A(u)A(v) - \delta(u-v)K(\bar{A}(u)A(v) + A(u)\bar{A}(v))\Big) \\
&= \partial_z \int_u \left(\bar{\partial}_u \frac{1}{\bar{u}-\bar{v}}\bar{k}A(u) + K\delta(z-u)\bar{A}(u)\right) \\
&\quad - \bar{\partial}_z \int_u \left(\partial_u \frac{1}{u-z}k\bar{A}(u) + K\delta(z-u)A(u)\right) \\
&= -\bar{k}\bar{\partial}A(z) + K\partial\bar{A}(z) + k\partial\bar{A}(z) - K\bar{\partial}A(z),
\end{aligned} \tag{166}
$$

where we used that $\langle d^\dagger \mathcal{J}(z)\rangle_{0,0} = 0$. If we were to do this for $d^\dagger J$ instead of $d^\dagger\mathcal{J}$ we would have gotten

$$\langle d^\dagger J(z)\rangle_{A,\bar{A}} = \bar{k}\bar{\partial}A(z) - K\partial\bar{A}(z) + k\partial\bar{A}(z) - K\bar{\partial}A(z), \tag{167}$$

meaning that as we saw before, we need $K = k = \bar{k}$ for conservation of the vector current. So making this choice, we get (remember we are not being careful with factors of 2 and stuff)

$$\langle d^\dagger \mathcal{J}(z)\rangle_{A,\bar{A}} = k(\partial\bar{A}(z) - \bar{\partial}A(z)) = k \star F(z). \tag{168}$$

Thus we have completed our jillionth derivation of the chiral anomaly.

We have only been working up to the one-loop level. However, we know that this result is one-loop exact for the standard reason: by gauge invariance the only possibility for $d^\dagger\mathcal{J}$ is $f(e^2)\star F$, where e is the gauge coupling. But $\int d^\dagger\mathcal{J} \in \mathbb{Z}$ and $\frac{1}{2\pi}\int F \in \mathbb{Z}$, and so $f(e^2)$ cannot continuously depend on e. Thus it must be independent of e, barring super pathological counterexamples. Since the answer for $d^\dagger\mathcal{J}$ is thus independent of the gauge coupling, it is one-loop exact (the gauge coupling appears where $\hbar$ appears in the action, and diagrams with l loops go as $\hbar^{-1+l}$ by Euler characteristic reasons).

12 *Subsystem symmetries and CMW-like constraints* ✓

Today is a short one. We will be looking at the paper [?] and their proof of a generalized Elitzer's theorem for subsystem symmetries (there were plenty of motivating examples for this even in pre-fracton days, e.g. Bose liquids, sliding Luttinger liquids, etc.). We will explain the logic behind the argument in the above paper that in a system with finite-ranged interactions, SSB of a discrete symmetry is impossible if the charge operators are one-dimensional, and

SSB of a continuous symmetry is impossible if they are two-dimensional. Now of course this cannot strictly be true as stated, since we know that e.g. 1-form discrete symmetries can be broken in greater than or equal to two spacetime dimensions (at $T = 0$). In the course of the argument we will see what further qualifications are needed for the statement to hold.

The argument, which is basically a clever use of dimensional reduction plus the normal CMW theorem, goes as follows. Consider a q form symmetry, by which we mean a symmetry whose charge operators act on $D - q - 1$ manifolds. In this diary entry (and in this diary entry only) we will take this to be our definition of a q form symmetry — in particular, we will not stipulate that the charge operators be topological in the spacetime, only in time (since we are interested in generic, viz. non-relativistic, systems).

Consider a theory where the fields are schematically denoted by ϕ. Furthermore let $S[\phi]$ be some local action. Consider a charge operator $Q(M)$, where M is a closed $D - q - 1$ submanifold of spacetime X. Let us break up the fields as $\eta(x) = \phi(x)$ for $x \in M$ and $\bar{\eta}(x) = \phi(x)$ for $x \in X \setminus M$. Finally let $\mathcal{O}[\eta, \bar{\eta}]$ be some (not necessarily local) operator charged under $Q(M)$. Cheeckily rewrite the vev of $\mathcal{O}$ as

$$\langle \mathcal{O}[\eta, \bar{\eta}] \rangle = \frac{1}{Z} \int \mathcal{D}\bar{\eta}\, \mathcal{D}\eta\, \mathcal{O} e^{-S[\eta,\bar{\eta}]} = \int \mathcal{D}\bar{\eta} \left(\frac{\int \mathcal{D}\eta\, \mathcal{O}\, e^{-S[\eta,\bar{\eta}]}}{\int \mathcal{D}\eta\, e^{-S[\eta,\bar{\eta}]}} \right) \frac{\int \mathcal{D}\eta\, e^{-S[\eta,\bar{\eta}]}}{Z}, \tag{169}$$

As usual, to get a vev for $\mathcal{O}$ we need to either add a symmetry-breaking field or fix boundary conditions appropriately — we will employ the latter approach.

Now we can bound the magnitude of the vev of $\mathcal{O}$ by

$$|\langle \mathcal{O}[\eta, \bar{\eta}] \rangle| \leq \int \mathcal{D}\bar{\eta} \left| \frac{\int \mathcal{D}\eta\, \mathcal{O}[\eta, \bar{\eta}_m]\, e^{-S[\eta,\bar{\eta}_m]}}{\int \mathcal{D}\eta\, e^{-S[\eta,\bar{\eta}_m]}} \right| \frac{\int \mathcal{D}\eta\, e^{-S[\eta,\bar{\eta}]}}{Z}, \tag{170}$$

where $\bar{\eta}_m$ is the value of the field configuration on $X \setminus M$ such that the absolute value of the expression in parenthesis in (169) is maximized (assuming that it's bounded). Then we see that we just have a factor of $Z/Z = 1$ in addition to the term evaluated at $\bar{\eta} = \bar{\eta}_m$, so

$$|\langle \mathcal{O}[\eta, \bar{\eta}] \rangle| \leq |\langle \mathcal{O}[\eta, \bar{\eta}_m] \rangle_{S[\eta,\bar{\eta}_m]}| \tag{171}$$

where on the RHS the expectation value is taken with respect to the η fields while working in the fixed $\bar{\eta}_m$ background. This means that we can bound the expectation value of $\mathcal{O}$ by the expectation value it takes on in the presence of a fixed field configuration for the fields living on $X \setminus M$. Note that the $\bar{\eta}$ fields are simply frozen to $\bar{\eta}_m$ rather than integrated out, and so if the effective action for the η fields is local if the original full D-dimensional action is regardless of what the gap in the spectrum looks like. Likewise, the effective action for η will have short-ranged interactions if the original D-dimensional action does as well.

Now comes the catch: since $Q(M)$ acts as a global symmetry on M, the RHS of the above equation is essentially the expectation value that an operator in a $D - q - 1$ dimensional theory has in the presence of a global 0-form symmetry generated by $Q(M)$. Thus we can apply the regular CMW theorem (assuming short-ranged interactions) for 0-form symmetries to conclude that $|\langle \mathcal{O}[\eta, \bar{\eta}] \rangle| = 0$ if M is $d \leq 2$ dimensional and it generates a continuous symmetry, or if M is $d \leq 1$ dimensional and it generates a discrete symmetry.

Now we know examples of higher symmetries that are spontaneously broken, and yet have charge operators whose dimension comes into conflict with the above result. So what gives? The point is that these counter examples all occur (to my knowledge) when the q-form symmetry in question arises from a gauge theory. This means that the decomposition $\int \mathcal{D}\phi \to \int \mathcal{D}\eta\mathcal{D}\bar{\eta}$ is impossible, since the Hilbert space does not factorize as $\mathcal{H}_M \otimes \mathcal{H}_{X\setminus M}$. This is not just an issue of M not being "smooth" in X: we could thicken it up into a D-dimensional submanifold, or we could try to smoothly interpolate between η and $\bar{\eta}$: nothing we could do would let us do the field decomposition in this way. The dimensional reduction approach that this method uses doesn't work for gauge theories, since their nonlocal-ness means that the degrees of freedom in different directions are all inter-related and can't get separated in the way they would need to be to make this argument work.[23]

13 *Yet another way to derive the chiral anomaly in two dimensions* ✓

Today is a short one: we'll be calculating the chiral anomaly / ABJ anomaly / mixed 't Hooft anomaly between vector and axial fields. We'll do this by looking at a Ward identity that gives the conservation of the axial current, which superficially (but only superficially) is a slightly different way compared to any that I've seen in books.

Since we already know the answer and have indeed derived it several times in previous diary entries, we won't worry too much about keeping numerical factors correct. The effective Euclidean action for the background vector field is

$$Z[A] = \int \mathcal{D}\psi\mathcal{D}\bar{\psi} \exp\left(-\int [\bar{\psi}i\not{D}\psi + J_\mu A^\mu]\right). \tag{172}$$

In two dimensions, we have (using the gamma matrices $\gamma^0 = X, \gamma^1 = Y, \bar{\gamma} = Z$) for a fermion $\psi = (L, R)^T$

$$J_0 = n_L + n_R, J_1 = i(n_L - n_R), \qquad \mathcal{J}_0 = n_L - n_R, \mathcal{J}_1 = i(n_L + n_R), \tag{173}$$

where $\mathcal{J}$ is the axial current. This means that in two dimensions we have

$$\mathcal{J}_\mu = -i\epsilon_{\mu\nu}J^\nu \implies \star d^\dagger \mathcal{J} = -idJ \tag{174}$$

where the i is a Euclidean signature artifact.

We can use this in a ward identity as follows: suppose we shift the gauge field by $\delta A = \star d\lambda$, where λ is a 0-form (with compact support). Then

$$\int \delta(J \wedge \star A) = \int dJ \wedge \lambda = i\int \star d^\dagger \mathcal{J} \wedge \lambda, \tag{175}$$

[23] This splittability issue isn't helped by adding charged matter unless that matter explicitly breaks the higher symmetry. For example we can try adding electric matter to pure Maxwell, but this wouldn't help us make arguments about the $U(1)_m$ 1-form symmetry since the objects charged under it, viz. the 't Hooft lines, are non-local in the electric variables.

and so evidently we have, taking λ to be infinitesimal so that we can expand to first order in λ,

$$Z[A + \star d\lambda] \approx Z[A]\left(1 - i\int \lambda \wedge \star\langle d^\dagger \mathcal{J}\rangle_A\right). \tag{176}$$

On the other hand, we can get an explicit expression for the lower orders in the expansion for $Z[A]$. Putting the det $\not{D}_A$ in the exponent in the usual way, we get the usual representation of $Z[A]$ as a sum of bubbles with A lines sticking out of them. The first order tadpole graph gives zero, while the second gives the usual polarization bubble. One can evaluate this explicitly, or use gauge invariance to write down the answer (up to the coefficient). So, to second order,

$$Z[A] \approx \exp\left[-\frac{1}{2\pi}\int F \wedge \star\left(\frac{1}{\Box}F\right)\right]. \tag{177}$$

This is the unique gauge-invariant dimension-2 thing we can build that's quadratic in A. Another way to write it uses

$$F \wedge \star(\Box^{-1}F) = A \wedge \star\frac{d^\dagger d}{\Box}A \to A_\mu(g^{\mu\nu} - q^\mu q^\nu/q^2)A_\nu, \tag{178}$$

which is the usual projector onto the transverse modes. Anyway, varying this to first order in λ, gives

$$Z[A + \star d\lambda] \approx Z[A]\left(1 - \frac{1}{\pi}\int F \wedge \star\Box^{-1}dd^\dagger \star \lambda\right). \tag{179}$$

Since it is acting on an exact form, $\Box^{-1} = (dd^\dagger)^{-1} = (d^\dagger)^{-1}d^{-1}$, and so after integrating by parts,

$$Z[A + \star d\lambda] \approx Z[A]\left(1 - \frac{1}{\pi}\int \lambda \wedge F\right). \tag{180}$$

Now we can match up the two ways of calculating the partition function to obtain

$$\langle d^\dagger \mathcal{J}\rangle_A = -\frac{i}{\pi}\star F, \tag{181}$$

which is the anomaly we wanted to show. The i is from our choice of Euclidean signature, and the $1/\pi$ (instead of $1/2\pi$) ensures that $\int \star d^\dagger \mathcal{J} \in 2\mathbb{Z}$ regardless of the A background, which is consistent with overall fermion number conservation and required since the $-1 \in U(1)_A$ is also the $-1 \in U(1)_V$, the latter of which we know is conserved in this approach.

14 *C, R, T, and fermions in three dimensions* ✓

Today's diary entry is a careful compendium of various facts about fermions and their symmetries in three spacetime dimensions. This has to a large extent been superseded by the diary entry on general fermion spacetime symmetry actions in general dimensions, but the present diary entry goes into a fair bit more detail and hence has been kept.

In this diary entry we will be in 2+1 dimensions, in $\mathbb{R}$ time. We will use the Weyl basis for the γ matrices:

$$\gamma_0 = iY, \quad \gamma_1 = X, \quad \gamma_2 = Z. \tag{182}$$

This has the advantage that all of the γ_μ's are real, which simplifies calculations with T. From the commutation relations of the γ's, we see that this choice works provided we use mostly positive signature.

Now we'll set conventions for what we mean by C and T. In QFT, T is an antiunitary operator that sends t to $-t$.[24] However, we have many options for what we mean by T, since we can compose T with any unitary transformation that commutes with the Lorentz group and still get something satisfying our definition of a T transformation. For a given situation, some of these choices for T will be symmetries, while others will not. In the following, by T, we will mean the antiunitary operator that acts on a Dirac fermion $\psi = \psi_1 + i\psi_2$ (here both ψ_1, ψ_2 are real Majorana fermions, and $\psi_i = (\psi_{i,L},\ \psi_{i,R})^T$) as

$$T : \psi(t,x) \mapsto \gamma_0\psi(-t,x), \quad \psi_1(t,x) \mapsto \gamma_0\psi_1(-t,x), \quad \psi_2(t,x) \mapsto -\gamma_0\psi_2(-t,x). \tag{183}$$

The γ_0 here switches L and R movers, which is something we want T to do. We will often write transformations like this as e.g. $T : \psi \mapsto \gamma_0\psi$, with the reversal of the time coordinate left implicit.

We could have also chosen to not put the minus sign in the transformation of ψ_2, and then we'd get a map $T : \psi \to \psi^\dagger$. This transformation will usually be denoted by CT, since we will define

$$C : \psi_i(t,x) \mapsto (-1)^{i+1}\psi_i(t,x), \tag{184}$$

which flips the sign of the imaginary part of ψ. Finally we have a reflection, which we take to act on the x coordinate as

$$R : \psi_i(t,x,y) \mapsto \gamma_1\psi_i(t,-x,y). \tag{185}$$

Note here we are being sloppy and writing x for either one spatial coordinate, or as shorthand for both spatial coordinates — context should prevent this from being unduly confusing.[25].

The final symmetry we'll be thinking about is the regular vector $U(1)$ symmetry. Written out explicitly, the $U(1)$ symmetry acts as a rotation on the Majorana fermions:

$$R_\alpha : \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \tag{186}$$

The current $J = \bar\psi\gamma_\mu\psi\, dx^\mu$ is odd under T, so that J_0 is even while J_1 is odd. For example,

$$T(\bar\psi\gamma_1\psi) = \bar\psi\gamma_0^T\gamma_0\gamma_1\gamma_0\psi = \bar\psi\gamma_0^2\gamma_0\gamma_1\psi = -\bar\psi\gamma_1\psi. \tag{187}$$

[24] By which we mean it acts on fields as $\phi(t) \mapsto \phi^*(-t)$. It doesn't actually act on t, which is just an integration variable in the action.

[25] These assignments for the actions of C, R, T are specific to our choice of signature $(-,+,+)$. Unfortunately, the representation theory of the various pin groups means that these choices do not carry over to other signatures. For example, suppose we chose the signature $(+,-,-)$, with γ matrices (X, iY, iZ). Now suppose T acted as $T = T_U K$, with K complex conjugation and T_U unitary. Then in order to preserve $\bar\psi i\partial\!\!\!/\psi$, we need to have $T_U = \gamma^2$! Note that the Hermitian term $m\bar\psi\psi$ is T-odd. Similarly, for R to preserve the kinetic term, we can again take it to act as γ^2, so that T and R only differ by complex conjugation. What a mess!

This means the charge operator $Q = \int J_0$ is even under T. Since $C(\psi) = \psi^\dagger$, J is odd under C, and so is Q. J_1 is odd under R while the other components are even, so $P(J) = J$ as a differential form. Anyway, from these definitions we see that we have the algebra

$$T^2 = (CT)^2 = \gamma_0^2 = (-1)^F, \quad C^2 = P^2 = \mathbf{1}, \quad Te^{iQ} = e^{-iQ}T, \quad Ce^{iQ} = e^{-iQ}C. \quad Re^{iQ} = e^{iQ}R. \tag{188}$$

There are three types of masses we will consider for the fermions. They are defined as

$$\begin{aligned} \bar{\psi} M_D \psi &\equiv i m_D (\bar{\psi}_1 \psi_1 + \bar{\psi}_2 \psi_2) \\ \bar{\psi} M_- \psi &\equiv i m_- (\bar{\psi}_1 \psi_1 - \bar{\psi}_2 \psi_2) \\ \bar{\psi} M_m \psi &\equiv i m_m \bar{\psi}_1 \psi_2. \end{aligned} \tag{189}$$

The second two break $U(1)$, and as we will see, are related by a $U(1)$ rotation to one another; hence they are not strictly speaking independent. We will look at each of the three masses in turn. As we will show in a second, they preserve or break the symmetries defined above in the following way:

M_D

First for the regular Dirac mass. One useful fact is that $\bar{\psi}_i \psi_i$ is even under T:

$$T(\bar{\psi}_i \psi_i) = \psi_i \gamma_0^T \gamma_0 \gamma_0 \psi_i = \bar{\psi}_i \psi_i. \tag{190}$$

Thus

$$T(\bar{\psi} M_D \psi) = -\bar{\psi} M_D \psi, \tag{191}$$

so that M_D is odd under T. Similarly, one shows that the Dirac mass is odd under R. On the other hand it preserves $U(1)$, since as a bilinear form for the vector (ψ_1, ψ_2), it is the identity. That it preserves $U(1)$ can also be checked explicitly, using the fact that $\bar{\psi}_1 \psi_2 = \bar{\psi}_2 \psi_1$ (there is no minus sign here, because of a minus sign picked up from the definition of γ_0). Finally, it also preserves C, since it is quadratic in ψ_2. Since it preserves $U(1)$ and C, there can be no anomalies involving combinations of just these two symmetries.

One perspective on why m_D breaks reflection is the following. Consider solving the Dirac equation in 2+1D: writing the Dirac spinor as $\psi = (\psi_+, \psi_-)^T$ (no more Majoranas until the next subsection), we can consider going into the $\mathbf{k} = 0$ rest frame, wherein we have

$$\partial_t \psi_+ = -m_D \psi_-, \qquad \partial_t \psi_- = m_D \psi_+. \tag{192}$$

Solutions to this are $\psi_+ = e^{\pm i m_D t}$, with the $\pm$ sign free to be chosen at will. We need to fix a convention, and will choose the + sign. This gives the solution $\psi = (\psi_+, \psi_-)^T = (e^{i m_D t}, i e^{i m_D t})^T$, which is a + eigenvector of J. What spin does this have? Spatial rotations are implemented in Spin(3) by $i[\gamma^1, \gamma^2]/4 = J/2$, and so we see that ψ has spin $1/2$. Now consider changing the sign of m_D: this is equivalent to changing our convention about which sign to choose in $e^{\pm i m_D t}$, which changes the eigenvalue of ψ under J, and means that ψ now has spin $-1/2$. Now while the choice of spin $\pm 1/2$ is a convention, after fixing a convention, the difference in spins between positive and negative m_D is not. Since a definite spin is picked out for m_D nonzero, T and R must be broken by a nonzero Dirac mass.

M_m

Now for the Majorana mass. Since $\bar{\psi}_i\psi_i$ is even under T, $\bar{\psi}_1\psi_2$ is odd. Thus

$$T(\bar{\psi}M_m\psi) = +\bar{\psi}M_m\psi. \tag{193}$$

However, since the Majoranna mass is linear in ψ_2, it is odd under C. By CRT symmetry it is thus odd under R as well (which is easily checked).

The Majorana mass also breaks $U(1)$, as is easily checked (as a bilinear form it is the matrix X, which has determinant -1 and thus can't transform in the trivial representation of $U(1)$). One also checks that under repeated applications of conjugation by the matrix representing a rotation $\pi/4$,

$$M_m \mapsto M_- \mapsto -M_m \to -M_- \to M_m. \tag{194}$$

Since M_m goes to minus itself under a $\pi/2$ rotation, M_m transforms in the charge 2 representation of $U(1)$. Thus it breaks the $U(1)$ symmetry down to the $\mathbb{Z}_2$ of $(-1)^F$ symmetry, which can never be broken since $(-1)^F$ is part of the Loretnz group ($(-1)^F$ is the generator of the center of $SU(2) = \text{Spin}(3)$).

However, saying that M_m breaks charge conjugation is a little bit hasty. As mentioned earlier, we are free to modify any of the symmetry operators by the action of a unitary operator which commutes with the Lortenz group — our usual example of such an operator will be a rotation which performs the $U(1)$ symmetry. To this end, define a new charge conjugation operator by

$$C_m \equiv Ce^{i\pi Q/2}, \quad C_m : \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \mapsto -\begin{pmatrix}\psi_2\\ \psi_1\end{pmatrix}. \tag{195}$$

With this definition we still have $C_m^2 = \mathbf{1}$, but now C_m no longer commutes with T (and so instead of $(C_mT)^2 = T^2 = (-1)^F$, we have $(C_mT)^2 = T^2e^{i\pi Q}C^2 = (-1)^{2F} = \mathbf{1}$). The point of doing this is that M_m is even under this charge conjugation, since

$$C_mM_mC_m = Ce^{i\pi/2}M_me^{-i\pi Q/2}C = -CM_mC = M_m, \tag{196}$$

since as a bilinear form between $\bar{\psi}$ and ψ, M_m is X while C is Z.

CRT means that we must also be able to define an R that is preserved by M_m, since it preserves T and a C as well (CRT just means that there exists *a* choice of C, R, and T such that their product acts as the identity on the terms in the Lagrangian—a generic choice of such symmetry operators will not always have a product which acts as the identity). In this case, since the charge operator commutes with R, we define

$$R_m \equiv Re^{-i\pi Q/2}, \tag{197}$$

which means that C_mR_mT is a symmetry of the M_m mass. The price of realizing these symmetries is that we get more complicated relations among the symmetry generators, e.g. how now neither the reflection nor the charge conjugation operators commute with T.

M_-

Finally we turn to M_m, which is related to M_m by a $\pi/4$ $U(1)$ rotation, as we just saw. (thus it also has charge 2 under the $U(1)$). It is like the reverse of M_m: it breaks T (since the $\bar{\psi}_i\psi_i$ terms are T-invariant), but not C (since it is biliear in ψ_2). Even though it is odd under T, saying that it breaks time reversal is a bit hasty. Indeed, consider the time reversal operator

$$T_- \equiv T e^{i\pi Q/2}. \tag{198}$$

It acts on the Majoranas as

$$T_- : \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \mapsto -\gamma_0 \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix}. \tag{199}$$

Now since conjugating with $e^{i\pi Q/2}$ sends M_- to $-M_-$ (as M_- transforms in the charge-2 representation of $U(1)$), and since M_D is odd under T, we see that M_- is preserved by T_-. It's also easy to check that $T_-^2 = T^2 = (-1)^F$, and that

$$T_- C = C T_- (-1)^F \implies (CT_-)^2 = \mathbf{1}. \tag{200}$$

As it stands M_- repsects a time reversal and a charge conjugation, but not a reflection. Thus by CRT we can find some new definition of R such that R is preserved. Indeed, we take

$$R_- \equiv R e^{i\pi Q/2}, \qquad R_- : \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \mapsto \gamma_1 \begin{pmatrix} -\psi_2 \\ \psi_1 \end{pmatrix}. \tag{201}$$

Then since M_- is in the charge-2 rep of $U(1)$, conjugating it by the $e^{i\pi Q/2}$ factor (which commutes with R) gives a factor of $e^{2\pi i/2} = -1$, which cancels its oddness under R. Thus M_- is even under R_-, and so is preserved by CR_-T_-.

Thus we conclude that identifying a mass term as odd or even under "time reversal" or "charge conjugation" or "reflection" is a bit subtle, since we have to specify exactly how these symmetries act. For some (legitimate) choices the symmetries may be broken, while for other (equally legitimate!) choices they may not be (nothing could be done to preserve a choice of each of C, R, and T in the case of the Dirac mass, though: since R commutes with e^{iQ} and since M_D is $U(1)$ invariant, nothing can change the fact that R is broken, and hence by CRT nothing can change the fact that at least one of C and T is broken).

15 *Free energy singularities, contact terms, and anomalies* ✓

Today we will try to motivate an anomaly in a theory with symmetry group G implies singularities in the free energy $\mathcal{F}[A]$ of the theory to background gauge fields A for G, and will then try to understand how the various ways of saturating anomalies in the IR reproduce these singularities. We will also go over the physical meaning of background field counterterms, prompted by wanting to understand some remarks by Zohar Komargodski at the 2019 Jerusalem winter school.

First let us briefly recapitulate what a ('t Hooft) anomaly is. Imo a particularly good modern introduction to this is in [**?**].

An anomaly for a symmetry group G, *in the context of this diary entry*, will mean that if the partition function of the theory in the presence of background fields A for the symmetry G is $Z[A]$, then $Z[A]$ is not just a function of the gauge equivalence class of A, viz. $Z[A] \neq Z[{}^U A]$ for some gauge transformation $U(x)$. Here A is a schematic notation which stands collectively for all the background fields involved — we will use A^a to refer to the field for a particular generator a of G, with current j^a. Specifically, if the theory is anomalous,[26] then we will have

$$Z[{}^U A] = e^{i\beta(A,U)} Z[A] \tag{202}$$

where β is a *local functional* of A and U. In this diary we will be assuming that $\beta(A,\gamma) \in \mathbb{R}$, so that the anomaly only affects the imaginary part of the free energy. This is true for every $\mathbb{R}$-time anomalous theory that I know about (provably true for chiral anomalies and makes sense, since we want the vev of the current, which comes from differentiating $\mathcal{F}[A]$ with an extra factor of i, to be real), but when one goes to $\mathbb{R}$ time (which we will be trying to avoid) one can get anomalies in the $\mathbb{R}$ part of $\mathcal{F}$, e.g. the trace anomaly / Weyl anomaly (although c.f. the diary entry on the meaning of the phrase "Weyl anomaly").

For the purposes of this diary entry, we will think of the global parts of the gauge transformations (for which $U(x)$ is a constant[27]) as being genuinely gauged (e.g. by working on a compact space), so that constant U is allowed in the above formula. In particular, this means that we do not have to turn on nontrivial backgrounds for all the factors in the symmetry group, e.g. if $G = H \times K$ we can turn on background fields for H only and then act with a generator of K — if a nontrivial β appeared here, it would signal a mixed anomaly between H and K. In this case, the anomaly comes from the fact that the background fields for H break the K symmetry, which can be seen without turning on a background for K (this approach is useful if K is some symmetry like T for which turning on backgrounds is annoying).

Since the gauge non-invariance occurs in the phase of the partition function, it is helpful to write things in terms of the free energy, in order to isolate this. In what follows we will specify to the case of Abelian G for simplicity (this is just for notational reasons, not physical ones — also, for discrete symmetries mentally replace $d, \wedge \mapsto \delta, \cup$, etc.), and write the gauge transformations as $A \mapsto A + d\gamma$. Let us write the free energy in terms of its real and imaginary parts as $F = \mathsf{F}[A] + i\mathcal{F}[A]$. Then we have[28]

$$\mathcal{F}[A + d\gamma] - \mathcal{F}[A] = -\beta(A,\gamma). \tag{203}$$

We can be slightly more specific about what β is. First, it has to vanish when A is pure gauge. This is because when A is pure gauge it can be canceled by a change of variables

[26] Again, under the local definition of "anomaly" — more general responses to changes in the background fields are possible (e.g. they could induce operator-valued changes), but we will not discuss them here.

[27] We will also specifying to 0-form symmetries when the notation forces us to commit to a particular co-dimension for the charge generators.

