Lectures on Generalized Symmetries and Phases of Gauge Theory

Thomas T. Dumitrescu

June 26, 2024

[ROUGH DRAFT VERSION: Please report typos/errors/misconceptions to tdumitrescu@physics.ucla.edu]

Lectures at IHES Summer School (June/July 2024) on Symmetries and Anomalies

xxx

Contents

1 Exercise Sheet (45 min/lecture)

1.1 Lecture 1

1.1.1 Exercise: Mean Field Ising Model

Much intuition about phases and transitions can be gleaned from mean-field theory. Let us consider the mean-field (i.e. semiclassical) dynamics of a real scalar field ϕ with a $\mathbb{Z}_2$ Ising symmetry. Since we are working at leading order in the semiclassical expansion, we just minimize the potential and ignore loop corrections. Thus the discussion applies in any spacetime dimension D. (Whether or not this is a good description depends on D.)

- Analyze the vacuum structure as a function of the mass $m^2 \in \mathbb{R}$ given a quartic potential of the form

$$V(\phi) = m^2\phi^2 + \lambda_4\phi^4, \qquad \lambda_4 > 0 \ . \tag{1.1}$$

  In particular discuss the order of the transition at $m^2 = 0$. (In applications to the classical, finite-temperature Ising model $m^2 \sim T - T_c$, but the discussion also applies to quantum phase transitions at zero temperature in the Ising universality class, in which case m^2 is some coupling in the Hamiltonian.)

- Show that the transition can be made 1st order by breaking the $\mathbb{Z}_2$ symmetry via a linear perturbation $\Delta V = h\phi$. In the Ising model $h \in \mathbb{R}$ is an external magnetic field. Sketch the phase diagram as a function of m^2, h. Argue that generically the only way for a line of 1st order phase transition to genuinely end (rather than turn into some other lines(s) of transitions) is in a 2nd order point.

- Consider the Ising model with $\mathbb{Z}_2$ symmetry and a sextic potential,

$$V(\phi) = m^2\phi^2 + \lambda_4\phi^4 + \lambda_6\phi^6 \ . \tag{1.2}$$

  Imagine that $\lambda_6 > 0$, so that the potential is stable, but that $m^2, \lambda_4 \in \mathbb{R}$ can have either sign. Analyze the phase diagram and show that the sign of λ_4 controls the order of the phase transition as we dial m^2. The point $m^2 = \lambda_4 = 0$ at which the order of the phase transition changes is called a multi-critical point. Here it is also called a tri-critical point since we are dialing two parameters (rather than the single parameter to reach a generic critical point). Is the tri-critical point described by the same physics as the line of second-order Ising transitions at $\lambda_4 > 0$?

- The previous point shows that a first order line can change into a second order line at a multi-critical point. Are there other possible behaviors for a 1st order line other than this and ending in a second order point? Hint: think of the phase diagram of water.

1.1.2 Exercise: Abelian Duality in Diverse Dimensions

Generalize the derivation of electric-magnetic duality in $D = 4$ reviewed above to the following settings. For low D these occur in many QFT applications . The case $D > 4$ is interesting in the context of string theory, supergravity, holography etc.

- Start with ordinary $U(1)$ Maxwell theory with field strength $f^{(2)} = da^{(1)}$ as above, but now work in D spacetime dimensions (with $D \geq 3$). Carry out the duality explicitly and show that the dual gauge field $\widetilde{a}^{(D-3)}$ that must be introduced as a Lagrange multiplier is a $(D-3)$-form gauge field with $(D-2)$-form field strength $\widetilde{f}^{(D-2)} = d\widetilde{a}^{(D-3)}$. Spell out explicitly the gauge transformations and flux quantization rule for $\widetilde{a}^{(D-3)}$. Hint: if the general case is confusing, first do $D = 3$.

- Given a compact boson $\chi \sim \chi + 2\pi$ in any dimension D, show how to dualize it into a $D-1$-form gauge field. Hint: this case overlaps with the $D = 3$ limit of the pervious point. In $D \geq 3$ such a a compact boson is necessarily a Nambu-Goldstone boson for its broken shift symmetry, while in $D = 2$ the compact boson is not a Goldstone boson (in agreement with the Coleman-Mermin-Wagner theorem on the absence of spontaneously continuous symmetry breaking in 2d).

- Both Maxwell theory and a compact boson are examples of a p-form gauge field. In general, a p-form gauge field $a^{(p)}$ has field strenght $f^{(p+1)} = da^{(p)}$ and gauge transformations

$$a^{(p)} \to a^{(p)} + d\lambda^{(p-1)} . \tag{1.3}$$

Here $\lambda^{(p-1)}$ is itself a $(p-1)$-form gauge field (defined recursively in p). Thus $\lambda^{(p-1)}$ has integer fluxes on $(p-1)$-cycles,

$$\frac{1}{2\pi}\int_{\Sigma_{p-1}} \lambda^{(p-1)} \in \mathbb{Z} , \tag{1.4}$$

and applying the Dirac argument in this case we learn that $f^{(p+1)}$ has integer fluxes on $(p+1)$-cycles,

$$\frac{1}{2\pi}\int_{\Sigma_{p+1}} f^{(p+1)} \in \mathbb{Z} . \tag{1.5}$$

We take the action to be of generalized Maxwell type:

$$S = \frac{1}{2e^2}\int_{\mathcal{M}_D} f^{(p+1)} \wedge * f^{(p+1)} . \tag{1.6}$$

What is the mass dimension of the coupling e^2? Show that $a^{(p)}$ can be dualized into a $(D-p-2)$-form gauge field $\widetilde{a}^{(D-p-2)}$ with dual field strength $\widetilde{f}^{(D-p-1)} = d\widetilde{a}^{(D-p-2)}$.

1.2 Lecture 2

1.2.1 Derivation of BF Theory

Start with the Abelian Higgs model deep in the Higgs phase ($m^2 \ll 0$), where both the vector boson and the radial mode ρ of the Higgs field $h = \rho e^{i\chi}$ are very massive. The only light mode is the compact scalar $\chi \sim \chi + 2\pi$. As in the problems for Lecture 1, apply Abelian duality to χ to replace it by a dynamical 2-form gauge field $b^{(2)}$. By throwing out all the massive modes, derive the description of the low-energy $\mathbb{Z}_{q_e}$ gauge theory as a BF theory that was introduced in lecture.

1.2.2 Gauging Subgroups of $U(1)^{(1)}_{e,m}$.

A useful application of the BF description of $\mathbb{Z}_{q_e}$ gauge theory that it allows us perform various discrete gauging in a simple way. Work through the following steps to see how this is done:

Let us first consider the following question: what happens if we gauge a $\mathbb{Z}^{(1)}_n \subset U(1)^{(1)}_m$ in the AHM (Note that this is anomaly free.)? To do this we simply promote $B^{(2)}_m \to b^{(2)}_m$ to be dynamical and add a BF term with coefficient n and a new 1-form gauge field $c^{(1)}$,

$$S \supset \frac{i}{2\pi}\int_{\mathcal{M}_4} b_m^{(2)} \wedge f^{(2)} + \frac{in}{2\pi}\int_{\mathcal{M}_4} b_m^{(2)} \wedge dc^{(1)} \ . \tag{1.7}$$

Integrating out $b^{(2)}$ we find that

$$f^{(2)} = n dc^{(1)} \ . \tag{1.8}$$

This shows that the fluxes of $f^{(2)}$ are valued in $2\pi n\mathbb{Z}$, rather then $2\pi\mathbb{Z}$. It is more convenient to solve for $a^{(1)} = nc^{(1)}$, up to gauge transformations, so that the action becomes

$$S[c^{(1)}, h] = \frac{n^2}{2e^2}\int_{\mathcal{M}_4} dc^{(1)} \wedge * dc^{(1)} + \int_{\mathcal{M}_4} \left(|Dh|^2 + V(|h|)\right) \ , \qquad D_\mu = \partial_\mu - inqc_\mu \ . \tag{1.9}$$

Thus we see there are two effects:

- the gauge coupling changes $e \to \frac{e}{n}$.
- The charge of the scalar h changes: $q \to nq$. As a result the electric 1-form symmetry enhances to $\mathbb{Z}_{nq}^{(1)}$ while the magnetic 1-form symmetry remains $U(1)_m^{(1)}$.

Now let us consider the inverse process where we gauge the full $\mathbb{Z}_{q_e}^{(1)}$ electric symmetry (we could also do it for a subgroup). This is done by promoting $B_e^{(2)} \to b_e^{(2)}$ to be dynamical and adding the BF term with coefficient q_e,

$$\frac{iq_e}{2\pi}\int_{\mathcal{M}_4} b_e^{(2)} \wedge dc^{(1)} \ . \tag{1.10}$$

The path integral over the conventionally normalized $U(1)$ gauge field $c^{(1)}$ restricts $b_e^{(2)}$ to be a flat $\mathbb{Z}_{q_e}$ gauge field with holonomies in $\frac{2\pi\mathbb{Z}}{q_e}$. We can therefore write $f^{(2)} - b_e^{(2)} = \frac{1}{q_e} dk^{(1)}$ with $k^{(1)}$ a standard $U(1)$ gauge field. We can then also replace $qa^{(1)} \to k^{(1)}$ in the covariant derivative, leading to the action

$$S[k^{(1)}, h] = \frac{1}{2q^2e^2}\int_{\mathcal{M}_4} dk^{(1)} \wedge * dk^{(1)} + \int_{\mathcal{M}_4} \left(|Dh|^2 + V(|h|)\right) \ , \qquad D_\mu = \partial_\mu - ik_\mu \ . \tag{1.11}$$

Thus we have $e \to qe$ and the charge of the scalar reduces from q to 1, resulting in a theory with no electric 1-form symmetry, but with a magnetic 1-form symmetry. Thus this operation is the exact inverse of the one described before.

There are several important lessons here:

- We can multiply or divide the charge q_e of h by an integer n by gauging a $\mathbb{Z}_n^{(1)}$ subgroup of $U(1)_m^{(1)}$ or $\mathbb{Z}_{q_e}^{(1)}$ electric respectively. This allows us to relate different models of interest.

- gauging a discrete symmetry (0-form or 1-form) commutes with RG flow. Equivalently, discrete gauging does not alter the theory locally, only globally, and the RG flow is about the local dynamics from UV to IR. A more sophisticated argument involves placing the 4d theory on the boundary of a 5d bulk TQFT and implementing the different discrete gaugings via suitable topological boundary conditions in the 5d TQFT.

  The commutativity wit the RG flow means we only have to determine the IR dynamics for one model, e.g. for $q_e = 1$. This model is completely Higgsed in the IR, with no TQFT. There is no electric 1-form symmetry, and the magnetic $U(1)_m^{(1)}$ is unbroken. In fact it is important that the IR also has no non-trivial SPT for the background field $B_m^{(2)}$, since there is no gauge-invariant 4d action that one can write for such a 2-form gauge field. Later we will encounter situations where there are non-trivial SPTs.

  If we now gauge a $\mathbb{Z}_q^{(1)} \subset U(1)_m^{(1)}$ then in the UV we get the ABH model with a charge q scalar, with $\mathbb{Z}_q^{(1)}$ symmetry. In the IR, we take the fully gapped, trivial theory (no SPT for $B_m^{(2)}$, and gauge $\mathbb{Z}_q^{(1)}$ there. This involves making $B_m^{(2)}$ dynamical and adding a BF term with coefficient q. This precisely engineers the $\mathbb{Z}_q$ TQFT that arises in the IR of the charge-q AHM.

- The two gauging procedures above are inverse operations. Thus no essential information is lost as we change the charge q, even though the size of the discrete electric 1-form symmetry changes. (Roughly, this is because the magnetic symmetry $U(1)_m$ is infinite. In discrete cases the magnetic symmetry grows as the electric one shrinks.) This is rather different from gauging continuous symmetry, which is a much more dramatic modification of a theory because it changes the local dynamics. One also loose global symmetries (at least the global symmetry that was gauged and possibly more). Sometimes one gains continuous ones (e.g. gauging $U(1)$ symmetry of complex scalar destroys $U(1)$ global symmetry but leads to $U(1)_m$ 1-form symmetry.)

1.2.3 Gravitational 2-Group

Consider a triangle anomaly involving 2-Stress-tensors and one photon in an abelian gauge theory. Argue that it leads to 2-group where $U(1)_m^{(1)}$ is extended by the Poincaré symmetry. Find an example of a $U(1)$ gauge theory with fermions ψ_i of charge q_i that is gauge-anomaly-free but has a non-trivial mixed anomaly/2-group of this kind.

2 References

- Gaiotto, Kapustin, Seiberg, Willett : "Generalized Global Symmetries"
- McGreevy: "Generalized Symmetries in Condensed Matter" (Review)
- Cordova, Dumitrescu, Intriligator, Shao: "Snowmass White Paper: Generalized Symmetries in Quantum Field Theory and Beyond" (Review)
- Tong: "Lectures on Gauge Theory" (for lots of background on gauge theory and anomalies for 0-form symmetries that I will not cover in detail here).

3 Symmetries, Phases, Landau Paradigm

Symmetries play an important role in physics. Broadly, we encounter two things commonly referred to as symmetries:

1. **Global symmetries:** These act non-trivially on something physical/observable,

   e.g. rotations act on the position operator $\vec{x}$ in quantum mechanics. The Hamiltonian H (equivalently Lagrangian) of the theory may or may not be rotationally symmetric, depending on whether we allow symmetry-violating terms such as $\vec{E} \cdot \vec{x}$ (with a fixed c-number background electric field $\vec{E}$) in H. There is nothing wrong with breaking rotational symmetry, but if it is preserved it has consequences: selection rules for matrix elements, Hilbert space organizes into symmetry multiplets, etc.

   Even approximate symmetries, where the breaking is suppressed by a small parameter, are useful, e.g. they lead to approximate degeneracies/multiplets. Many examples in atomic and nuclear physics. In particle physics, the proton (938 MeV) and the neutron (940 MeV) have approximately the same mass, because they are related to an approximate isospin symmetry. The small mass difference comes from explicit isospin breaking (up-down quark mass difference, electromagnetic contributions to the proton mass) and is responsible for many important phenomena in nuclear physics (finely tuned).

   Things get even more interesting if we do quantum mechanics with an infinite number of degrees of freedom (i.e. QFT), in which case we can have things like spontaneous symmetry breaking, where the symmetry is there and has consequences, just not the naive ones familiar from quantum mechanics.

Another things that happens in QFT is that there is a notion of locality in space and time. This greatly enriches the notion of both operators/observables and symmetries, e.g. we can have standard local operators/fields, e.g. a real scalar field $\phi(x = t, \vec{x})$, which can be acted on by conventional global symmetries, e.g. $\phi \to -\phi$. (In modern parlance these are called zero-form symmetries.) But we can also have operators/defects that live on higher-dimensional submanifolds (lines, surfaces, ...) of spacetime. Correspondingly we have generalized symmetries that act on such extended operators. The simplest such generalized symmetries are called higher-form symmetries (but there are more exotic generalizations than that).

2. **Gauge Symmetries:** These appear as an useful – one might say essential – tool in the construction/description of many theories of physical interest, e.g. it is essential to describe the long-range electromagnetic force in nature in a way that is compatible with unitarity (i.e. quantum mechanics) and locality (i.e. special relativity). They also naturally arise in many condensed matter systems, where they are typically emergent.

   Whenever the notion of a gauge symmetry is sensible, it refers to an exact redundancy of the system we are trying to describe (i.e. we quotient by gauge transformations in the Feynman path integral), not an actually physical symmetry acting on observables. In fact the hallmark of a gauge symmetry is that it does not act on any physical observables, which must all be gauge invariant!

   The fact that gauge symmetries don't act on anything makes it impossible to determine "what the gauge group is" in the abstract. A notion of gauge group make sense in a particular semi-classical limit, but it does not make sense for theories that are strongly quantum mechanical and/or coupled. This fact underlines the rich set of gauge-theory dualities that have emerged over the past decades in field theory, string theory, and condensed matter physics. These dualities show that e.g. the same theory can be described by gauge theories with **different gauge groups**, and some gauge theories have dual descriptions that do not involve any gauge fields at all. (Simple example: Chiral Lagrangian for Pions in QCD.)

So even though physically, we are interested in gauge theories, the tool that will help us understand them better are the global (non-gauge) symmetries, roughly for two (related) reasons:

- Symmetries (and the long list of things they can do) are very rigid structures that can often be tracked form weak to strong coupling. Thus they give insight into strongly coupled physics even if the strong-coupling problem cannot be solved. (If it can be that is very special, e.g. 2d Ising model, integrable models, typically in low dimensions, though planar 4d N=4 SYM is an example in 4d).

- Symmetries characterize phases and phase transitions: Landau's paradigm of symmetry breaking.

  Prototypical example: ising model (lattice), equivalently ϕ^4 theory, with effective Lagrangian

$$\mathscr{L} \sim (\partial\phi)^2 - V(\phi) \; , \qquad V(\phi) = m^2\phi^2 + \lambda\phi^4(\lambda > 0) \; . \tag{3.1}$$

  Note that there is a unitary, internal (i.e. non-spacetime) $\mathbb{Z}_2$ symmetry sending

$$\mathbb{Z}_2 : \phi(x) \to -\phi(x) \; . \tag{3.2}$$

  This is the spin-flip symmetry of the Ising model. Discuss phase diagram as function of $m^2 \in \mathbb{R}$: two phases, one with unique vacua and $\mathbb{Z}_2$ unbroken. In that case $\mathbb{Z}_2$ acts on ϕ and is realized "linearly" a la Wigner (selection rules etc), e.g. $< \phi^{\text{odd}} >= 0$. One with two vacua and spontaneously broken $\mathbb{Z}_2$ symmetry. Which vacuum we are in depends on vev $< \phi >= \pm v$. Now the selection rules don't apply ("non-linear realization of symmetry"), but there is something else that results from the broken symmetry: a finite-tension domain wall separating the two vacua. So the symmetry has consequences even in the broken phase. Note that the regime with $\mathbb{Z}_2$ breaking and with unbroken $\mathbb{Z}_2$ are inevitably separated by a phase transition: Landau paradigm. Note that we cannot in general predict the order (could be 1st order or continuous), but the fact that there is a transition is inescapable.

4 The $O(2)$ Model

Before we move to our first gauge theory example, let us first revisit some basic facts about the $O(2)$ model.

The $O(2)$ model in D spacetime dimensions is describe by a complex-valued scalar field $\phi(x)$ with Lagrangian

$$\mathscr{L} = \partial_\mu\overline{\phi}\partial^\mu\phi - V(|\phi|) \; , \qquad V(|\phi|) = m^2|\phi|^2 + \lambda|\phi|^4 \; , \qquad \lambda > 0 \; . \tag{4.1}$$

Here $m^2 \in \mathbb{R}$ can be any real number, as in the Ising model.

What are the global symmetries of the model?

- There is a continuous $U(1)$ flavor symmetry (aka a zero-form symmetry) under which ϕ has charge $Q = 1$ and thus rotates by phases,

$$e^{i\alpha Q} : \phi(x) \to e^{i\alpha}\phi(x) \ . \tag{4.2}$$

  Noether's theorem gives the associated conserved current:

$$j_\mu = i\overline{\phi}\partial_\mu\phi - i\partial_\mu\overline{\phi}\,\phi \ . \tag{4.3}$$

  One useful consequence of this is that we can couple the theory to a background $U(1)$ gauge field (which is as of yet non-dynamical):

$$\Delta\mathscr{L} = A^\mu j_\mu \ , \qquad \partial_\mu \to D_\mu = \partial_\mu - iqA_\mu \ . \tag{4.4}$$

  The coupling to background gauge fields for global symmetries will be a useful tool for us, e.g. it allows us to talk about **anomalies and SPTs**.

- There is a $\mathbb{Z}_2$ charge-conjugation symmetry

$$C : \phi(x) \to \overline{\phi}(x) \ . \tag{4.5}$$

  This is also a unitary zero-form symmetry. Together with $U(1)$, this make the symmetry

$$O(2) = U(1) \rtimes C \ . \tag{4.6}$$

  Hence the name $O(2)$ model.

- There is the full Poincare symmetry, associated with a conserved symmetric stress-energy tensor $T_{\mu\nu}$, and also discrete spacetime symmetries: parity (unitary) and time-reversal (anti-unitary). Will have more to say about them in the future. Note that the existence of $T_{\mu\nu}$ allows us to couple the theory to a non-dynamical background metric $g_{\mu\nu}$, and in fact to study it on arbitrary spacetime manifolds (Euclidean or Lorentzian). This is a very useful tool.

Sketch the phase diagram as a function of m^2:

- When $m^2 > 0$ there is a unique vacuum, all the symmetries are unbroken and linearly realized. the massive ϕ particle transforms in a representation of these. The conserved charge $Q = \int d^3x j_0$ is a well-defined operator that annihilates the vacuum and generates the unbroken $U(1)$ symmetry.

- When $m^2 < 0$ is sufficiently negative, then $< \phi >= v \in \mathbb{C}^*$ gets a vev, which spontaneously breaks the $U(1)$ symmetry to nothing, leading to a circle of vacua and a massless Nambu-Goldstone Boson (NBG) χ. At long distances,

$$\phi(x) \simeq ve^{i\chi(x)} . \tag{4.7}$$

  This means that the current flows to $j_\mu \sim v\partial_\mu\chi$, so that acting with the current on the vacuum produces a single, well-defined NGB state (note a 2-particle or multi-particle state). This is a hallmark of broken currents for internal/flavor symmetries. Note that the NBG $\chi \sim \chi+2\pi$ is a compact scalar field. This follows from the fact that the broken symmetry is $U(1)$, which is parameterized by a compact angle. The compactness of χ will have important consequences for us later.

  Note that integrating the charge-density j_0 over a spatial slice gives the conserved charge Q, but in the symmetry breaking phase this charge correspond to a goldstone bosons of exactly zero-momentum/frequency, which is not a normalizable state in the Hilbert space. Thus the charge operator Q does not exist in the broken phase. However commutators of Q with other operators, as well as the current j_μ exist and are interesting objects to analyze ("current algebra").

Again we learn that the $m^2 = \pm\infty$ phases must be separated by a transition, since one breaks the $U(1)$ symmetry while the other does not.

5 Ordinary (0-Form) Symmetry Basics

The most familiar/ubiquitous kind is an ordinary (0-form) symmetry:

- Continuous flavor symmetry with conserved Noether current j_μ leading to codimension-1 charge defects. The simplest version of such a defect is the standard conserved charge Q obtained by integrating the current over a fixed time-slice:

$$Q = \int d^3x \, j_0 . \tag{5.1}$$

Here current conservation implies that $[Q, H] \sim \dot{Q} = 0$. A useful generalization appears in Euclidean signature, where we can integrate the normal component of the current over an arbitrary closed three-manifolds Σ_3,

$$Q[\Sigma_3] = \int_{\Sigma_3} j_\perp \ . \tag{5.2}$$

Now current conservation and Gauss' theorem implies that $Q[\Sigma_3]$ is independent of Σ_3, i.e. we can deform Σ_3 by a small amount, and as long as we do not encounter other operators/defects this does not change the charge. Thus $Q[\Sigma_3]$ is topological, in the sense that only the topology of Σ_3 matters.

Using the charge $Q[\Sigma_3]$ one can defined an exponentiated topological defect for every element $g = e^{i\alpha} \in U(1)$ of the symmetry group:

$$U(g, \Sigma_3) = \exp\left(i\alpha Q[\Sigma_3]\right) \ . \tag{5.3}$$

Note that $U(g, \Sigma_3)U(g', \Sigma_3) = U(gg', \Sigma_3)$ so that these defects obey the group law. Here we are considering an abelian symmetry for simplicity, but the statement is true in general.

If a local operator $\mathcal{O}(x)$ carries charge $q \in \mathbb{Z}$ under the symmetry, this charge can be detected in various ways:

- By evaluating the commutator $[Q, \mathcal{O}(x)] = iq\mathcal{O}(x)$ in Lorentzian signature.
- By encircling $\mathcal{O}(x)$ with the topological surface operator $Q[\Sigma_3]$ in Euclidean signature. If Σ_3 links the point x where $\mathcal{O}$ resides exactly once (e.g. take Σ_3 to be a standard three-sphere with center at x) then

$$Q[\Sigma_3]\mathcal{O}(x) = q\mathcal{O}(x) \ . \tag{5.4}$$

- By studying the contact term in the OPE $\partial^\mu j_\mu(x)\mathcal{O}(0) \sim q\delta(x)\mathcal{O}(0)$, which in turn integrates up to a non-contact term in the OPE $j_\mu(x)\mathcal{O}(x) \sim \frac{q}{x^{d-1}}\mathcal{O}(0)$ (here I am being sloppy about the tensor structure and am only indicating the scaling with powers of x).

If we act with a local operator $\mathcal{O}(x)$ (suitably smeared) on the vacuum $|0\rangle$, and we assume that $Q|0\rangle = 0$ with well-defined Q, so that the symmetry is not spontaneously broken, then we create a normalizable state of charge q living in the Hilbert space of the theory. To detect this charge we must act with Q on the state, i.e. we must integrate over an entire time-slice, since charge can in principle reside at every point in space. In simple cases the state $\mathcal{O}|0\rangle$ will be a single-particle state. If we think of the 1d worldline of the particle so created, then the spatial integral needed to evaluate Q cuts it transversely.

- Continuous spacetime symmetry with conserved stress tensor $T_{\mu\nu}$.
- Discrete symmetries (no currents, but codimension-1 defects)

Given the utility of 0-form symmetries, there have been many attempts at generalizations:

- Higher spin currents $j_{(\mu\nu\rho\cdots)}$, generalizing $j_\mu, T_{(\mu\nu)}$. In $d > 2$ spacetime dimensions this is highly constrained by the Coleman-Mandula theorem. Interesting exceptions: CFTs and SUSY theories.
- Higher-form symmetries. These do not act on local operators/particles but rather on line defects/strings (evade Coleman-Mandula theorem). This is why they naturally and frequently arise in gauge theories!
- There are also non-invertible/categoricaal versions of 0-form and higher form symmetries (other lectures). I will focus on the invertible case: higher-form and higher-group symmetries – in particular the first non-trivial case, i.e. 1-form and 2-group symmetries.

Goal: study 4d gauge theory examples of increasing complexity through the lens of these global symmetries and see what we can learn.

6 Free Maxwell Theory

Prototypical example of gauge theory with 1-form symmetries.

Many close analogies with the compact ($c = 1$) boson in 2d. Roughly, 4d Maxwell theory is to 1-form symmetries as the 2d compact boson is to 0-form symmetries.

We will work in Euclidean signature. Therefore without specific mention to the contrary, we will consider Euclidean path integrals with suitable insertions/defects along points, lines, surfaces etc., without thinking about states and operators in Hilbert space. Nevertheless we will be sloppy and refer to such insertions as operators. This sloppiness is standard in the context of Lorentz-invariant theories, because it does not cause any trouble there: every operator can be a defect and vice versa, depending on how we orient it in spacetime.