[28] We're working in $\mathbb{R}$ time, but also in conventions where $F = -\ln Z$, without the i. In retrospect this may not have been the most aesthetically pleasing choice; oh well.

in the path integral (at least perhaps up to local contact terms; these will be irrelevant for determining the anomaly anyway), and hence $Z[A = d\alpha]$ cannot have any α dependence. Since the expectation value of the current (i.e. $\delta S/\delta A$; not necessarily a continuous object) in the background field is $\langle j\rangle = i\delta F[A]/\delta A$, the functional β controls the expectation value of the divergence of the current as (taking the variation to be $\delta A = d\gamma$)

$$\langle d^\dagger j^a(x)\rangle = d^\dagger i\frac{\delta}{\delta d\gamma}(F[A+d\gamma] - F[A])|_{\gamma=0} = \frac{\delta}{\delta\gamma^a(x)}\beta(A,\gamma)|_{\gamma=0}. \tag{204}$$

However, this being non-vanishing does not always mean there is anomaly, and indeed a nonzero β does not necessarily imply an anomaly — it may just be because we haven't defined the current properly. In particular, we might be able to re-define our current by some terms only involving the background field such that the RHS of the above, and hence the anomaly, vanishes. To explain this, we need a short digression on the role of contact terms in QFT.

<digression>

Normally in QFT one is free to add local counterterms in the background gauge field A at will to the action (if we don't plan on making the background fields dynamical, the counterterms don't even have to gauge invariant). This is because when they are non-dynamical such background fields are merely devices for computing correlation functions of the dynamical fields they couple to, and such local functionals of A don't affect the results of computing any of the physical correlation functions that we care about. Adding a local term $\alpha\int f(A)$ for some local functional f leads to a theory with α dependence which only appears when evaluating correlation functions of currents at coincident points. Since we normally don't care about such correlation functions (they are some non-universal UV stuff that we aren't trying to capture with the free energy), we can provisionally regard local functionals of the background fields as physically unimportant. This means that if we take the free energy $\mathcal{F}[A]$ for the field A written in momentum space and expand in the field momentum q, *only* the terms that go as negative integer or non-integer powers of q will be regarded as physical; all the positive-integer powers are beyond the purview of QFT. QFT is for the most part the study of things that are regulator-independent (well, expect that anomalies have to do with regulators, which is kind of the point), and almost by definition does not capture purely UV things.

Saying that all counterterms built from background gauge fields are unphysical is maybe going too far, though. After all, the CS term $A\wedge dA$ determines the Hall conductivity (here A is a *background* EM field, not the dynamical field which is integrated over in the field theory description of the Hall effect), which is physical and well-defined (that the CS term only affects correlation functions at coincident points can also be understood from the fact that CS theory has no radiation: the equations of motions have no derivatives, and the field strengths can be solved for as a local function of the sources). But how can the Hall conductivity be a universal thing to compute on the field theory level, if the Hall response is determined by a contact term in the background gauge fields? From a QFT perspective, we would say that only the Hall conductance mod 1 is universal information about the response of a given QFT. One way to argue on the contrary though is that counterterms that we add to change our regularization prescription should be able to be added locally, that is, we should be able to add them with a spatially varying coefficient. Of course, adding the term $\int \alpha(x) A \wedge dA$ is not allowed because it breaks gauge invariance unless α is a constant integer multiple of $1/4\pi$, and so the CS term is not a trivial change in regularization scheme in the same sense that e.g. $\int F \wedge \star F$ would be (which gives a very singular $\Box\delta(x)$ modification to the current-current correlation function).

</digression>

With these comments in mind, we see that since local parts of the free energy density are non-universal we are free to modify $\mathcal{F}[A]$ by any local functional $-i \int f(A)$, which may or may not be gauge-invariant (with the exception of those whose coefficients are quantized, as mentioned above — this issue won't actually come up until the last example, though). In particular if it is the latter, then we might be able to choose f such that the anomalous phase β is canceled, since now

$$i\delta\mathcal{F} = i(\mathcal{F}[A + d\gamma] - \mathcal{F}[A]) = \beta(A, \gamma) - [f(A + d\gamma) - f(A)], \tag{205}$$

so that the variation of the free energy is $i\delta\mathcal{F} = \beta - \delta f$. Therefore if $\beta(A, \gamma)$ is a total variation, the free-energy can be made gauge-invariant by adjusting it with a local counterterm — doing so also re-defines the current (via $j = \delta_A S$), which gives us a new current which is conserved.

Anyway, from the above formula we see that we thus have a sort of cohomology problem: a nontrivial anomaly comes when β cannot be expressed as the gauge variation of a local functional of A, even though it is itself is a local functional. A good discussion of this is in [?].

We now attempt to give a more geometric picture to this cohomology problem, something that I first saw the idea of in [?]. Mathematically, if $\mathcal{A}$ is the moduli space of all possible gauge field configurations and if $\mathcal{G}$ is the group of gauge transformations,[29] then $Z[A]$ is a section of a line bundle over $\mathcal{A}/\mathcal{G}$, while $\mathcal{F}$ is a section of a $U(1)$ bundle over $\mathcal{A}/\mathcal{G}$.[30] $Z[A]$ is always a global section, but the bundle over $\mathcal{A}/\mathcal{G}$ may still be nontrivial, and its non-triviality will indicate an anomaly. Indeed, if there were no anomaly then we could add an appropriate counterterm $\delta^{-1}\beta$ to make $Z[A]$ a genuine function on $\mathcal{A}/\mathcal{G}$ and hence the bundle for the partition function (free energy) would be the trivial product $\mathcal{A}/\mathcal{G} \times U(1)$ ($\mathcal{A}/\mathcal{G} \times U(1)$). Taking the contrapositive means that a nontrivial bundle $\implies$ an anomaly. We will assume the converse is true in what follows; it is true at least for all the examples we will discuss.[31]

There are multiple ways in which the the bundle E of which $\mathcal{F}[A]$ is a section could be nontrivial. First, it may be nontrivial because it has a connection with curvature which

[29] $\mathcal{G}$ includes all gauge transformations, viz. large ones as well, which are gauge transformations by functions $U(x)$ which cannot be continuously deformed to $\mathbf{1}$. An example of where we need to take into account such transformations is Witten's $SU(2)$ anomaly, where gauge transformations by $U(x)$ for $U(x)$ homotopic to the generator of $\pi_4(SU(2)) = \mathbb{Z}_2$ leave the partition function invariant if the theory involves an odd number of chiral fermions coupled to A.

[30] Thinking like this is probably useful only when the anomaly can be probed by turning on continuous background fields.

[31] As in [?], the catch is that the bundle being trivial just means that one can find a bundle automorphism that takes the bundle to the product bundle; the fact that this bundle automorphism may be non-local but that we are only working modulo local counterterms is the source of the difficulties.

gives invariants like Chern numbers that are nonzero. Secondly, it may be flat but with a connection that has nontrivial holonomy, so that the transition functions can not all be trivialized. We will see examples of both in a second.

We now want to elaborate on why anomalies imply singularities in the (imaginary part of) the free energy $\mathcal{F}$. The basic reason is from the issue of bundle topology above: $\mathcal{F}[A]$ is not a global section of E, and so when we write it in a particular coordinate system, it has branch cuts, which manifest themselves as singularities (this doesn't mean that correlation functions of the currents are necessarily singular, since the variational derivative is always taken within a single coordinate patch on $\mathcal{A}/\mathcal{G}$). While these branch cuts are most clearly illustrated in the case with both discrete and continuous symmetries (parity anomaly; towards the end), we will first look at what this means for just $U(1)$ symmetries.

First, we will see that any theory with an anomaly involving products of $U(1)$s will necessarily be massless. This intuitively obvious fact (how could you saturate an anomaly coming from a diagram that can be derived perturbatively from a gapped theory which in the IR has only finitely many dof?) can be proven by looking at the formula for the divergence of the current. Indeed, write the vev of $d^\dagger j^a$ as $\langle d^\dagger j^a \rangle = P(F_A)$, where $P(F_A) = \delta_{\gamma^a}\beta(A,\gamma)|_{\gamma=0}$ is some functional of the field strength.[32] For example, consider a chiral anomaly-like case in D dimensions, where

$$\langle d^\dagger j^a \rangle = \alpha \varepsilon^a{}_b \star (dA^b)^{\wedge(D/2)}, \tag{206}$$

with α a constant. Then, taking $D/2$ functional derivatives wrt A, we have

$$\varepsilon^{ab}\delta\left(\sum_{i=0}^{D/2} q_i\right) q_0^{\mu_0}\left(\prod_{i=1}^{D/2}\frac{\delta}{\delta A^b_{\mu_i}}\right)\langle j_a^{\mu_0}(q)\rangle \sim \varepsilon^{\mu_1\dots\mu_{D/2}}{}_{\nu_1\dots\nu_{D/2}}\prod_{i=1}^{D/2} q_i^{\nu_i}. \tag{207}$$

Since the functional derivatives acting on the current produce more currents inside the expectation value, this becomes

$$\varepsilon^{ab}\delta\left(\sum_{i=0}^{D/2} q_i\right)\left\langle j_a^{\mu_0}(q)\prod_{i=1}^{D/2} j_b^{\mu_i}(q_i)\right\rangle \sim \varepsilon^{\mu_1\dots\mu_{D/2}}{}_{\nu_1\dots\nu_{D/2}}\prod_{i=1}^{D/2} q_i^{\nu_i}(q^{\mu_0})^{-1} \tag{208}$$

Therefore the anomaly tells us that this $D/2+1$ point correlation function of the currents is long-ranged in $\mathbb{R}$ space — hence the theory cannot be massive, no matter how the anomaly gets saturated in the IR.

Vis-a-vis singularities in $\mathcal{F}[A]$: as we said above, the anomaly here is saying not just that the free energy in not invariant under local gauge transformations, but that a nontrivial background field actually breaks the global symmetry. Indeed, we can explicitly show by e.g. Fujikawa that the partition function changes under the action of a *global* symmetry (e.g. the fermion measure changes even when we perform a chiral rotation that is everywhere constant in spacetime). Getting a nonzero result for $\delta A = d\lambda$ in the limit where λ is a constant means there must be some sort of singularity coming from the inverse of a differential operator which allows the answer for $\lambda = const$ to be nonzero.

[32] It can only depend on F_A in these examples since it must be local and must vanish when A is locally exact as we said above ethan: *come back and talk about $A_1 \wedge F_2$ and stuff* .

Let's look at one example in a little more detail, which of course is something we've seen a jillion times already, viz. the mixed anomaly between $U(1)^A$ and $U(1)^V$ for fermions in even dimensions. We know that e.g. in 1+1D, a $U(1)^A$ gauge transformation $\delta A^A = d\lambda$ produces, in the usual regularization scheme, a term

$$\delta\mathcal{F}[A^A, A^V] = \frac{i}{\pi}\int \lambda F^V = \delta_\lambda \frac{i}{2\pi}\int (d^{-1}A^A)F^V, \tag{209}$$

which is the variation of a term that is gauge invariant under $U(1)^V$, but which is singular at zero momentum (here by $d^{-1}A^A$ we just mean A^A_μ/q_μ). This means that in this regularization scheme,

$$\mathcal{F}[A^A, A^V] = \frac{i}{2\pi}\int A^A_\mu A^V_\nu \varepsilon^{\nu\lambda} q_\lambda/q_\mu + \cdots = \frac{i}{2\pi}\int (A^A_0 A^V_0 (q_1/q_0) - A^A_1 A^V_1 (q_0/q_1) + A^A_1 A^V_0 - A^A_0 A^V_1) + \ldots, \tag{210}$$

where the dots are gauge-invariant (and in general non-local, i.e. involving field strengths and $\Box^{-1}$s) or built from local counterterms. We can always add $A^V \wedge A^A$ to get rid of the last two terms in parenthesis, but the claim is that the singular parts (which are responsible for the gauge non-invariance) will always be there, in any regularization scheme.

We can show this for example by looking at what happens to the integration measure under gauge transformations of both gauge fields. For $A^a \mapsto A^a + d\lambda^a$, standard manipulations with the Jacobian show that free energy changes as

$$\delta\mathcal{F}[A^A, A^V] \propto i \lim_{M^2\to\infty} \mathrm{Tr}\left[e^{(\not{D}_{A^V+\bar{\gamma}A^A})^2/M^2}(\lambda^A + \bar{\gamma}\lambda^V)\right], \tag{211}$$

where we have chosen a heat-kernel regulator that is invariant under both gauge transformations (if we don't do this then we definitely won't be able to make the gauge fields dynamical). Now we write the square of the Dirac operator as (see another diary entry for the proof)

$$(\not{D}_{A^V+\bar{\gamma}A^A})^2 = D^2_{A^V+\bar{\gamma}A^A} + \frac{\gamma^\mu\gamma^\nu}{2}(F^V_{\mu\nu} + \varepsilon_\mu{}^\lambda F^A_{\lambda\nu}). \tag{212}$$

When acting on the gauge-invariant Hilbert space the square of D just produces $-k^2$, where k is the momentum. When we do the integral over all k, we get a factor of M^2, which means in the limit that only the first term in the expansion of the exponential of the field strength survives. Then taking the trace over the spin indices, we find

$$\delta\mathcal{F}[A^A, A^V] = i\frac{1}{\pi}\int (\lambda^A F^V + \lambda^V F^A). \tag{213}$$

This means that the general expression for the free energy is

$$\mathcal{F}[A^A, A^V] = \frac{i}{2\pi}\int (d^{-1}A^A F^V + d^{-1}A^V F^A + C A^A \wedge A^V) + \ldots, \tag{214}$$

where the $\ldots$ are gauge invariant and C is some arbitrary coefficient.[33] By looking at this one sees that taking $C = \pm 1$ will render the theory gauge invariant under one of the two

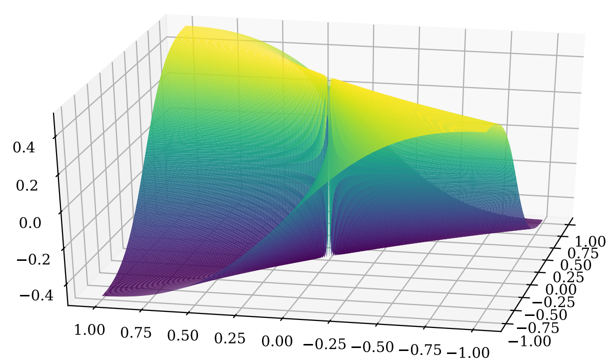
[33] An $A^A \wedge \star A^V$ counterterm obviously doesn't help us.

symmetries, but getting something that is invariant under both is impossible (if the relative sign between the F^A and F^V terms had turned out differently this would not be the case, and we would have had something non-anomalous).

How are the singularities reproduced when SSB occurs? In the way one would expect — they are produced by integrating out the Goldstones. However since the Goldstones couple differently to the gauge field than the UV fields do, we'll briefly discuss how this works explicitly. To illustrate this example we again look at an anomalous $U(1)$ symmetry (for notation's sake). Here a typical IR action includes schematically $(d\phi - A)^2$, with ϕ the goldstone. Now we know that $\mathcal{F}[A] = 0$ when A is pure gauge, while at the same time we know that for A a constant, the free energy must be $\int A \wedge \star A$, since the $d\phi \wedge \star A$ terms die after IBP (the nonzero free energy coming from the "velocity boost" of the SSBken state). These two facts are only compatible if there are some ∂^{-1}s in the free energy. Indeed, one can quickly check that the relevant part of the free energy is

$$\mathcal{F}[A] \sim \int F_A \wedge \star \Box^{-1} F_A = \int A_\mu (g^{\mu\nu} - q^\mu q^\nu / q^2) A_\nu. \tag{215}$$

This is singular at zero momentum because of the second term, which cares about the way that we approach zero momentum; this singularity allows $\mathcal{F}[A]$ to go to zero for pure gauge configurations but still be nonzero when the gauge fields are constant. Just for fun, here's a plot of what the second term looks like (both the diagonal and off-diagonal parts are singular in the same way):



(216)

One example where we have continuous gauge fields but an anomaly for a discrete symmetry is the mixed anomaly between T and $U(1)$ for Dirac fermions in odd spacetime dimensions. For example, in $D = 2 + 1$, we show in another diary on the parity anomaly that the partition function is

$$Z[A] = |Z[A]| \exp\left(\frac{ik}{4\pi} \int A \wedge dA + \frac{\pi i}{2} \eta(A)\right) \tag{217}$$

with the definition

$$\eta(A) = \left(\frac{1}{4\pi^2} \int A \wedge dA + 1\right)_2 - 1, \tag{218}$$

with the subscript meaning mod 2. Here k is a IQH response counterterm, which we will regard as physically meaningful in keeping with the comments above. From this we see that $\mathcal{F}[A]$ is a nontrivial $U(1)$ bundle over $\mathcal{A}/\mathcal{G}$. For example, when we're on the usual $S^2 \times S^1$ with $2\pi n$ flux through the S^2, we have (setting $k=0$ for now)

$$\mathcal{F}[A] = \frac{n\pi i}{2}\left(\left(\frac{h}{2\pi}+1\right)_2 + 1\right), \tag{219}$$

where h is the holonomy of A around the S^1. The important part here is that as A winds around a closed (non-contractible) loop in the base space $\mathcal{A}/\mathcal{G}$, viz as h goes from 0 to 2π, we have $\mathcal{F}[A]$ go from 0 to $n\pi$, and then when $h = 2\pi + \varepsilon$, we suddenly have a jump to $\mathcal{F}[A] = -n\pi$. Now when $n \in 2\mathbb{Z}$, we see that there exists a choice of k such that $\mathcal{F}[A]$ is made to vanish. In a more hep-th way of thinking, where we regard the IQH responses as trivial counterterms which can be added at will to the free energy, this means that when $n \in 2\mathbb{Z}$ there is no anomaly. Our perspective will be slightly different, viz that when $n \in 2\mathbb{Z}$ we have a collection of theories parameterized by k, with one of them being anomaly free.[34] Regardless, when $n \in 2\mathbb{Z}+1$, the holonomy of $\mathcal{F}[A]$ as we move around the loop in $\mathcal{A}/\mathcal{G}$ is only in $\pi(2\mathbb{Z}+1)$, and can't be canceled by any local counterterm — hence, the "parity" anomaly, with $\mathcal{F}[A]$ giving a flat bundle over $\mathcal{A}/\mathcal{G}$ which is rendered non-trivial by the non-existence of a global section coming from the nontrivial transition function at the point where the sign of $\mathcal{F}$ flips (which, as it has to be, is also the point at which $Z[A]=0$). This transition function becomes a branch cut singularity when we work in a single coordinate patch on $\mathfrak{a}/\mathcal{G}$. Thus finite sums can still produce singularities in the free energy (and in this case as well as the last, it is the zero modes which are responsible for them).

16 *The parity anomaly revisited* ✓

Today's diary entry is a look at a way of understanding the parity anomaly, focused on looking rather explicitly at the zero modes realized in certain gauge field backgrounds. This diary entry was prompted by wanting to understand in detail some of the content presented by by Seiberg during his 2019 Jerusalem lectures.

The starting point is of course a Dirac fermion in 2+1D coupled to a background $U(1)$ field A. Beyond the usual minimal coupling, we can add any local gauge-invariant counterterms that are functions only of A, since such terms do not affect any fermion correlation functions beyond modifying contact terms. In 2+1D this means in particular that we can include a (properly quantized) CS counterterm in the background fields. So the action under consideration is (in $\mathbb{R}$ time)

$$S = \int \left[\bar{\psi} i \not{D}_A \psi + \frac{k}{4\pi} A \wedge dA\right], \tag{220}$$

where the CS term is to be thought of as a counterterm that we're allowed to modify the free energy by. We can get the anomaly by specifying to a certain gauge field configuration.

[34] In no perspective are we allowed to mod out by $\frac{1}{4\pi k}\int A \wedge dA$ FQH responses.

We will take space to be an S^2 with $\int_{S^2} F_A = 2\pi$, and will suppose that the spatial gauge fields are time-independent. The CS term then integrates to

$$\frac{k}{2\pi}\int A_0 F_{xy} = k\int dt A_0. \tag{221}$$

Now the monopole background means the Hamiltonian for the fermions has a zero mode (see a previous diary entry; a monopole background of flux $2\pi n$ supports n zero modes). Since the Hamiltonian for the fermions is the spatial part of $\bar\psi i \not{D}_A \psi$, the action for this zero mode is

$$S_0 = \int dt\ \left(\bar\psi(i\partial_t + A_0)\psi + kA_0\right). \tag{222}$$

Since the zero mode on the monopole is a two-level system, this is exactly equivalent to the problem of a single free fermion in quantum mechanics, coupled to a background $U(1)$ field.

We have already analyzed this in an earlier diary entry, where we saw that it had a mixed anomaly between $U(1)$ and C (or T). There are two degenerate states, viz. $|k\rangle, |k+1\rangle$. Time reversal doesn't exchange them, while we define C to act as

$$C : |l\rangle \mapsto |-l+2k+1\rangle, \tag{223}$$

so that C interchanges $|k\rangle$ and $|k+1\rangle$.

Now we want to compute the partition function, as a matrix element of e^{-iHt} between initial and final states. To see an anomaly in C or T we will want to take symmetric boundary conditions. We will do the simplest thing, viz. summing over $|k\rangle$, $|k+1\rangle$ by working with periodic boundary conditions in time (we will avoid going to $i\mathbb{R}$ time though, since we want to think about T). We can then easily compute the partition function for this system in the deep IR, since we just have to sum over the two degenerate ground states.[35] The Hamiltonian is just $QA_0 = (\bar\psi\psi + k)A_0$, and so

$$Z[A] = e^{ik\oint A_0}(1 + e^{i\oint A_0}). \tag{224}$$

This is not T or C invariant unless k is the (disallowed) value of $-1/2$, or if A is a special T-preserving configuration, like $A = 0$ or $A_0 = \pi/|S^1_t|$.[36]

Just to be pedantic: even though A is currently a classical background field, we are considering a transformation where A is in fact acted on by T, C. Technically this is a bit unfair, since A is just a real c-number function (proportional to $\mathbf{1}$ when acting on the Hilbert space), and hence cannot really be acted on by any symmetry transformation. A more mathematically precise way of doing things would be to say that we have a symmetry if the partition function $Z[A]$ is covariant, namely that after applying the symmetry action

[35] Here "deep IR" means at energy scales below the gap between the zero modes and the rest of the spectrum, which is set by the curvature of the spatial S^2. Going to the deep IR like this is just done to simplify the formula for $Z[A]$ — including the contributions to $Z[A]$ from all the higher energy modes will just modify the real part of $Z[A]$ in a way that doesn't affect our conclusions about things related to anomalies (and indeed, the anomaly is an RG invariant, so ignoring everything going on above the degenerate ground states can be done wolog if we just want to know about the anomaly).

[36] In order for there to be no anomaly we must be able to maintain the symmetries while summing over all A, and so the fact that $Z[A]$ is symmetric for some values of A is irrelevant.

for a group element g we get $Z[A] \mapsto Z[{}^g A]$ for some appropriate g-action. However this is equivalent to saying that $Z[A]$ is invariant under the symmetry as well as a combined action by g^{-1} on A, and since this is what we'd be doing anyway if A was made dynamical, we will be continuing to use this perspective. For example, T acts on the partition function as $\langle \phi_f | e^{-iHt} | \phi_i \rangle \mapsto \langle T\phi_f | e^{-iHt} | T\phi_i \rangle$, in addition to acting on A in the appropriate way. When we trotterize this to get the path integral, the symmetry action has the effect of replacing all states in the resolutions of the $\mathbf{1}$ by their T-reversed images: thus the action of T can be taken into account by replacing $\phi \mapsto T^\dagger \phi T$ in the action, *without* conjugating the i in e^{iS}. Now the appropriate action on the background field to maintain T symmetry at the classical level is $T : A \mapsto -A$ as forms (after an implicit change of variables). We see from the above expression that T is then broken[37] (the same arguments apply for $C : A \mapsto -A$), and this breaking of T and C is of course the mixed anomaly between $U(1)$ and T and C, which for historical reasons we call the parity anomaly.[38]

Now we write the partition function as

$$Z[A] = e^{ik \oint A_0} e^{i\pi\eta(A)/2} |Z[A]|, \qquad |Z[A]| = e^{i\pi\eta(A)/2} + e^{-i\pi\eta(A)/2}, \tag{226}$$

where we have defined

$$\eta(A) \equiv \left(\frac{1}{\pi} \oint A_0 + 1 \right)_2 - 1, \tag{227}$$

where the subscript means mod 2. This expression is gauge invariant and well-defined under changing transition functions on patches of the S^1 by things in $2\pi\mathbb{Z}$ (because of the mod 2), and so $Z[A]$ is manifestly gauge invariant. This way of defining the eta invariant means that $\eta(A)$ takes values in $[-1, 1)$, so that $\cos(\pi\eta(A)/2) = |Z[A]|$ is always positive.

Note that $\eta(A)$ has a discontinuity of 2 when the holonomy of A_0 crosses π. This means that the function $e^{i\pi\eta(A)/2}$ has a discontinuity at $\oint A_0 = \pi$, where it goes from i to $-i$. Note that when we square $e^{i\pi\eta(A)/2}$ we get the level-1 CS term, which is always well-defined — therefore the gauge invariant $e^{i\pi\eta(A)/2}$ is the correct way of writing the more schematic $e^{i\frac{1}{8\pi}\int A}$. Anyway, the point of writing things this way is that the (imaginary part of the) free energy now has a singular contribution as a function of the background fields, and as discussed in an earlier diary entry, such a singular behavior is a requirement for the existence of an anomaly.[39] Despite this singularity, the partition function $Z[A]$ is still continuous. This

[37]This might actually make $Z[A]$ look T invariant as written since $A_0(t) \mapsto A_0(-t)$ and since $\oint dt$ is even under the change of variables, but remember that this form of the action came from using $\int F_A = 2\pi$. The time-reversed version of this is $T(\int F_A) = -2\pi$, which means that an extra minus sign appears in the T transformation of $i \oint A_0$.

[38]We've been looking at a path integral perspective, but this can also be seen in Hamiltonian language by noting that the algebra is actually (here R_α is the $U(1)$ rotation by α)

$$TR_\alpha = R_{-\alpha} T e^{-i(2k+1)}, \tag{225}$$

which is only the expected algebra if $k = -1/2$ (same applies for C).

[39]Note how this singularity in the imaginary part of the free energy comes from the phase of the partition function and that in particular, it comes from a discrete sum over two degenerate states. Therefore we do not have to integrate out a massless degree of freedom coupled to A in order to produce a singularity in the free energy $F[A]$, just summing over a finite number of degenerate states is sufficient. This is how singular free energies can arise from summations over degenerate ground states even in the context of thinking about

is because the singularity in the free energy happens exactly when $|Z| = 0$ (so the singularity is just in the derivative of $Z[A]$ wrt A, i.e. in the current correlation functions), due to a zero mode of the Dirac operator which occurs at the point where $\oint A_0 = \pi$.

That $|Z[A]|$ vanishes at this point is obvious from the above expression we wrote for it, but we can also see it by looking at the fermions. This goes as follows. On the S^1 of time, we work with the usual $\psi(t) = -\psi(t+\beta)$ boundary condition. Then the frequency is modded in $\frac{2\pi}{\beta}(l-1/2), l \in \mathbb{Z}$, and we can decompose

$$\psi(t) = \sum_{l\in\mathbb{Z}} e^{\frac{2\pi i}{\beta}(l-1/2)t}\psi_k, \tag{228}$$

and so the condition for ψ_k to be a zero mode of $i\not{D}_A$ is that $-2\pi(l-1/2) = kA_0$. This happens if $k\int A_0 = \pi$, since then we can fix a gauge in which $A_0 = \pi/k\beta$. Basically, the π flux around the S^1 that is threaded in when $\oint A_0 = \pi$ "cancels" the AP boundary conditions for fermions, giving them effectively periodic boundary conditions and hence a zero mode. A Ramond spin structure would be handled by inserting $(-1)^F$ into the trace in the computation of $Z[A]$, giving

$$Z_R[A] = e^{i\pi\eta(A)/2 + ik\oint A_0}\sin(\pi\eta(A)/2), \tag{229}$$

and so the zero mode of $\not{D}_A$ exists if $\oint A_0 \in 2\pi\mathbb{Z}$ (which one can also see from decomposing $\psi(t)$ in frequency modes).

Finally, note that the T and C non-invariance of the partition function depended on the boundary conditions we chose. For example, if we computed the partition function for a C breaking scenario where our boundary conditions had a definite particle number (e.g. either $|k\rangle$ or $|k+1\rangle$ entered Z, but not both), then the anomaly would be "gone", in the sense that the partition function could be rendered C and T symmetric (in fact, just $Z[A] = 1$) by an appropriate choice of counterterm k.

Anyway, let's return to a more general problem where we can't so easily take advantage of dimensional reduction to get a quantum mechanics problem. The general expression for the partition function has to still be expressed in terms of the eta invariant, by topological arguments. The general definition of $\eta(A)$ is such that it agrees with the one above after taking the specific case of a 2π flux through a spatial S^2: therefore we must have

$$\eta(A) = \left(\frac{1}{4\pi^2}\int A\wedge dA + 1\right)_2 - 1, \tag{230}$$

and the partition function is analogously

$$Z[A] = |Z[A]|\exp\left(\frac{ik}{4\pi}\int A\wedge dA + \frac{\pi i}{2}\eta(A)\right). \tag{231}$$

Note that if we had taken the flux to be $2\pi n$ instead, we would get n zero modes. The sum over these in the IR limit of the partition function would then produce $e^{\pi i n\eta(A)/2}$, which

e.g. anomalies in TQFTs where the theory in question is gapped and the sum over ground states is finite. In these scenarios the real part of the free energy is non-singular (it's a sum of a finite number of smooth exponentials), but as we have seen this does not preclude the imaginary part from being singular.

can be killed by a local counterterm if n is even; hence there is no T-breaking in these backgrounds. However, when we say "gauge A", we (in this diary entry) include a sum over all flux sectors (as far as I can tell, this is not for sure required), and so in any case there is still a mixed anomaly, because if A is made dynamical then T and C are broken.

17 T, CT, and dualities ✓

Today's diary entry is an elaboration on an exercise that Nati Seiberg assigned to the students at the 2018 / 2019 Jerusalem winter school on QFT. The problem was to explain why, in dualities, T and CT symmetries are often exchanged.

In the following, we will use a notation where $\mathcal{T}$, $\mathcal{CT}$ are the "duals" of T and CT under some "duality map" $\mathcal{D}$. They are defined by

$$\mathcal{D}[T\mathcal{O}] = \mathcal{T}\mathcal{D}[\mathcal{O}], \quad \mathcal{D}[CT\mathcal{O}] = \mathcal{C}\mathcal{T}\mathcal{D}[\mathcal{O}], \tag{232}$$

where $\mathcal{O}$ is any field that has an image under $\mathcal{D}$. The claim is that the usual story for dualities is $\mathcal{T} = CT, \mathcal{CT} = T$.

Particle on a ring

The first, simplest possible example is that of the duality (just a Fourier transform) between p and q for a particle on a ring, with Lagrangian

$$\mathcal{L}[q] = (\partial_t q)^2 - q^2. \tag{233}$$

We define the symmetries T and C to act as

$$T : q(t) \mapsto q(-t), \qquad CT : q(t) \mapsto -q(t). \tag{234}$$

From here on, the time dependence of variables will mostly be kept implicit.

The conjugate variable to q is p. When we write the path integral in the Hamiltonian formulation, we have the Berry phase term $S \supset \int dt\, \dot{q}p$. Under T, this goes to $-\int dt\, \dot{q}T(p(-t))$, while under CT we get $+\int dt\, \dot{q}CT(p(-t))$. Therefore invariance of this term under the symmetries tells us that p transforms as

$$T : p \mapsto -p, \qquad CT : p \mapsto p. \tag{235}$$

Now duality here (alias Fourier transform) is

$$\mathcal{D} : q \mapsto p, \quad p \mapsto -q. \tag{236}$$

Here the minus sign, which says that $\mathcal{D}^2 = -\mathbf{1}$, can be seen in several ways. One is that we require the symplectic form $dq \wedge dp$ to be invariant, with the antisymmetry of the $\wedge$ product necessitating the minus sign. Another way to see this is to note that the square of the Fourier transform is an inversion. That is, letting $\mathcal{F}$ be the Fourier transform,

$$\mathcal{F}^2[f(t)] = \mathcal{F} \int dt\, e^{i\omega t} f(t) = \int d\omega \int dt\, e^{i\omega t'} e^{i\omega t} f(t) = f(-t). \tag{237}$$

This is just because of the fact that the Fourier transform performs a $\pi/2$ rotation in frequency-time space, with a π rotation then corresponding to a reversal of the time coordinate. Since dualities are often performed by a Fourier transform, and the one in the present context indeed is, we expect $\mathcal{D}^2 : q \mapsto -q$, which it does. As we will see, this holds even for more advanced kinds of dualities like Electromagnetic duality, which again is basically just a Fourier transform.

We can summarize the (now obvious) fact that $\mathcal{D} : T \leftrightarrow CT$ by drawing the following commutative diagram:

$$\begin{tikzcd} q \arrow[r, "\mathcal{D}"] \arrow[d, "T"] & p \arrow[d, "\mathcal{T}"] \\ q \arrow[r, "\mathcal{D}"] & p \end{tikzcd}, \tag{238}$$

which tells us that $\mathcal{T} = CT$. A similar diagram shows that $\mathcal{CT} = T$.

Electromagnetic duality in four dimensions

A slightly more sophisticated example is electromagnetic duality in 3+1 dimensions. We will work with conventions where $T, C : A \mapsto -A$ (treating A as a 1-form and again suppressing the t argument). Thus the vector components E^i are even under time reversal while those of B^i are odd, with both E^i and B^i odd under C. This convention is subideal for many contexts, but we will stick with it due to it being the one most common in the literature.

As we have seen several times in previous diary entries, electromagnetic duality is (up to a constant of proportionality involving the gauge coupling), implemented by Hodge duality: $\mathcal{D} : F \mapsto \star F$. Recall how this works: we implement $F = dA$ by the Lagrange multiplier term

$$S \supset \frac{i}{2\pi} \int F \wedge d\widetilde{A}, \tag{239}$$

and then integrate out F by doing a shift of F by something proportional to $\star d\widetilde{A}$. If we were to insert F into the path integral, since $\langle F \rangle = 0$ when integrating out F, after the shift to eliminate the Lagrange multiplier term we'd be left with a path integral containing just an insertion of $\star d\widetilde{A}$; hence why $\mathcal{D}$ is basically Hodge duality.

On the components of the field strength, the duality is

$$\mathcal{D} : E \mapsto B, B \mapsto -E. \tag{240}$$

The minus sign in the second map is a Lorentzian minus sign coming from lowering a time index on F, and ensures that $\mathcal{D}^2 = -\mathbf{1}$ (this is just because $\star^2 = (-1)^{1+p(D-p)}$ on p-forms in D-dimensional Minkowski space. In Euclidean signature this minus sign is still picked up since the proportionality constant between $\mathcal{D}(F)$ and $\star F$ is imaginary).

Now we can draw the following commutative diagram:

$$\begin{tikzcd} E \arrow[r, "\mathcal{D}"] \arrow[d, "T"] & B \arrow[d, "\mathcal{T}"] \\ E \arrow[r, "\mathcal{D}"] & B \end{tikzcd}. \tag{241}$$

Thus we conclude that $\mathcal{T}$ must act trivially on B, from which we can make the identification $\mathcal{T} = CT$. Similarly,

$$\begin{tikzcd} E \arrow[r, "\mathcal{D}"] \arrow[d, "CT"] & B \arrow[d, "\mathcal{C}\mathcal{T}"] \\ -E \arrow[r, "\mathcal{D}"] & -B \end{tikzcd}, \tag{242}$$

so that $\mathcal{C}\mathcal{T} = T$ since B is odd under $\mathcal{C}\mathcal{T}$. Thus duality exchanges T and CT. Another way to see this is to couple the theory to a background field for the $U(1)$ 1-form symmetry, and to use the transformation properties of this background field under T and CT to fix the identification of the symmetries on both sides. If H is the 2-form background field, we have (this was derived in a previous diary entry)

$$i\int ||F_A - H||^2 \leftrightarrow i\int (||F_{\widetilde{A}}||^2 + F_{\widetilde{A}} \wedge H), \tag{243}$$

where we have omitted real constants (and are in $\mathbb{R}$ time). On the LHS, we see that H transforms in the same way under T, CT as F, and hence as A, does. On the RHS, because of the i, we see that H transforms in the opposite way as $F_{\widetilde{A}}$. Thus since the field strengths are odd under T and even under CT, duality exchanges C and CT. Using this type of $iF_{\widetilde{A}} \wedge H$ term (think of this as a field theory version of $ik \cdot x$) to show that $T \leftrightarrow CT$ is a typical strategy.

3d particle-vortex duality

We can also look at a slightly more complicated example, that of the duality between the 2+1 dimensional WF scalar and another 2+1 WF scalar coupled to a gauge field (and deformed away from their critical points by mass terms, with the mass terms defined to vanish at the critical point). The duality is (adding a background field A to both theories)

$$|D_A\phi|^2 + r|\phi|^2 + u|\phi|^4 \leftrightarrow |D_a\widetilde{\phi}|^2 + \widetilde{r}|\widetilde{\phi}^2| + \widetilde{u}|\widetilde{\phi}|^4 + \frac{i}{2\pi} a \wedge F_A. \tag{244}$$

Since we are choosing $T(A) = -A$, we need $T(\phi) = \phi$. Then from the T-invariance of the CS term on the RHS (this is not a functional δ function since A is not integrated over and a is not a Lagrange multiplier, so the sign is meaningful), we see that $T(a) = a$, i.e. that T acts as CT on the dual gauge field. From the invariance of $|D_a\widetilde{\phi}|^2$, we get $T(\widetilde{\phi}) = \widetilde{\phi}^\dagger$, in keeping with T acting as CT on the dual fields.

18 *T, CT, their anomalies, and bosonization in 1+1D* ✓

Today's diary entry is similar to yesterday's, and concerns the realization of the T and CT symmetries when doing bosonization of fermions coupled to a $U(1)$ gauge field in 1+1D. We will also show that when the fermions are Kramers doublets T and CT are anomalous, something that I hadn't seen in the literature before.

Since we have fermions and spacetime symmetries involved, we can already anticipate a panoply of different choices for how the symmetries act. To identify how the symmetries act, it is helpful to add a classical background field to both sides, which lets us set the standards for T, CT on both sides of the duality.

First we have to get the conventions straight for where the currents go under duality. We will work in $\mathbb{R}$ time, with conventions where $\mathcal{D}[\psi_\pm] = e^{-i\phi_\pm}$. These are the conventions in which the Dirac mass maps to $\cos(\theta)$, where $\theta = (\phi_+ - \phi_-)/2$ is the dual field to $\phi = (\phi_+ + \phi_-)/2$. In these conventions, the vector current for the fermions maps to the vector current for the bosons (*not* the topological current for the bosons), so that in the usual notation

$$\mathcal{D}[J_\pm] = \frac{1}{\pi}\partial_\pm \phi. \tag{245}$$

Note that to write this we technically have had to use the equations of motion for the θ and ϕ fields (this is exact though since they appear quadratically in the action). More on this point and the (crucial!) factor of $1/\pi$ can be found in the diary entry on bosonizatoin conventions. Anyway, using this we find

$$\mathcal{D}\left[\int J \wedge \star A\right] = \mathcal{D}\left[\int (J_+ A_- + J_- A_+)\right] = \frac{1}{\pi}\int d\phi \wedge \star A = \frac{1}{\pi}\int \theta\, F_A, \tag{246}$$

where we used $d\phi \overset{\text{eom}}{=} \star d\theta$, which again is legit since θ just appears quadratically in the action (the justification of the integration by parts and the well-definedness of these terms is in the diary entry on DB cohomology).

Summarizing, the duality map $\mathcal{D}$ thus does (for why the radius is $R = \sqrt{2}$, see the diary entry on bosonization conventions)

$$\mathcal{D} : \frac{1}{2\pi}\bar{\psi} i \not{D}_A \psi \leftrightarrow \frac{1}{2\pi} d\phi \wedge \star d\phi + \frac{1}{\pi}\theta F_A. \tag{247}$$

If we had used the convention where $\mathcal{D}[\psi_\pm] = e^{\mp i\phi_\pm}$ we would have obtained the T-dual image of this action.

As a sanity check, the coefficient of $1/\pi$ is correct since a chiral transformation $\psi \mapsto e^{i\bar{\gamma}\alpha}\psi$ does $\theta \mapsto \theta + \alpha$. This reproduces the shift in the action of $(\alpha/\pi)\int F_A$ that comes from the chiral anomaly on the fermion side — it is α/π and not $\alpha/2\pi$ since integrating this term gives the divergence of the chiral current, and since the $\mathbb{Z}_2$ subgroup of $U(1)_A$ is preserved (as it is also part of $U(1)_V$), the integral of the divergence must be in $2\mathbb{Z}$.

Now for the symmetries. For posterity's sake, we record the easily proved facts that for any real differential form B,

$$T[B] = \zeta_B B \implies T[dB] = \zeta_B dB, \quad T[\star B] = -\zeta_B dB, \qquad \zeta_B = \pm 1, \tag{248}$$

and likewise for T replaced by a reflection R. Here the notation again keeps the spacetime arguments implicit, and also tacitly includes a change of variables so that the spacetime arguments of the fields are unchanged at the end. For example, on a 1-form A, we have

$$T[A] = -\zeta_A A_t(-t,x)\, dt + \zeta_A A_i(-t,x)\, dx^i \to \zeta_A A, \tag{249}$$

where the $\to$ means doing the change of variables $t \to -t$ (all differential forms will always appear under an integral sign so this can always be done). Thus in components, $A \mapsto \zeta_A A$ means that $A_t \mapsto -\zeta_A A_t$ while $A_i \mapsto \zeta_A A_i$. Because of this tacit change of variables, when working out the transformation properties of integrals, we can take $\int$ to be odd under any orientation-reversing spacetime symmetry, since the change of variables means the domain of integration gets its orientation reversed. Therefore we have e.g.

$$T: \int_X A \wedge B = -\zeta_A \zeta_B \int A \wedge B. \tag{250}$$

We will fix conventions where C, R, T act on a $U(1)$ gauge field and the various components of its field strength as (from the above, we see that the differential form dA transforms in the same way as A does)

$$\begin{aligned} T &: A \mapsto -A, \quad E^i \mapsto E^i \quad B^i \mapsto -B^i \\ C &: A \mapsto A, \quad E^i \mapsto -E^i, \quad B^i \mapsto -B^i \\ R &: A \mapsto A, \quad E^i \mapsto -E^i, \quad B^i \mapsto B^i \end{aligned} \tag{251}$$

These are the old-school conventions and ideally T and CT would be swapped — we'll stick with these though since they're mostly what's used in the literature.

When coupling to the fermions, we need the $\int J_\psi \wedge \star A$ term to be T-invariant. Since $T(A) = -A$, we need (again as differential forms, so that $J_\psi = \bar{\psi}\gamma_\mu \psi dx^\mu$)

$$T(\star J_\psi) = \star J_\psi \implies T(J_\psi) = -J_\psi, \tag{252}$$

so that the transformation of J_ψ correctly matches that of A. This transformation rule makes sense because T always takes $\psi_\pm^\dagger \psi_\pm \mapsto \psi_\mp^\dagger \psi_\mp$, and hence sends $J_\psi^0 \mapsto J_\psi^0, J_\psi^1 \mapsto -J_\psi^1$. Charge conjugation C sends $A \mapsto -A$, so that $CT(A) = A$, and likewise $CT(J_\psi) = J_\psi$. In $\mathbb{R}$ time we have $\mathcal{J}_\psi = \star J_\psi$ where $\mathcal{J}$ is the axial current, and so $T(\mathcal{J}) = \mathcal{J}, CT(\mathcal{J}) = -\mathcal{J}$.

Now we have to fix conventions for how the fermions transform under C, P, T. We will fix the signature to be $(+,-)$, with γ matrices $\gamma^0 = X, \gamma^1 = -iY = J$.[40] For time reversal, both choices $T = Y\mathcal{K}$ or $T = X\mathcal{K}$ are consistent: the former gives $T^2 = (-1)^F$ while the later gives $T^2 = \mathbf{1}$.[41] We will denote these two choices by T_- and T_+, respectively. We will take charge conjugation to be performed by $C : \psi \mapsto \bar{\psi}C^\dagger, \bar{\psi} \mapsto -C\psi$, which is a symmetry of the action if $C = Y$.[42] Component-wise, this is

$$C : \psi_\pm \mapsto \pm i\psi_\pm^\dagger. \tag{253}$$

[40]If we change the signature, we will be able to change whether $P^2 = \mathbf{1}$ or $-\mathbf{1}$ on fermions: since we are mostly focused on time reversal, we won't worry about exploring all these options, and will just choose $(+,-)$ signature, in which $P^2 = \pm\mathbf{1}$ is determined from the choice of $T^2 = \pm\mathbf{1}$ and CPT invariance (this just means that we are working in conventions where we take P to be the thing appearing in CPT — this is of course not however the only option for P; other choices are consistent but less canonical).

[41]This is because for $T = U_T\mathcal{K}$, we need $U_T^\dagger Z U_T = -Z$.

[42]Note that the free action would also be preserved by $C : \psi_\pm \mapsto \psi_\pm^\dagger$. However, in this case, the Dirac mass $m_D\bar{\psi}\psi$ would be C-odd, which we don't want.

In our conventions P is the operation that appears in CPT (one could argue that this is philosophically not a very good approach to take, but my 2019 self thought otherwise), and hence our choices of C and T force us to choose $P = X$ if we work with T_+, and $P = J$ if we work with T_- (if we had chosen $(-,+)$ signature, the two actions of P would be reversed). More on the method to this madness is explained on the diary entry on fermions and spacetime symmetries. Working out the action on the field components, we see that

$$(CT_+)^2 = (-1)^F, \qquad (CT_-)^2 = +\mathbf{1}, \tag{254}$$

so that the square of $CT_\pm$ is opposite to the square of $T_\pm$.

T_+: Consider first the case when fermions are Kramers singlets. Then applying T_+ to $\mathcal{D}[\psi_\pm] = e^{-i\phi_\pm}$, we get

$$\begin{tikzcd} \psi_\pm \arrow[r, "\mathcal{D}"] \arrow[d, "T_+"] & e^{-i\phi_\pm} \arrow[d, "\mathcal{T}_+"] \\ \psi_\mp \arrow[r, "\mathcal{D}"] & e^{-i\phi_\mp} \end{tikzcd} . \tag{255}$$

Thus we have[43]

$$\mathcal{T}_+ : \phi_\pm \mapsto -\phi_\mp, \qquad \phi \mapsto -\phi, \quad \theta \mapsto \theta. \tag{256}$$

Charge conjugation evidently acts on the boson side as $\mathcal{C} : \phi_\pm \mapsto -\phi_\pm \mp \pi/2$. However, one must be careful in deriving its action on θ and ϕ. θ and ϕ are fermion-parity even, which leaves room for a minus sign to enter in the action of charge conjugation which doesn't appear when C acts on single fermion operators. So for example,

$$C : \psi_+^\dagger \psi_- \mapsto -\psi_+ \psi_-^\dagger = +\psi_-^\dagger \psi_+ \implies \mathcal{C} : e^{i\theta} \mapsto e^{-i\theta}. \tag{257}$$

If we had naively applied $\mathcal{C}$ to $\phi_+ - \phi_-$, we would have obtained $\mathcal{C} : \theta \mapsto -\theta - \pi$ instead. The difference between this transformation and the correct result comes down to the fact that when we bosonize, we only apply the mosonization map $\mathcal{D}$ to things that are normal-ordered, and so we have to keep track of possible normal-ordering signs that appear. Similarly, considering the action of C on $\psi_+^\dagger \psi_-^\dagger$, one finds the action of $\mathcal{C}$ on ϕ, so that

$$\mathcal{C} : \phi \mapsto -\phi, \quad \theta \mapsto -\theta. \tag{258}$$

This means that the Dirac mass $\bar\psi\psi \to 2\cos\theta$ is $\mathcal{C}$-even while the chiral mass $i\bar\psi Z \psi \to 2\sin\theta$ is $\mathcal{C}$-odd, as expected for the present conventions.

Putting these together,

$$\mathcal{C}\mathcal{T}_+ : \phi_\pm \mapsto \phi_\mp \mp \pi/2, \qquad \phi \mapsto \phi, \quad \theta \mapsto -\theta. \tag{259}$$

A sanity check is that $(\mathcal{C}\mathcal{T}_+)^2 = -\mathbf{1}$ when acting on $\phi_\pm$ (we will always have $(\mathcal{C}\mathcal{T}_\pm)^2 = \mathbf{1}$ when acting on ϕ or θ since these variables are fermion-parity even, so we just need to check

[43] When we look at the action of symmetries on ϕ, θ, which are related to fermion bilinears, we need to remember Klein factors, which are acted on by the γ matrices. For example, $\mathcal{D}[\psi_+^\dagger \psi_-] = \kappa_+ \kappa_- e^{i\theta}$, $\mathcal{D}[\psi_-^\dagger \psi_+] = \kappa_- \kappa_+ e^{-i\theta}$, and so $\mathcal{T}_+ : \theta \mapsto \theta$ only makes sense if $\mathcal{T}_+ : \kappa_\pm \to \kappa_\mp$.

the action of $(\mathcal{C}\mathcal{T}_\pm)^2$ on $\phi_\pm$). Indeed, keeping in mind that in this representation the complex conjugation in $\mathcal{T}_+$ sends scalars to minus themselves,

$$(\mathcal{C}\mathcal{T}_+)^2 : \phi_\pm \to -\phi_\mp \to -(-\phi_\mp \pm \pi/2) \to -\phi_\pm \pm \pi/2 \to \phi_\pm \pm \pi. \tag{260}$$

Finally, with this choice of T_+, we have $P_+ = X$ on the fermions; thus

$$\mathcal{P}_+ : \phi_\pm \mapsto \phi_\mp, \qquad \phi \mapsto \phi, \quad \theta \mapsto -\theta. \tag{261}$$

T_-**:** Now consider the case when the fermions are Kramers doublets. We take $T_- = Y\mathcal{K}$, so that $T : \psi_\pm \mapsto \pm i\psi_\mp$. Then we have

$$\begin{array}{ccc} \psi_\pm & \xrightarrow{\quad \mathcal{D} \quad} & e^{-i\phi_\pm} \\ \big\downarrow{\scriptstyle T_-} & & \big\downarrow{\scriptstyle \mathcal{T}_-} \\ e^{\pm i\pi/2}\psi_\mp & \xrightarrow{\quad \mathcal{D} \quad} & e^{-i\phi_\mp \pm i\pi/2} \end{array} . \tag{262}$$

This identifies the bosonized action of T_- as

$$\mathcal{T}_- : \phi_\pm \mapsto -\phi_\mp \mp \pi/2, \quad \phi \mapsto -\phi, \quad \theta \mapsto \theta - \pi. \tag{263}$$

Consequently,

$$\mathcal{C}\mathcal{T}_- : \phi_\pm \mapsto \phi_\mp \pm \pi, \qquad \phi \mapsto \phi, \quad \theta \mapsto -\theta + \pi. \tag{264}$$

Lastly, we have parity, which acts as $P_- = J$ on the fermions. Thus

$$\mathcal{P}_- : \phi_\pm \mapsto \phi_\mp + (1 \mp 1)\pi/2, \qquad \phi \mapsto \phi + \pi, \quad \theta \mapsto -\theta - \pi. \tag{265}$$

We can summarize what we've derived so far in the following table:

	C	P_+	T_+	CT_+	P_-	T_-	CT_-
ϕ	$-$	$+$	$-$	$+$	$\phi+\pi$	$-$	$+$
θ	$-$	$-$	$+$	$-$	$-\theta-\pi$	$\theta-\pi$	$-\theta+\pi$
m_D	$+$	$+$	$+$	$+$	$-$	$-$	$-$
m_5	$-$	$-$	$+$	$-$	$+$	$-$	$+$
$\int \theta F_A$	$+$	$+$	$+$	$+$	$\frac{1}{2}\int F_A$	$-\frac{1}{2}\int F_A$	$-\frac{1}{2}\int F_A$

(266)

Here the $\pm$ signs mean the fields get multiplied by ± 1, while the other entries indicate the amount to which each term shifts (e.g. ϕ shifts by π under $P-$). Note that $CP_\pm T_\pm$ acts trivially on θ and $\int \theta F_A$, as required.[44]

We see from the table that in the Kramers doublet case, all of the symmetries P_-, T_-, and CT_- are anomalous. Indeed, we can see this from the fact that neither of the fermion mass terms (which bosonize to $\cos\theta, \sin\theta$) are P_-, T_-, or CT_- invariant! Therefore we cannot do PV regularization without breaking these symmetries; hence the anomalies. On the other hand, the Kramers singlet case is non-anomalous (with the Dirac mass being a symmetry-preserving mass to give the PV regulator).

[44] CP_-T_- doesn't act trivially on ϕ but that's okay since $e^{i\phi} \sim \psi_+^\dagger \psi_-^\dagger$ won't appear in by itself in a Lorentz-invariant theory (it will only appear as $\partial\phi$ and $\bar{\partial}\phi$). (note to self — come back to this!)

19 *1-form anomalies in CS theory* ✓

Today we will examine the anomalous nature of the 1-form symmetries present in various CS theories.

$U(1)_k$

Let's start with the simplest example, viz. $U(1)_k$. The 1-form symmetry acts as

$$\mathbb{Z}_k^{(1)} : A \mapsto A + \lambda, \qquad k\lambda \in 2\pi H^1(X;\mathbb{Z}). \tag{267}$$

The charge operators are of course the Wilson lines. We can see that this is a symmetry by e.g. computing the spectrum of operators in the theory, but for posterity's sake let's see how it works from the action. Since λ is flat, a naive approach tells us that $\delta S = \frac{1}{4\pi}\int (k\lambda)\wedge F_A$ under the symmetry, which is only in $\frac{1}{2}\overline{\mathbb{Z}}$ (we are using the notation $\overline{\mathbb{Z}} \equiv 2\pi\mathbb{Z}$). Note that we cannot integrate this by parts to get zero by the flatness of λ, due to A not being strictly a well-defined form[45]. As usual, the confusion can be ameliorated by writing things in terms of the field strengths by using a bounding 4-manifold M. Then

$$\delta S = \frac{k}{2\pi}\int_M F_\lambda \wedge F_A + \frac{k}{4\pi}\int_M F_\lambda \wedge F_\lambda, \tag{268}$$

where F_λ is the field strength of the extension of λ into the bulk 4-manifold M. Note that since the holonomy of λ may be nontrivial, although it is flat on the boundary, it will not be flat in M, and a priori, it will not be globally an exterior derivative, i.e. we may not have $F_\lambda = d\lambda$ globally on M.

Warning: the following paragraph will be slightly pedantic. Now we need to integrate by parts: we will get only boundary terms, since $dF_A = d(d\lambda) = 0$. However, doing so is slightly subtle, since λ might not be a globally well-defined form. Thus we cannot write e.g. $\int d\lambda \wedge B = \int \lambda \wedge dB + \int_{\partial M} \lambda \wedge B$ for a 2-form B (the sign is correct because of the supercommutativity of d). However, since λ is flat, we know that λ is a well-defined form on ∂M. Thus in the bulk, we may write

$$\lambda = \Lambda + B, \qquad F_B \in 2\pi H^2(M,\partial M;\mathbb{Z}), \tag{269}$$

where Λ is a $U(1)$ gauge field which is globally well-defined so that $[F_\Lambda] = [d\Lambda] = 0$ in $2\pi H^2(M;\mathbb{Z})$, and B is a non-globally-well-defined part which vanishes on ∂M since $\lambda|_{\partial M}$ is globally well-defined (thus $\lambda|_{\partial M} = \Lambda|_{\partial M}$). Thus we can write

$$\delta S = \frac{k}{2\pi}\int_M [(d\Lambda + F_B)\wedge F_A + d\Lambda \wedge F_B] + \frac{k}{4\pi}\int_M (d\Lambda \wedge d\Lambda + F_B \wedge F_B). \tag{270}$$

[45] A similarly hasty use of integration by parts on the CS action leads to confusion in the usual way of showing that $k \in \mathbb{Z}$ in the CS action, namely by e.g. placing the theory on $S^1 \times S^2$ with $\int F = 2\pi$ around the S^2. In the usual story one integrates by parts to get $S = \frac{k}{2\pi}\oint A_t \int_{S^2} F_{xy} = k\oint A_t$ (there is a factor of 2 here from the IBP), which says that $k \in \mathbb{Z}$ for invariance under large gauge transformations around the S^1. But what if we first did the large gauge transformation, and then did the integration by parts? Since the field strength of the large gauge transformation vanishes, the IBP fails to pick up a factor of 2, and we conclude that the change in the action is instead $(k/4\pi)\oint \lambda \int_{S^2} F_{xy}$ for $\oint \lambda = 2\pi$, which seems to imply that $k \in 2\mathbb{Z}$ is required. So, it is best to only integrate by parts when we really know that it is legit.