6.1 Maxwell Basics

The action in flat spacetime $\mathcal{M}_4 = \mathbb{R}^4$ is

$$S[a^{(1)}] = \frac{1}{2e^2} \int_{\mathcal{M}_4} f^{(2)} \wedge * f^{(2)} = \frac{1}{4e^2} \int d^4x \, f^{\mu\nu} f_{\mu\nu} \ . \tag{6.1}$$

Comments:

- Superscript (p) denotes a p-form, e.g. gauge field is a 1-form,

  $$a^{(1)} = a_\mu dx^\mu \ , \tag{6.2}$$

  and its field strength is a 2-form,

  $$f^{(2)} = da^{(1)} \ , \qquad f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu \ . \tag{6.3}$$

  Theen $f^{(2)}$ is closed (Bianchi identity),

  $$df^{(2)} = 0 \ . \tag{6.4}$$

- The other two Maxwell equations come from varying $a^{(1)}$,

  $$d * f^{(2)} = 0 \ . \tag{6.5}$$

- Gauge "symmetry" – more precisely gauge redundancy (does not act on any physical quantities),

  $$a^{(1)} \to a^{(1)} + d\lambda \ , \qquad \lambda \sim \lambda + 2\pi \ . \tag{6.6}$$

  Here $\lambda(x)$ is a circle-valued gauge parameter (function of spacetime). With these gauge transformations, a_μ is a conventional $U(1)$ gauge field (or connection). The standard Dirac argument links the periodicity of λ to the allowed fluxes:

  $$\frac{1}{2\pi} \int_{\Sigma_2} f^{(2)} \in \mathbb{Z} \ . \tag{6.7}$$

  Here Σ_2 is any closed, oriented surface (e.g. S^2). In the path integral we integrate over distinct gauge orbits of a_μ, i.e. we divide out by gauge transformations.

- We cannot rescale a_μ without modifying the periodicity of λ or the integrality of the fluxes. This means that the gauge coupling e is meaningful even in the free theory (c.f. radius of the compact boson in 2d).

- 0-form symmetries of Maxwell theory:
  - Spactime translations and $SO(4)$ rotations (bosonic theory, no local operators in spinor representations of $Spin(4)$). Continuous symmetries with conserved current $T_{\mu\nu}$.
  - $\mathbb{Z}_2$ charge conjugation
$$\mathsf{C} : a_\mu \to -a_\mu \ . \tag{6.8}$$
  - In Lorentzian signature there are parity P (unitary) and time-reversal T (anti-unitary) symmetries. By the CPT theorem these are not independent. This is because in Euclidean signature, both amount to a reflection symmetry R, which enhances $SO(4)$ to $O(4)$.
- We can couple Maxwell theory to any Riemannian metric $g_{\mu\nu}$ and take $\mathcal{M}_4$ to be any (sufficiently smooth) 4-manifold. If we only use $SO(4)$ symmetry then $\mathcal{M}_4$ must be oriented. Using the $O(4)$ symmetry of Maxwell theory (which involves parity/time-reversal) we can even generalize to non-orientable manifolds. One should think of $\mathcal{M}_4$ and $g_{\mu\nu}$ as background fields for the spacetime symmetries, just as bundles and connections are background fields for internal symmetries. Unless stated otherwise we assume that $\mathcal{M}_4$ is oriented.

6.2 Electric-Magnetic Duality

Free Maxwell theory has many equivalent presentations. An important/useful one is related to the original presentation by electric-magnetic duality or S-duality (analogous to $R \to \frac{1}{R}$ duality of the compact boson). To derive it we start with[1]

$$S = \frac{1}{2e^2} \int_{\mathcal{M}_4} f^{(2)} \wedge * f^{(2)} \ , \tag{6.9}$$

but instead of trivilaizing the constraints

$$df^{(2)} = 0 \ , \qquad \frac{1}{2\pi} \int_{\Sigma_2} f^{(2)} \in \mathbb{Z} \ , \tag{6.10}$$

[1] Equivalently, we can start with a formulation of the theory (e.g. the one due to Sen and Schwarz) in which duality is manifest. This is analogous to the Luttinger liquid presentation of the compact boson in 2d.

by writing $f^{(2)} = da^{(1)}$ in terms of a $U(1)$ connection, we impose these constraints (in an equally local way) using a Lagrange multiplier $\widetilde{a}^{(1)}$,

$$S[f^{(2)}, \widetilde{a}^{(1)}] = \frac{1}{2e^2} \int_{\mathcal{M}_4} f^{(2)} \wedge * f^{(2)} + \frac{i}{2\pi} \int_{\mathcal{M}_4} d\widetilde{a}^{(1)} \wedge f^{(2)} \ . \tag{6.11}$$

Now $f^{(2)}$ is an unconstrained 2-form, while $\widetilde{a}^{(1)}$ is the Lagrange multiplier field, which is itself a conventionally normalized $U(1)$ connection,

$$\widetilde{a}^{(1)} \to \widetilde{a}^{(1)} + d\widetilde{\lambda}, \qquad \widetilde{\lambda} \sim \widetilde{\lambda} + 2\pi \ , \qquad \frac{1}{2\pi} \int_{\Sigma_2} \widetilde{f}^{(2)} \in \mathbb{Z} \ . \tag{6.12}$$

The path integral over $\widetilde{a}^{(1)}$ can be done exactly: it is a path-integral Fourier representation for a δ-function that sets $df^{(2)} = 0$ (this comes from the $\widetilde{a}^{(1)}$ equation of motion, which are enforced by path integrating over its Gaussian fluctuations in every flux sector) and restricts the fluxes of $f^{(2)}$ to lie in $2\pi\mathbb{Z}$ (this comes from summing over all possible fluxes of $\widetilde{a}^{(1)}$). A simple example to keep in mind is $\mathcal{M}_4 = S^2 \times S^2$. There the $\widetilde{a}^{(1)}$ flux sum over one S^2 enforces $f^{(2)}$ flux quantization on the other (Poincaré dual) sphere, but it works on any (oriented) four-manifold. Thus integrating over $\widetilde{a}^{(1)}$ leads to the original presentation of Maxwell theory.

The magnetic dual description is obtained by integrating out the unconstrained $f^{(2)}$,

$$\frac{i}{e^2} * f^{(2)} = \frac{1}{2\pi} \widetilde{f}^{(2)} \ , \qquad \widetilde{f}^{(2)} = d\widetilde{a}^{(1)} \ . \tag{6.13}$$

The regrettable but essential factor of i comes about because we are in Euclidean signature, where electric fields are naturally imaginary and magnetic fields naturally real. It is absent in Lorentzian signature, where the same equation reads $\star_L f^{(2)} \sim \widetilde{f}^{(2)}$, which implies that the conventional, real, Lorentzian electromagnetic fields are related via

$$E_i \sim \widetilde{B}_i \ , \qquad B_i \sim -\widetilde{E}_i \ . \tag{6.14}$$

We can now eliminate $f^{(2)}$ from the action to find

$$S[\widetilde{a}^{(1)}] = \frac{1}{2\widetilde{e}^2} \int \widetilde{f}^{(2)} \wedge * \widetilde{f}^{(2)} \ , \qquad \widetilde{e}^2 = \frac{4\pi^2}{e^2} \ . \tag{6.15}$$

For generic values of e the S-duality operation is a change of presentation, not a symmetry. At the self-dual value $e^2 = \widetilde{e}^2 = 2\pi$ the model is invariant under S, which is therefore a symmetry. It is a 0-form symmetry that acts on the local field-strength operator,[2]

$$S : f^{(2)} \quad \rightarrow \quad \widetilde{f}^{(2)} = i * f^{(2)} \ . \tag{6.16}$$

Note that it is not a $\mathbb{Z}_2$ symmetry, but rather a $\mathbb{Z}_4$ symmetry, since $S^2(f^{(2)}) = - *^2 f^{(2)} = -f^{(2)}$. In particular, $S^2 = \mathsf{C}$. (Similarly, $R \to \frac{1}{R}$ duality of the compact boson is also a $\mathbb{Z}_4$ symmetry at the self-dual radius.)

6.3 1-form Symmetries

A continuous p-form global symmetry is associated with a conserved $(p+1)$-form current,

$$d * J^{(p+1)} = 0 \qquad \Longleftrightarrow \qquad \partial^\mu J_{[\mu\nu_2\cdots\nu_{p+1}]} = 0 \ . \tag{6.17}$$

We can then integrate over a co-dimension $(p+1)$-cycle Σ_{d-p-1} to construct the corresponding p-form charge,

$$Q^{(p)} = \int_{\Sigma_{d-p-1}} *J^{(p+1)} \ . \tag{6.18}$$

The conservation equation implies that the dependence on Σ_{d-p-1} in spacetime is topological. In particular, this means that $Q^{(p)}$ is invariant under time-translations and hence conserved.

Let us focus on the case $p = 1$, i.e. 1-form symmetries, which are relevant for Maxwell theory, which has two closed 2-form operators,

$$d * f^{(2)} = df^{(2)} = 0 \ . \tag{6.19}$$

The correctly normalized 2-form currents are given by

$$* J_e^{(2)} = \frac{i}{e^2} * f^{(2)} \ , \qquad *J_m^{(2)} = \frac{1}{2\pi} f^{(2)} \ , \tag{6.20}$$

which allows us to define two conserved 1-form charges,

$$Q_e^{(1)} = \frac{i}{e^2} \int_{\Sigma_2} *f^{(2)} \ , \qquad Q_m^{(1)} = \frac{1}{2\pi} \int_{\Sigma_2} f^{(2)} \ . \tag{6.21}$$

In our (Euclidean) conventions, these are precisely the electric and magnetic Gaussian flux integrals over Σ_2, which measure the total amount of charge enclosed. (We will confirm this

[2] In Lorentzian signature $S(f^{(2)}) = *_L f^{(2)}$, where $*_L$ is the Lorentzian Hodge star operator that satisfies $*_L^2 = -1$ on 2-forms.

below.) Note that S-duality acts via

$$S \ : \ Q_e^{(1)} \to \widetilde{Q}_m^{(1)} \ , \qquad Q_b^{(1)} \to -\widetilde{Q}_e^{(1)} \ . \tag{6.22}$$

In $U(1)$ gauge theory the charges $Q_{e,m}^{(1)}$ above are integers, and hence they generate compact $U(1)_{e,m}^{(1)}$ symmetries, which are generated by the following codimension-2 topological defects (sometimes called symmetry defects),

$$\begin{aligned} U_e^{(1)}(\theta, \Sigma_2) &= \exp\left(-i\theta Q_e^{(1)}[\Sigma_2]\right) = \exp\left(\frac{\theta}{e^2}\int_{\Sigma_2} *f^{(2)}\right) \ , \\ U_m^{(1)}(\theta, \Sigma_2) &= \exp\left(-i\theta Q_m^{(1)}[\Sigma_2]\right) = \exp\left(-\frac{i\theta}{2\pi}\int_{\Sigma_2} f^{(2)}\right) \ . \end{aligned} \tag{6.23}$$

Here $\theta \sim \theta + 2\pi$ is an angle which parameterizes a group element in $U(1)_{e,m}^{(1)}$ respectively.

The structure worked out above is only when we have continuous 1-form symmetries with associated 2-form currents. However, the symmetry defects also exist for discrete 1-form symmetries, with θ replaced by an element of the discrete group.

Comments:

- By virtue of their codimension, 1-form symmetry defects can always moved past each other, so that the corresponding 1-form symmetries commute.
- In gauge theories one typically finds electric 1-form symmetries associated with the center of the gauge group, sometimes called center symmetries. In the case of free Maxwell theory this is the $U(1)_e^{(1)}$ symmetry. Note that the notion of center symmetry depends on the presentation of the theory, e.g. it is not invariant under duality (c.f. momentum and winding symmetry in the 2d compact boson).

6.4 Wilson and 't Hooft Lines

In Euclidean signature, topological defects such as $Q_{e,m}^{(1)}$ or $U_{e,m}^{(1)}(\theta, \Sigma_2)$ act on other operators/defects by linking. For instance, codimension-1 charge defects associated with a 0-form symmetry link local (point) operators in spacetime.

By contrast the 1-form charges can link with line-defects, but not with local operators. The former can therefore be charged under 1-form symmetries, while the latter cannot. (This has important consequences that we will return to.)

The line defects that are charged under $Q_{e,m}^{(1)}$ are electrically charged Wilson lines and magnetically charged 't Hooft lines:

- A Wilson line of charge q_e along a closed curve C is defined as follows,

$$W_{q_e}(C) = \exp\left(iq_e \int_C a^{(1)} \right) \ . \tag{6.24}$$

Note that this can be thought of as the worldline of a charged particle traversing C. Gauge invariance under 2π-periodic gauge transformations requiers $q_e \in \mathbb{Z}$. It can be checked by direct computation[3] that

$$Q_e^{(1)}\left(W_{q_e}(C)\right) = q_e W_{q_e}(C) \ , \qquad Q_m^{(1)}\left(W_{q_e}(C)\right) = 0 \ . \tag{6.25}$$

Here we assume that the surface Σ_2 over which we evaluate $Q_{e,m}^{(1)}$ is a small sphere linking C exactly once.

We will soon obtain this result by simpler, but less direct route. The Wilson line thus represents an infinitely massive, non-dynamical probe particle of electric charge q_e traversing C in spacetime.

- 't Hooft lines $H_{q_m}(C)$ along C represent non-dynamical probe particles of magnetic charge q_m traversing C. They are defined by imposing the condition that at each point on C the magnetic field has a Dirac monopole singularity of charge q_m. A more careful regularization involves drilling a small hole (e.g. in the shape of a ball B^3, with boundary $\partial B^3 = S^2$) around the Dirac singularity and imposing the condition that the magnetic flux across the boundary of the hole is q_m. This is precisely the quantity measured by $Q_m^{(1)}$, so that

$$Q_m^{(1)}\left(H_{q_m}(C)\right) = q_m H_{q_m}(C) \ , \qquad Q_e^{(1)}\left(H_{q_m}(C)\right) = 0 \ . \tag{6.26}$$

Note that Wilson and 't Hooft lines are exchanged under electric-magnetic duality, so that the 't Hooft lines $H_{q_n}(C)$ can be thought of as Wilson lines for the magnetic dual gauge field $\widetilde{a}^{(1)}$.