Assuming we choose M to be spin if k is odd, the last term vanishes modulo $\overline{\mathbb{Z}}$. Since Λ is globally well-defined, the $d\Lambda \wedge F_B$ term vanishes on account of the flatness of F_B and the fact that $F_B|_{\partial M} = 0$. Likewise the $F_B \wedge F_A$ part vanishes mod $\overline{\mathbb{Z}}$: we can see this by decomposing A in the same way that we decomposed Λ, and using that $\frac{1}{2\pi}\int F_C \wedge F_B \in \overline{\mathbb{Z}}$ for $F_C \in 2\pi H^2(M, \partial M; \mathbb{Z})$. So finally, we integrate the remaining two terms by parts and get

$$\delta S = \frac{1}{2\pi}\int (k\lambda) \wedge F_A, \tag{271}$$

since $\Lambda|_{\partial M} = \lambda|_{\partial M}$ and since $d\lambda|_{\partial M}$ is flat. But since $k\lambda$ has periods in $\overline{\mathbb{Z}}$, we see that $\delta S \in \overline{\mathbb{Z}}$, and so indeed, the 1-form transformation $\delta A = \lambda$ *is* a symmetry of the action.

To gauge this symmetry, we want the "split" symmetry operators $U(g; C)$ (not only the full charge operators) to act trivially on the Hilbert space, where the split symmetry operators are defined on *open* submanifolds $C : \partial C \neq 0$ and implement a transformation by the group element g (here $g \in \mathbb{Z}_k$). Requiring the global (unsplit) charge operators to act trivially is equivalent to projecting onto the singlet sector of the Hilbert space, which can be done by inserting the operator

$$\Pi_{\mathbf{1}} = \sum_{C \in H_1(X;\mathbb{Z}_k)} e^{i\int_C A} \tag{272}$$

into the path integral. This is orbifolding. This is not quite how we want to think of gauging, since we haven't made the symmetry local in any way, we've just projected onto a particular (trivial) value of the charge (when we gauge we want to work with genuinely gauge-invariant states, not ones that are in some particular gauge fixing). A hint of the anomaly can be seen by noticing that the split symmetry operator $U(q, C)$, which naively includes $e^{iq\int_C A}$ since the current for the 1-form symmetry is $j = \star A$, is not gauge invariant (under both the 0-form and 1-form gauge transformations) if $\partial C \neq 0$. This is a problem since after including the 2-form gauge field we expect $U(q, C)$ to be modified by the canonical momentum for the 2-form gauge field (think of $\nabla \cdot \mathbf{E}$), which we don't expect to transform in a way that would fix this issue, and hence the gauge variance of $U(q, C)$ will continue to be problematic.

Let's now see how this plays out. We will let B be the $\mathbb{Z}_k$ 2-form gauge field. To enforce the quantization of B's periods, we will as usual add to the action the BF term (we are in $\mathbb{R}$ time, so no factors of i are included)

$$S \supset \frac{k}{2\pi}\int B \wedge d\phi, \tag{273}$$

where ϕ is a 2π-periodic scalar. Now we can fix the variance under 0-form gauge transformations of $e^{i\int_C \star j} = e^{i\int_C A}$ when $\partial C \neq 0$ by writing the operator $U(q, C)$ which implements the gauge transformation as

$$U(q, C) = e^{iq\int_C (A + d\phi)}, \tag{274}$$

provided that under $A \mapsto A + d\gamma$ we have $\phi \mapsto \phi - \gamma$ (this preserves the 2π-periodicity of ϕ, since γ is itself a 2π-periodic scalar). This makes sense, since ϕ is the canonical momentum for B, and so this is exactly what we normally do when gauging the symmetry operators: the operators which perform the gauge transformations are the original charge operators defined on open submanifolds, with the canonical momentum for the gauge field integrated along their boundaries (again as an example, the generator of gauge tforms in QED is the integral of the matter current over an open volume, together with the integral of $\star F$, the canonical momentum for the gauge field, over the boundary of the volume).

Now by design, if $D \in C_1(X;\mathbb{Z})$ is such that $C \cap D \neq 0$, then $W(D) = e^{i\int_D A}$ is not gauge invariant under the $\mathbb{Z}_k^{(1)}$ gauge transformations, since it does not commute with $U(q,C)$. Note that no matter what D is, we can always find a C such that $W(D)$ is not invariant under $U(q,C)$: this is true even when $[D]=0$ in $H_1(X;\mathbb{Z})$, in which case $W(D)$ is actually neutral under the original 1-form global symmetry.

We can make $W(D)$ gauge invariant by attaching a surface operator built out of B to it: if $[D]=0$ in $H_1(X,\partial X;\mathbb{Z})$ (so that D either bounds a disk, is a linear combination of nontrivial classes in $H_1(X;\mathbb{Z})$ with total "charge" zero so that it bounds some other surface, or together with a submanifold of the boundary of spacetime bounds a surface) we can find some M such that $\partial M \setminus (\partial M \cap \partial X) = D$ (here X is spacetime, and gauge transformations always vanish at ∂X). The operator

$$\widetilde{W}(M) = \exp\left(i\int_D A + i\int_M B\right) \tag{275}$$

is then gauge-invariant. Why? Because when we compute its commutation relation with $U(q,C)$ (with e.g. $C \cap D = 1$), we get one factor of $e^{2\pi i q/k}$ from the $[A,A] \sim i/k$ commutation relation, and another from the $[\phi, B] \sim i/k$ commutation relation, which occurs from the contact term between the ϕ inserted at the end of C and the B integrated over M. If $[D] \neq 0$ in $H_1(X;\mathbb{Z})$ then $W(D)$ can't be made gauge-invariant, and its vev vanishes (although this was true before gauging, since $\langle W(D)\rangle$ can then be shifted by a change in integration variables which doesn't affect the boundary conditions on A).

As we hinted at above, the anomaly is seen very simply from the fact that the operators $U(q,C)$ which perform the $\mathbb{Z}_k^{(1)}$ gauge transformations are not themselves invariant under the same $\mathbb{Z}_k^{(1)}$ transformations (although as we have seen they are at least invariant under the 0-form $U(1)$ gauge transformations on A). That is, they don't commute with themselves (because of $[A,A] \sim i/k$). Since $\partial C \neq 0$, it is impossible to attach a B surface to render $U(q,C)$ gauge invariant. Thus the $\mathbb{Z}_k^{(1)}$ symmetry can't actually be gauged.

We can also see this from the action. Basically, while F_A can be made gauge invariant by $F_A \mapsto F_A - B$, the CS term cannot be made gauge invariant since it involves more than just F_A. Indeed, let us write the variation of A under the 1-form gauge transformation as

$$\delta A = \lambda, \qquad \lambda \in \frac{1}{k} Z^1_O(X;\mathbb{Z}). \tag{276}$$

Here we have defined $Z^1_O(X;\mathbb{Z})$ as the set of 1-forms such that their spatial Poincare duals are *open* codimension-1 submanifolds of space, which have integral intersection number with every element in $C_1(X;\mathbb{Z})$ that intersects them transversely. Thus the elements in $Z^*_O(X;\mathbb{Z})$ are not closed, but they are not closed in a very specific way (we are not letting λ be an arbitrary 1-form since we are gauging a $\mathbb{Z}_k$ 1-form symmetry, and not a $U(1)$ one). Another way to say this is that

$$\lambda \in \frac{1}{k} Z^1_O(X;\mathbb{Z}) \implies \int_C \lambda \in \frac{1}{k}\mathbb{Z} \quad \forall C \in C_1(X;\mathbb{Z}), \tag{277}$$

where the value for the integral will generically depend on the exact choice of C, and not just its homotopy equivalence class. Connecting this with our earlier notation vis-a-vis the $U(q,C)$ gauge transformation operators, we would say that $U(q,C)$ shifts A by $\lambda = \frac{q}{k}\widehat{C}$, where $\widehat{C}$ is the spatial Poincare dual of $C \in C_1(X;\mathbb{Z})$.

Anyway, the CS term varies as

$$\delta \int A \wedge dA = 2 \int A \wedge d\lambda + \int \lambda \wedge d\lambda. \tag{278}$$

To cancel at least the first term we can try to introduce a $\mathbb{Z}_k$ gauge field B to the action and add

$$S \supset -\frac{1}{2\pi} \int A \wedge B, \tag{279}$$

with $\delta B = d\lambda$ under the gauge transformation. However this a) cannot cancel the term in δS quadratic in λ and b) produces an extra piece linear in B. So, after adding this coupling, the total variation of S is

$$\delta S = \frac{k}{4\pi}\delta \int (A \wedge dA - 2A \wedge B) = \frac{k}{4\pi} \int (2\lambda \wedge B + \lambda \wedge d\lambda). \tag{280}$$

Of course, we can make the action gauge invariant by letting B live in four dimensions, at the price of picking up an explicit dependence on a bounding 4-manifold M. This is just because F_A can always be made gauge-invariant, and we can write the terms in our modified action involving A as

$$S \supset \frac{k}{4\pi} \int_M (F_A - B) \wedge (F_A - B), \tag{281}$$

which is manifestly gauge invariant (and still only depends on $A|_{\partial M}$). However, and this is where the anomaly comes in, it depends on the choice of M, since

$$\frac{k}{4\pi} \int_{N_4 | \partial N_4 = \emptyset} (F_A - B) \wedge (F_A - B) \in \frac{1}{k}\overline{\mathbb{Z}}, \tag{282}$$

which is not valued in $\overline{\mathbb{Z}}$ except in the trivial case $k=1$ where there is no symmetry to begin with (here we have used the fact that the periods of B are valued in $k^{-1}\overline{\mathbb{Z}}$ — the periods of F_A are still in $\overline{\mathbb{Z}}$ though, since the 1-form gauge transformations only change A by forms which are globally well-defined up to elements in $2\pi H^2(N_4;\mathbb{Z})$ (or are they always globally well-defined?)). Since k copies of this bulk action integrate to something in $\overline{\mathbb{Z}}$ over all closed 4-manifolds, we have a $\mathbb{Z}_k$ anomaly.

To write the full gauged action for the four-dimensional B, we just need to include the term which makes B into a $\mathbb{Z}_N$ gauge field. Since B lives in four dimensions, the appropriate BF term is $(k/2\pi)\int_M B \wedge F_\Phi$, where Φ is a 1-form $U(1)$ gauge field. But this term changes as $(k/2\pi)\int_M d\lambda \wedge F_\Phi = (k/2\pi)\int_{\partial M} \lambda \wedge F_\Phi$ under the 1-form gauge transformation , which is problematic. The way to get around this is to include a $-(k/2\pi)\int_{\partial M} B \wedge \Phi$ boundary term in the action. Together with the boundary term, the full part of the action involving Φ is $-(k/2\pi)\int_M F_B \wedge \Phi$, which is manifestly invariant under the 1-form gauge transformation . Recapitulating, the full action is

$$S = \frac{k}{4\pi} \int_{\partial M} (A \wedge F_A - 2A \wedge B - 2B \wedge \Phi) + \frac{k}{4\pi} \int_M (B \wedge B + 2B \wedge F_\Phi). \tag{283}$$

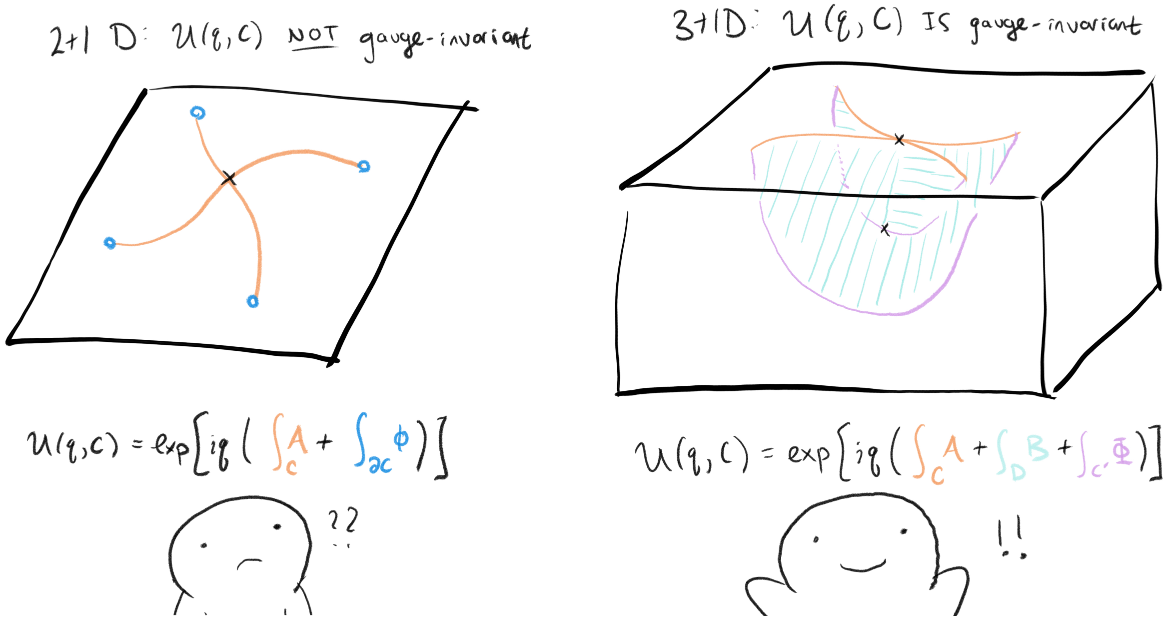


Figure 1: The generators of gauge transformations for the 1-form gauge symmetry in gauged $U(1)_k$. Contact terms that contribute to the commutator of the charge operators are marked with black x's. In 2+1D the $U(q, C)$ are charged and the symmetry can't be gauged, while in 3+1D, with B surfaces extending into the bulk, it can be.

The full generator of gauge transformations is now

$$U(q, C) = \exp\left(iq\left[\int_C A + \int_D B + \int_{C'} \Phi\right]\right), \tag{284}$$

where D is a disk with $\partial D = C \cup C'$, $\partial C = \partial C'$, and where C' is entirely contained within the four-dimensional bulk. The $U(q, C)$'s commute with one another: the contact term between the A's on the surface is canceled between a contact term where the $\int_{C'} \Phi$ line intersects the $\int_D B$ surface. This is illustrated in Figure 1. On the left we show the $U(q, C)$ operators in the strictly 2+1D theory, which are not gauge invariant. On the right we show how, after attaching B surfaces and Φ lines to them, they become gauge invariant.

Twisted $\mathbb{Z}_N$ gauge theory

We now look at the DW theory which we will call $DW_{p,q}$, namely

$$\mathcal{L} = \frac{p}{4\pi} a \wedge da + \frac{q}{2\pi} a \wedge db. \tag{285}$$

What are the global symmetries? First, there is clearly a $\mathbb{Z}_N^{(1)}$ symmetry that shifts b by $1/N$ times a large gauge transformation. Similarly, there is also a $l \equiv \gcd(p, q)$ symmetry shifting a: this is the best we can do, as e.g. the coupling between a and b means we don't have the full $\mathbb{Z}_p^{(1)}$ symmetry of the first term, unless p divides q, and we don't have the full $\mathbb{Z}_q^{(1)}$ symmetry from shifts on a in the second term, unless q divides p.

`<diversion>`

Let's pause for a momentum to discuss the spin and statistics of the lines in this theory. A naive reading on this would be as follows: the canonical momentum for a is $pa + qb$ and the canonical momentum for b is qa. Thus b lines commute with each other, while a lines have a self-linking phase determined by $1/p$. The mutual statistics of a and b is nonzero because they do not commute with each other, and is determined by $1/q$.

This naive reading is incorrect: even if the canonical momentum for a field ϕ as read off from $\partial\mathcal{L}/(\partial\partial\phi)$ does not involve ϕ itself, ϕ may still have nontrivial statistical interactions with itself. Indeed, the correct way to determine the commutation relations between Wilson lines is by using the inverse of the K matrix. Let's quickly remind ourselves of why: for $i \in \mathbb{Z}_{\dim K}$ and letting $\star q^\alpha \cdot J^\alpha = \star q_i^\alpha \cdot J_i^\alpha$ be the 2-form Poincare dual to a support of a particular configuration of Wilson lines $\prod_\alpha W_\alpha = \prod_\alpha e^{i\sum_j \oint_{C_\alpha} A_j}$, we have

$$\langle \prod_\alpha W_\alpha \rangle = \frac{1}{Z[J=0]} \int \prod_i \mathcal{D}A_i \exp\left(\frac{i}{4\pi}\int A_i[K]_{ij} \wedge dA_j + i\sum_{\alpha,i} q_i^\alpha \int A_i \wedge \star J_i^\alpha\right). \quad (286)$$

Shifting A to kill off the AJ coupling, we get

$$\langle \prod_\alpha W_\alpha \rangle = \exp\left(2\pi i \sum_{\alpha,\beta} q_i^\alpha q_j^\beta \int \star J_i^\alpha [K^{-1}]_{ij} \wedge d^{-1} \star J_j^\beta\right). \quad (287)$$

Taking all the Wilson loops to be supported on the boundaries of disks means that the $\star J^\alpha$ are not in $\ker(d)$, and so the above formula makes sense. Anyway, taking two linked loops, one with a unit charge for A_i and another with a unit charge for A_j (and taking the framing of each loop to be trivialized so that the diagonal in α terms in the above formula do not contribute) gives us the braiding matrix

$$[S]_{ij} = \exp\left(2\pi i [K^{-1}]_{ij}\right). \quad (288)$$

This can also be derived just by looking at $[A_i, \pi_{A_m}] = \sum_j [A_i, \bar{K}_{mj} A_j] = i\delta_{im}$. Here spacetime indices are kept implicitly, with $[A_i, A_j] = A_i \wedge A_j - A_j \wedge A_i$. Also, $\bar{K} = K/2\pi$. Anyway, multiplying by $[\bar{K}^{-1}]_{mk}$ and summing over m:

$$\sum_j [A_i, A_j]\delta_{k,j} = i \sum_m \delta_{im}[\bar{K}^{-1}]_{mk} \implies [A_i, A_j] = i[\bar{K}^{-1}]_{ij}. \quad (289)$$

Using this commutation relation to unlink any loops that are linked together in $\prod_\alpha W_\alpha$, one recovers the above expression for the S matrix (after choosing a framing).

In the present $DW_{p,q}$ example, the K matrix and its inverse are

$$K = \begin{pmatrix} p & q \\ q & 0 \end{pmatrix}, \qquad K^{-1} = \begin{pmatrix} 0 & 1/q \\ 1/q & -p/q^2 \end{pmatrix}. \quad (290)$$

Thus even though a appears in the canonical momentum for b, we see that b still fails to commute with itself. So we see that the b line is *not* a boson, despite the fact that its canonical momentum does not involve itself. In fact, it has spin $-p/2q^2$! And similarly, despite the self-CS term for a, we see that a is actually a boson! Physically, what's going on here is that b lines carry flux for a, which by the self-CS term for a have nontrivial braiding with themselves, since this term tells us that a flux also carries a charge. This allows b lines to not commute with themselves. Likewise, a lines carry a flux, which makes them seem like they would not commute with themselves. But a fluxes also carry b charge, and b charge carries a flux, and this all works out in such away that the a lines actually carry net zero a flux.

A particularly transparent example of when this happens is the case when $q = p$. In that case, we can diagonalize the K matrix by something in $SL(2,\mathbb{Z})$ via

$$K \mapsto \Lambda^T K \Lambda = qZ, \qquad \Lambda = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{291}$$

This means that in terms of the variables b and $c = a + b$, the Lagrangian is that of $U(1)_q \otimes U(1)_{-q}$. In this formulation, it is clear that b has spin $-1/(2q)$ (from the $U(1)_{-q}$ factor), while a has spin 0 mod 1, since $a+b$ has spin $+1/(2q)$ and

$$e^{2\pi i s(a)} = e^{2\pi i s(c-b)} \sqrt{[S]_{c-b,c-b}} = [S]_{c,-b} \sqrt{[S]_{c,c}[S]_{-b,-b}} = 1 \cdot \sqrt{e^{2\pi i/q} e^{-2\pi i/q}} = 1 \implies s(a) =_1 0. \tag{292}$$

</diversion>

Now let's continue to look at the symmetries of the theory. The $\mathbb{Z}_q^{(1)}$ symmetry which shifts b by $d\phi/q$ is easy to identify: it is generated by the operator

$$\mathbb{Z}_q^{(1)} = \langle e^{i \oint a} \rangle, \tag{293}$$

which has the right qth root of unity phase linking with the $e^{i\oint b}$ line needed to generate the symmetry. Since $e^{i\oint a}$ has trivial self-linking, this symmetry is not anomalous.

Now for the $\mathbb{Z}_l^{(1)}$ symmetry that shifts a lines by $e^{2\pi i/l}$ (recall $l \equiv \gcd(p,q)$). Since a has trivial self-linking, the operator generating this symmetry should include $\exp(iq/l \oint b)$, since the linking of a and b lines gives a phase $e^{2\pi i/q}$. But this operator also shifts b lines, which is bad since b lines are neutral under the $\mathbb{Z}_l^{(1)}$ symmetry. If we tack on a line $e^{i\beta \oint a}$ to the symmetry generator, imposing that the generator link trivially with b tells us that

$$-\frac{pq}{q^2 l} + \frac{\beta}{q} = 0 \implies \beta = p/l. \tag{294}$$

This means that the $\mathbb{Z}_l^{(1)}$ symmetry is generated by the line

$$\mathbb{Z}_l^{(1)} = \langle \exp\left(i\frac{q}{l} \oint b + i\frac{p}{l} \oint a \right) \rangle. \tag{295}$$

What is the anomaly of this symmetry? To find out, we need the self-linking phase of the charge operator. This phase determines the anomaly as

$$\text{Anomaly} = \frac{1}{2}\left(-\frac{p}{q}\left(\frac{q}{l}\right)^2 + 2\frac{1}{q}\frac{qp}{l^2} \right) = \frac{p}{2l^2} \quad \text{mod } 1, \tag{296}$$

where the first term is the self-linking of b and the second is the a-b mutual phase (the factor of $1/2$ is because we want the spin of the charge operator. On spin manifolds, we should take this mod $1/2$ and not mod 1). This is indeed an anomaly appropriate for a $\mathbb{Z}_l^{(1)}$ symmetry, since it is a $\mathbb{Z}_l$ effect, in that $l(p/l^2) = p/l \in \frac{1}{2}\mathbb{Z}$ indicates that l copies of the charge operator is either trivial, or a transparent fermion. One special case that shows up often is when $p = -rq$ and the theory has two $\mathbb{Z}_q^{(1)}$ symmetries. In this case, the anomaly of the $\mathbb{Z}_q^{(1)}$ symmetry that shifts a is $-r/q$.

Finally, note that there's a mixed anomaly, of a $\mathbb{Z}_l$ character, between the two symmetries. This is just due to the fact that the generators for the $\mathbb{Z}_q^{(1)}$ and $\mathbb{Z}_l^{(1)}$ symmetries don't commute: the phase between them is $e^{2\pi i/l}$ (which is trivial if we take l copies of either generator, as it should be).

This conclusions can be corroborated by just going in and trying to gauge the symmetry directly. The symmetry that shifts b is clearly non-self-anomalous, since b only appears in the action by way of its field strength and we can just make the replacement $F_b \mapsto F_b - B_b$, where B_b is the background field for the $\mathbb{Z}_q^{(1)}$ symmetry. However, since the generator for the symmetry that shifts b carries charge under the $\mathbb{Z}_l^{(1)}$ symmetry , adding the B_b field will break the $\mathbb{Z}_l^{(1)}$ symmetry. Indeed, after adding the B_b field the action shifts by the following term under $a \mapsto a + \frac{1}{l}d\phi$:

$$\delta S = \frac{q/l}{2\pi}\int d\phi \wedge B_b \in \frac{1}{l}\overline{\mathbb{Z}}. \tag{297}$$

Thus we recover the $\mathbb{Z}_l$ mixed anomaly between the two 1-form symmetries.

Basically because of the self-CS term for a, the $\mathbb{Z}_l^{(1)}$ symmetry shifting a has a self-anomaly. To find the appropriate characterization of the anomaly, we start from the gauge-invariant bulk action (omitting the Lagrange multipliers that make B_a, B_b quantized appropriately for simplicity)

$$\begin{aligned} S &= \frac{p}{4\pi}\int_M (F_a - B_a)\wedge(F_a - B_a) + \frac{q}{2\pi}\int_M (F_a - B_a)\wedge(F_b - B_b) \\ &= S_{\partial M} + S_{bulk}, \end{aligned} \tag{298}$$

where M is some bounding 4-manifold, and

$$\begin{aligned} S_{\partial M} &= S_{DW_{p,q}} - \frac{1}{2\pi}\int_{\partial M}\left[a\wedge(pB_a + qB_b) + qb\wedge B_a\right], \\ S_{bulk} &= \frac{1}{4\pi}\int_{\partial M}\left[pB_a\wedge B_a + 2qB_a\wedge B_b\right]. \end{aligned} \tag{299}$$

The second line in the above equation parametrizes the anomaly. If we consider the dependence on the choice of M by integrating S_{bulk} over a closed 4-manifold, we see that the first term is valued in $p\overline{\mathbb{Z}}/2l^2$ on a non-spin manifold, and $p\overline{\mathbb{Z}}/2l^2$ on a spin manifold, while the second term is valued in $\overline{\mathbb{Z}}/l$. The quantization of the second term confirms the $\mathbb{Z}_l$ nature of the mixed anomaly, while the quantization of the first term confirms our result for the anomaly of the $\mathbb{Z}_l^{(1)}$ symmetry.

$U(N)_{k,q}$

Our conventions will be such that $U(N)_{k,q}$ is defined though

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[\mathcal{A}\wedge d\mathcal{A} - \frac{2i}{3}\mathcal{A}\wedge\mathcal{A}\wedge\mathcal{A}\right] + \frac{q-k}{4\pi N}\mathrm{Tr}[\mathcal{A}]\wedge d\mathrm{Tr}[\mathcal{A}]. \tag{300}$$

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. The reason why the effective $U(1)$ level is qN can be seen by starting with the decomposition

$$U(N)_{k,q} \cong [SU(N)_k \times U(1)_{qN}]/\mathbb{Z}_N, \tag{301}$$

where the quotient identifies the center of $SU(N)$ with the appropriate Nth roots of unity in $U(1)$. Since the quotient here says that we can freely change transition functions in the $U(1)$ bundle to make the cocycle condition fail by Nth roots of unity so long as we change the transition functions in the $SU(N)$ bundle in the opposite way, the $\mathbb{Z}_N$ quotient is equivalent to gauging the diagonal $\mathbb{Z}_N^{(1)}$ symmetry which acts on both $SU(N)$ and $U(1)$ fields; for the $U(1)$ part to be well-defined its level then needs to be in $N\mathbb{Z}$, as indicated above.

At the level of manipulating actions, we start by decomposing the $U(N)$ field as

$$\mathcal{A} = A + \boldsymbol{\mathcal{A}}\mathbf{1}, \tag{302}$$

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—hence $N\boldsymbol{\mathcal{A}}$ is a properly-quantized $U(1)$ field, and $N\int F_{\boldsymbol{\mathcal{A}}} \in \overline{\mathbb{Z}}$. The quotient comes from the correlation of the transition functions between A and $\boldsymbol{\mathcal{A}}$ (more on this when we talk about $SU(N)_k$ in the next subsection). In terms of these fields, we have

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[A\wedge dA - \frac{2i}{3}A\wedge A\wedge A\right] + \frac{qN}{4\pi}\boldsymbol{\mathcal{A}}\wedge d\boldsymbol{\mathcal{A}}, \tag{303}$$

so that qN is indeed the "effective $U(1)$ level" as claimed above. To get this we've used that A is traceless and that

$$\mathrm{Tr}[A\wedge A\wedge\boldsymbol{\mathcal{A}}] = \boldsymbol{\mathcal{A}}\wedge\mathrm{Tr}[A\wedge A] = 0 \tag{304}$$

on account of $\mathrm{Tr}[X\wedge Y] = (-1)^{|X||Y|}\mathrm{Tr}[Y\wedge X]$.

Now the $U(1)$ part started out with a $\mathbb{Z}_{qN}^{(1)}$ symmetry pre-gauging. After we gauge to perform the $\mathbb{Z}_N$ quotient though, the quantization condition on $\boldsymbol{\mathcal{A}}$ is modified, so that only $NF_{\boldsymbol{\mathcal{A}}}$ has periods in $\overline{\mathbb{Z}}$. Now let us shift $\boldsymbol{\mathcal{A}}$ by λ, with $d\lambda = 0$. The action changes by

$$\delta S = \frac{q}{2\pi}\int \lambda\wedge(NF_{\boldsymbol{\mathcal{A}}}). \tag{305}$$

Since $NF_{\boldsymbol{\mathcal{A}}}$ is quantized in $\overline{\mathbb{Z}}$, we see that $\delta S \in \overline{\mathbb{Z}}$ provided that $\lambda = \frac{1}{q}d\phi$. Thus we see that the $U(N)_{k,q}$ theory has a $\mathbb{Z}_q^{(1)}$ symmetry, that acts by shifting $\boldsymbol{\mathcal{A}}$.

Is it anomalous? Yes: the charge operator for the remaining $\mathbb{Z}_q^{(1)}$ symmetry is

$$U(p,C) = e^{iNp\oint_C \boldsymbol{\mathcal{A}}}, \qquad p\in\mathbb{Z}_q, \tag{306}$$

with the factor of N needed to perform the shift correctly, and ensures invariance under the gauged diagonal $\mathbb{Z}_N^{(1)}$ symmetry. Computing the braiding phase of the charge operator with itself, we find a phase of $N^2/(Nq)$ since Nq is the effective $U(1)$ level. Thus the anomaly is measured by $N/q \mod 1$. This means in particular that there is no anomaly if $q = N$ (in order for the the theory to be well-defined $q = N$ means $k \in N\mathbb{Z}$). Note that the anomaly of $U(N)_{k,q}$ is the same as the anomaly of N copies of $U(1)_q$.

$SU(N)_k$

Now we look at $SU(N)_k$ CS theory. For all k, this theory has a $\mathbb{Z}_N$ 1-form symmetry, coming from the center of the gauge group. What is the anomaly of the $\mathbb{Z}_N^{(1)}$ symmetry?

Four-dimensional perspective

The easiest way of figuring this out is probably by using what we know about regular four-dimensional pure YM at various values of θ. We know that $\exp(ik\int \mathcal{L}_{CS}[A]/4\pi)$ is the operator which implements the $\theta \mapsto \theta + 2\pi k$ similarity transformation in $SU(N)$ YM, where θ is 2π-periodic, and so if we know what the $\theta \mapsto \theta + 2\pi k$ shift does in the $PSU(N)$ theory, where the 1-form symmetry has been gauged, we'll be able to say something about the anomaly of the gauged CS theory.

Let us now go partway towards turning the theory into a $PSU(N)$ gauge theory by adding a background $\mathbb{Z}_N$ 2-form field B (we'd get the full $PSU(N)$ theory by path integrating over B). We went over how to do this in last year's diary, but I think the discussion there was a bit confused and long-winded. Here's how it works: we first write the $SU(N)$ theory as

$$\mathcal{L} = \frac{\theta}{8\pi^2}\left(\mathrm{Tr}[F_{\mathsf{A}} \wedge F_{\mathsf{A}}] - \mathrm{Tr}[F_{\mathsf{A}}] \wedge \mathrm{Tr}[F_{\mathsf{A}}]\right) + \frac{1}{2\pi} F_Y \wedge \mathrm{Tr}[\mathsf{A}]. \tag{307}$$

Here Y is a Lagrange multiplier field, and A is a $U(N)$ gauge field.[46]

As in the last subsection, we will find it helpful to decompose A as

$$\mathsf{A} = A + \mathcal{A}\mathbf{1}. \tag{308}$$

Here A is $\mathfrak{su}(N)$-valued and $\mathcal{A}$ is $\mathfrak{u}(1)$-valued. However, A is not a connection on a $SU(N)$ bundle, and $\mathcal{A}$ is not a connection on a $U(1)$ bundle: rather, the transition functions g_A and $g_{\mathcal{A}}$ satisfy

$$\delta g_A \delta g_{\mathcal{A}} = \mathbf{1}, \qquad \delta g_A, \delta g_{\mathcal{A}} \in \mathbf{1}e^{\frac{2\pi i}{N}\mathbb{Z}}. \tag{309}$$

In this description, we have a gauge transformation whereby the transition functions g_A and $g_{\mathcal{A}}$ change by opposite roots of unity. Note that this means that only $NF_{\mathcal{A}}$ is a legit $U(1)$ field strength.

Anyway, let's see why this is equivalent to the $SU(N)$ theory. We just integrate out Y: this sets $F_{\mathcal{A}} = 0$ and the sum over $[F_Y] \in H^2(X;\mathbb{Z})$ tells us that we can set $\mathcal{A} = \frac{1}{N}d\phi$,

[46] The second term in the parenthesis ensures that, because the full term in parenthesis is the second Chern class of a $U(N)$ field, we have $\theta \sim \theta + 2\pi$ identically, without having to first integrate out Y. This is desired because $\theta \sim \theta + 2\pi$ in the $SU(N)$ theory (the $SU(N)_k$ CS theory is not spin; more on this later), while if the second term in parenthesis were absent we might not have such a periodicity.

for $d\phi$ a large gauge transformation. The flatness constraint tells us that we will always have $\delta g_{\mathcal{A}} = \mathbf{1}$ (since a nontrivial $\delta g_{\mathcal{A}}$ would contribute to the 1st Chern class), and hence $\delta g_A = \mathbf{1}$ as well: now both $SU(N)$ and $U(1)$ factors are legitimate bundles. Additionally, such an $\mathcal{A}$ can be completely absorbed into a change of the $g_{\mathcal{A}}$ transition functions by Nth roots of unity (the transition functions change by *constants* on each double-overlap). These transition functions can then be absorbed into the g_A transition functions, and so the $\mathcal{A}$ field completely disappears, leaving us with an $SU(N)$ action, as required.

The theory has a $\mathbb{Z}_N^{(1)}$ symmetry that comes from twisting the transition functions in the $SU(N)$ bundle by Nth roots of unity. In our $U(N)$ formulation, this is equivalent to shifting the $g_{\mathcal{A}}$ by Nth roots of unity, which in turn is equivalent to keeping the $g_{\mathcal{A}}$ transition functions fixed, but making a shift $\mathcal{A} \mapsto \mathcal{A} + \frac{1}{N} d\phi$. To gauge this symmetry then, we should make the replacement $F_{\mathcal{A}} \mapsto F_{\mathcal{A}} - B\mathbf{1}$. In what follows we will take B to be some fixed background field with periods in $2\pi/N$ around all closed 2-cycles. This gives us the Lagrangian

$$\mathcal{L} = \frac{\theta}{8\pi^2} \left(\mathrm{Tr}[(F_{\mathcal{A}} - B\mathbf{1})^{\wedge 2}] - (\mathrm{Tr}[F_{\mathcal{A}} - B\mathbf{1}])^{\wedge 2} \right) + \frac{1}{2\pi} Y \wedge \mathrm{Tr}[F_{\mathcal{A}} - B\mathbf{1}]. \tag{310}$$

Consider integrating out Y. This sets $F_{\mathcal{A}} = B$, which means that $F_{\mathcal{A}}$ becomes quantized in periods of $2\pi/N$. Because of the connection between the transition functions of the $SU(N)$ and $U(1)$ bundles, we then erase $F_{\mathcal{A}} - B\mathbf{1}$ from the action and get

$$\mathcal{L} = \frac{\theta}{8\pi^2} \mathrm{Tr}[F_A \wedge F_A], \qquad w_2(E_{PSU(N)}) = \frac{N}{2\pi} B \in H^2(X; \mathbb{Z}_N), \tag{311}$$

where A is now a connection on a $PSU(N)$ bundle $E_{PSU(N)}$. Thus we see the role of B is to turn the $SU(N)$ connection into a $PSU(N)$ connection, with the topological class of the $PSU(N)$ bundle controlled by the cohomology class of B. When B gets integrated over, we perform a sum over all $PSU(N)$ bundles, and obtain a genuine $PSU(N)$ gauge theory.

In a previous diary entry we saw that the 2π periodicity in θ is lost in the $PSU(N)$ theory, and instead that changing θ by 2π induces a shift in the action given by a counterterm in B. Indeed, we can write $\mathcal{L}$ as

$$\mathcal{L} = \theta c_2(E_{U(N)}) + \frac{\theta}{4\pi} \left(-\mathrm{Tr}[F_{\mathcal{A}} \wedge B\mathbf{1}] + N \mathrm{Tr}[F_{\mathcal{A}}] \wedge B + \frac{N - N^2}{2} B \wedge B \right) + \frac{1}{2\pi} Y \wedge \mathrm{Tr}[F_{\mathcal{A}} - B\mathbf{1}]. \tag{312}$$

The advantage of writing it this way is that the $SU(N)$ part of $F_{\mathcal{A}}$ has completely disappeared into the second Chern class of the $U(N)$ bundle. Now integrating out Y, we have (using $\mathrm{Tr}[F_A \wedge B\mathbf{1}] = 0$)

$$\mathcal{L} = \theta c_2(E_{U(N)}) + \frac{\theta}{4\pi} \left(-N + N^2 + \frac{N - N^2}{2} \right) B \wedge B, \tag{313}$$

where now the $U(N)$ bundle $E_{U(N)}$ is constrained to have first Chern class $c_1(E_{U(N)}) = \frac{N}{2\pi} B \in H^2(X; \mathbb{Z})$. Since the second Chern class is integral, under $\theta \mapsto \theta + 2\pi$ the action shifts as (modulo elements of $2\pi\mathbb{Z}$)

$$S_\theta \mapsto S_\theta + \frac{1}{4\pi}(N^2 - N) \int B \wedge B, \tag{314}$$

This is nontrivial, since $\int B \wedge B \in \mathbb{Z}/N^2$ ($\in 2\mathbb{Z}/N^2$) on generic (spin) closed 4-manifolds, and hence θ is actually not 2π periodic.

Anyway, the point is the following: consider a domain wall where θ jumps by 2π. We know that such a domain wall can be created by inserting $\exp(i\int_X \mathcal{L}_{CS}[A]/4\pi)$ into the path integral, where X is a 3-manifold defining the domain wall. By the above discussion, we know that the action differs on the two sides of the domain wall by a $B \wedge B$ counterterm in the background field. However, integrating $B \wedge B$ over an open submanifold of spacetime is not a gauge-invariant thing to do! Doing a gauge transformation on B produces an anomalous term, consisting of an integral over the codimension-1 submanifold X:

$$\delta S = \frac{i}{4\pi}(N-1)\int_X \mathrm{Tr}[2B \wedge \lambda + \lambda \wedge d\lambda], \tag{315}$$

for $\delta B = d\lambda$ (and we are tacitly writing e.g. B for $\mathbf{1}B$). Since we know that $PSU(N)$ gauge theory in four dimensions is self-consistent, this anomaly must be canceled by an anomaly of the $SU(N)_1$ CS theory.

The anomaly is determined by looking at how the shift in S_θ depends on the bounding 4-manifold. Integrating it over a closed 4-manifold tells us that $e^{i\delta S_\theta} = e^{2\pi l \frac{N-1}{2N}}$ for some $l \in \mathbb{Z}_N$. Thus we can conclude that the CS theory $SU(N)_1$ has anomaly $(N-1)/2N$ mod 1. The anomaly for $SU(N)_k$ must then be $k(N-1)/(2N) \mod 1$, since $SU(N)_k$ is the theory defined by the similarity transform on the codimension-1 slice where the $\delta\theta = 2\pi k$ domain wall happens, and the gauge-non-invariance of the bulk action in the presence of the domain wall is exactly k times the result when the θ angle jumps by 2π. So, the theory has an anomaly given by

$$\text{Anomaly} = \frac{k(N-1)}{2N} \mod 1 \quad (\mathrm{mod}\, 1/2 \text{ if spin}). \tag{316}$$

Here the reduced anomaly for the spin case comes from the fact that the intersection form is then even, which limits the phases that δS_θ in (314) can take when integrated over closed 4-manifolds. Actually, we can do a bit better: if $k \in 2\mathbb{Z}$ then the N^2 part of (315) is trivial on all manifolds, and so we can effectively say that the anomaly is just $-k/N$ if $k \in 2\mathbb{Z}$.

Three-dimensional perspective

Now let's look at this from the three-dimensional perspective directly. One naive way to write the $SU(N)_k$ theory is to write

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[\mathcal{A} \wedge d\mathcal{A} + \frac{2i}{3}\mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A}\right] + \frac{1}{2\pi} y \wedge d\mathrm{Tr}[\mathcal{A}], \tag{317}$$

where y is a Lagrange multiplier that roughly speaking turns the $U(N)$ field $\mathcal{A}$ (a $\mathfrak{u}(N)$-valued form) into an $\mathfrak{su}(N)$-valued 1-form. This is not completely correct, however, since if k is odd this theory is spin, while we know that $SU(N)_k$ is non-spin for any value of k (because the $SU(N)$ instanton number is equal to $2\pi \int \mathrm{c}_2(E)$ where $\mathrm{c}_2(E)$ is the second Chern class, which is integral on all closed manifolds, spin or not).

To fix this, we will add a $U(1)_p$ term using the $U(1)$ field $\mathrm{Tr}[\mathcal{A}]$. Note that we are free to shift the definition of the Lagrange multiplier field by

$$y \mapsto y \pm \mathrm{Tr}[\mathcal{A}] \tag{318}$$

(since $\mathrm{Tr}[\mathcal{A}]$ is a properly quantized $U(1)$ field), which changes p by ± 2. So, to find out how to render the theory non-spin, we just need to find out the correct parity to use for p.

Anyway, to get the answer for the correct non-spin theory, we write the full Lagrangian as

$$\mathcal{L} = \frac{k}{4\pi}\mathrm{Tr}\left[\mathcal{A}\wedge d\mathcal{A} - \frac{2i}{3}\mathcal{A}\wedge\mathcal{A}\wedge\mathcal{A}\right] + \frac{\eta_k}{4\pi}\mathrm{Tr}[\mathcal{A}]\wedge d\mathrm{Tr}[\mathcal{A}] + \frac{1}{2\pi}y\wedge d\mathrm{Tr}[\mathcal{A}], \tag{319}$$

where η_k is to be determined. The integral needing to be done to check the quantization condition on η_k is

$$I = \frac{2\pi}{8\pi^2}\int_{M_4}\left(k\mathrm{Tr}[F_{\mathcal{A}}\wedge F_{\mathcal{A}}] + \eta_k \mathrm{Tr}[F_{\mathcal{A}}]\wedge\mathrm{Tr}[F_{\mathcal{A}}] + 2dy\wedge d\mathrm{Tr}[\mathcal{A}]\right) \tag{320}$$

for some closed 4-manifold M_4. The last term is always in $\overline{\mathbb{Z}}$, while the first two can be written as

$$I = 2\pi\int c_2(E_{U(N)}) + \pi(k+\eta_k)\int \mathrm{Tr}[F_{\mathcal{A}}/2\pi]\wedge\mathrm{Tr}[F_{\mathcal{A}}/2\pi], \tag{321}$$

where $\mathrm{c}_2(E_{U(N)})$ is the second Chern class of the $U(N)$ bundle. Since this is always an integral class regardless of the base space of the bundle, we conclude that we need $k+\eta_k$ to be even. Thus we can take e.g. $\eta_k = 0$ if $k \in 2\mathbb{Z}$, and $\eta_k = -1$ if $k \in 2\mathbb{Z}+1$. Another (simpler) choice (and the one we will adopt) is to simply set $\eta_k = -k$, which as we mentioned above is equivalent since η_k and $\eta_k \pm 2$ define equivalent theories. Adopting this choice, we have

$$\mathcal{L}_{SU(N)_k}[A] = \mathcal{L}_{U(N)_{k,k(1-N)}}[\mathcal{A}] + \frac{1}{2\pi}y\wedge d\mathrm{Tr}[\mathcal{A}]. \tag{322}$$

Thus $SU(N)_k$ is realized as a constrained version of a $U(N)_{k,q}$ theory at $q = k(1-N)$. The freedom to shift y by $\pm\mathrm{Tr}[\mathcal{A}]$ manifests itself in the equivalence $q \sim 2N$.

As we saw previously, we can split up $\mathcal{A}$ into an $SU(N)$ part A and a diagonal part $\boldsymbol{\mathcal{A}}$, provided that the cocycle conditions for the A and $\boldsymbol{\mathcal{A}}$ parts fail in canceling ways. Recall the decomposition $\mathcal{A} = A + \boldsymbol{\mathcal{A}}\mathbf{1}$. With this decomposition, in order to implement the matching-cocycle-conditions property, we require that the diagonal transformation shifting the transition functions for both A and $\boldsymbol{\mathcal{A}}$ by opposite Nth roots of unity be a gauge transformation. Note that we can do such a shift while keeping A traceless, since we are only changing the transition functions by constants: the change in transition functions is done at the level of the glueing data between patches, not at the level of the 1-forms A defined on single patches. By contrast, when we perform such a shift on $\boldsymbol{\mathcal{A}}$, we will do it by directly taking $\delta\boldsymbol{\mathcal{A}} = \frac{1}{N}d\phi$ (ϕ as usual is 2π-periodic), without changing the transition functions for the $\boldsymbol{\mathcal{A}}$ bundle. Either way we do it, the effect of this identification is to gauge a diagonal $\mathbb{Z}_N^{(1)}$ symmetry that shifts both A and $\boldsymbol{\mathcal{A}}$. The transformation acts nontrivially on A Wilson lines since they are defined by $\mathrm{Tr}[e^{i\int_{U_\alpha}A}e^{i\Lambda_{\alpha\beta}}e^{i\int_{U_\beta}A}\dots]$, with $\Lambda_{\alpha\beta}$ the transition functions between patches, and since the transformation shifts the $\Lambda_{\alpha\beta}$'s. Note that this gauge transformation, while not changing the field strength $F_{\mathcal{A}}$, *does* change the field strengths of A and $\mathcal{A}$: if we make the cocycle condition fail by an Nth root of unity on a given triple overlap of patches, then this induces fractional flux in both A and $\mathcal{A}$.

Now we can get a more precise understanding of what the Lagrange multiplier y is doing. Integrating out y tells us that $d\mathcal{A} = 0$, and that $\int \mathcal{A} \in \frac{1}{N}\overline{\mathbb{Z}}$ around all closed 1-manifolds. Thus we may write $\mathcal{A} = \frac{1}{N}d\phi$. But we see that this is gauge-equivalent to $\mathcal{A} = 0$ under the 1-form gauge symmetry. So, integrating out y leaves us with just the $SU(N)_k$ part of the action, which is what we want.

Anyway, returning to $\mathcal{L}$, we have

$$\mathcal{L} = \frac{k}{4\pi}\left(\mathrm{Tr}\left[A\wedge dA - \frac{2i}{3}A\wedge A\wedge A\right] + (N-N^2)\mathcal{A}\wedge d\mathcal{A}\right) + \frac{N}{2\pi}y\wedge d\mathcal{A}, \tag{323}$$

again using the tracelessenss of A and the antisymmetry to kill the $A\wedge A\wedge\mathcal{A}$ contribution. Note the $k(N-N^2)$ level of the $\mathcal{A}$ CS term: peeking back at the analysis of the bulk gauge theory, we see that this is exactly the right number needed to cancel the bulk anomaly, and is a hint that we're on the right track.

Let's pause to figure out what the symmetry is. We started with a pure $SU(N)_k$ CS term, which as we know has a $\mathbb{Z}_N^{(1)}$ symmetry. We then wrote it in terms of a $U(N)_{k,q}$ theory plus a Lagrange multiplier, where for us we chose $q = k(1-N)$. As we saw earlier, the $U(N)_{k,q}$ theory by itself has a $\mathbb{Z}_q^{(1)}$ global symmetry. This symmetry is generically broken by the Lagrange multiplier term, since under it we have

$$\delta S = \frac{N}{2\pi k(1-N)}\int F_y \wedge d\phi \notin \overline{\mathbb{Z}}. \tag{324}$$

So, does this mean that we have no 1-form symmetry? This would be a problem if so. But actually, the $\mathbb{Z}_N^{(1)}$ symmetry that we need to be there does exist. To see how it works, consider shifting $\mathcal{A}$ by some flat 1-form λ. The action changes as

$$\delta S = \frac{k(1-N)}{2\pi}\int \lambda\wedge(NF_{\mathcal{A}}) + \frac{N}{2\pi}\int F_y\wedge\lambda. \tag{325}$$

In order for the last term to be in $\overline{\mathbb{Z}}$, we see that we need to take $\lambda = d\phi/N$. Then the variation in S is

$$\delta S = \frac{k(1-N)/N}{2\pi}\int \lambda\wedge(NF_{\mathcal{A}}). \tag{326}$$

This is in general nontrivial, but we see that we can cancel it, if we take the symmetry transformation to involve an appropriate shift in y as well. This gives us a genuine $\mathbb{Z}_N^{(1)}$ symmetry, under which we have

$$\mathbb{Z}_N^{(1)} : \mathcal{A} \mapsto \mathcal{A} + \frac{1}{N}d\phi, \qquad y \mapsto y - \frac{k(1-N)}{N}d\phi. \tag{327}$$

If this is the right symmetry, it should shift fundamental Wilson lines by Nth roots of unity. And indeed it does:

$$\mathbb{Z}_N^{(1)} : W_f(C) = \mathrm{Tr}_f[e^{i\oint_C (A+\mathcal{A}\mathbf{1})}] \mapsto e^{i\frac{1}{N}\oint_C d\phi}W_f(C) \tag{328}$$

(note that $e^{i\oint \mathcal{A}}$ is not a gauge-invariant operator to consider the transformation properties of). Note that $e^{i\oint y}$ also shifts under the symmetry, so that it must also be electrically charged. More on this in a bit.

What is the operator which generates this symmetry? It turns out to be $\exp(i\oint y)$. This is rather surprising, since looking at the action one might be forgiven for thinking that the y line was bosonic.

To find the statistics of the y line, it is helpful to Higgs the theory down to $\mathbb{Z}_N$. In terms of the $SU(N)$ variables, the effect of the Higgsing is to leave the theory with only $\mathbb{Z}_N$ transition functions as degrees of freedom. In the continuum, it's easier to deal with this condition by writing the transition functions instead as diagonal $\mathbb{Z}_N$ 1-form matrices, with trivial transition functions. So to that end, Higgsing for us at the computational level means taking $A = 0$ and $\mathcal{A} = a$, with a a $\mathbb{Z}_N$ field. Since $A = 0$ and A has trivial transition functions, the cocycle condition will be satisfied exactly for a, and the flux of F_a will be quantized in the regular way. So, upon doing this, we get the $DW_{p,q}$ theory with $p = kN(1-N)$, $q = N$:

$$SU(N)_k \xrightarrow{Higgs} \frac{kN(1-N)}{4\pi} ada + \frac{N}{2\pi} y \wedge da. \tag{329}$$

Note that in addition to the $\mathbb{Z}^{(1)}_{\gcd(kN(1-N),N)} = \mathbb{Z}^{(1)}_N$ symmetry, we also have a symmetry that shifts y by a $\mathbb{Z}_N$ gauge field. The appearance of this magnetic symmetry is expected after we move from $SU(N)$ (which has no t'Hooft line operators since $\pi_1(SU(N)) = 0$) to $\mathbb{Z}_N$ (which does have magnetic operators since we can have $\mathbb{Z}_N$ branch cuts in the transition functions).

We've already been through this theory in lots of detail, and we learned that the mutual statistics between the a and y lines are[47]

$$[S]_{a,a} = 1, \quad [S]_{a,y} = e^{2\pi i/N}, \quad [S]_{y,y} = e^{-2\pi i k(1-N)/N}. \tag{330}$$

Recall from a ways back that we could perform a change of variables on y that shifted $k(1-N) \mapsto k(1-N) \pm 2N$. We see that this leaves the braiding phases invariant (and because of the factor of 2, it also leaves the spins invariant), and so reassuringly the shift indeed acts trivially on the modular data of the theory.

From the above entries of the S matrix, we see that the line $e^{i\oint y}$ generates the $\mathbb{Z}^{(1)}_N$ symmetry of $SU(N)_k$, since these braiding phases mean that wrapping lines with the line $e^{i\oint_C y}$ is equivalent to performing the shift (327) (where $d\phi$ is determined by the topology of C).

We can now easily figure out the anomaly: from taking the square root of $[S]_{y,y}$ to get the spin of the generating line, we read off the anomaly as $k(1-N)/2N \mod 1$. If we are on a spin manifold then having the generating line be a fermion is okay, and so in that case the anomaly is $k(1-N)/2N \mod 1/2$. Note that this is exactly the right anomaly to cancel the bulk anomaly that we derived earlier in (316)! Nice. Note that the anomaly of $SU(N)_k$ is the same as that of $[SU(N)_1]^{\otimes k}$, because of the constant k prefactor. Also note that since $N-1$ is coprime to N, we will only have a non-anomalous theory if $k \in 2N\mathbb{Z}$ (or $k \in N\mathbb{Z}$ if spin).

[47]The N/N factor in the $[S]_{y,y}$ matrix element is important, since when square-rooted it contributes to the spin of the y line. However, it does not affect how y lines transform under the $\mathbb{Z}^{(1)}_N$ symmetry.

Can we say anything about this line in the $SU(N)$ context? Yes: under the $\mathbb{Z}_N^{(1)}$ symmetry we have

$$e^{i\oint y} \mapsto e^{2\pi i k/N} e^{i\oint y}. \tag{331}$$

Since Wilson lines in the fundamental transform with a $e^{2\pi i/N}$ phase, this tells us that the generator $e^{i\oint y}$ can be identified with a Wilson line in a k index symmetric $SU(N)$ representation. This makes sense because, as noted in [**?**], $e^{i\oint y}$ is the operator we get when slicing open the 2-dimensional surface operator which implements the $\mathbb{Z}_N^{(1)}$ symmetry in the 3+1 D theory. Now the $SU(N)_k$ theory lives at an interface where the bulk θ angle changes by $2\pi k$. The Witten effect means that the t'Hooft operators on both sides of the surface (which are not genuine line operators) have electric charges differing by k. This k difference in electric charges is realized by the fact that the charge operator on the interface, namely $e^{i\oint y}$, carries electric charge k.

This is a manifestation of the mixed anomaly between the $\mathbb{Z}_N^{(1)}$ symmetry and time reversal at $\theta \in \pi(2\mathbb{Z}+1)$. Indeed, consider a 2π domain wall for θ, where θ jumps from $-\pi$ to π. The operator which inserts this domain wall is the charge operator for T, since it interpolates between the two ground states (which differ by $\theta \mapsto -\theta$). The mixed anomaly comes from the fact that this domain wall operator and the surface operator which implements the $\mathbb{Z}_N^{(1)}$ symmetry don't commute: indeed, they do not commute because of a contact term, and their lack of commutativity can be seen from the fact that along their intersection is a fundamental Wilson line (since we are in four dimensions, a 3-manifold and a 2-manifold intersect at a 1-manifold).

As we have seen, if we try to gauge the $\mathbb{Z}_N^{(1)}$ symmetry in the 2+1D theory, we run in to problems since the operators which perform the gauge transformations (the fundamental Wilson lines) do not commute with each other. This can be fixed by using the same procedure as in the $U(1)_k$ case. First, we write the action for the theory as

$$\begin{aligned} S = \frac{k}{4\pi}\int_{\partial X} \mathrm{Tr}\left[\mathcal{A}\wedge d\mathcal{A} + \frac{2i}{3}\mathcal{A}^3\right] - \frac{k}{4\pi}\int_{\partial X}\mathrm{Tr}[\mathcal{A}]\wedge \mathrm{Tr}[F_{\mathcal{A}}] + \frac{k(N-1)}{2\pi}\int_{\partial X} B\wedge \mathrm{Tr}[\mathcal{A}] \\ + \frac{1}{2\pi}\int_{\partial X}(y\wedge \mathrm{Tr}[F_{\mathcal{A}} - B\mathbf{1}] + N\Phi\wedge B) - \frac{N}{2\pi}\int_X F_\Phi \wedge B + \frac{k(N-N^2)}{4\pi}\int B\wedge B. \end{aligned} \tag{332}$$

To get this, we just took the four-dimensional gauged action (with Φ a 1-form Lagrange multiplier to make B a $\mathbb{Z}_N$ field), and integrated by parts. The extra $N\Phi\wedge B$ term is needed to make things gauge invariant, as we will see.

The transformation rules for the fields are as follows. First, we have a gauge transformation under which $\delta y = \delta\Phi = d\alpha$. Next, we have a $\mathbb{Z}_N^{(1)}$ gauge transformation, generated by $e^{i\oint_C y}$, which shifts

$$\mathcal{A}\mapsto \mathcal{A} + \frac{2\pi}{N}\widehat{C}, \quad B\mapsto B + \frac{2\pi}{N}d\widehat{C}, \quad y\mapsto y + \frac{2\pi k}{N}\widehat{C}, \quad \Phi\mapsto \Phi + \frac{2\pi k}{N}\widehat{C}. \tag{333}$$

Here $\widehat{C}$ is the Poincare to a possibly open curve in ∂X, with the Poincare dual having some arbitrary extension into the bulk X. Here $\widehat{C}$ is such that $\int_{C'}\widehat{C}\in\mathbb{Z}$ for all $C'\in C_1(\partial X;\mathbb{Z})$, but where the value for the integral may depend on more than just the homotopy class of

C'. One can check that the action is invariant up to the term $-\frac{k}{2\pi}\int \widehat{C}\wedge F_B$, which is in $\overline{\mathbb{Z}}$ because of the quantization on F_B. Thus, the whole action is gauge-invariant.