- Fusing $W_{q_e}(C)$ and $H_{q_m}(C)$ leads to a general dyonic line of electromagnetic/1-form charges (q_e, q_m). Note that while pure electric or magnetic lines are naturally bosonic, dyons can be fermions if the Dirac angular momentum $\frac{q_e q_m}{2}$ is half-integral. For instance this is the case for the minimal $(1,1)$ dyon that results from fusing the fundamental $q_e = q_m = 1$ Wilson and 't Hooft lines. As we will explain in more detail below, it is really

[3] The theory is free, so all correlators can be evaluated by saddle point.

the statistics of a line (i.e. whether it is a boson or a fermion, correlated with whether its spin is in $\mathbb{Z}$ or $\mathbb{Z}+\frac{1}{2}$) that is a meaningful, scheme-independent property of the line.

[KEY FACT ABOUT 1-FORM SYMMETRIES:]

The fact that the electric and the magnetic flux surrounding the lines is conserved implies that the charged lines that are sourcing this flux cannot end. For otherwise one could slide the topological flux surface past the endpoint of the line, where there is no more flux. **[DRAW PICTURE. 1-form symmetries are flux conservation symmetries!]**

We will soon see that the presence of electrically or magnetically charged matter fields will enable certain lines to end, and this will break some or all of the 1-form symmetry of Maxwell theory.

[END OF LECTURE 1]

6.5 Coulomb Phase as Spontaneous 1-Form Symmetry Breaking

One common way to characterize phases of gauge theory is via the potential energy $V_{Q\overline{Q}}(r)$ between a charge Q and its CPT conjugate $\overline{Q}$. As we will see, the qualitative behavior of this potential is related to the realization of the 1-form symmetry in the IR, i.e. whether it is spontaneously broken or not. Note that here Q may be a dynamical charge in the theory (e.g. an electron or positron in QED), or it could be an external, non-dynamical, infinitely heavy probe charge.

The corresponding concept for 0-form symmetries is central to the study of phases and phase transitions (Landau paradigm): to phases with different realizations of the symmetry must be separated by a phase transition. Any spontaneously broken 0-form symmetry leads to massless Nambu-Goldstone bosons (NBGs), which are created by the 1-form currents associated with the broken symmetry generators. We will here see the analgoue of this for 1-form symmetries.

Let us first discuss the static $Q\overline{Q}$ potential for electric and magnetic charges,

$$V_{q_e\bar{q}_e}(r) \sim -\frac{e^2 q_e^2}{r} , \qquad V_{q_m\bar{q}_m}(r) \sim -\frac{\widetilde{e}^2 q_m^2}{r} \sim -\frac{q_m^2}{e^2 r} . \tag{6.27}$$

Here the computation for the magnetic case is best done in the magnetic dual description. We recognize the standard electric and magnetic Coulomb potentials. For this reason free Maxwell theory is said to be in a Coulomb phase.

Let us now switch to the seemingly unrelated question of whether or not the $U(1)^{(1)}_{e,m}$ symmetries of Maxwell theory are spontaneously broken or not. (We will soon see that these two questions are in fact related.) We will argue from two point of view (one fast and less so) that the 1-form symmetries are in fact spontaneously broken:

1.) The fast way to see this is to note that the charges $Q^{(1)}_{e,m}$ are linear in the free photon field $f^{(2)}$. Thus they create 1-photon states (of zero-momentum and hence not normalizable), and thus they do not annihilated the vacuum. This shows that they are spontaneously broken. (In principle this leaves open the possibility that ad discrete subgroup might remain unbroken. We will rule this out momentarily.)

2.) Another way to check whether a symmetry is spontaneously broken in a given vacuum is to exhibit a charged operator $\mathcal{O}$ that has a vacuum expectation value (vev). Rather than examining $\langle \mathcal{O} \rangle$ directly, it is often more convenient to examine the correlator $\langle \mathcal{O}^\dagger \mathcal{O} \rangle$ as we separate the operators by a large distance. If the result is non-zero, then by cluster decomposition $\langle \mathcal{O} \rangle \neq 0$.

   Example: in the $O(2)$ model we could take $\mathcal{O} = \phi$ to be the fundamental scalar field. Then looking at $\langle \phi \rangle$ is one way to determine whether the symmetry is broken or note. But we can also instead look at $\langle \phi^\dagger(x)\phi(0) \rangle$ in the limit $x^\mu \to \infty$ and see whether the correlation function decays to zero or not, the latter case implying SSB.

   For line defects the analogue of the $\mathcal{O}^\dagger \mathcal{O}$ correlator is a line defect $L(C)$ along a large loop C,

$$\langle L(C) \rangle \ , \qquad C \to \infty \ . \tag{6.28}$$

   Here it does not matter precisely in which way the loop is taken to infinity (e.g. we could uniformly scale up the contour C), much as it does not matter precisely how $\mathcal{O}$ and $\mathcal{O}^\dagger$ are taken to infinity either. If $\langle L(C \to \infty) \rangle \neq 0$, then the line defect L has a non-zero vev.

   An important subtlety is that a loop operator can be modified by local counterterms along the loop (much like a local operator $\mathcal{O}$ can be rescaled by a wavefunction renormalization constant, e.g. $\phi \to Z\phi$). The simplest such counterterm is the length of the loop, also known as its perimeter $P(C) = \int_C ds$. Other counterterms involve the extrinsic curvature of the loop and can be shown to be irrelevant for our conclusions below.[4]

[4] A term involving a single power of the extrinsic curvature κ is marginal since κds is dimensionless. However it can be shown that this is only fully covariant local counterterm in 2d, roughly because lines are codimension-1 there. (In general the extrinsic curvature K^i_{ab} is valued in the normal bundle of the curve, which is $d-1$ dimensional, thus no fully local counterterm containing a single power of K^i_{ab} can be

We can thus redefine

$$L(C) \to e^{-MP(C)} L(C) \ . \tag{6.29}$$

Here M is a (finite or infinite, positive or negative) constant with dimensions of mass. It can be thought of as a mass renormalization for the (infinitely) heavy probe particle represented by $L(C)$. Thus as long as a loop operator behaves as

$$\langle L(C) \rangle \sim (\text{const.}) e^{-MP(C)} \ , \qquad C \to \infty \ , \tag{6.30}$$

then $L(C)$ has a non-zero vev after a suitable redefinition of the operator.

By contrast, if $\langle L(C) \rangle$ decays faster than perimeter law for large loops C, we say that $L(C)$ has zero vev.

Let us relate this to our previous discussion of the $Q\overline{Q}$ effective potential $V_{Q\overline{Q}}(r)$, which can be extracted from a large rectangular Q-loop of width r and height T = euclidean time, with vev

$$\langle \Box(r,T) \rangle = \exp\left(-TV_{Q\overline{Q}}(r) - MP(\Box)\right) \tag{6.31}$$

Here we have already accounted for a possibly perimeter counterterm. Note that the section of the perimeter along the T-direction shifts $\Delta V \sim M$ by a constant, reflecting the ambiguous additive scale of the potential.

Since the Coulomb potentials between static charges vanish, so that $V_{Q\overline{Q}} \to$ const., we see that all loops in Maxwell theory obey perimeter law, i.e. they have a vev.

A similar (and very general) result holds for circular Wilson loops: consider any conformal field theory (CFT), of which free Maxwell theory is an example. Then conformal symmetry forces the $Q\overline{Q}$ potential associated with any line defect L to be Coulomb like (e.g. this also happens in $\mathcal{N} = 4$ SYM),

$$V(r) = \frac{\alpha}{r} \ , \qquad \alpha = \text{const.} \ . \tag{6.32}$$

Up to scheme-dependent perimeter terms, the vev of the circular loop $L(C)$ (the radius

constructed.) Moreover, CPT requires that orientation reversal complex conjugates the loop and this implies that the extrinsic curvature counterterm is purely imaginary, i.e. it would only affect the phase, not the magnitude of the loop. Finally, terms $\sim \kappa^n ds$ with $n > 1$ vanish for large loops and can be ignored in our discussion below.

does not matter by conformal symmetry) is

$$\langle L(C) \rangle = \exp(-\alpha) \ , \tag{6.33}$$

Recall that in Maxwell theory $\alpha \sim e^2 q_e^2$ for charge-q_e Wilson loops and $\alpha \sim \widetilde{e}^2 q_m^2$ for charge-q_m 't Hooft loops.

The discussion above shows that all loops in Maxwell theory have a vev, and hence both $U(1)^{(1)}_{e,m}$ symmetries of the theory are completely spontaneously broken, i.e. there is no unbroken subgroup.

6.6 Background Fields and Anomalies

An essential tool in the study of 0-form symmetries and their anomalies is the introduction of suitable background gauge fields. For instance, to study a standard $U(1)$ current j_μ it is useful to couple it to a background $U(1)$ gauge field A_μ, via

$$\Delta \mathscr{L} \sim A_\mu j^\mu + \cdots \ . \tag{6.34}$$

The resulting partition function $Z[A]$ generates correlators of j_μ, and it is a useful tool for detecting anomalies and other subtle aspects of those correlators.

The appropriate generalization of A_μ for a higher-form symmetry is a totally anti-symemtric abelian gauge field $B_{[\mu_1 \cdots \mu_n]}$. Such fields are prominent in string theory. In $D < 5$ dynamical fields of this kind can always be dualized into ordinary scalars and Maxwell fields, which is why they are less common there. However, duality is only a statement about dynamical gauge fields, i.e. it is an on-shell statement. It does not apply to background fields. Thus there is an interesting role for background B-fields even in $D = 2, 3, 4$.

For 1-form symmetries such background fields are 2-form gauge fields,

$$B^{(2)}_{e,m} \to B^{(2)}_{e,m} + d\Lambda^{(1)}_{e,m} \ . \tag{6.35}$$

Here $\Lambda^{(1)}_{e,m}$ are themselves (locally defined) 1-form gauge fields, with standard $U(1)$ gauge transformations and fluxes,

$$\frac{1}{2\pi} \int_{\Sigma_2} d\Lambda^{(1)}_{e,m} \in \mathbb{Z} \ . \tag{6.36}$$

2-form gauge fields such as $B^{(2)}_{e,m}$ are familiar from string theory and supergravity, but there they are typically dynamical (unless we take a decoupling limit, e.g. to engineer a standard

QFT in string theory). Here they are (at least for now) non-dynamical background fields.

We claim that the way to couple $B_e^{(2)}$ and $B_m^{(2)}$ to Maxwell theory is as follows,

$$S[a^{(1)}, B_e^{(2)}, B_m^{(2)}] = \frac{1}{2e^2}\int_{\mathcal{M}_4} (f^{(2)} - B_e^{(2)}) \wedge *(f^{(2)} - B_e^{(2)}) + \frac{i}{2\pi}\int_{\mathcal{M}_4} B_m^{(2)} \wedge f^{(2)} \ . \tag{6.37}$$

Note that the terms linear in the background fields precisely couple to the 2-form currents via

$$S\Big|_{\mathcal{O}(B)} = i\int_{\mathcal{M}_4} \left(B_e^{(2)} \wedge *J_e^{(2)} + B_m^{(2)} \wedge *J_m^{(2)}\right) \ . \tag{6.38}$$

The quadratic counterterm $\sim B_e^{(2)} \wedge *B_e^{(2)}$ (sometimes known as a seagull term) ensures that the first term in the action is invariant under $\Lambda_e^{(1)}$ gauge transformations if we also transform

$$a^{(1)} \to a^{(1)} + \Lambda_e^{(1)} \ . \tag{6.39}$$

Note that this is consistent since both $a^{(1)}$ and $\Lambda_e^{(1)}$ are conventionally quantized $U(1)$ gauge fields. Note that the actual global symmetry, which leaves the background field $B_e^{(2)}$ invariant, corresponds to a flat 1-form gauge parameter $d\Lambda_e^{(1)} = 0$. Clearly $f^{(2)}$ is also invariant under such a shift, while the Wilson loop transforms as follows,

$$W_{q_e}(C) \to \exp\left(iq_e\int_C \Lambda_e^{(1)}\right) W_{q_e}(C) \ . \tag{6.40}$$

Since $\Lambda_e^{(1)}$ is flat, this phase can only be non-trivial if C is non-contractible, either because it is a non-trivial cycle in the geometry of $\mathcal{M}_4$, or because it links some other defect.

We now consider the invariance of the second term in the action above,

$$S_{BF} = \frac{i}{2\pi}\int B_m^{(2)} \wedge f^{(2)} \ . \tag{6.41}$$

This is a so-called BF term; it is somewhat analogous to a Chern-Simons term. It is invariant under $B_m^{(2)} \to B_m^{(2)} + d\Lambda_m^{(1)}$, since both $d\Lambda_m^{(1)}$ and $f^{(2)}$ have fluxes in $2\pi\mathbb{Z}$. (In general, gauge invariance requires this term to have an integer coefficient, but we normalize $B_m^{(2)}$ so that this coefficient is 1.)

Consider an 't Hooft line $H_{q_m}(C)$ of charge q_m. Then the flux integral $\int f^{(2)}$ on a small sphere linking C is $2\pi q_m$, so that the BF term above collapses to

$$S_{BF} = iq_m\int_{\Sigma_2} B_m^{(2)} \ , \qquad \partial\Sigma_2 = C \ . \tag{6.42}$$

This shows that under a gauge-transformation $B_m^{(2)} \to B_m^{(2)} + d\Lambda_m^{(1)}$, the 't Hooft line transforms as follows,

$$H_{q_m}(C) \to \exp\left(-iq_m \int_C \Lambda_m^{(1)}\right) H_{q_m}(C) \ . \tag{6.43}$$

In particular this means that the magnetic dual gauge fields shifts as

$$\widetilde{a}^{(1)} \to \widetilde{a}^{(1)} + \Lambda_m^{(1)} \ . \tag{6.44}$$

Since S_{BF} is invariant under magnetic background gauge transformations, and the Maxwell kinetic term is invariant under electric ones, it follows that the full variation of the action is entirely due to the electric background gauge transformation of the BF term,

$$S[a^{(1)} + \Lambda_e^{(1)}, B_e^{(2)} + d\Lambda_e^{(1)}, B_m^{(2)} + d\Lambda_m^{(1)}] = S[a^{(1)}, B_{e,m}^{(2)}] + \frac{i}{2\pi}\int_{\mathcal{M}_4} B_m^{(2)} \wedge d\Lambda_e^{(1)} \ . \tag{6.45}$$

Thus the action is invariant, up to a c-number shift that only depends on the background fields. This is the hallmark of an 't Hooft anomaly for the global electric and magnetic 1-form symmetries (c.f. momentum-winding mixed anomaly of the compact boson in 2d).