Anyway, these transformation laws let us write down the correct, gauge invariant, generator of gauge transformations for the gauged $\mathbb{Z}_N^{(1)}$ symmetry. It is

$$U(q,C) = \exp\left(iq\left[\int_C y + \int_{C'} \Phi + k\int_D B\right]\right). \tag{334}$$

Here $C \cup C' = \partial D$, with C' only in the bulk; see Figure 1 for a similar setup. Do these operators commute with each other? Yes! $U(q,C)$ and $U(p,\widetilde{C})$, with C, $\widetilde{C}$ two intersecting curves in ∂X will have a contribution to their commutator of the form $e^{2\pi i qpk/N}$, which comes from the commutator of the two y lines. However, they will also have a compensating contribution from the commutator between the $\int_{\widetilde{C}'}\Phi$ line and the $k\int_D B$ surface (which intersect in the bulk), since Φ and B have a braiding phase of $e^{2\pi i/N}$. Thus the $U(q,C)$ are indeed legit generators of the $\mathbb{Z}_N^{(1)}$ gauge transformations.

Summary

Since this diary entry has kind of exploded, let's make a summary table. The theories that we've looked at are

$$\begin{aligned}
&U(1)_k : \frac{k}{4\pi} A\wedge dA \\
&DW_{p,q} : \frac{p}{4\pi} a\wedge da + \frac{q}{2\pi} a\wedge db \\
&U(N)_{k,q} : \frac{k}{4\pi}\mathrm{Tr}[\mathcal{A}\wedge d\mathcal{A} + 2i/3\mathcal{A}^3] + \frac{q-k}{4\pi N}\mathrm{Tr}[\mathcal{A}]\wedge d\mathrm{Tr}[\mathcal{A}] \qquad (k-q)\in N\mathbb{Z} \\
&SU(N)_k : U(N)_{k,k(1-N)} + \frac{1}{2\pi} y\wedge d\mathrm{Tr}[\mathcal{A}].
\end{aligned} \tag{335}$$

The symmetries and anomalies (on general, non spin manifolds, provided that the theory is not spin) are

	1-form symmetry	Anomaly (mod 1)	Spin?
$U(1)_k$	$\mathbb{Z}_k$	$1/k$	if $k\in 2\mathbb{Z}+1$
$DW_{p,q}$	$\mathbb{Z}_q$ on b, $\mathbb{Z}_{\gcd(p,q)}$ on a	0,$p/\gcd(p,q)$,$1/\gcd(p,q)$ (mixed)	if $p\in 2\mathbb{Z}+1$
$U(N)_{k,q}$,	$\mathbb{Z}_q$	N/q	if $k+(q-k)/N\in 2\mathbb{Z}+1$
$SU(N)_k$	$\mathbb{Z}_N$	$(k-Nk)/2N$	No

(336)

Here the anomaly is determined by taking the mod 1 residue of the entry in the third column. In the last column we have indicated when the theories are spin, which will be determined in the subsequent diary entry.

One interesting thing is to check how this is compatible with known level-rank dualities. For example, consider the duality $U(1)_N \leftrightarrow SU(N)_1$ (it is usually $U(1)_{-N}$, but in these conventions the anomalies are such that we write it as $U(1)_N$). This duality hold holds as spin TQFTs. Indeed, while they have the same $\mathbb{Z}_N^{(1)}$ symmetry, let's compare their anomalies: for $U(1)_{-N}$ we have $1/N$, while for $SU(N)_1$ we have $(1-N)/2N$. These are of course not the same. But, on a spin manifold, the anomaly of $SU(N)_k$ is actually $(k-Nk)/N \mod 1$ since the generator of the $\mathbb{Z}_N^{(1)}$ symmetry is allowed to be a fermion. Setting $k=1$ the anomaly becomes $1/N$, which matches that of the $U(1)_N$ theory.

20 *The chiral anomaly in 2 dimensions and the fermion bubble* ✓

Today is a very basic calculation that can likely be found in a nonzero number of QFT texbooks but which was a good brainwarmer and which I figured couldn't hurt to keep around for posterity. We will be considering massless fermions coupled to a $U(1)$ gauge field in 1+1D and computing the two-dimensional analogue of the triangle diagram in order to derive the divergence of the chiral current.

The relevant diagram to compute when examining $\partial_\mu j_5^\mu$ is a fermion bubble with one insertion of j_5 and one gauge field leg. Thus to get $\partial_\mu j_5^\mu$ we need to compute the polarization bubble $\Pi_{\mu\nu}(q^2)$ with a $\bar{\gamma}$ inserted in the trace, and then contract it with a q_μ and a A_ν. For external momentum q, and making sure to order the matrices in the numerator correctly, the graph gives (the is from the propagators kill the minus sign from the fermion loop)

$$\text{(bubble graph)}(q^2) = e^2 \int_p \text{Tr}\left[\gamma^\mu \bar{\gamma} \frac{i\not{p}}{p^2} \gamma^\nu \frac{\not{p}+\not{q}}{(p+q)^2}\right]. \tag{337}$$

Now use

$$\not{p}\gamma^\nu = 2p^\nu - \gamma^\nu \not{p}, \qquad \gamma^\mu \bar{\gamma} = \epsilon_{\mu\nu}\gamma^\nu, \tag{338}$$

where we are in $\mathbb{R}$ time and have chosen X, iY as γ matrices. We also need

$$\text{Tr}[\gamma^\mu\gamma^\nu] = 2g^{\mu\nu}, \qquad \text{Tr}[\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma] = 2(g^{\mu\nu}g^{\lambda\sigma} - g^{\mu\lambda}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\lambda}) \tag{339}$$

to simplify the trace. One gets

$$\text{(bubble graph)}(q^2) = 2e^2 \int_p \frac{p^\nu(p+q)^\mu + p^\mu(p+q)^\nu - g^{\mu\nu}p^\sigma(p+q)_\sigma}{p^2(p+q)^2}, \tag{340}$$

which we evaluate with Feynman parameters in the usual way. We then use dimensional regularization to do the integrals and renormalize with minimal subtraction, yielding

$$i\Pi^{\mu\nu}(q^2) = \frac{ie^2}{\pi}\left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right), \tag{341}$$

which is in the form required by gauge invariance. The divergence of the axial current to one-loop order is then

$$\begin{aligned} \partial_\mu j_5^\mu \to q_\mu j^\mu &= q_\mu \epsilon^{\mu\nu}\frac{e^2}{\pi}\left(g_{\nu\lambda} - \frac{q_\nu q_\lambda}{q^2}\right)A^\lambda \\ &= \frac{e^2}{\pi} q_\mu \epsilon^{\mu\lambda} A_\lambda. \end{aligned} \tag{342}$$

In $\mathbb{R}$ space then,

$$\partial_\mu j_5^\mu = \frac{e^2}{2\pi} F_{\mu\nu}\epsilon^{\mu\nu} = 2e^2 \mathrm{c}_1, \tag{343}$$

where c_1 is the first Chern class. The fact that the divergence of the axial current is *twice* the first Chern class is essential, since without the factor of 2 we'd conclude that the A_μ background field could lead to a non-conservation of fermion number (as c_1 is an integral class). This factor of two comes from the fact that each charge-1 chiral fermion contributes $e^2\mathrm{c}_1$ to the anomaly, and a single Dirac fermion has two such chiral fermions. As another sanity check, we see that if we didn't have the ϵ symbol (corresponding to computing the divergence of the vector current), we'd get a divergence of $d^\dagger \Box^{-1} d^\dagger dA = \Box^{-1} d^\dagger d^\dagger dA = 0$.

Note that we also could have regulated by imposing a hard cutoff in momentum space. The $q^\mu q^\nu$ term would be unchanged since it has no UV divergences, but the $g^{\mu\nu}q^2$ piece would pick up a $\ln\Lambda$ divergence, which we would regulate away. The form of the resulting Π would then imply that gauge invariance is violated (the divergence of the regular vector current $\bar\psi\gamma^\mu\psi$ is non-zero). Thus, courtesy of the mixed anomaly between $U(1)_A$ and $U(1)_V$, we may gauge one of them, but doing so breaks the other.

21 *Topological defects in axion electrodynamics and anomaly inflow* ✓

Today's diary entry comes from wanting to understand in detail constructions involving axions and anomaly inflow as sketched in e.g. Jeffery Harvey's TASI lectures on anomalies.

We will be considering QED coupled chirally to a scalar with a $U(1)$-breaking potential:

$$\mathcal{L} = -\frac{1}{2}F\wedge\star F + i\bar\psi \not{D}_A\psi - \frac{1}{2}(\partial\sigma)^2 + \bar\psi(\mathrm{Re}(\sigma) + i\bar\gamma\mathrm{Im}(\sigma))\psi - V(\sigma), \tag{344}$$

where $V(\sigma)$ has a classical minimum at $m \neq 0$. When the scalar gets a vev, it induces a mass term for the fermions, which in addition to the regular $m\bar\psi\psi$ mass term also has a chiral part.

We will show that when σ is Higgsed (viz. $\sigma \to me^{i\theta}$), the fermions are responsible for generating a spacetime-dependent theta angle for the gauge field. We will then consider a field configuration with a topological string defect in θ, and show that the action for such a field configuration suffers from a gauge anomaly by computing the divergence of the gauge current (this can be done with a one-loop calculation). By solving the Dirac equation on the string defect, we will show that the defect hosts chiral zero modes whose own gauge anomaly renders the full theory gauge invariant.

The anomaly

Throughout we will assume we are working in the symmetry-broken phase, where σ gets a vev, thereby giving a mass to the fermions. We assume the potential $V(\sigma)$ is such that the classical vacua form an S^1, and thus it makes sense to consider defect codimension-2 objects (strings) around which the phase of σ winds by something in $2\pi\mathbb{Z}$.

One (schematic) argument for why the fermions generate a θ term is as follows: for $\sigma = me^{i\theta}$, we can eliminate the $\bar{\gamma}$ coupling in $\mathcal{L}$ by performing the shift

$$\psi \mapsto e^{-i\theta\bar{\gamma}/2}\psi \tag{345}$$

in the path integral. Then because of the chiral anomaly, the action should shift by

$$S \mapsto S + \frac{1}{8\pi^2}\int \theta F \wedge F, \tag{346}$$

thus generating a (spacetime-dependent) θ angle (the shift is the integral of θ against the second Chern character and not twice the second Chern character since we've rotated by $\theta/2$). This is subtle for defect string configurations though, since θ is not well-defined on its own, which makes doing the usual Fujikawa method kind of tricky. It will turn out that the right answer is the integrated-by-parts version of this, namely

$$\frac{1}{8\pi^2}\int d\theta \wedge F \wedge A. \tag{347}$$

To get this, we compute the divergence of the gauge current in the limit of large m and obtain the above expression. To leading order the divergence in the gauge current is caused by a diagram consisting of a single ψ loop with an external J_μ source, an A_ν gauge field leg, and an external σ leg. To evaluate this diagram and get a non-zero answer, we need to correctly take into account the fermion mass. This is done most easily by choosing a particular symmetry-breaking state to evaluate the diagram in. Let $\sigma = \sigma_1 + i\sigma_2$ with σ_i real. We will take the state where the vev of σ is real, with $i\sigma_2$ the Goldstone. We thus write $\langle\sigma\rangle = \langle\sigma_1\rangle = m$ for the fermion mass, and so the chiral coupling between σ and the fermions gives the fermions a mass and leaves them with a $\bar{\psi}i\bar{\gamma}\sigma\psi$ coupling. In this minimum, the contribution to the gauge current from this diagram on scales much larger than m^{-1} is

$$J^\mu(q+k) = e\int_p \mathrm{Tr}\left[\gamma^\mu \frac{\not{p}+m}{p^2-m^2}\gamma^\nu \frac{\not{p}-\not{q}+m}{(p-q)^2-m^2}(i\bar{\gamma}\sigma_2(k))\frac{\not{p}-\not{q}-\not{k}+m}{(p-q-k)^2-m^2}\right]A_\nu(q). \tag{348}$$

In this expression, q is the momentum of the external photon, and k is the momentum of the axion field (the σ field, but in the chosen symmetry-breaking state only the imaginary part σ_2 is fluctuating).

We can simply this a fair bit by noting that terms with an odd number of γ matrices will vanish after being traced over, since $\mathrm{Tr}(\gamma^\mu) = 0$ for all μ. In order to survive the trace with $\bar{\gamma}$, we need exactly four compensating γ matrices, so only terms linear in m survive in the numerator. We then use (in minkowski, mostly negative signature)

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma\bar{\gamma}] = -4i\epsilon^{\mu\nu\lambda\sigma}. \tag{349}$$

Any numerator odd in the momentum p which is being traced out will vanish by symmetry, and because of the ϵ symbol in the above trace, no terms involving $\not{p}$ will survive. Thus the only term which survives is

$$J^\mu(q+k) = 4me\epsilon^{\mu\nu\lambda\sigma}\int_p \frac{\sigma_2(k)q_\lambda(q_\sigma+k_\sigma)}{(p^2-m^2)((p+q)^2-m^2)((p+q+k)^2-m^2)}A_\nu(q). \tag{350}$$

We then simplify the denominator with Feynman parameters, and shift p to complete the square (which happily doesn't affect the numerator). This gives

$$J^\mu(q+k) = 4me\epsilon^{\mu\nu\lambda\sigma}\int_{x,y}\sigma_2(k)q_\lambda(q_\sigma+k_\sigma)A_\nu(q)\int_p\frac{1}{(p^2-\Delta)^3}, \tag{351}$$

where Δ is a gross function of q, k, and the Feynman parameters x, y. We can then do the integral no problem, and we get

$$J^\mu(q+k) = -\frac{ime\epsilon^{\mu\nu\lambda\sigma}\sigma_2(k)q_\lambda(q_\sigma+k_\sigma)}{8\pi^2}\int_{x,y}\frac{1}{\Delta}. \tag{352}$$

Now we have been assuming that the fermion mass m is much bigger than the momentum scales of interest, namely q and k. With this approximation Δ turns out to simply be m^2. We can also drop the quadratic in q term, since the gauge field is assumed to be smooth so that $[\partial_\mu, \partial_\nu]A = 0$, and so we get

$$J^\mu(q+k) = -\frac{ie\epsilon^{\mu\nu\lambda\sigma}\sigma_2(k)q_\lambda k_\sigma}{4\pi^2 m}A_\nu(q). \tag{353}$$

In real space, we could write this as

$$J^\mu = -\frac{ie\epsilon^{\mu\nu\lambda\sigma}F_{\nu\lambda}\sigma_1\partial_\sigma\sigma_2}{8\pi^2\sigma_1}. \tag{354}$$

Of course, this expression for J^μ is not general since we've chosen a particular vacuum to evaluate the diagram in. Getting the general expression is easy though: we just write down the appropriate expression which reduces to the above answer when $\langle\sigma\rangle$ is real and which is invariant under the chiral rotation $(\sigma_1, \sigma_2)^T \mapsto R_\alpha(\sigma_1, \sigma_2)^T$ (where $R_\alpha \in O(2)$ is a rotation by α under which $\psi \mapsto e^{-i\alpha\bar{\gamma}/2}$). Such a rotation preserves the matrix $iY\partial$, and so we write the general J^μ as

$$J^\mu = \frac{ie\epsilon^{\mu\nu\lambda\sigma}F_{\nu\lambda}}{8\pi^2|\sigma|^2}(\sigma_2\partial_\sigma\sigma_1 - \sigma_1\partial_\sigma\sigma_2), \tag{355}$$

which indeed reduces to our previous expression when we choose the $\langle\sigma\rangle = \langle\sigma_1\rangle$ vacuum. One can also check that if we go back and work about the $\langle\sigma\rangle = i\langle\sigma_2\rangle$ vacuum with an $i\bar{\gamma}\sigma_2$ mass term for the fermions and a $\bar{\psi}\psi\sigma_1$ interaction vertex, then we get the other term in the above equation. Finally, parametrizing $\sigma = me^{i\theta}$, we find that the gauge current is

$$J^\mu = \frac{e}{8\pi^2}\epsilon^{\mu\nu\lambda\sigma}\partial_\nu\theta F_{\lambda\theta}, \tag{356}$$

which ends up being independent of the fermion mass and matches the $d\theta \wedge F \wedge A$ answer we had guessed earlier for the shift in the action, since $J \propto \star(d\theta \wedge F)$. Since $dF = 0$, The divergence in the gauge current is

$$d^\dagger J = \frac{e}{8\pi^2}\star(d^2\theta \wedge F). \tag{357}$$

This is not zero, since θ is not a well-defined function when we consider a string defect configuration, only $d\theta$ is. Since

$$\int_\gamma d\theta = 2\pi \tag{358}$$

for any curve γ which links the defect (assuming the defect has 2π winding),

$$d^2\theta = 2\pi\widehat{s}, \tag{359}$$

where s is a 2-chain parametrizing the string and $\widehat{s}$ is its Poincare dual. This works because

$$\int_\gamma d\theta = \int_D d^2\theta = 2\pi \int_D \widehat{s} = 2\pi \mathrm{int}(D, s) = 2\pi, \tag{360}$$

where D is a disc bounded by γ and $\mathrm{int}(D, s) = 1$ is the intersection number. For example, if s lies along the z-axis in space and doesn't move in time, $d\theta = \frac{1}{r}\hat{\phi}$ in cylindrical coordinates. Thus the divergence in the gauge current

$$d^\dagger J = \frac{e}{4\pi} \star (\widehat{s} \wedge F) \tag{361}$$

is non-zero only on the string defect, and so under gauge transformations $A \mapsto A - d\alpha$, the action shifts as

$$S \mapsto S + \frac{e^2}{4\pi} \int_s F\alpha, \tag{362}$$

where the integral is over the string defect. In order for this theory to make sense, this anomaly needs to be canceled by something living on the string.

Anomaly cancellation

We expect that an anomaly of the above form will be canceled by a gauge anomaly from chiral fermions living on the string, and this is indeed the case. Suppose the string lives along the z axis, and work in cylindrical coordinates. The Dirac equation is

$$(i\not{D}_A + m(r)e^{i\bar{\gamma}\theta})\psi = 0, \tag{363}$$

where $m(r)$ goes to zero at $r = 0$ and goes to the minimum of $V(\sigma)$ at $r \to \infty$. Since $\not{D}_A$ anticommutes with $\bar{\gamma}$, we can write

$$i\not{D}_A\psi_\mp + m(r)e^{\pm\theta}\psi_\pm = 0, \tag{364}$$

where $\psi_\pm$ are eigenspinors of $\bar{\gamma}$. For simplicity, let us take only A_z, A_t to be non-zero (we just want to know the form of the fermion solution for a particular gauge field configuration). This still allows for non-zero field strength on the string (i.e. non-zero E_z), which gives us an F for which the anomaly is non-zero. Our ansatz for ψ_- is

$$\psi_- = \eta f(r), \tag{365}$$

where η is a zero-mode of the Dirac operator restricted to the string:

$$i(\gamma^t \partial_t + \gamma^z \partial_z - ie(A_t + A_z))\eta = 0. \tag{366}$$

Since the string worldsheet is two-dimensional, we can assign the zero-mode η a define parity, and thus it has the potential to contribute to a cancellation of the gauge anomaly (we will determine its parity shortly). Thus we need to solve

$$\eta i \not{D}_A^{\perp} f(r) = -m(r) e^{i\theta} \psi_+, \tag{367}$$

where $\not{D}_A^{\perp}$ is the Dirac operator on the coordinates orthogonal to the string. Solving this is easy since A is zero for these coordinates. Thus we have

$$i(\cos\theta \gamma^x + \sin\theta \gamma^y)\eta \partial_r f(r) = -m(r) e^{i\theta} \psi_+. \tag{368}$$

We can take care of the $m(r)$ factor with an $f(r)$ which is exponentially localized to the string. We write $f(r) = \exp(-\int_0^r dr' m(r'))$, thus

$$i(\cos\theta \gamma^x + \sin\theta \gamma^y)\eta \exp\left(-\int_0^r dr' m(r')\right) = e^{i\theta}\psi_+. \tag{369}$$

We can cleverly re-write this as

$$i\gamma^x e^{i\theta \bar{\gamma}_{\text{ext}}} \eta \exp\left(-\int_0^r dr' m(r')\right) = e^{i\theta}\psi_+, \tag{370}$$

where $\bar{\gamma}_{\text{ext}} = -i\gamma^x\gamma^y$ is the chirality operator on the components orthogonal to the string. To write the Dirac equation in this form, we have taken η to be an eigenspinor of $\bar{\gamma}_{\text{ext}}$, this is possible since $\bar{\gamma}$ and $\bar{\gamma}_{\text{ext}}$ commute. Thus can solve the problem by choosing η to be the positive-chirality eigenstate of $\bar{\gamma}_{\text{ext}}$. Since η came from ψ_- which is a negative chirality eigenstate of $\bar{\gamma}$, the fact that $\bar{\gamma}_{\text{ext}}\eta = +\eta$ implies $\bar{\gamma}_{\text{int}}\eta = -\eta$, so that η is a negative-chirality zero-mode on the string. Since this also determines ψ_+, we have

$$\psi_- = \eta \exp\left(-\int_0^r dr' m(r')\right), \qquad \psi_+ = i\gamma^x \eta \psi_-, \tag{371}$$

where again, $\bar{\gamma}_{\text{int}}\eta = -\eta$. Since γ^x commutes with $\bar{\gamma}_{\text{int}}$, we importantly have that the chirality of the zero mode components of both the ψ_+ and ψ_- is *the same* (while the chirality of the components orthogonal to the string are opposite).

Recapitulating, the presence of the string leads to a pair of zero modes propagating along the string with the same chirality. Since they are chiral with negative chirality, we can write their contribution to the Lagrangian as

$$i\bar{\eta} \not{D}_A^s \frac{(1 - \bar{\gamma}_{\text{int}})}{2} \eta, \tag{372}$$

where $\not{D}_A^s$ is the Dirac operator restricted to the string. Thus we have a coupling between the gauge field and the chiral current, which leads to a gauge anomaly. When we do a gauge transformation, this shifts by the usual gauge anomaly in two dimensions, determined by the index theorem. For $A \mapsto A - d\alpha$, the action for the string zero modes changes by

$$\delta S_s = -\int_s \frac{e^2}{4\pi} F\alpha, \tag{373}$$

which exactly cancels the anomaly in the gauge current caused by the current-fermion-gauge-field diagram we computed earlier! Thus anomaly inflow from the bulk onto the string defect renders the whole theory self-consistent and anomaly-free.

22 *Flavor symmetries of fermions and $Sp(N)$ gauge theories* ✓

In this diary entry we will discuss global flavor symmetries of fermions—both in general terms and in the specific example of fermions coupled to an $SU(2)$ gauge field (inspired by wanting to understand the construction in [**?**]).

First, some notational housekeeping. In the following, we will let

$$J \equiv (-iY) \otimes \mathbf{1}_N \tag{374}$$

be the symplectic form preserved by elements in $Sp(2n;\mathbb{K})$, where $\mathbb{K}$ is some field. The compact subgroup of $Sp(2n;\mathbb{C})$ will be denoted

$$Sp(n) \equiv U(2n) \cap Sp(2n;\mathbb{C}). \tag{375}$$

The question that motivated this entry was the following. Consider

$$\mathcal{L} = \sum_{i=1}^{N} \bar{\psi}_i \not{D}_A \psi_i, \tag{376}$$

where A is the connection on some gauge group (which may be trivial). What is the global internal symmetry group of the above theory? Naively the answer is $U(N)$ (plus a possible $\mathbb{Z}_2^C$ depending on the gauge group), but this is a little bit hasty. To elucidate what the full symmetry group is, break ψ apart in terms of two real fields via

$$\psi_i = \chi_i + i\eta_i, \tag{377}$$

where χ_i and η_i are majoranas, and define

$$\Psi^T \equiv (\chi_1, \chi_2, \ldots, \chi_N, \eta_1, \ldots, \eta_N)^T. \tag{378}$$

Then since the action of $O(2N)$ preserves the commutation relations of the Majorannas and leaves $\mathcal{L}$ invariant, the global flavor symmetry is clearly $O(2N)$. Since $U(N) \subset O(2N)$, the full symmetry group is bigger than the naive $U(N)$.

What sorts of constraints can break the $O(2N)$ down to the naive $U(N)$? As far as mass terms go, the Dirac mass is $\bar{\psi}\psi = \bar{\chi}\chi + \bar{\eta}\eta$, since $\bar{\eta}\chi = \bar{\chi}\eta$. Thus the Dirac mass is invariant under $O(2N)$ and hence also under $U(N)$. The fermion number operator however is $\psi^\dagger\psi = 2 + 2i\chi^T\eta$, which up to a constant is $\Psi^T J\Psi$, and therefore is *not* preserved by the full $O(2N)$. This term is of course preserved by the diagonal $U(1)$, since the action of $U(1)$ is by $\Psi \mapsto U\Psi$, with $U = S \otimes \mathbf{1}_{N\times N}$ and $S \in SO(2)$. Since $-iY \in SO(2)$ and $SO(2)$ is Abelian, we have

$$U^T JU = (S^T \otimes \mathbf{1})(-iY \otimes \mathbf{1})(S \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1})(-iY \otimes \mathbf{1})(S^T S \otimes \mathbf{1}) = J. \tag{379}$$

Now while $\psi^\dagger\psi$ is not preserved by $O(2N)$, it is preserved by the full $U(N)$ (as should be obvious from how it acts on the complex fermions). Moreover, $U(N) \subset O(2N)$ is the maximal subgroup that preserves $\psi^\dagger\psi$. Indeed, preserving $\psi^\dagger\psi$ means preserving $\Psi^T J\Psi$, which means that if $R \in O(2N)$ is to preserve $\Psi^T J\Psi$, we need $R^T JR = J \implies R \in Sp(2N;\mathbb{R})$. Thus the group of transformations that preserve complex fermion number is

$$O(2N) \cap Sp(2N;\mathbb{R}) \cong U(N). \tag{380}$$

Here the last equality is a manifestation of the 2-in-3 property, namely that the intersection

$$O(2N) \cap GL(N;\mathbb{C}) \cap Sp(2N;\mathbb{R}) = U(N), \tag{381}$$

and that actually $U(N)$ is equal to the intersection of any two of the three groups on the LHS. Why is this? Let $V \in Sp(2n;\mathbb{R})$. Then $V^T JV = J$. Alternatively, let $V \in GL(N;\mathbb{C})$. Then when viewed as a $2N \times 2N$ real matrix, in order to have a legit complex structure, we need V to commute with some matrix i, such that $i^2 = -\mathbf{1}$ and $Vi = iV$. If the complex structure and symplectic structure being considered are compatible,[48] then we need to take $i = J$. Finally, if $V \in O(2N)$, then $V^T V = \mathbf{1}$. Thus if $V \in O(2N) \cap GL(N;\mathbb{C}) \cap Sp(2N;\mathbb{R})$, then $V^T = V^{-1}, V^T JV = J$, and $V^{-1}JV = J$. Then we see that any two of these properties implies the third; hence the 2-in-3 property. We can realize the matrices in $U(N)$ in this way by using our knowledge of the Lie algebra of $Sp(N)$, and taking only the real part. So we claim that all the elements in $U(N)$ can be written as

$$O(2N) \cap Sp(2N;\mathbb{R}) = U(N) = \{\exp(\mathbf{1} \otimes A + iY \otimes S) \,:\, A^T = -A,\ B^T = B\}. \tag{382}$$

It's easy to check that the above matrices are orthogonal and preserve J. Have we missed any? No, let's count dimensions: there are $(N^2 - N)/2$ choices for A and $(N^2 + N)/2$ choices for S, and all of these choices give distinct elements in $O(2N) \cap Sp(2N;\mathbb{R})$. This adds up to N^2 total elements, which is the same as the number of generators for $U(N)$. So indeed, all the elements in $U(N)$ can be written as real matrices in this way.

[48] A complex structure on a vector space $\mathcal{W}$ means we can realize it as a direct sum $\mathcal{W} \cong V \oplus V$, where V is real (so e.g. the tangent bundle TM has an [almost] complex structure). A symplectic structure on $\mathcal{W}$ means that we can decompose it as $\mathcal{W} \cong W \oplus W^*$ for W real (this means choosing coordinates and momentum; e.g. T^*M has a symplectic structure with the dimensions coming from M being the coordinates and those coming from the fiber being the momenta), with the symplectic form being given by $\omega(v \oplus f, u \oplus g) = f(u) - g(v)$. The compatibility of the complex and symplectic structures means that we can choose V (known as the real subspace) and W (known as the Lagrangian subspace) to be equal, with the symplectic form corresponding to multiplication by i.

Anyway, enough with that digression. Returning to the fermion problem, we see that $O(2N)$ contains elements which do not preserve $\psi^\dagger\psi$. Thus if we restrict to transformations that preserve the fermion number, we get that the flavor part of the symmetry group is the naive $U(N)$. As example, consider the matrix $Z \otimes \mathbf{1} \in O(2N)$. In the Ψ basis this sends all the χ's to themselves, and it multiplies the η's by minus signs. Thus $Z \otimes \mathbf{1} : \psi_i \mapsto \psi_i^\dagger$, and so $Z \otimes \mathbf{1}$ is charge conjugation. This doesn't preserve $\psi^\dagger\psi$, and indeed, while $Z \otimes \mathbf{1} \in O(2N)$, $Z \otimes \mathbf{1} \notin Sp(2N;\mathbb{R})$ and so $Z \otimes \mathbf{1} \notin U(N)$.

Another way to understand how the $O(2N) \to U(N)$ restriction of the symmetry group can come about is to remember that complex numbers are not simply two copies of $\mathbb{R}$: there is a complex structure that relates the two copies. Consider multiplication by i, $\psi_i \to i\psi_i$. We see that in the Ψ basis, this acts as J. Thus $i = J$ when acting on the Majorana fermions. Now if our flavor symmetry transformation R does not involve complex conjugation, then $Ri\psi = iR\psi$. But when written in terms of Majoranas, this means that $RJ = JR$, and so from the orthogonality of R, we have $R \in Sp(2N;\mathbb{R})$, and thus from the 2-in-3 property we know that $R \in U(N)$ (another way to say this is that $RJ = JR$ is the requirement of the existence of a complex structure, and tells us that $R \in GL(N;\mathbb{C})$. But from the 2-in-3 property, the orthogonality of R then implies that R is in $Sp(2N;\mathbb{R})$ as well. Since R is then both orthogonal and symplectic / complex structure preserving, must have $R \in U(N)$). So if we want R to preserve the complex structure, i.e. for R to not be anti-unitary, then R must be in $U(N)$ (which admittedly sounds kind of tautological).

$SU(2)$ gauge theory with N Dirac fermions

We now try to understand the global symmetry of N Dirac fermions, all coupled to an $SU(2)$ gauge field in the fundamental. This again comes from wanting to understand the construction in [**?**]. As in the previous subsection, the naive guess for what the internal part of the global symmetry should be, namely $U(N)/\mathbb{Z}_2$ (or maybe $[U(N)/\mathbb{Z}_2] \rtimes \mathbb{Z}_2$ for charge conjugation) is not correct. In fact the internal part of the global symmetry is actually $PSp(N)$!

Let's see how this comes about. Let $\psi_i = (\psi_{i\uparrow}, \psi_{i\downarrow})^T$ be one of the Dirac fermions in the fundamental of $SU(2)$. A single fermion in $SU(2)_f$ can be built from four majoranas. We can build it as a matrix field as follows:

$$\mathcal{X}_i = \frac{1}{\sqrt{2}}(\chi_i^1 \mathbf{1} + i\chi_i^a \sigma^a), \tag{383}$$

with $a \in \{x, y, z\}$. We build the constituent complex fermions from the Majoranas so that (dropping the flavor index for simplicity)

$$\mathcal{X} = \frac{1}{\sqrt{2}} \begin{pmatrix} \chi^1 + i\chi^z & i\chi^x + \chi^y \\ i\chi^x - \chi^y & \chi^1 - i\chi^z \end{pmatrix} = \begin{pmatrix} \psi_\uparrow & \psi_\downarrow \\ -\psi_\downarrow^\dagger & \psi_\uparrow^\dagger \end{pmatrix}. \tag{384}$$

With this one can check that $\mathrm{Tr}[\bar{\mathcal{X}} \not{D}_A \mathcal{X}]$ gives the correct Dirac Lagrangian, with the $SU(2)$ gauge field A acting on the right in the covariant derivative. The mass term $\mathrm{Tr}[\bar{\mathcal{X}}\mathcal{X}]$ is $\sum_\alpha \bar{\chi}^\alpha \chi^\alpha$, as expected.

Consider the right action on $\mathcal{X}$ by $SU(2)$. Right multiplication by e.g. $e^{i\alpha Z}$ does

$$\mathcal{X} \mapsto \mathcal{X} e^{i\alpha Z} = \begin{pmatrix} e^{i\alpha}\psi_\uparrow & e^{-i\alpha}\psi_\downarrow \\ -e^{i\alpha}\psi_\downarrow^\dagger & e^{-i\alpha}\psi_\uparrow^\dagger \end{pmatrix}, \tag{385}$$

which is just what a gauge rotation about the z axis in $SU(2)$ should do. So, we see that the $SU(2)$ we want to gauge is the right action on $\mathcal{X}$ by $SU(2)$.

The left action then parametrizes the system's global flavor symmetry. In order for $\bar{\mathcal{X}} \not{D}_A \mathcal{X}$ to be left invariant, the U in $\mathcal{X} \mapsto U\mathcal{X}$ must be unitary, and since there are N flavors of Dirac fermions, $U \in U(2N)$. However, there is an additional restriction. Indeed, consider the fact that

$$\mathcal{X}^\dagger = (Y \otimes \mathbf{1})\mathcal{X}^T (Y \otimes \mathbf{1}). \tag{386}$$

Now take $\mathcal{X} \mapsto U\mathcal{X}$. Then we need

$$\mathcal{X}^\dagger U^\dagger = (Y\otimes\mathbf{1})\mathcal{X}^T U^T (Y\otimes\mathbf{1}) \implies \mathcal{X}^\dagger = (Y\otimes\mathbf{1})\mathcal{X}^T (Y\otimes\mathbf{1})^2 U^T (Y\otimes\mathbf{1})U = \mathcal{X}^\dagger (Y\otimes\mathbf{1}) U^T (Y\otimes\mathbf{1})U. \tag{387}$$

In particular, this means that

$$U^T J U = J \implies U \in Sp(N) = U(2N) \cap Sp(2N; \mathbb{C}). \tag{388}$$

So, the global symmetry on the left action is tentatively identifiable with $Sp(N)$.

Actually this is not completely true, since it may happen that elements of the global symmetry acting from the left act in the same way as elements of the gauge group acting from the left. Clearly this is true for $-\mathbf{1}$, which acts the same both as a $Sp(N)$ element from the left and an $SU(2)$ element from the right. In fact, this is the only common element shared by the two actions. Indeed, consider a given element of $Sp(N)$ acting from the left, and ask if it is equivalent to an element of $SU(2)$ acting from the right. Since the $SU(2)$ acts in the same way on each Dirac fermion, we just need to look for elements of $Sp(N)$ that are diagonal on the flavor index, and so we can restrict ourselves to a single flavor wolog, and take the left action to be that of $Sp(1) = SU(2)$. Then consider the $U(1)$ rotation $e^{i\alpha Z}$ acting from the left. This multiplies both $\psi_\uparrow$ and $\psi_\downarrow$ by the same phase. This can never be done by an element of $SU(2)$ acting on the left: the only element which just multiplies $\psi_\uparrow$ and $\psi_\downarrow$ by phases does so in a gauge-invariant way, namely by multiplying $\psi_\uparrow$ by $e^{i\alpha}$ and $\psi_\downarrow$ by $e^{-i\alpha}$. So the left action by $e^{i\alpha Z}$ is only equivalent by the right action of something in $SU(2)$ if $e^{i\alpha Z} = -\mathbf{1}$. Since every element in the left $SU(2)$ can be written as $e^{i\alpha Z}$ in the right choice of basis, every element of the left $SU(2)$ action (except $-\mathbf{1}$) must also not be expressible as the action of some $SU(2)$ element from the right. Thus only the $-\mathbf{1}$ gets modded out, and the global symmetry acting on the left is in fact $PSp(N) = Sp(N)/\mathbb{Z}_2$.

The fact that the global symmetry is so big is kind of surprising! For example, take $N = 1$: the internal part of the global symmetry is then $PSp(1) = SU(2)/\mathbb{Z}_2 = SO(3)$. If we just looked at the Lagrangian $\bar{\psi}\not{D}_A\psi$, we might have thought that the global internal symmetry was $U(1)$, or maybe $O(2) = U(1) \rtimes \mathbb{Z}_2$ after including charge conjugation. But in fact the real global symmetry is bigger! This is because the conclusion that the symmetry is $U(1)$ came from requiring the global symmetry to act identically on both of the components of the $SU(2)$ doublet. This is a natural thing to do, since the global symmetry has to commute with the action of the gauge group. But we see from this example that we can actually have the global symmetry act nontrivially on the different components in the $SU(2)$ doublet! For example, consider the left action by $e^{i\alpha Z}$. This is the diagonal $U(1)$ that we would have guessed to be the naive global symmetry. But what about the left action by $e^{i\alpha X}$? One checks that this sends e.g $\psi_\uparrow \mapsto i\psi_\downarrow^\dagger, \psi_\downarrow \mapsto i\psi_\uparrow^\dagger$: so it mixes the two components of the doublet, but also charge-conjugates them; this allows it to commute with gauge transformations. Thus the action of charge conjugation is built in to the $PSp(N)$ symmetry. Or more precisely, it's mixed up between the $PSp(N)$ and $SU(2)$ actions.

More abstractly, the fact that we get a $PSp(N)$ global symmetry can be understood through the decomposition

$$SO(4N) \supset \frac{SU(2) \times Sp(N)}{\mathbb{Z}_2}. \tag{389}$$

The relevance of this is that N Dirac fermions in the fundamental of $SU(2)$ can be written as $4N$ Majorannas, which are acted on by $SO(4N)$. The $Sp(N)$ factor in the above decomposition is the largest subgroup which commutes with the $SU(2)$, and so after gauging the $SU(2)$ we are left with an $Sp(N)/\mathbb{Z}_2$'s worth of global symmetry.

Note that this inclusion is not an equality in general, as we check by computing dimensions: as a Lie algebra, $\dim \mathfrak{so}(4N) = (16N^2 - 4N)/2 = 8N^2 - 2N$, while

$$\dim[\mathfrak{su}(2) \times \mathfrak{sp}(N)] = 3 + (N^2 - N)/2 + 3(N^2 + N)/2 = 2N^2 + N + 3 \leq \dim \mathfrak{so}(4N). \tag{390}$$

In fact the equality does hold when $N = 1$ for which both Lie algebras are 6-dimensional, which is just a manifestation of

$$SO(4) = \frac{SU(2) \times Sp(1)}{\mathbb{Z}_2}, \tag{391}$$

since $Sp(1)$ has alias $SU(2)$.[49]

Let's see how this decomposition works explicitly, at the level of Lie algebra generators. $Sp(N)$ is complex, and so in order to embed it in $SO(4N)$, we will need to send $i \mapsto J_2$, where

[49] Here's a way to see $SO(4) = [SU(2) \times SU(2)]/\mathbb{Z}_2$: let $r^2 = x^2 + y^2 + z^2 + t^2 = 1$ define S^4, and consider the matrix $U = t\mathbf{1} + i(xX + yY + zZ)$. We will think of the coordinates $v = t + iz, w = y + ix$ as equivalent coordinates for the sphere in $\mathbb{C}^2$. With this,

$$U = \begin{pmatrix} v & w \\ -\bar{w} & \bar{v} \end{pmatrix} \tag{392}$$

Consider $[SU(2)_L \times SU(2)_R]/\mathbb{Z}_2$, with the first factor acting on U on the left and the second factor acting on the right. A generic element of $SU(2)$ is conjugate to $e^{i\theta Z}$. Acting on the left, these actions send

$$e^{i\theta Z_L} : v \mapsto e^{i\theta v}, \quad w \mapsto e^{i\theta w}, \qquad e^{i\theta Z_R} : v \mapsto e^{i\theta v}, \quad w \mapsto e^{-i\theta w}. \tag{393}$$

This means that the left $SU(2)$ action, after being conjugated so that it only has a Z part, rotates the tz plane and the xy plane by equal angles, while the right $SU(2)$ action rotates them by opposite angles. Since both actions preserve $\det U = 1$, they are symmetries of S^4. Furthermore since they only rotate planes (pairs of basis vectors) instead of individual basis vectors, they are orientation preserving (okay, also since they have $\det = 1$). We can generate all of $SO(4)$ by rotating arbitrary planes, and so after modding out by the common $-\mathbf{1}$ to both actions, we get the desired isomorphism.

now we are using the notation $J_2 \equiv -iY$. Recalling a diary entry from last year wherein the generators for $\mathfrak{sp}(N)$ were written down, we have that when embedded in $\mathfrak{so}(4N)$, the $\mathfrak{sp}(N)$ generators are

$$\mathfrak{sp}(N) \ni (A_N \otimes \mathbf{1}_2 + S_N^y \otimes iY) \otimes \mathbf{1}_2 + (S_N^x \otimes X + S_N^z \otimes Z) \otimes J_2, \tag{394}$$

where A_N is an antisymmetric $N \times N$ matrix, the S^a are symmetric $N \times N$ matrices, and $\mathbf{1}_2$ is the 2×2 unit matrix.

What about the $SU(2)$ factor? To embed this in $SO(4N)$, we need to turn the symmetric generators X, Z into antisymmetric matrices. We do this by tensoring with J_2. Since the exponential map is $\exp(\mathfrak{so}(4N)) = SO(4N)$ (no factor of i), we also want the $SU(2)$ generators to be i times the normal physicist-convention $SU(2)$ generators. One sees that the following three generators

$$\sigma^1 = \mathbf{1}_N \otimes J_2 \otimes X, \quad \sigma^2 = \mathbf{1}_N \otimes \mathbf{1}_2 \otimes iY, \quad \sigma^3 = \mathbf{1}_N \otimes J_2 \otimes Z \tag{395}$$

obey $[\sigma^a, \sigma^b] = \epsilon^{abc}\sigma^c$, and hence generate an $SU(2) \subset SO(4)$ subgroup. Furthermore, we see that all of these generators commute with the $\mathfrak{sp}(N)$ generators, telling us that indeed, $[SU(2) \times Sp(N)]/\mathbb{Z}_2 \subset SO(4N)$, where the quotient is because $-\mathbf{1}_{2N}$ is in both factors. Furthermore we see that $Sp(N)$ is the largest subgroup that commutes with $SU(2)$, so that when e.g. $SU(2)$ is gauged, $PSp(N)$ is the global symmetry that remains.

23 *Global symmetry that remains after gauging a subgroup* ✓

Suppose we have some fields transforming under a global symmetry group $\mathcal{G}$ (which includes spacetime symmetries, e.g. some pin group if they are fermions), and we gauge a subgroup $G \subset \mathcal{G}$. What is the surviving global symmetry group? Not $\mathcal{G}/G$: G may not be normal, and so it may "take out" more of $\mathcal{G}$ than just itself. Today we will answer the question, and discuss a few illustrative examples.

We can find the resulting global symmetry group by examining how various charge operators for the symmetries in $\mathcal{G}$ commute with the generator of gauge transformations in G. If U_θ generates the gauge transformation $\psi \mapsto e^{\theta^a T^a_G}\psi$ for the G gauge group (T^a are the generators of G; the case of G discrete is basically the same) and if e^{iQ_h} is the charge operator for a global symmetry acting with an element $h \in \mathcal{G}$, then we require that

$$e^{iQ_h} U_\theta = U_{\theta'} e^{iQ_h}, \tag{396}$$

since then the action of e^{iQ_h} is well-defined when acting on physical states, for which U_θ acts as $\mathbf{1}$ for all choices of θ.

Let the surviving global symmetry group be denoted by $\mathcal{G}'$. Then the above means that if $h \in \mathcal{H}'$ then for any $g \in G$ we must have $h^{-1}gh = g'$ for some $g' \in G$. The means that h must be in the normalizer[50] of G with respect to $\mathcal{G}$, quotiented by G itself (which is trivially in the normalizer):

$$\mathcal{G}' = N_{\mathcal{G}}(G)/G. \tag{397}$$

[50]Recall that $N_{\mathcal{G}}(G) = \{h \in \mathcal{G} \,|\, hG = Gh\}$.

That $N_{\mathcal{G}}(G)/G$ is a subgroup[51] of $\mathcal{G}$ is easy to check (easy to see that the normalizer is a subgroup, and G is by definition normal in $N_{\mathcal{G}}(G)$, so we can consistently take the quotient). The simplest example is of course $\mathcal{G} = H \times G$, for which $N_{\mathcal{G}}(G) = \mathcal{G}$ and hence $\mathcal{G}' = \mathcal{G}/G = H$.

Gauging $U(1) \subset O(2n)$

As a slightly nontrivial example, think of $\mathbb{R}$ fermions and let $\mathcal{G} = O(2n)$ be a flavor symmetry with G the diagonal $U(1)$, a given element of which takes the form $R_\theta^{\oplus n}$, where R_θ is a 2×2 rotation matrix. Now $[R_\theta, x\mathbf{1} + yJ] = 0$ for all x, y, and therefore one can show that $U(n) \subset O(2n)$ commutes with the diagonal $U(1)$ (we embed $U(n)$ into $O(2n)$ by writing each complex entry $z = x + iy$ as $x\mathbf{1} + yJ$), and so we at least have $U(n) \subset N_{O(2n)}(U(1))$.

What about the reflection that extends $SO(2n)$ to $O(2n)$? If we take this reflection to be the generator R of $\mathbb{Z}_2^R$ such that $R = Z \oplus \mathbf{1}_{2n-2}$, then we get something that doesn't commute with the $U(1)$, since conjugation by R does

$$R_\theta^{\oplus n} \mapsto R R_\theta^{\oplus n} R = R_{-\theta} \oplus R_\theta^{n-1} \notin U(1). \tag{398}$$

Hence, $\mathbb{Z}_2^R \notin N_{O(2n)}(U(1))$. However, consider the action of $Z^{\oplus n}$, which reflects every other axis. This performs the action that we would usually associate with charge conjugation, viz. $Z^{\oplus n} R_\theta^{\oplus n} Z^{\oplus n} = R_{-\theta}^{\oplus n} \in U(1)$. Therefore we define the charge conjugation matrix C by

$$C \equiv Z^{\oplus n} \in N_{O(2n)}(U(1)). \tag{399}$$

Suppose $n \in 2\mathbb{Z}$. Then C has determinant 1, and is in fact part of $SO(2n)$—it is not an outer automorphism extending $SO(2n)$ to $O(2n)$. However, if $n \in 2\mathbb{Z}+1$, $\det Z^{\oplus n} = -1$, and it is a reflection outer automorphism that extends to $O(2n)$. Regardless of whether it is a reflection or not, it is not in $U(n)$, since when represented as a matrix in $O(2n)$, the only diagonal matrix in $U(n)$ is the identity. Now one checks that C is a good moniker for $Z^{\oplus n}$ by noting that $C^\dagger U C = U^*$ for any $U \in U(n)$. Therefore we have at least $N_{O(2n)}(U(1)) \supset U(n) \rtimes \mathbb{Z}_2^C$. In fact the normalizer is exactly $U(n) \rtimes \mathbb{Z}_2^C$: $U(n) \rtimes \mathbb{Z}_2^C$ is a maximal subgroup of $O(2n)$ (which can be worked out from the material in [**?**]), and since the normalizer is a proper subgroup of $O(2n)$, it must be $U(n) \rtimes \mathbb{Z}_2^C$. Taking the quotient by $U(1)$, we then get that the remaining global symmetry is

$$\mathcal{G}' = PSU(n) \rtimes \mathbb{Z}_2^C. \tag{400}$$

In this case the $\mathbb{Z}_2^C$ really is a charge conjugation symmtey, since it corresponds to the group $\mathrm{Out}(PSU(n)) = \mathbb{Z}_2$.[52]

Gauging $SU(2) \subset O(4n)$

As another example relevant for fermions, considering gauging an $SU(2)$ subgroup of $O(4n)$, where the $SU(2)$ subgroup acts in a block-diagonal way, with each block a 4×4 orthogonal

[51] We might call this the Weyl group of G in $\mathcal{G}$, or something like that, since the Weyl group in the context of Lie theory is defined as $N_{\mathcal{G}}(T)/T$, where T is a maximal torus.

[52] We are tacitly assuming $n \neq 2$.

matrix. We will write basis vectors in $\mathbb{R}^{4n}$ suggestively as $v \equiv (\chi_1^\uparrow, \eta_1^\uparrow, \chi_1^\downarrow, \eta_1^\downarrow, \chi_2^\uparrow, \dots)^T$, where we think of the variables as Majorana fermions coming from n complex fermions in the fundamental of $SU(2)$; e.g. $\psi_{\sigma i} = \chi_i^\sigma + i\eta_i^\sigma$. The $SU(2)$ we're gauging is then realized as (the subscript g appears when needed to distinguish the gauge group from other groups floating around)

$$SU(2)_g \ni \begin{pmatrix} x\mathbf{1} + yJ & w\mathbf{1} + zJ \\ -w\mathbf{1} + zJ & x\mathbf{1} - yJ \end{pmatrix}^{\oplus n}, \qquad (x, y, w, z) \in S^3. \tag{401}$$

Let's first work out the simple case of $n = 1$. We claim that the normalizer includes $Sp(1) \cong SU(2)$, which is realized as matrices of the form

$$Sp(1) \ni x\mathbf{1} \otimes \mathbf{1} + w\mathbf{1} \otimes J + yJ \otimes X + zJ \otimes Z, \qquad (x, y, w, z) \in S^3. \tag{402}$$

Here the first tensor factors keep track of the way of representing complex numbers with real matrices, while the second factors keep track of the Pauli matrix structure of the group elements. Indeed, the above matrices can be seen to commute with the gauged $SU(2)_g$, which is generated by matrices of the form

$$SU(2)_g \ni x\mathbf{1} \otimes \mathbf{1} + yJ \otimes \mathbf{1} + wZ \otimes J + zX \otimes J, \qquad (x, y, w, z) \in S^3. \tag{403}$$

Here by contrast it is the second tensor factors that keep track of the representation of $1, i$ in terms of real matrices. The antisymmetric form preserved by the $Sp(1)$ in the normalizer is $J \otimes \mathbf{1}$. Therefore the normalizer is at least $N_{O(4)}(SU(2)) = [Sp(1) \times SU(2)]/\mathbb{Z}_2 = SO(4)$, with the quotient coming from $-\mathbf{1}$ being in both groups.[53]

We claim that the reflection that extends $SO(4)$ to $O(4)$ is not in the normalizer. Abstractly, this is because this reflection generates $\mathrm{Out}(SO(4)) = \mathbb{Z}_2$, which from experience we know exchanges the two $SU(2)$s in $SO(4) \cong (SU(2) \times SU(2))/\mathbb{Z}_2$—therefore conjugation by the reflection should not be in the normalizer of a single $SU(2)$ factor. Let's check this with an example: the reflection can be taken to be $R = (-Z) \oplus \mathbf{1}$. When we act on the matrix $J \otimes X \in SU(2)_g$ in the gauged $SU(2)$ we get

$$R^{-1}(J \otimes X)R = -(\mathbf{1} \otimes J) \in Sp(1), \tag{404}$$

so that conjugation by the reflection indeed swaps $SU(2)_g$ and $Sp(1)$, meaning that the $Sp(1)$ is the reflected image of the gauge group. Therefore since the normalizer is a proper subgroup of $O(4)$ and contains $SO(4)$, it must be $SO(4)$. Another way of saying this is that since $SO(4)$ is a maximal subgroup of $O(4)$, $SO(4)$ must be the whole normalizer, with the global symmetry group thus being $\mathcal{G}' = PSp(1) = PSU(2)$.

[53] A rather highbrow way of saying why the $Sp(1)$ commutes with the $SU(2)_g$ is that it uses the isomorphism coming from pseudoreality of the fundamental rep of $SU(2)$ together with complex conjugation to create a trivial action on the $SU(2)_g$. The matrices $\mathbf{1} \otimes \mathbf{1}$ and $\mathbf{1} \otimes J$ in $Sp(1)$ form a $U(1)$ that is the obvious one which commutes with $SU(2)_g$ (the diagonal particle-number $U(1)$ symmetry in the action). The matrices $\mathbf{1} \otimes X, \mathbf{1} \otimes Z$ both have the effect of complex-conjugating the $SU(2)_g$, since they both anti-commute with the way we've chosen to represent the number i in that group, namely as $\mathbf{1} \otimes J$. They can then be combined with the matrix $J \otimes \mathbf{1}$, which is the isomorphism establishing the pseudoreality of $SU(2)$, to produce $J \otimes X$ and $J \otimes Z$, which complete the $Sp(1)$ part of the normalizer.

This means "charge conjugation" is already included in the normalizer: indeed, if we let it act as $\chi^\sigma + i\eta^\sigma \mapsto \chi^\sigma - i\eta^\sigma$ then $C = \mathbf{1} \otimes Z$ is already included, and if we let it exchange spins then it is $C = J \otimes Z$, which is also included. But really, the point is that we shouldn't be calling such a thing charge conjugation: $\mathrm{Out}(PSU(2)) = \mathbb{Z}_1$ and $PSU(2)$ isn't a semi-direct product, and so there can't possibly be any type of charge conjugation symmetry remaining in the theory after the gauging occurs.

Now we return to the case of arbitrary n. It is easier in this case to make a change of basis and package the vector of χs and ηs as a $2n \times 2$ matrix with complex entries. Thought of as a n-component column vector V with matrix-valued entries, the ith entry is the "quaternonic fermion" matrix

$$V_i = \begin{pmatrix} \chi_i^\uparrow + i\eta_i^\uparrow & \chi_i^\downarrow + i\eta_i^\downarrow \\ -\chi_i^\downarrow + i\eta_i^\downarrow & \chi_i^\uparrow - i\eta_i^\uparrow \end{pmatrix}. \tag{405}$$

The $SU(2)$ we are gauging is then realized as the right action on the above matrices, with the elements in $SU(2)$ written as 2×2 complex matrices, rather than 4×4 real matrices. The length of the vector v is determined by

$$|v|^2 = \frac{1}{2}\mathrm{Tr}[V^\dagger V], \tag{406}$$

and as such is properly preserved by the $SU(2)$ right action. This form also makes it clear that the length is preserved by a left $SU(2n)$ action, which by construction commutes with the right $SU(2)$. However, the left $SU(2n)$ action is too big: the structure of the is in the above form of V_i needs to be preserved by any putative left action, so that the right $SU(2)$ acts properly. The structure of the is (the $\sqrt{-1}s$, not the flavor labels) is encapsulated in the relation $(\mathbf{1}_n \otimes J)^\dagger V J = V^*$, which needs to be preserved by the left action. Therefore if $U \in SU(2n)$ then we need

$$(\mathbf{1}_n \otimes J)^\dagger U (\mathbf{1}_n \otimes J)(\mathbf{1}_n \otimes J)^\dagger V J = (UV)^* \implies (\mathbf{1}_n \otimes J)^\dagger U (\mathbf{1}_n \otimes J) = U^*, \tag{407}$$

which since U is unitary means $U^T(\mathbf{1} \otimes J)U = \mathbf{1} \otimes J$, and so in fact we must have $U \in Sp(n)$. If this is a bit too slick, one can also make the following (still rather slick) argument: whatever the global symmetry group that remains is, the full symmetry group (gauged $SU(2)$ + global) better have only real representations, since we started off with an $O(4n)$ symmetry. Since we gauged an $SU(2)$ acting in the fundamental, which is pseudoreal, the global symmetry group must also act via a pseudoreal representation, since the $\otimes$ of two ps$\mathbb{R}$ reps is $\mathbb{R}$. Therefore the global symmetry can't be $U(2n)$—$Sp(n)$ works though, since it acts in a ps$\mathbb{R}$ way.

Anyway, we now know that the normalizer is at least $N_{O(4n)} \subset [Sp(n) \times SU(2)]/\mathbb{Z}_2$. Is this the whole normalizer? This is in fact the whole normalizer, since a math fact [**?**] is that $(Sp(n) \times Sp(1))/\mathbb{Z}_2 = PSp(n) \times SU(2)$ is a maximal subgroup of $O(2n)$, and through the same reasoning as in the last example, this must be the full normalizer. Therefore the global symmetry remaining is found by taking a quotient by $SU(2)$, producing

$$\mathcal{G}' = PSp(n). \tag{408}$$

Note that we do not get $U(n) \rtimes \mathbb{Z}_2$ or similar, which we might have naively concluded based on thinking about complex fermions.

Finally, what about charge conjugation? There actually is no real charge conjugation symmetry in this case: $\mathrm{Out}(PSp(n)) = \mathbb{Z}_1$ and so there's no type of charge conjugation that we're missing, and $PSp(n)$ can't be written as a semidirect product involving a $\mathbb{Z}_2$ factor,[54] and so there is no $\mathbb{Z}_2^C$ symmetry hiding in the $PSp(n)$. If we were to think about charge conjugation as sending $\chi_i^\sigma \mapsto \chi_i^\sigma$ and $\eta_i^\sigma \mapsto -\eta_i^\sigma$ then it is already included in $(Sp(n) \times SU(2))/\mathbb{Z}_2$, while if we were to have it acting with J in on the $SU(2)$ factor then it'd already be included in the $Sp(n)$ factor—but either of these actions wouldn't really be a charge conjugation symmetry, since neither of them are outer automorphisms. So, while our un-gauged symmetry group $O(4n)$ includes a charge conjugation symmetry since we can write it as $SO(4n) \rtimes \mathbb{Z}_2^C$, when we gauge $SU(2)$, the existence of a charge conjugation symmetry goes away.