Comments (standard for anomalies):

- Even though the electric and magnetic symmetries appear on equal footing – they are exchanged by duality – they anomaly breaks this symmetry. In our current presentation the anomaly appears when we turn on non-zero magnetic $B_m^{(2)}$ and also perform an electric background gauge transformation. This is because of the way we coupled the background fields to Maxwell theory. We can add to the action S a local counterterm $S_{\text{c.t.}}$, which only depends on the background fields,

$$S_{\text{c.t.}} = \frac{iK}{2\pi}\int B_e^{(2)} \wedge B_m^{(2)} \ , \qquad K \in \mathbb{R} \ . \tag{6.46}$$

  Note that this term modifies the two-point function

$$\langle J_e^{(2)}(x) J_m^{(2)}(y)\rangle \sim \langle f^{(2)}(x) * f^{(2)}(y)\rangle \ . \tag{6.47}$$

  by contact terms involving (derivatives of) $\delta^{(4)}(x-y)$, but only at coincident points. By dialing the constant K, we can choose these contact terms so that either $J_e^{(2)}$ or $J_m^{(2)}$ or neither are conserved at coincident points. For instance, if we choose $K = -1$ then S is invariant under electric gauge transformations and $J_e^{(2)}$ is conserved at coincident points. However the price to pay is that now the action is not invariant under $B_m^{(2)}$

gauge transformations, and hence $J_m^{(2)}$ is not conserved at coincident points.

The statement of the anomaly is that there is no choice of local counter terms such that both currents are conserved at coincident points, or equivalently that S is completely invariant under both $B_e^{(2)}$ and $B_m^{(2)}$ background gauge transformations. Thus the anomaly is a mixed anomaly between the two symmetries. Each symmetry on its own is not anomalous.

The anomaly of course implies that the theory under discussion cannot be gapped and trivial – of course it is not, since it has a free massless field, the photon (this is essential a direct consequence of this mixed anomaly).

- Below we will discuss what it means to gauge the 1-form symmetries by promoting $B_{e,m}^{(1)}$ to suitable dynamical fields. Clearly the mixed anomaly prevents us from gauging both, but we can gauge anomaly-free subgroups.

- It is very useful to characterize the anomaly by inflow from a 5d SPT with action,

$$S_{5d} = \frac{i}{2\pi} \int_{\mathcal{M}_5} B_e^{(2)} \wedge dB_m^{(2)} \ . \tag{6.48}$$

This is a 5d Chern-Simons/BF term for the background fields. It has no dynamical fields, e.g. its Hilbert space on any closed 4-manifold has precisely one state. Such a theory is also called invertible (in math) or a symmetry protected topological phase (SPT) in condensed matter physics. As long as $\mathcal{M}_5$ is closed S_{5d} is fully gauge invariant. It is also symmetric in $B_e^{(2)} \leftrightarrow B_m^{(2)}$ since we can integrate by parts. If $\mathcal{M}_5$ has a boundary $\partial\mathcal{M}_5 = \mathcal{M}_4$ then neither statement is true: integrating by parts induces a local counterterm on $\mathcal{M}_4$, and background gauge transformations lead to a boundary anomaly of the kind described above.

Note: Here we have taken the sign of S_{5d} to reproduce the boundary anomaly found above. This is often done in QFT/higher-energy physics, where the 5d bulk is viewed as a fictitious machine whose only job it is to summarize the anomaly.

In situations where there are no net anomalies to begin with (e.g. in most condensed matter systems, or if the symmetries are gauged, like in string theory, supergravity, or holography) we can add the bulk as a physical extra dimensions, and flip the sign of S_{5d}. Then the 5d bulk cancels the anomaly of the 4d boundary and the coupled 4d/5d system is anomaly free. In this case there is also no obstruction to gauging the coupled bulk-boundary system.

6.7 Comments on Flat Backgrounds

Because the 1-form symmetries are continuous, the gauge fields $B^{(1)}_{e,m}$ need not be flat. However, various simplifications happen when we assume that they are flat. The flat case is also more similar to the discrete case we will encounter soon.

If $B^{(2)}_{e,m}$ are flat, then their holonomies on 2-cycles are topological and 2π-periodic,

$$\theta_{e,m}[\Sigma_2] = \int_{\Sigma_2} B^{(2)}_{e,m} \sim \theta_{e,m}[\Sigma_2] + 2\pi \ . \tag{6.49}$$

We can therefore think of them as angles $\theta_{e,m}$. Recall that the coupling of the background fields to the theory looks as follows,

$$S\big|_{\mathcal{O}(B)} = i \int_{\mathcal{M}_4} \left(B^{(2)}_e \wedge * J^{(2)}_e + B^{(2)}_m \wedge * J^{(2)}_m \right) \ . \tag{6.50}$$

Now we see that fixing a particular Σ_2, as well as well as holonomies $\theta_{e,m}[\Sigma_2]$ for the background fields, inserts into the path integral symmetry defects located on the Poincare-dual 2-cycle $\mathrm{PD}[\Sigma_2]$,

$$U^{(1)}_e\left(\theta_e, \mathrm{PD}[\Sigma_2]\right) U^{(1)}_m\left(\theta_m, \mathrm{PD}[\Sigma_2]\right) \ . \tag{6.51}$$

Another thing we can do when $B^{(2)}_e$ is to trade the conventionally quantized $U(1)$ gauge field $a^{(1)}$ used above for a new twisted gauge field $c^{(1)}$ which has the property that

$$dc^{(1)} = f^{(2)} - B^{(2)}_e \ . \tag{6.52}$$

In particular this means that the fluxes of $c^{(1)}$ are not integers, because they are shifted by the holonomies of $B^{(2)}_e$,

$$\int_{\Sigma_2} dc^{(1)} \in 2\pi\mathbb{Z} + \theta_e[\Sigma_2] \ . \tag{6.53}$$

This is only possibly if the periodicity of the gauge transformations of $c^{(1)}$ are modified in a way that is compatible with this equation. This is called a twisted cocycle condition. Perhaps the most well-known example of such a twisted connection is a Spin_c connection, which arises naturally in theories like QED, where all fermions have odd electric charges.

7 The Abelian Higgs Model

7.1 1-Form Symmetry Breaking by a Charged Scalar

Let us generalize our discussion of Maxwell theory by adding a single complex scalar h of electric charge $q_e \in \mathbb{Z}$. The resulting model has the following (Euclidean) Lagrangian

$$\mathscr{L} = \frac{1}{4e^2} f^{\mu\nu} f_{\mu\nu} + |D_\mu h|^2 + V(|h|) \ , \qquad D_\mu = \partial_\mu - i q_e a_\mu \ . \tag{7.1}$$

Here $V(|h|)$ is any gauge-invariant potential for h that we can tune to engineer different phases. The coupling to the background fields $B^{(2)}_{e,m}$ is as before, i.e. we substitute $f^{(2)} \to f^{(2)} - B^{(2)}_e$ and add $S_{BF} = \frac{i}{2\pi} \int_{\mathcal{M}_4} B^{(2)}_m \wedge f^{(2)}$.

Since the charged scalar does not modify the Bianchi identity, $df^{(2)}$, the magnetic symmetry is completely intact.

However, the electric symmetry is explicitly broken. This can be seen by examining Maxwell's equation,

$$d * J^{(2)}_e = \frac{i}{e^2} d * f^{(2)} = *j^{(1)}_e \neq 0 \ . \tag{7.2}$$

Thus the electric 2-form current is not conserved, and hence the electric flux integral $Q^{(1)}_e[\Sigma_2]$ is no longer topological: consider a cobordism $\mathcal{M}_3$ from Σ_2 to Σ_2' so that $\partial\mathcal{M}_3 = \Sigma_2 - \Sigma_2'$. Then

$$Q^{(1)}_e[\Sigma_2] - Q^{(1)}_e[\Sigma_2'] = \int_{\mathcal{M}_3} *j^{(1)}_e \ . \tag{7.3}$$

The operator on the right-hand side measures the total electric charge enclosed in $\mathcal{M}_3$. Since the fundamental unit of charge is q_e it is natural to suspect that the violation of $Q^{(1)}_e$ is also a multiple of q_e. If this is true, the exponentiated generators,

$$U_{e,k} = \exp\left(\frac{2\pi i k}{q_e} \int_{\Sigma_2} *J^{(2)}_e\right) \tag{7.4}$$

are still topological and generate a $\mathbb{Z}^{(1)}_{q_e}$ 1-form symmetry.

Thus adding a charge-q_e scalar explicitly breaks

$$U(1)^{(1)}_e \quad \longrightarrow \quad \mathbb{Z}^{(1)}_{q_e} \ . \tag{7.5}$$

A quick way to see this is to note that the field h must be attached to a Wilson line of charge q_e to make a gauge-invariant operator. Thus such a line can end. (We also say that such a line can be screened.) And since all local field constructed out of h have charge that is a multiple of q_e, no Wilson line of charge less than q_e can end. Thus the electric 1-form symmetry is broken to its order q_e cyclic subgroup.

Let us now show that we can modify the coupling of Maxwell theory to $B_e^{(2)}$ to exhibit this symmetry. Under a 1-form gauge transformation,

$$a^{(1)} \to a^{(1)} + \Lambda_e^{(1)} \ , \qquad D_\mu \to (D_\mu - iq_e\Lambda_{e,\mu}) \ . \tag{7.6}$$

The only chance we have of absorbing the $q_e\Lambda_e^{(1)}$ term in the covariant derivative of h is to reduce $B_e^{(2)}$ to a flat gauge field,

$$dB_e^{(2)} = 0 \ . \tag{7.7}$$

This is only correct as a statement about differential forms. In integer cohomology it is enough that the class $\frac{1}{2\pi}H^{(3)} = \frac{1}{2\pi}dB_e^{(2)}$ is N-torsion, i.e. that $\frac{N}{2\pi}H^{(3)} = 0$ in $H^{(3)}(\mathcal{M}_4, \mathbb{Z})$.

As we have already discussed, all gauge-invariant information in the flat $B_e^{(2)}$ is contained in the holonomies on 2-cycles,

$$\theta_e[\Sigma_2] = \int_{\Sigma_2} B_e^{(2)} \sim \theta_e[\Sigma_2] + 2\pi \ , \tag{7.8}$$

and the dependence on Σ_2 is topological.

We can restrict to a $\mathbb{Z}_{q_e} \subset U(1)$ by choosing all holonomies to satisfy

$$\theta_e \in \frac{2\pi}{q_e}\mathbb{Z}_{q_e} \ . \tag{7.9}$$

Up to gauge transformations, we therefore have

$$B_e^{(2)} = \frac{2\pi}{q_e} b_e^{(2)} \ , \qquad b_e^{(2)} \in H^2(\mathcal{M}_4, \mathbb{Z}_{q_e}) \ . \tag{7.10}$$

A change of cohomology representative, $b_e^{(2)} \to b_e^{(2)} + \delta\lambda_e^{(1)}$ with $\lambda_e^{(1)} \in C^1(\mathcal{M}_4, \mathbb{Z}_{q_e})$, can similarly be identified with a 1-form gauge transformation of the form

$$B_e^{(2)} \to B_e^{(2)} + d\Lambda_e^{(1)} \ , \qquad \Lambda_e^{(1)} = \frac{2\pi}{q_e}\lambda_e^{(1)} \ . \tag{7.11}$$

Returning back to our original problem of determining the 1-form symmetry of Maxwell theory with a charge-q_e scalar, we find that the covariant derivative of the scalar shifts as follows under an electric $\mathbb{Z}_{q_e}^{(1)}$ background gauge transformation,

$$\left(\partial_\mu - iq_e a_\mu\right) \qquad \longrightarrow \qquad \left(\partial_\mu - iq_e a_\mu - 2\pi i \lambda_e^{(1)}\right) \ . \tag{7.12}$$

Now $2\pi\lambda_e^{(1)}$ is a 1-cochain that whose periods over any 1-chain lie in $2\pi\mathbb{Z}_{q_e} \subset 2\pi\mathbb{Z}$. Locally we can therefore find a periodic scalar $\chi \sim \chi + 2\pi$ such that $2\pi\lambda_e^{(1)} = d\chi$. Now we can remove this term by shifting $h \to he^{i\chi}$. Note that this is single-valued and that it leaves all other terms in the Lagrangian invariant.

Note that the actually symmetry, which leaves $b_e^{(2)}$ invariant consists of closed chains, which can therefore be identified with cohomology classes,

$$\delta\lambda_e^{(1)} \qquad \Longrightarrow \qquad \lambda_e^{(1)} \in H^1(\mathcal{M}_4, \mathbb{Z}_{q_e}) \ , \tag{7.13}$$

or equivalently flat $\mathbb{Z}_{q_e}$ 1-form gauge fields. For instance, a Wilson loop of charge n transforms as

$$W_n(C) \to \exp\left(\frac{2\pi i n}{q_e}\lambda_e^{(1)}(C)\right) W_n(C) \ . \tag{7.14}$$

We can also reduce the anomaly modulo q_e by noting that $\frac{1}{2\pi}dB_m^{(2)}$ defines an integer cohomology class in $H^3(\mathcal{M}_4, \mathbb{Z})$, which can be reduced modulo q_e. Thus the anomaly inflow action reads,

$$S_5[b_e, B_m] = \frac{2\pi i}{q_e}\int_{\mathcal{M}_5} b_e^{(2)} \cup \left[\frac{1}{2\pi}dB_m^{(2)}\right]_{q_e} \ , \tag{7.15}$$

where $[\cdots]_q$ implies reduction modulo q. Thus the anomaly is still there and must be matched in every phase of the theory.

7.2 Phases of the Abelian Higgs Model

Let us illustrate this for the simple case of the potential

$$V(|h|) = m^2|h|^2 + \lambda_4|h|^2 \ . \tag{7.16}$$

It is known that this model has two phases, separated by a first-order phase transition at some critical m_*^2, which we renormalize to be at $m_*^2 = 0$. The transition is first order because of the Coleman-Weinberg mechanism., i.e. due to radiative corrections that become important near the transition. Let us discuss instead the physics in the different phases. In particular we would like to understand the realization of the 1-form symmetries.

7.2.1 Coulomb Phase

When $m^2 > 0$ and sufficiently large, we can reliably integrate out the scalar. This leads to a IR effective action which consists of free Maxwell theory deformed by irrelevant operators (sometimes called Euler-Heisenberg effective action),

$$\mathscr{L}_{\rm IR} = \frac{1}{4e^2} f^{\mu\nu} f_{\mu\nu} + \mathcal{O}\left(\frac{f^4}{m^4}\right) \cdots \ . \tag{7.17}$$

Thus the long-distance theory is in a Coulomb phase, with its leading IR behavior described by free Maxwell theory, i.e. all loops have perimeter law. In particular $U(1)_m^{(1)}$ is spontaneously broken and the photon is the corresponding NGB. The discrete $\mathbb{Z}_q^{(1)}$ symmetry is also spontaneously broken. In fact, in the deep IR it enhances to an accidental/emergent $U(1)_e^{(1)}$. This can be seen explicitly because Maxwell's equations now read

$$d\left(*f^{(2)} + \mathcal{O}\left(\frac{f^3}{m^4}\right) + \cdots\right) = 0 \ , \tag{7.18}$$

so there appears to be an exactly conserved electric 1-form current at long distances, even though the symmetry is emergent and violated by the massive h particles at sufficiently high energies. This is unlike the case of emergent 0-form symmetries, which are generically violated by irrelevant operators at long distances. The reason this does not happen for emergent 1-form symmetries is that there are no local operators that are charged under them, so it is not possible to capture the breaking at the level of the IR effective action. Remembering that this effective action is a power-series expansion in the small parameter $\varepsilon = \frac{E_{\rm IR}}{m}$, with $E_{\rm IR}$ a low energy scale, this means that explicit violations of emergent 0-form symmetries are typically power suppressed at long distance (think neutrino masses from dimension 5 operators $\mathcal{O}(\varepsilon)$ or proton decay from dimension 6 operators $\mathcal{O}(\varepsilon^2)$ in the standard model). By contrast, the violations of emergent 1-form symmetries are exponential small and non-analytic in ε,

$$e^{-1/\varepsilon} \sim e^{-m/E} \sim e^{-mR_{\rm IR}} \ , \qquad R_{\rm IR} \sim \frac{1}{E_{\rm IR}} \gg m^{-1} \ . \tag{7.19}$$

The factor $e^{-mR_{\rm IR}}$ can be thought of an instanton effect where the heavy particle travels some distance $R_{\rm IR}$, which is very suppressed at long distances.

7.2.2 Higgs Phase

The model for $m^2 < 0$ is famously the Landau-Ginzburg effective description of a superconductor with a charge-q_e Cooper pair of electrons (usually $q_e = 2$).

The theory is in a Higgs phase, i.e. the charge-q scalar field gets a vev $\langle h \rangle \neq 0$. This Higgses the $U(1)$ gauge group to its $\mathbb{Z}_{q_e}$ subgroup (it is common, but imprecise, to say that the gauge symmetry is broken to $\mathbb{Z}_{q_e}$). The theory is fully gapped: the photon eats the phase of h to make a massive vector boson, and the radial Higgs mode of h gets a mass from the potential. However, it is not trivial at long distances: the is a discrete $\mathbb{Z}_{q_e}$ topological gauge theory at long distance (i.e. there is topological order).

In this phase the $U(1)^{(1)}_m$ symmetry is unbroken, and the 't Hooft lines H_{q_m} charged under that symmetry create from the vacuum Abrikosov-Nieleson-Oleson (ANO) magnetic vortices of integer magnetic flux/vorticity q_m. All of these have finite tension $\Sigma(q_m)$. If we repeat the calculation of the static potential for monopoles, the 't Hooft line will create an 't Hooft line connecting the monopoles from the vacuum. This will lead to a static potential $V_{q_m\bar{q}_m}(r) = \sigma(q_m)r$, i.e. a linearly confining potential for the monopoles. Plugging into the expectation value for a large rectangular loop of width r and height T, we find that

$$\langle H_{q_m}(\Box \to \infty)\rangle \sim \exp\left(-\sigma(q_m)Tr\right) \sim \exp\left(-\sigma(q_m)A(C)\right) = 0 \ . \tag{7.20}$$

Here $A(C) = rT$ is the area bounded by the rectangular loop. This is the famous area law for loops. Since the area is much larger than the perimeter for large loops, the loop expectation value vanishes. This is the statement that the $U(1)^{(1)}_m$ symmetry is unbroken.

In fact, because of the unbroken $\mathbb{Z}_2$ gauge symmetry, there are also fractional vortices with vorticity $\in \frac{1}{q}\mathbb{Z}$. (In a superconductor with $q_e = 2$, a half-vortex is known as a π-flux).) The fractional vortices are not created by genuine 't Hooft lines, but rather by open topological surfaces ending on a fractional 't Hooft line. The fractionalized vortices are a potential source of an anomaly, as we will elaborate soon. **[General theme: fractionalization can lead to anomalies.]**

However the $\mathbb{Z}^{(1)}_q$ symmetry is spontaneously broken and this matches the 't Hooft anomaly from above. One way to see this is to note that in the Higgs phase the electric field is screened and $V_{q_e\bar{q}_e}(r) \sim e^{-Kr} + \text{const.}$, leading to perimeter law for the Wilson loops, and hence a vev.

To see this explicitly, let us consider an explicit Lagrangian describing the low-energy $\mathbb{Z}_{q_e}$

TQFT,

$$S_{\rm TQFT}=\frac{iq_e}{2\pi}\int_{\mathcal{M}_4} b^{(2)}\wedge\left(f^{(2)}-B_e^{(2)}\right)+\frac{i}{2\pi}\int_{\mathcal{M}_4} B_m^{(2)}\wedge f^{(2)}~. \tag{7.21}$$

Here $f^{(2)}=da^{(1)}$ is the original Maxwell gauge field from above, while $b^{(2)}$ is a new dynamical $U(1)$ 2-form gauge field, with gauge transformations

$$b^{(2)}\to b^{(2)}+d\lambda^{(1)}~,\qquad \frac{1}{2\pi}\int_{\Sigma_2} d\lambda^{(1)}\in\mathbb{Z}~. \tag{7.22}$$

The field $b^{(2)}$ can be thought of as arising from the compact scalar $\chi\sim\chi+2\pi$ that represents the phase of the dynamical scalar field $h=\rho e^{i\chi}$ via an electric-magnetic duality transformations (much like the one we studied in Maxwell theory itself).

Let us first turn off the background fields $B^{(2)}_{e,m}$. Then the equations of motion imply that $b^{(2)}$ and $a^{(1)}$ are flat, with holonomies in $\mathbb{Z}_{q_e}$. **[More precisely, doing the path integral over $f^{(2)}$ implies that the flux $db_e^{(2)}$ is q_e-torsion, $\frac{q_e}{2\pi}db^{(2)}=0$ in $H^3(\mathcal{M}_4,\mathbb{Z})$. Similarly, doing the path integral over $b_e^{(2)}$ implies that $f^{(2)}$ is q_e-torsion.]** Integrating out $b^{(2)}$ we find a $\mathbb{Z}_{q_e}$ 1-form gauge theory, while integrating out $a^{(2)}$ leads to a $\mathbb{Z}_{q_e}$ 2-form gauge theory. These are simply dual description of the same TQFT. The topological operators are Wilson lines of $a^{(1)}$ and Wilson surfaces of $b^{(2)}$, which have non-trivial correlation function (explicitly computable from the Gaussian path integral)

$$\langle\exp\left(im\int_C a^{(1)}\right)\exp\left(in\int_{\Sigma_2} b^{(2)}\right)\rangle=\exp\left(\frac{2\pi imn}{q_e}{\rm Link}(\Sigma_2,C)\right)~. \tag{7.23}$$

Here ${\rm Link}(\Sigma_2,C)$ is the oriented Linking number. Note that this answer is periodic in $n,m\sim n,m+q_e$, as is appropriate for $\mathbb{Z}_{q_e}$ charges.

Thus the Wilson surface of $b^{(2)}$ is the topological defect implementing the $\mathbb{Z}_{q_e}^{(1)}$ symmetry acting on the Wilson lines of $a^{(1)}$. Conversely, the Wilson lines of $a^{(1)}$ are generators of $\mathbb{Z}_{q_e}^{(2)}$ symmetry acting on the Wilson surfaces of $b^{(2)}$. This symmetry was not present in the UV: it emergent in the IR. The fact that the symmetry defects link non-trivially is an indicator of the mixed 't Hooft anomaly between them.

Note that all of these topological operators have non-zero expectation values. This can be seen by taking large loops or surfaces and then shrinking them to a point, leading to expectation value 1 modulo perimeter or surface counterterms. Thus both the $\mathbb{Z}_{q_e}^{(1)}$ and the $\mathbb{Z}_{q_e}^{(2)}$ symmetry are spontaneously broken.

Note that no mention was made here of the unbroken $U(1)_m$ symmetry. How is the mixed anomaly of this symmetry with $\mathbb{Z}_{q_e}^{(1)}$ matched in this phase? Here the emergent $\mathbb{Z}_{q_e}^{(2)}$

symmetry is key. Let us explain this by re-activating the background fields:

$$S_{\mathrm{TQFT}} = \frac{iq_e}{2\pi}\int_{\mathcal{M}_4} b^{(2)} \wedge \left(f^{(2)} - B_e^{(2)}\right) + \frac{i}{2\pi}\int_{\mathcal{M}_4} B_m^{(2)} \wedge f^{(2)} \ . \tag{7.24}$$

This is manifestly gauge-invariant under $a^{(1)} \to a^{(1)} + \lambda^{(0)}$, but to ensure invariance under $b^{(2)} \to b^{(2)} + d\lambda^{(1)}$ we must require

$$dB_e^{(2)} = 0 \ , \qquad \int_{\Sigma_2} B_e^{(2)} \in \frac{2\pi}{q_e}\mathbb{Z} \ . \tag{7.25}$$

This is precisely the statement that $B_e^{(2)}$ is a background field for the discrete $\mathbb{Z}_{q_e}^{(1)}$ subgroup of $U(1)_e^{(1)}$. We can now investigate background gauge transformations,

$$a^{(1)} \to a^{(1)} + \Lambda_e^{(1)} \ , \qquad B_{e,m}^{(2)} \to B_{e,m}^{(2)} + d\Lambda_{e,m}^{(1)} \ . \tag{7.26}$$

As in the UV, all terms except the BF term are invariant, and this term exactly spits out the same anomaly as in the UV.

Note that the $\mathbb{Z}_{q_e}^{(1)}$ symmetry with background $B_e^{(2)}$ is intrinsic to the IR TQFT: it acts on the topological Wilson loops, which spontaneously break the symmetry. By contrast the 't Hooft loops which are charged under $U(1)_m^{(1)}$ with background $B_m^{(2)}$ are all confined and disappear from theory. It is thus superficially confusing how the anomaly (which involves B_m) is matched in the IR.

The answer involves fractionalization of the $U(1)_m^{(1)}$ magnetic symmetry. Genuine 't Hooft lines are confined, but the theory has vortices of fractional vorticity $\frac{n}{q_e}$. The worldsheet Σ_2 of these fractional vortices is represented at long distances by the charge-n Wilson surface of b_2 along Σ_2,

$$\exp\left(in\int_{\Sigma_2} b^{(2)}\right) \ . \tag{7.27}$$

Inserting this into the path integral and integrating out $b^{(2)}$ we find that

$$f^{(2)} = \frac{2\pi n}{q_e}\mathrm{PD}[\Sigma_2] + B_e^{(2)} \ , \tag{7.28}$$

where PD indicates the closed 2-form that is δ-function localized along the PD cycle of Σ_2, i.e. the cycle linking the worldsheet. Inserting this into the path integral we find that the BF term gives rise to

$$\frac{i}{2\pi}\int_{\mathcal{M}_4} B_m^{(2)} \wedge f^{(2)} = \frac{in}{q_e}\int_{\Sigma_2} B_m^{(2)} \ . \tag{7.29}$$

This is precisely the statement that the vortex carries fractional charge $\frac{n}{q_e}$. Thus charge fractionalization for the unbroken $U(1)^{(1)}_m$ symmetry gives a non-trivial coupling of $B^{(2)}_m$ to the low-energy TQFT that matches the anomaly.

Note that $B^{(2)}_m$ is not intrinsic to the low-energy theory. Rather, the low-energy theory has a $\mathbb{Z}^{(2)}_q$ symmetry generated by the topological Wilson lines of $a^{(1)}$. The appropriate background field is a flat 3-form $C^{(3)}$ with $\mathbb{Z}_{q_e}$ periods,

$$dC^{(2)} = 0 \ , \qquad \int_{\Sigma_3} C^{(3)} \in \frac{2\pi}{q_e}\mathbb{Z} \ . \tag{7.30}$$

This couples to the BF theory as follows,

$$\Delta S = \frac{iq_e}{2\pi}\int_{\mathcal{M}_4} a^{(1)} \wedge C^{(3)} \ . \tag{7.31}$$

Note that the restriction of $C^{(3)}$ to a flat $\mathbb{Z}_{q_e}$ gauge field is needed for $a^{(1)}$ gauge invariance. We thus see that the anomalous shift of the action under $a^{(1)} \to a^{(1)} + \Lambda^{(1)}_e$ is

$$S \to S + \frac{iq_e}{2\pi}\int \Lambda^{(1)}_e \wedge C^{(3)} \ . \tag{7.32}$$

Then we can embed $B^{(2)}_m$ in $C^{(3)}$ via

$$C^{(3)} = \frac{1}{q_e} dB^{(2)}_m \ , \tag{7.33}$$

which indeed has holonomies in $\frac{2\pi}{q_e}\mathbb{Z}$.