24 *Charge conjugation, outer automorphisms, and the best definition of time reversal* ✓

In today's diary, we'll briefly go over the reasons for why what the definitions the community at large uses for T and for CT should really be swapped, and will try to elucidate the meaning of CPT in simpler terms.

This discussion should be prefaced by mentioning that the philosophy here is that symmetries aren't less of some fundamental thing that a theory is defined by; rather they are more like tools that we use to study a given theory. Consequently, the exact definition of T and CT and so on is rather subjective. We define symmetries in a way that suits our needs for doing calculations, rather than having them handed down to us from on high. Therefore as there is no real "natural" definition of these symmetries, one should try to work with conventions that are as pleasant as possible — hence the present diary entry.

"Problems" with the historical definition of T

Conceptually, I think that the most natural definition of a time-reversal symmetry is one which is antiunitary and reverses time, and does nothing else pertaining to other possible internal symmetries. In particular, T should be a part of the Lorentz group, in the sense that its action (modulo the $\mathbb{C}$ conjugation part) should be restricted to act only via an action of the Lorentz group on the Lorentz indices of the fields in the theory.[55]

To see an example of why the historically-used T isn't necessarily part of the Lorentz group and hence fails this criterion, consider fermions coupled to a gauge field A,[56] with the action containing the term $\bar{\psi}\not{A}^a T^a \psi$ for A_μ^a some real vector fields. We will focus on this term for clarity and since it is conceptually the simplest thing to look at (all other irreps of Lorentz can be obtained from $\otimes$s of (s)pinor reps; hence the focus on fermions). Anyway,

[54] Because $PSp(n)$ is connected.

[55] Terminology reminder: the Lorentz group is not connected, and only the **1** component of the Lorentz group needs to be a symmetry in relativistic QFT. So e.g. reflections are part of the Lorentz group, but may or may not be symmetries.

[56] This field may or may not be dynamical; we are just introducing it in order to be able to talk about the action of symmetries in a convenient way.

this term is definitely something that from experience we "expect" to be T-invariant. Since A^a_μ has a vector index the natural action of time reversal would be something like

$$T : A_\mu \mapsto U_T^\dagger A_\mu U_T = (-1)^{\delta_{\mu,0}} A_\mu, \tag{409}$$

for some unitary U_T—note that there is no action on the gauge index of A^a_μ. However, this does not lead to T-invariance in general, because of the $\mathbb{C}$ conjugation. The general condition for the T invariance of this coupling is (since we are physicists, the T^as are Hermitian)

$$U_T^\dagger A^a_\mu (T^a)^* U_T = (-1)^{\delta_{\mu,0}+1} A^a_\mu T^a, \tag{410}$$

which will not be satisfied for a general gauge group and representation.

Consider first the case when the fermions transform in a real representation. Then $(T^a)^* = -T^a$ on account of the group representations being $e^{i\theta^a T^a}$, and so in this case the expected transformation law for T, viz. (409), leaves $\bar\psi \not{A} \psi$ invariant.[57]

Now consider the case when the fermions are in a pseudoreal representation R, with J the antisymmetric matrix relating R and R^* through conjugation. Then since our generators are Hermitian, $J^\dagger T^a J = -(T^a)^*$. Therefore we see that we can take A to transform with the sign in (409) along with conjugation by J, so that $U_T = J$ when acting on the gauge indices.[58] This means that the action of time reversal actually involves a nontrivial action on the gauge indices, because of the conjugation by J. Therefore the action of T does not just depend on the Lorentz indices of the field in question, which in my opinion is not ideal for a definition of T.

The situation is even worse for fermions in a complex representation. Since there is no longer an isomorphism connecting R and R^*, we won't be able to find a choice of U_T that will work, and we have to map each of the different A^a individually with a different sign: we need to act on the generator index with a transformation via $A^a_\mu \mapsto R^{ab} A^b_\mu$. For example, for the fundamental of $SU(3)$ with the usual basis for T^a, we would need T to map A^a for $a = 2, 5, 7$ with one sign, and A^a for $a = 1, 3, 4, 6, 8$ with the other.

Anyway, the point here is that this definition of T necessitates adding in an action on the gauge indices which cancels the complex conjugation performed by T, except in the case where the fermions transform in a real representation. This means that this T does not generically act solely on Lorentz indices. In other words, if we write $U_T = U_{T,I} \otimes U_{T,L}$ where the first factor acts on the internal indices and the second factor acts on the Lorentz indices, then $U_{T,I} \neq \mathbf{1}$.[59] In fact the precise requirements for the two factors are that (assuming the

[57] In this case there may not be any nontrivial notion of charge conjugation anyway, and so T and CT might not be distinguishable to begin with.

[58] We can either have

$$T : A^a_\mu \mapsto (-1)^{\delta_{\mu,0}} J^\dagger A^a_\mu J, \tag{411}$$

or we can strip off the Js and modify the transformation of the fermions ψ under T by an action of J on the flavor indices. Either way, $\bar\psi \not{A} \psi$ is invariant.

[59] We are as always ignoring weird things like SUSY where the internal and spacetime dof mix.

Dirac adjoint is used)[60]

$$U_{T,I}^\dagger (T^a)^* U_{T,I} = -T^a, \quad U_{T,L}^\dagger \mathcal{K}[i\gamma^0(-\gamma^0\partial_0 + \gamma^i\partial_i)]\mathcal{K} U_{T,L} = i\partial\!\!\!/. \tag{412}$$

With this transformation, $A\!\!\!/$ transforms in the same way as $i\partial\!\!\!/$, and the fermion action is consequently invariant.

Why CT is better

First, recall what we mean by C: it is a $\mathbb{Z}_2$ outer automorphism of the symmetry group G,[61] which acts on a field ψ in a representation R of G as $\psi \mapsto C\psi$, with $C\psi$ transforming in the dual representation R^*. Sometimes we would write this as $\psi \mapsto \psi^*$ in the case of e.g. a $U(1)$ symmetry for a single complex fermion, but this could always be re-written as $\psi \mapsto C\psi$ where we think of ψ as two real fermions, and of C as acting via the matrix Z.

There are two possible conventions for the action of C: it can dualize the representation of only the *internal* symmetries involved, or it can include a dualization of the spacetime symmetry representation as well. For example, in the former definition, C would map a field which annihilates left-handed neutrinos to a field which annihilates left-handed antineutrinos, while in the latter definition it would map to a field which annihilates right-handed antineutrinos. Here an antiparticle is one whose *internal* symmetry quantum numbers are all the duals of the quantum numbers of the particle in question—therefore in the former definition C sends particles to antiparticles, while in the latter definition it does this plus an action on the Lorentz indices (usually by parity). Now the former definition of C may not even be a legit operation to perform in a given QFT (right-handed neutrinos [not antineutrinos] may not even exist!), while with the second definition, C is always a legit thing to do. Therefore, we will work with the later definition, as we have done throughout most of the diary.

In any case, in the discussion of C, P, and T (or better, C, R, T), C always feels like a bit of a misfit, since it's not part of the Lorentz group. In fact, with our definition of T, we have already seen that T is not always part of the Lorentz group either! However, we will now argue that the product CT always *is* part of the Lorentz group.[62]

If we include the action of C in the definition of time reversal, the problems found above for the case of a ps$\mathbb{R}$ or $\mathbb{C}$ representation go away. Indeed, assuming that the theory possesses a C symmetry and writing the charge conjugation matrix as $C_I \otimes C_L$, we see that C-invariance requires[63]

$$C_L^\dagger \gamma_\mu C_L = -\gamma_\mu^T, \qquad C_I^\dagger T^a C_I = -[T^a]^T. \tag{413}$$

[60] Pedantic detail: remember that the sign in front of ∂_0 doesn't come from $U_{T,L}^\dagger \partial_0 U_{T,L} = -\partial_0$, since in the present way of thinking about things we are never *actually* acting on spacetime; the action of symmetries is entirely on the fields, and not on numbers like t (conceptually, I think it's best to always think of the action as being entirely performed through conjugating second-quantized operators). The minus sign instead comes when when change variables $t \mapsto -t$ in the action.

[61] Where pedantically the "full" symmetry group is G', with $G'/\mathbb{Z}_2^C = G$, or $G' \cong G \rtimes \mathbb{Z}_2^C$.

[62] Again, by "part of the Loretnz group", we mean that its action on a field is determined entirely by the Lorentz indices of the field—it acts trivially on all other indices, like flavor indices. It still complex conjugates fields though, so the action is not solely through the action of the Loretnz group.

[63] We are using the same conventions as in the long diary entry on pinors and representation theory, where $\bar\psi M\psi \mapsto \bar\psi C^\dagger M^T C\psi$ under C.

Now recall that we are in conventions where the T^a are Hermitian; this means that the equation for C_I can be written

$$C_I[T^a]^T C_I^\dagger = -T^a \implies C_I[T^a]^* C_I^\dagger = -T^a. \tag{414}$$

But we see that this is the same as the equation for $U_{T,I}$, just with †s in different places! Hence we may in fact set $U_{T,I} = C_I^\dagger$, meaning that when we put C and T together as $CT = (CT)_I \otimes (CT)_L \mathcal{K}$, we have $(CT)_I = U_{T,I} C_I = C_I^\dagger C_I = \mathbf{1}$, so that in fact CT acts only on Lorentz indices as claimed.

Particles, antiparticles, and CPT

The exact transformation that "takes particles to antiparticles" is, judging from the internet, a source of great confusion. On account of QFT being unitary, one fuzzily expects there to be some sort of "particles to antiparticles / Hermitian conjugation" transformation that does some sort of charge conjugation-y thing and is a symmetry of all relativistic QFTs (this of course turns out to be CPT). But then what exactly "exchanges particles and antiparticles"? Some people say C, others say some perverse version where C is anti-unitary, etc. In cmt CT often acts as particle-hole, so is it CT? But when learning about leptogenesis / early universe cosmology, the matter / anti-matter asymmetry is usually couched in terms of CP violation, so is it CP? Making things worse, Green, Schwartz and Witten say that it is in fact CPT that exchanges particles and antiparticles. Agrh! What's going on?

As usual with this kind of stuff, it's a bit of a conceptual / terminological minefield. To me, the best discussions of this stuff are in Weinberg vol I (this is really the definitive reference I feel), Sidney Coleman's QFT book chapter 22, and Haag.

Basically, in situations where there are some particles and antiparticles which are distinct, C interchanges particles and antiparticles (again, an antiparticle only has opposite *internal* quantum numbers), and C invariance of a theory means that e.g. particles can be substituted for antiparticles in any process without changing the amplitude for that process to occur (this does not mean that C can be defined to act trivially on particles that are their own antiparticles if there are also other things in the theory—e.g. in order for C to be a symmetry in QED, we must define it so that the photon transforms with C-parity -1).[64] Just exchanging particles and antiparticles is not the same as doing the nebulous "Hermitian conjugation thing" that we expect to be a symmetry of all relativistic QFTs, and indeed of course not all physical theories are invariant under C. Consider e.g. the weak interaction. The coupling of the leptons to the gauge field looks like (ignoring prefactors)

$$\mathcal{L}_W \ni \bar{\Psi} \not{W} (1 - \bar{\gamma}) \Psi + h.c. = (j_V^{a\mu} - j_A^{a\mu}) W_\mu^a + h.c., \tag{415}$$

where e.g. $\Psi = (e, \nu_e)^T$ for the first generation. Since j_A and j_V transform with different signs in $d \in 4\mathbb{Z}$,[65] there is no way to render this term C-invariant, and so the weak interactions respect no type of charge conjugation symmetry.

[64] And remember that we really are *defining* the action of C, *not* deriving it.

[65] When $d \in 4\mathbb{Z}$, we have $C^\dagger \gamma_\mu^T C = -\gamma_\mu$ so that the vector current is odd, while

$$C^\dagger (\gamma_\mu \bar{\gamma})^T C = -C^\dagger \bar{\gamma} C \gamma_\mu = -\bar{\gamma}\gamma_\mu = +\gamma_\mu \bar{\gamma}, \tag{416}$$

and so the chiral current is even.

On the other hand, this term does respect CP symmetry, and sometimes one hears that it is really CP which exchanges particles and antiparticles, e.g. as in discussions of cosmology (again, CP may really be C depending on how C is defined). CP defined with our definition of C is of course also not always a symmetry, and the real-life example is the weak coupling to the quarks provides an example with broken CP:

$$\mathcal{L}_W \ni \bar{d}_i(1-\bar{\gamma})\not{W} u_j V^{ij} + h.c. \tag{417}$$

where V is the CKM matrix, $d_i = \begin{pmatrix} d & s & b \end{pmatrix}_i^T$ are the down-like quarks, and $u_i = \begin{pmatrix} u & c & t \end{pmatrix}_i^T$ are the up-like quarks. It turns out that in our universe the CKM matrix is such that CP is violated.

The real "Hermitian conjugation thing" that is always a symmetry in relativistic QFT is of course CRT (i.e, we can always choose a C, R, and T such that CRT is a symmetry). Actually, if we use $\mathcal{T} = CT$ for the "correct" time reversal transformation which only acts on Lorentz indices, then CRT becomes $R\mathcal{T}$; this is obviously always a symmetry, as can be seen by analytically continuing into Euclidean time (see Coleman's book for a good explanation of why analytically continuing Feynman diagrams is always legit) and noting that $R\mathcal{T}$ just does a rotation. Basically, $R\mathcal{T}$ takes a spatial slice and "flips it over", which involves reversing time and changing an odd number of spatial coordinates (since this operation preserves the spatial slice it has a well-defined action on the Hilbert space of the theory). Anyway, this is why some sources refer to CRT as the thing which establishes "a correspondence between particles and antiparticles". Whether or not this is the best terminology depends on whether antiparticles have opposite quantum numbers under spacetime symmetries as well as internal ones, but with our present definition of C (dualizes the whole representation, not just the internal part), I think this terminology is fair.

25 *Real, complex, and chiral Majoranas—disambiguation* ✓

Today we will try to explain exactly what the classifier "Majorana" means when applied to fermions. In some contexts (usually CMT), a Majorana fermion is taken to mean a $\mathbb{R}$ fermion, and in others (usually particle physics) it's taken to mean a (perhaps $\mathbb{C}$) fermion which admits a Lorentz-invariant pairing with itself. The confusion between these definitions is particularly onerous in the many review articles out there that talk about Majoranas as e.g. neutrinos side-by-side with Majoranas as realized in solid-state physics—the two notions are different, but I've never really seen this clearly explained anywhere. Distinguishing between the various cases is what we'll do today.

Majoranas as real fermions

The most common definition (or at least the definition we will take as canonical) of a Majorana is a fermion which transforms in a real representation of the total symmetry group (spacetime + internal) of the theory.

Let's think about how this could arise. In general, the mode expansion for a fermion field is

$$\psi_\alpha(x) = \int_{\mathbf{p}} \left(c_{\mathbf{p}\sigma} u_\alpha(\mathbf{p},\sigma) e^{ip\cdot x} + d^\dagger_{\mathbf{p}\sigma} v_\alpha(\mathbf{p},\sigma) e^{-ip\cdot x} \right), \tag{418}$$

with α the spinor index and σ an (implicitly summed over) spin index determined according to the representation theory of the little group of $\mathbf{p}$. In order for ψ to transform in a well-defined way under the symmetry groups involved, we must have that if c annihilates particles transforming under the internal symmetry group in a representation R, then d annihilates particles transforming under R^*.

The u and v vectors are there to do the following job: when we perform a Lorentz transformation, we do it by acting on the creation / annihilation operators as (ignoring translations)

$$U(\Lambda) c_{\mathbf{p}\sigma} U^{-1}(\Lambda) = \sqrt{(\Lambda p)^0/p^0}\, D_{\sigma\rho}(\Lambda, p) c_{\mathbf{p}_\Lambda \rho}, \tag{419}$$

with D the representation matrix acting on spin (it depends on Λ and p since it's the representations of the little group of $\mathbf{p}$ that are relevant here, so e.g. in the massive case a boost into the particle's rest frame is involved). The u and v coefficients are "intertwiners" that convert this action into an action that results in a homogeneous transformation of the ψ field as a whole. That is, they allow us to write the transformed version of $\psi(x)$, after substituting in the RHS of the above expression into the mode expansion (and likewise for $d^\dagger$), as $R_\Lambda \psi(\Lambda^{-1}x)$, with R_Λ some chosen (s)pinor representation of the Lorentz transformation. In order for ψ to have a well-defined transformation under the spacetime symmetry group, it must therefore annihilate particles in the appropriate (s)pinor representation, and create particles in the dual of this representation (e.g. if it creates left-handed electrons, it must annihilate right-handed positrons).

If the full (spacetime + internal) representation that ψ transforms under is isomorphic to its dual,[66] then c and d destroy particles with the same quantum numbers, and it is possible to in fact cut down on the degrees of freedom and have $c = d$. The conjugate of ψ in this case is

$$\psi^* = \int_{\mathbf{p}} \left(c_{\mathbf{p}\sigma} v^*_\alpha(\mathbf{p},\sigma) e^{ip\cdot x} + c^\dagger_{\mathbf{p}\sigma} u^*_\alpha(\mathbf{p},\sigma) e^{-ip\cdot x} \right). \tag{420}$$

If we can find a unitary matrix C such that[67]

$$u^*(\mathbf{p},\sigma) = \gamma^0 C^T v(\mathbf{p},\sigma), \quad v^*(\mathbf{p},\sigma) = \gamma^0 C^T u(\mathbf{p},\sigma), \tag{421}$$

then we will have (I'm assuming mostly negative signature here for notational simplicity)

$$\psi^* = \gamma^0 C^T \psi \implies \bar\psi C^\dagger = \psi. \tag{422}$$

Looking back at the long diary entry on pinors and representation theory, we see that $C : \psi \mapsto \bar\psi C^\dagger$ is exactly how we defined charge conjugation, and so this condition reads $C : \psi \mapsto \psi$ — fermions which satisfy the above condition are their own charge-conjugates.

[66] We will avoid calling such a representation real or pseudoreal; see the next section for more detail why.

[67] The pesky γ^0s here are why conventions for charge conjugation very often differ by an action of γ^0, i.e. by a parity transformation. Therefore it is important to disambiguate between C and CP when reading stuff.

This condition is usually taken as the definition of a Majorana fermion in particle physics; see the next section for more discussion. However, the cond-mat (and math-ph) circles usually seem to define a Majorana fermion as a *real* field. The condition $\psi = \bar{\psi}C^\dagger$ is then not good enough, and we require the stronger condition of $\psi = \psi^*$, i.e. that the unitary matrix relating the conjugates of the u and v spinors is given precisely by $C = \gamma^0$ (equivalently, we require that the Dirac and Majorana adjoints agree).

This means that a Majorana, in the sense of a real fermion, must transform under a real representation of the full (spacetime + internal) symmetry group. When is this possible? Consulting the results in the big diary entry on the representation theory of pin groups, we see that we can have chiral Majorana spinors in spacetime dimension $d = 2$, Majorana spinors in $d = 3$, and Majorana spinors (but *not* chiral Majorana spinors) in $d = 4$. Similarly, we can have Majorana pinors in dimensions $d = 2, 3, 4$, but the existence of Majorana pinors depends on the choice of signature in $d = 3, 4$, e.g. we can have Majorana pinors that transform under $\mathrm{Pin}(2, 1)$ but not under $\mathrm{Pin}(1, 2)$, and under $\mathrm{Pin}(3, 1)$ but not under $\mathrm{Pin}(1, 3)$ (here the notation is $\mathrm{Pin}(s, t)$ with s positive signature and t negative signature).

Finally, note that second-quantized Majorana operators of the kind used in e.g. the Kitaev chain haven't shown up in the present discussion. The second-quantized operators appearing in the mode expansion for ψ are not these kinds of operators; they still create particles (and act on a Fock space of integer dimension), and even though the Majorana is in a real representation with $\psi^* = \psi$ and is hence its own antiparticle, $c^\dagger \neq c$.

Complex Majoranas

As we mentioned above, other people (usually particle physicists) use a more general definition of Majorana fermion, viz. a field which is self-dual in the sense of the last section: ψ^* is related to ψ through a unitary transformation built from the charge conjugation matrix C and γ^0. With this definition, a Majorana is any fermion that admits an invariant pairing with itself—I'll take this to mean any fermion such that $\chi^T C \not{\partial} \chi$ is invariant under Lorentz transformations and under any internal symmetry group action that the theory might come with; others might use the requirement that a Majorana mass term $\chi^T M_m \chi$ exist (these two definitions are slightly different; consider e.g. whether or not chiral Majorana spinors make sense in 1+1D, as $\nexists$ a mass term).

One might think that this implies that the χ can be taken to be real. Indeed, as we saw, it means that there must be a unitary transformation given by $\gamma^0 C$ relating the transposed matrices in the full representation R of the total symmetry group to their inverses. If the representation R were unitary, then this would mean that C provided a unitary transformation between R and R^*, since the conjugate would combine with the transpose in χ^T to produce the requisite inverse. Therefore if R were unitary, then we would conclude that Majoranas in the above sense would exist only for R a real representation.

However, being non-compact, the spacetime part $\mathrm{Spin}(d-1, 1)$ (or $\mathrm{Pin}(d-1, 1)$, or what have you) will *not* be represented with unitary matrices acting on the Lorentz indices![68] This means that while we may be able to construct an invariant pairing of the form above, this

[68]The discussion here is a bit sloppy, since of course everything in QFT transforms under unitary representations. Non-compact Lie groups don't have any finite dimensional unitary representations, and so the action of $\mathrm{Spin}(d-1, 1)$ or $\mathrm{Pin}(d-1, 1)$ on the spinor indices is non-unitary, since there simply aren't any unitary

pairing may not necessarily connect R with R^*, since a field transforming under $R^\dagger$ will not necessarily pair with one transforming under R to give something invariant.

For example, consider $\mathrm{Spin}(3,1)$. The matrices generating the action in the spinor representation can be chosen to be e.g.

$$S_{ij} = S^+_{ij} \oplus S^-_{ij} = \frac{1}{2}\varepsilon_{ijk}(\sigma^k \oplus \sigma^k), \qquad S_{0i} = S^+_{0i} \oplus S^-_{0i} = \frac{i}{2}(\sigma_i \oplus -\sigma_i). \tag{423}$$

The chiral spinor representations $S_\pm$ are obviously complex. However, since they are all built out of σ matrices, we have

$$\Lambda^\pm_{\theta,\beta} Y = [e^{i(\theta^{ij}S^\pm_{ij} + \beta^k S^\pm_{0k})}]^T Y = Y e^{-i(\theta^{ij}S^\pm_{ij} + \beta^k S^\pm_{0k})} = Y[\Lambda^\pm_{\theta,\beta}]^{-1}. \tag{424}$$

Therefore the term $\chi^T_\pm Y \chi_\pm$ is in fact $\mathrm{Spin}(3,1)_\pm$ invariant, for $\chi_\pm$ transforming in the $S_\pm$ representation. Note how this gives a mass between two *Weyl* fermions—the $\chi_\pm$ are "Majorana-Weyl fermions". Also note that we can build a Dirac fermion from two of these Majorana-Weyl fermions—this wouldn't be possible if the Majoranas were real, simply by counting dof.

Because of the absence of complex conjugation, we will only be able to create an invariant out of a $\chi_\pm$ bilinear if $\chi_\pm$ transforms in a $\mathbb{R}$ or ps$\mathbb{R}$ representation under all internal symmetries, since all internal symmetries will be represented unitarily on the Hilbert space (we normally don't discuss infinite-dimensional internal symmetries). This means for example that it is still impossible to have a Majoranna (in the present section's definition of the term) that transforms in a complex representation of an internal symmetry group, even though the Majoranna field itself is allowed to be complex. Basically, all the subtleties and distinctions between real Majoranas and the more general type of $\mathbb{C}$ Majoranas considered here originate from the action of the spacetime symmetry group.

Neutrinos

These complex Majoranas are the ones encountered in neutrino physics—we know from colloquium talks that neutrinos might be Majoranas, and since they are chiral and live in 3+1 dimensions, they definitely cannot be $\mathbb{R}$ fermions. Actually saying that neutrinos can be Majoranas is kind of a cheat, as we shall see.

First let's recall some hep-ph and see how this works. We can write the neutrino field in general as

$$\nu_L = \int_{\mathbf{p}} \left(c_{\mathbf{p}\sigma} u_L(\mathbf{p},\sigma) e^{ip\cdot x} + d^\dagger_{\mathbf{p}\sigma} v_R(\mathbf{p},\sigma) e^{-ip\cdot x} \right). \tag{425}$$

So, ν_L creates a left-handed neutrino and destroys a right-handed anti-neutrino. ν_L is part of an $SU(2)$ doublet along with the e_L field, and carries $U(1)_Y$ charge -1. How can neutrinos

representations on $\mathbb{C}^{2^{\lceil (d-1)/2 \rceil}}$. However, the actual (s)pinors aren't vectors in $\mathbb{C}^{2^{\lceil (d-1)/2 \rceil}}$—they carry position / momentum labels, and instead live in $\mathbb{C}^{2^{\lceil (d-1)/2 \rceil}} \otimes L^2(\mathbb{R}^{d-1})$. Since this space is infinite-dimensional, the Lorentz group can (and does) act on it in a unitary way—after all, the representation in (419), acting on the second quantized operators, *is* unitary. In our whole discussion we've completely forgot about position / momentum indices, and so for our purposes, the action of the Lorentz group appears non-unitary (and when we say that a term like $\chi^T C \partial\!\!\!/ \chi$ is "invariant", we mean as regarded in $\mathbb{C}^{2^{\lceil (d-1)/2 \rceil}}$, i.e. forgetting about position indices; of course the position indices of the χ fields are permuted by the Lorentz transform.)

get a mass term? They can't have a Dirac mass for the same reason as the other leptons, viz. the fact that the weak interactions couple only via left-handed components, while Dirac masses mix left- and right-chiralities (and there is no $\bar{\nu}_R$ field in the SM to pair with anyway!). We can't use the usual Yukawa coupling to the Higgs, either; the coupling between $\bar{e}_R$ and the doublet (e_L, ν_L) gets broken down under SSB so as to only give the electron a mass.

Suppose that the neutrino field is a Majorana field, so that $\nu_{L\alpha} = \bar{\nu}_{L\beta} C^\dagger_{\beta\alpha}$. Can we have a Majorana mass term

$$\mathcal{L}_{SM} \supset m_M \nu_L^T C \nu_L \tag{426}$$

in the SM? No unfortunately not; such a term violates both $SU(2)_W$ and $U(1)_Y$. This is because actually ν_L can *not* actually be a Majorana fermion—there's no way we can take $c = d$ in the ν_L field mode expansion while preserving $U(1)_Y$, since the representation is complex and unitary.

So, why do particle physicists talk about Majorana neutrinos? They don't seem to fit into the framework of complex neutrinos as outlined in the previous section. The reason that this terminology quasi makes sense is seen by considering what happens when $SU(2)_W \times U(1)_Y$ gets spontaneously broken. We may consider the term

$$\mathcal{L}_{SM} \supset \frac{1}{\Lambda}[\boldsymbol{\phi}^T J \boldsymbol{\nu}]^T C [\boldsymbol{\phi}^T J \boldsymbol{\nu}], \tag{427}$$

where $\boldsymbol{\nu} = (e_L,\, \nu_L)^T$, the matrix multiplication of the two square brackets is over spinor indices, and $\boldsymbol{\phi}$ is the Higgs, with Λ an energy scale needed because of the irrelevance of the operator in question (irrelevant in the full UV theory where the Higgs isn't condensed; after the Higgs gets a vev you might say that it is relevant like any other fermion mass term). The J makes the expression $SU(2)_W$ invariant, and the opposite $U(1)_Y$ charges for the Higgs and the leptons ensure that $U(1)_Y$ is okay too. When SSB happens the first component of $\boldsymbol{\phi}$ gets a vev v, and we get the Majorana mass term

$$\mathcal{L}_{SM} \supset \frac{v^2}{\Lambda} \nu_L^T C \nu_L. \tag{428}$$

So, *after* SSB occurs and the $SU(2)_W$ and $U(1)_Y$ symmetries disappear, we can indeed think of the ν_Ls as Majoranas, with a Majorana mass (this is made possible by their neutrality under $U(1)_{EM}$).

Anyway, from the above discussion, we can really see that neutrinos in cond-mat (real fermions) are rather conceptually distinct from neutrinos in hep-ph (complex fermions that are equal to their charge-conjugates; in the case of neutrinos only after SSB). This is what makes the "solid state physicsists realized Majorana's vision before particle physicists did" spiel so vexing—the two things are totally different.