7.3 Duality, Monopoles, and Confinement

Above we described everything for a scalar of electric charge q_e. We can use electric-magnetic (or S-) duality do map this to a system with no electric charges and a scalar of magnetic charge q_m. The whole discussion then plays out the same, with e and m switched:

- The model has an electric $U(1)^{(1)}_e$ symmetry and a $\mathbb{Z}^{(1)}_{q_m}$ magnetic symmetry.
- In the magnetic Higgs phase $U(1)^{(1)}_e$ is unbroken, there are electric flux/vortex strings that have finite tension and linearly confine all electric charges, i.e.

$$V_{q_e\bar{q}_e}(r) \sim \sigma(q_e) r \ . \tag{7.34}$$

Equivalently, all Wilson loops have area law. Magnetic charge is deconfined and screened modulo q_m, there is a $\mathbb{Z}_{q_m}$ TQFT whose Wilson lines are the UV 't Hooft lines mod q_m. All 't Hooft lines have perimeter law.

Thus a magnetic Higgs phase is tantamount to electric confinement: confinement and Higgsing are electric-magnetic dual. This key insight (due to Mandelstam and 't Hooft) is central to our modern understanding of confinement in non-Abelian gauge theories. This is because there are certain (mostly supersymmetric) models where confinement is described by an effective Abelian Higgs Model for magnetic monopoles.

If a theory has both electric and magnetic charges then we cannot always to to a purely electric duality frame. However, we can still use the screening picture to determined the 1-form symmetry group (which may be trivial).

If the charge lattice is not fully populated by dynamical charges, the theory has some 1-form symmetry that can be used to diagnose different phases, depending whether the loops charged under that symmetry satisfy perimeter or area law.

7.4 Higgs-Confinement Continuity

What happens when there is no 1-form symmetry whatsoever, e.g. because the charge lattice is fully populated? Then ever single line can be screened. In this situation there is no symmetry that can be used to distinguish e.g. Higgs and confining phases. It is then natural to expect that there is no phase transition in the model (c.f. breaking $\mathbb{Z}_2$ symmetry in the Ising model via magnetic field: phase transition can be avoided). This was first shown explicitly in certain lattice models, but has since been observed in many (e.g. supersymmetric) examples.

We will return to this point in the last lecture on Wednesday, when we will discuss QCD and QCD-like theories, i.e. non-Abelian Yang-Mills gauge theories with matter fields in the fundamental (e.g. quarks).

8 Multi-Flavor QED

So far we have seen 0-form symmetries and 1-form symmetries, but there has not been any particularly dramatic interplay between the two.

A simple example where such an interplay – in the form of a structure called 2-group symmetry – arises is massless QED with $N_f \geq 2$ flavors.

Reference: Cordova, Dumitrescu, Intriligator "Exploring 2-Group Global Symmmetry"

We are thus studying $U(1)$ gauge theory with N_f flavors of 4-component Dirac fermion Ψ^i_D (with flavor index $i = 1, \ldots, N_f$ of electric charge $q_e = 1$. We can write the QED Lagrangian in the standard Dirac form,

$$\mathscr{L} = -\frac{1}{4e^2} f^{\mu\nu} f_{\mu\nu} - i \overline{\Psi}_{Di} \gamma^\mu D_\mu \Psi^i_D \ . \tag{8.1}$$

For many purposes, but especially for discussing symmetries, it is useful to work with 2-component Weyl fermions, rather than 4-component Dirac fermions,

$$\Psi^i_D = \begin{pmatrix} \psi^i_\alpha \\ \overline{\chi}^{\dot\alpha}_{\tilde{i}} \end{pmatrix} \ , \qquad i = 1, \ldots, N_f \ . \tag{8.2}$$

Here $\alpha = 1, 2$ is a left-handed 2-component spinor index, and $\dot\alpha = \dot 1, \dot 2$ is a right-handed 2-component spinor index. For a detailed introduction to 2-component spinors see for instance the SUSY book by Wess and Bagger (whose conventions I typically follow) or the QFT book by Mark Srednicki. Note that Hermitian conjugation exchanges left- and right-handed spinors,

$$(\psi_\alpha)^\dagger = \overline{\psi}_{\dot\alpha} \ . \tag{8.3}$$

Spinor indices are raised and lowered from the left using the $SL(2, \mathbb{C})$ invariant Levi-Civitta symbols $\varepsilon^{\alpha\beta}$, $\varepsilon_{\alpha\beta}$ (and their dotted counterparts), which satisfy

$$\varepsilon^{12} = \varepsilon_{21} = 1 \ . \tag{8.4}$$

When spinor indices are suppressed they are contracted (like $X^\alpha{}_\alpha$ and $\overline{Y}_{\dot\alpha}{}^{\dot\alpha}$ for undotted and dotted spinors).

Then the charge-one Dirac fermions Ψ^i_D amounts to having N_f 2-component Weyl fermion ψ^i_α of charge 1 (sometimes referred to as the left-handed component) and one 2-component Weyl fermion $\chi^{\tilde{i}}_\alpha$ of charge -1 (sometimes called the right-handed component). In these variables, the QED Lagrangian reads

$$\mathscr{L} = -\frac{1}{4e^2} f^{\mu\nu} f_{\mu\nu} - i \overline{\psi}_i \overline{\sigma}^\mu (\partial_\mu - i a_\mu) \psi^i - i \overline{\chi}_{\tilde{i}} \overline{\sigma}^\mu (\partial_\mu + i a_\mu) \chi^{\tilde{i}} \ . \tag{8.5}$$

Note, crucially, that the flavor index i on the left-handed field and the flavor index $\tilde{i}$ on the right-handed field are completely independent. This will be important in our analysis of the symmetries below.

8.1 An Overly Naive Guess at the Symmetry

To figure out what the symmetry of the theory might be, let's start with the free fermions, prior to gauging the vector-like $U(1)$ symmetry. Then the symmetry is clearly

$$U(N_f)_L \times U(N_f)_R \ , \tag{8.6}$$

where the left symmetry only acts on the ψ^i_α, with i a fundamental index, and the right symmetry only acts on the $\chi^{\tilde{i}}_\alpha$ with $\tilde{i}$ a fundamental index.

Now let us gauge the $U(1)_V$ symmetry of QED under which the fermions have electric charge ± 1. We then know that three things happen:

- Since the $U(1)_V$ is now gauged, it is no longer a zero-form symmetry.
- We do not get any electric 1-form symmetry since our fields have electric charge ± 1.
- We do get a continuous magnetic 1-form symmetry $U(1)^{(1)}_m$ due to the fact that the Bianchi identity holds, $df^{(2)} = 0$.

  Very naively, this suggests that the global symmetry of the model is

$$G^{(0)} \times G^{(1)} \ , \qquad G^{(0)} = \frac{\big(U(N_f)_L \times U(N_f)_R\big)}{U(1)_V} \ , \qquad G^{(1)} = U(1)^{(1)}_m \ . \tag{8.7}$$

This is wrong for two reasons, both of which have to do with triangle anomalies:

- The $U(1)_A$ symmetry suffers from an Adler-Bell-Jackiw (ABJ) triangle anomaly **[DRAW DIAGRAM]** and as a result it is not a conventional 0-form symmetry. Depending on which observables you study, it may or may not be a symmetry (e.g. it is a symmetry in flat space, but not in the presence of generic monopoles/'t Hooft lines). This can be formulated more usefully by recasting $U(1)_A$ as a non-invertible symmetry – you'll hear more about it from other lecturers, and for today we'll simply ignore it.
- The second reason is that the remaining symmetry is **not** a direct product

$$G^{(0)} \times G^{(1)} \ , \qquad G^{(0)} = SU(N_f)_L \times SU(N_f)_R \ , \qquad G^{(1)} = U(1)^{(1)}_m \ . \tag{8.8}$$

  Rather the product is deformed into a non-trivial 2-group, wherein $G^{(0)}$ non-trivially extends $G^{(1)}$.

We will now explain how this works starting using background fields and triangle anomalies.

8.2 2-Group Global Symmetry from Triangle Anomalies

Triangle anomalies in QED involve a single fermion loop with three insertions, which could be either a $SU(N_f)_L$ current $j^a_{L,\mu}$, with a and adjoint index of $SU(N_f)_L$, or a $SU(N_f)_R$ current $j^{\widetilde{a}}_{R,\mu}$ with $\widetilde{a}$ an adjoint index of $SU(N_f)$, or a dynamical photon (equivalently, an insertion of the magnetic 1-form current $\sim f^{(2)}$).

An important fact is that the vector-like $SU(N_f)_V \subset SU(N_f)_L \times SU(N_f)_R$ is free of anomalies, because we can add a Dirac mass term that respect this symmetry. Recall that anomalies only receive contributions from massless fields.

For this reason it is sufficient to focus on e.g. $SU(N_f)_L$ symmetry with current $j^a_{L,\mu}$ for the purpose of discussing anomalies in this theory. We will call the $SU(N_f)_L$ background gauge field that couples to this current L^a_μ as a source.

Let is discuss the triangle anomalies of this theory in turn:

- You can consider a triangle diagram involving three external photons. **[DRAW DIAGRAM]** This vanishes in QED by charge-conjugation symmetry. This triangle diagram represents a genuine gauge anomaly, which would make the theory inconsistent. So its essential that it vanishes.

- There is a triangle diagram involving three insertions of $j^a_{L,\mu}$. **[DRAW DIAGRAM]**Schematically,

$$\partial^\mu_x \langle j^a_\mu(x) j^b_\nu(y) j^c_\rho(z) \rangle \neq 0 \ . \tag{8.9}$$

  More precisely, the right-hand side is a c-number contact term, which is only non-zero at coincident points $x = y = z$. While it can be written down in detail, it is easier to characterizing it by activating an $SU(N_f)_L$ non-dynamical background gauge field L^a_μ, which couples to $j^a_{\mu L}$ as a source.

  The the anomalous triangle can be expressed by the following equation:

$$D_\mu j^{\mu a}_L \sim d^{abc} \varepsilon^{\mu\nu\rho\lambda} \partial_\mu L^b_\nu \partial_\rho L^c_\lambda + \cdots \ .$$

  Here the dots mean additional non-Abelian terms, and d^{abc} is a (suitably normalized) totally symmetric invariant symbol of $SU(N_f)$, which exists as long as $N_f \geq 3$. The factor of N in front of the anomaly is due to the fact that the quarks ψ_α, which are in the fundamental of $SU(N_f)_L$, are also in the fundamental of $SU(N)$, so there are N fundamentals of $SU(N_f)$.

  Note that the right-hand side of this anomaly equation is not an operator: it vanishes when we turn off the background field L_μ^a. Thus the symmetry is not actually broken. Instead, the right-hand side in the presence of a background field means that the symmetry has an 't Hooft anomaly (with itself). We have previously discussed mixed ' t Hooft anomalies, e.g. between the electric and magnetic 1-form symmetries in Maxwell theory, but here it is an anomaly intrinsic to $SU(N_f)_L$.

Just as was the case there, this 't Hooft anomaly is rigid along RG flows and not renormalized. And in particular, the anomaly must be matched in any IR phase that the theory can flow to without explicitly breaking the $SU(N_f)_L$ symmetry. Of course this example is weakly coupled and so the matching is trivial.

- Finally, we can consider a triangle diagram involving a single photon and two $SU(N_f)_L$ currents, **[DRAW DIAGRAM]**. This leads to the following non-conservation equation for the $SU(N_f)_L$ current,

$$d * j_L^{(1)a} \sim k dL^{(1)a} \wedge f^{(2)} + \cdots \ , \tag{8.10}$$

where the ellipsis represents non-linear terms required by $SU(N_f)_L$ background gauge invariance. Her $k \in \mathbb{Z}$ is a quantized anomaly coefficient, which is minimal $k = 1$ in our QED example (since our electrons have unit electric charge), $L^{(1)a}$ is the $SU(N_f)_L$ background field, and $f^{(2)}$ is the dynamical photon field strength.

Note that this is somewhere in between an ABJ anomaly, where the right-hand side is an operator, and an 't Hooft anomaly, where the right-hand side is a c-number background. Note also that when we turn off the background, $L^{(1)a} = 0$, then the current is conserved so the $SU(N_f)_L$ symmetry is not broken by the anomaly.

What then is the effect of this mixed background-operator anomaly? The answer is that it deforms the symmetry into a non-trivial 2-group, where the $SU(N_f)_L^{(0)}$ 0-form symmetry extends the $U(1)_m^{(1)}$ magnetic 1-form symmetry.

The quickest way to see this is is to consider $SU(N_f)_L^{(0)}$ background gauge transformations,

$$L^{(1)a} \to L^{(1)a} + d\lambda^{(0)a} + \cdots \ , \tag{8.11}$$

under which the integrand of the Euclidean path integral measure transforms as

$$\exp\left(ik \int \lambda^{(0)a} dL^{(1)a} \wedge f^{(2)} + \cdots \right) \tag{8.12}$$

Note that this is not just some innocuous c-number phase that multiplies the partition function like in the case of an 't Hooft anomaly, because it contains the field/operator $f^{(2)}$. The fix for this is that $f^{(2)}$ is precisely the current for the $U(1)_m^{(1)}$ symmetry, which couples to its own background field $B_m^{(2)}$ via

$$S_m = \frac{i}{2\pi} \int B_m^{(2)} \wedge f^{(2)} \ . \tag{8.13}$$

We thus see that we can absorb the anomalous, operator-valued transformation of the path integral measure if we declare that $B_m^{(2)}$ transforms non-trivially under $SU(N_f)_L$ background gauge transformations – schematically (and up to factors):

$$B_m^{(2)} \to B_m^{(2)} - k\lambda^{(0)a} dL^{(1)a} + \cdots \ . \tag{8.14}$$

We recognize this as a version of the Green-Schwarz anomaly cancellation mechanism from string theory, but for background fields rather than dynamical ones. This is one of many ways of characterizing the fact that $SU(N_f)_L^{(0)}$ and $U(1)_m^{(1)}$ fuse to a non-trivial 2-group global symmetry.

The Green-Schwarz shift of $B_m^{(2)}$ under $SU(N_f)_L$ gauge transformation shows that $SU(N_f)_L$ is not a subgroup of the 2-group. By contrast $U(1)_m^{(1)}$ is a subgroup, and the two-group is a non-trivial extension of $U(1)_m^{(1)}$ by $SU(N_f)_L^{(0)}$. (The integer k coming from the mixed anomaly represents the obstruction class that prevents the extension from splitting.) This fact has many interesting consequences for the dynamics of theories with 2-group symmetry – you'll hear more about this in the lectures by Clay Cordova.

9 Yang-Mills Theory and QCD

9.1 1-Form Symmetries in Pure $SU(N)$ Gauge Theory

A primary motivation of these lectures is to better understand the dynamics (phases, phase transitions) of interesting 4d non-Abelian gauge theories, like QCD – an $SU(3)$ theory with quarks in the fundamental representation.

Recall: free Maxwell theory has both a $U(1)_e^{(1)}$ electric and a $U(1)_m^{(1)}$ magnetic 1-form symmetries, because it has neither dynamical electric nor magnetic charges. This means that Wilson and 't Hooft lines cannot end – flux conservation symmetry.

Adding electric/magnetic charges in multiples of n breaks $U(1)_{e,m}^{(1)} \to \mathbb{Z}_n^{(1)}$. Now Wilson and 't Hooft lines can end, but only in multiples of n. In other words flux is conserved modulo n.

What happens in pure YM theory, say with gauge group $SU(N)$? They key is the gluons themselves carry color charge: they transform in the adjoint representation of the gauge group.

This means that the adjoint Wilson line can always be physically screened and in particular it can end:

$$\mathrm{Tr}_{\mathrm{adj}}\, Pexp\left(i\int^{x} A_{\mathrm{adj}}\right) F(x) \tag{9.1}$$

This is gauge invariant. The same applies to any representation R of $SU(N)$ that occurs in the product of any number of adjoint representations.

What are these representations? They are precisely the ones described by Young Diagrams whose number of boxes p is divisible by N, i.e. p is an integer multiple of N.

What is algebraically specially about these representations? There is an important subgroup of the $SU(N)$ gauge group called the center: it consists of precisely those $SU(N)$ transformations that commute with all elements of the $SU(N)$ gauge group. For $SU(N)$ this central subgroup (also just called the "center") consists of the following matrices,

$$U_\omega = \omega \mathbb{1}_{N\times N}\ , \qquad \omega^N = 1\ , \tag{9.2}$$

i.e. ω is an N-th root of unity. Note that U_ω is clearly unitary, and that $\det U_\omega = \omega^N = 1$, so it is in fact an $SU(N)$ matrix. Thus the center of $SU(N)$ consists of the N-th roots of unity. This is a cyclic group of order N, i.e. it is $\mathbb{Z}_N$.

Every representation R of $SU(N)$ induces a representation of the $\mathbb{Z}_N$ center. since this group is abelian it suffices to specify a $\mathbb{Z}_N$ valued charge, in other words an integer mod N. For $SU(N)$ this mod-N charge is also called the N-ality. (For $SU(3)$, which is the case relevant for QCD, it is called triality.)

It can be shown that the N-ality of a representation R is precisely the number of boxes in its Young diagram. Simple examples:

- The $SU(N)$ fundamental representation $\square$ always has N-ality 1.
- The $SU(N)$ adjoint representation can be made by taking $N \otimes \overline{N}$ − trace, so it is a Young diagram with one fundamental $\square$ times one anti-fundamental $N-1$ vertical box-stack. **[DRAW FIGURE!]** It thus has N boxes and trivial N-ality.
- The symmetric and anti-symmetric representations both have N-ality 2.

Thus the Wilson lines that can be screened by, i.e. end on, gluon fields are precisely those that transform trivially under the $\mathbb{Z}_N$ center of $SU(N)$, or equivalently those that have vanishing N-ality. This is in one to one correspondence with saying that these are all representations of the the quotient group

$$PSU(N) = SU(N)/\mathbb{Z}_N \ . \tag{9.3}$$

As a simple example $PSU(2) = SU(2)/Z_2 = SO(3)$.

Let us discuss the flip side of this: consider any Wilson line $W_R(C)$ in some representation of $SU(N)$ that has non-trivial N-ality, the most obvious candidate being the fundamental representation $\square$ with N-ality 1, but also the symmetric or anti-symmetric representations (at least for sufficiently large N: for $N = 2$ the anti symmetric is the singlet and the symmetric the adjoint, while for $N = 3$ the symmetric and anti-symmetric representations have triality $2 \equiv -1(3)$. In fact the anti-symmetric of $SU(3)$ is exactly the anti-fundamental representation.)

Since the Gluons have trivial N-ality they cannot screen the N-ality of any such line. Thus the N-ality of a Wilson line is a meaningful, conserved notion of "electric flux", in other words it is a 1-from symmetry. One often denotes this one-form symmetry by $\mathbb{Z}_N^{(1)}$, with the superscript (1) emphasizing the fact that it is a 1-form symmetry.

One can explicitly show this by: a) constructing topological surface operators that implement the action of the symmetry on Wilson lines and b) by explicitly coupling pure YM theory to discrete $\mathbb{Z}_N^{(1)}$ background gauge fields. These are discrete versions of the electric B-field $B_e^{(2)}$ in Maxwell theory. I will not explain this construction in detail because it is somewhat technical. But if you work on such things it is essential to know it.

Note that $SU(N)$ YM theory does not have magnetic 1-form symmetries. Just like the electric 1-form symmetries are given by the center of gauge group $Z(G)$ for any gauge group G, with $Z(G) = \mathbb{Z}_N$ for $G = SU(N)$, the magnetic 1-form symmetries are given by (the Pontryagin dual of) $\pi_1(G)$. Since $SU(N)$ is simply connected, its $\pi_1(SU(N)) = 0$ and there are no magnetic 1-form symmetries.

In summary: pure $SU(N)$ YM theory has $\mathbb{Z}_N^{(1)}$ electric 1-form symmetry, called center symmetry, and no magnetic 1-form symmetry.

The $\mathbb{Z}_N^{(1)}$ center symmetry was discovered in the early days of QCD, by thinking about what the theory does at finite temperature. There one finds that the symmetry is unbroken at low temperatures, signaling confinement, and spontaneously broken at high temperatures. This also means that pure YM theory undergoes a genuine thermal phase transition as it is heated up. This is relevant to the early universe and physics at heavy ion colliders.

Let us discuss what unbroken and spontaneously broken $\mathbb{Z}_N^{(1)}$ symmetry implies, recalling the analogy with the $U(1)_m^{(1)}$ symmetry in the Abelian Higgs model. There we learned that:

- Spontaneously broken $U(1)_m^{(1)}$ means a massless photon, interpretable as a Nambu-Goldstone boson for the symmetry. 't Hooft lines charged under the symmetry have perimeter law.

- Unbroken $U(1)_m^{(1)}$ occurs in the gapped Higgs phase. There 't Hooft lines are confined – they obey area law, and there are finite tension Magnetic strings (ANO vortices) carrying magnetic flux, i.e. $U(1)_m^{(1)}$ charge.

In pure YM theory the 1-form symmetry is electric, i.e. acts on Wilson rather than 't Hooft lines, and it is a discrete $Z_N^{(1)}$ symmetry, rather than a continuous one. This means that if the symmetry is spontaneously broken it need not imply a massless particle (no Goldstone theorem!), so in principle there could still be a gap. But the IR could not be completely trivial, e.g. in the Ising model in the $\mathbb{Z}_2^{(0)}$ breaking phase there are two vacua, but the theory is gapped. Analogously, spontaneously breaking $Z_N^{(1)}$ can in principle lead to a gapped theory, but there must be a TQFT in the deep IR. This is similar to what happened in the Abelian Higgs model with electric charge $q_e > 1$. There the Higgs phase was gapped, but there is a $\mathbb{Z}_{q_e}$ TQFT in the deep IR. In any case the Wilson loops have perimeter law, i.e. they are deconfined.

What does unbroken $\mathbb{Z}_N^{(1)}$ symmetry mean? It means confinement, i.e. the expectation values of all Wilson loops with non-trivial N-ality (which cannot be completely screened by gluons) decay more rapidly than perimeter law, and thus have zero vev. In terms of static potentials, it means that the confining potential between a quark Q in some rep R and its anti-quark $\overline{Q}$ rises faster than a constant at long distances, thus confining the $Q\overline{Q}$ pair!Conventional wisdom says that in fact more is true:

- The system will develop a mass gap, and the only particles will be massive glueballs. Proving this is a million dollar Clay Millenium problem.

- The mechanism for confinement is the existence of finite-tension confining electric strings, that give rise to a linearly rising $Q\overline{Q}$ potential at long distances, and hence an area-law decay for all Wilson loops of non-trivial N-ality. This certainly implies unbroken $Z_N^{(1)}$ center symmetry, but is (at least at face value) a stronger statement. Note: there are examples of theories that are gapless and confine in the sense of having an unbroken 1-form center symmetry, but the confining potential is not linear and there are no finite tension strings. A simple example is 3d Maxwell theory, where the Coulomb potential $V(r) \sim \log r$ is already confining.

The main evidence that this is true comes from the lattice! It is hardwired into the idea of Mandelstam and 't Hooft that electric confinement in YM is somehow the electric-magnetic dual of magnetic confinement via ANO vortices/flux tubes/strings in a Higgs model. This has been made sharpest in supersymmetric theories, as explained for instance by Seiberg and Witten.

Something similar happens in pure YM at finite temperature, which amounts to compactifying euclidean time $\tau \sim \tau + \beta$ with the KK radius $\beta = 1/T$ the inverse temperature. As realized early on by Poyakov and Susskind, the natural thing to do is to wrap the Wilson loop on the circle, so it is topologically non-trivial. Sometimes such a compactified time-like Wilson loop is known as a Polyakov loop. So we have $S^1_\beta \times R^3$ with a wilson loop wrapped on S^1_β. This loop looks from the point of view of the 3d space $\mathbb{R}^3$ like a point operator – a standard local operator in 3d! And correspondingly the surface operator implementing the $\mathbb{Z}^1_N$ center symmetry in 4d reduces to a surface operator in 3d, which is codimension 1 and behaves like a standard 0-form symmetry in 3d. This is the conventional setting for symmetry breaking in 3d: we have a 0-form $\mathbb{Z}_N$ symmetry under which the Polyakov loop is charged (according to its N-ality) and a standard 0-form symmetry in 3d that can measure that N-ality.

Now we can ask whether the symmetry is broken or not: at low temperature $T \ll \lambda$, i.e. at very large β, we are nearly in 4d, and the fact that the symmetry is unbroken there means that it should also be unbroken on large circles. This means that the expectation value of the Polyakov loop on the S^1_β should be exactly zero, since it is charged under the center symmetry (assuming it has non-trivial N-ality, like the fundamental loop).

What happens on small circles, i.e. at large temperatures? The confining phase melts (the transition is belived to be second order for $SU(2)$ gauge theory, with Ising critical exponents, and 1-st order for $N \geq 3$, becoming strongly 1st order for largere N), and at very high temperatures one sees the deconfined quarks and gluons of the UV theory. In some very rough sense (which is not completely correct) high-temperatures should correspond to weak coupling $g(T) \ll 1$ because of asymptotic freedom, and this is mostly (but not completely) true. Using this one can say quite a bit about the high-T phase, which is a quark-gluon plasma. Now the Polyakov loop has an expectation value – this does not mean perimeter law, since we are not scaling the size of the loop, but simply that it has an expectation value when viewed as a 3d local operator.

The transition from low-T confined phase to large-T deconfined phase is a sharp thermodynamic phase transition. In fact it is more or less a standard Landau-style symmetry breaking transition, except that the order parameter is a Polyakov loop (local in 3d, but not in 4d) and the broken symmetry looks like a normal 3d 0-form symmetry, but actually descends from a 4d 1-form symmetry.

9.2 From YM to QCD

So far we have mostly discussed pure YM theory, possibly in with very heavy probe quarks. QCD contains light quarks. They are described by Dirac fermions Ψ_D in the fundamental representation R (or any other representation if we want to consider that) of the gauge group, with Lagrangian

$$\mathscr{L}_{\text{quark}} = i\overline{\Psi}_D\gamma^\mu \left(\partial_\mu - iA_\mu^a T^a(R)\right)\Psi_D - m_q\overline{\Psi}_D\Psi_D \ . \tag{9.4}$$

Here m_q is the quark mass. If there are several flavors of quarks, there is such a Lagrangian for each of them.

In QCD, the quarks are in the fundamental $R = \square$ of the $SU(N)$ group. We consider the situation where there are N_f such quarks, which we take either massless or light compared to Λ. In the real world we have $N = N_f = 3$, with the u, d, s quarks being rather light. There are also c, b, t quarks that are much heavier that we'll ignore in our discussion.

In this theory, the QCD coupling runs off to very strong values and can make the quarks of QCD form color-neutral bound states: either $Q\overline{Q}$ meson bound states, or $\wedge^N Q$ baryon bound states. Together these make up the Hadronic degrees of freedom we see in real world QCD, at least for $N = 3$ colors.

In common parlance one calls this regime of QCD "confined" but it is more correct to say that it is color neutral or color screened, i.e. all observed particles carry zero net color charge.

10 QCD Has No 1-form Symmetries; Higgs/Confinement Continuity

Above, I alluded to the fact that confinement in QCD is less sharply defined than in QCD. Let us elaborate on this.

Pure YM theory has an exact $\mathbb{Z}_N^{(1)}$ one-form center symmetry, which protects the strings. The conserved "chromoelectric" flux is the center charge (or N-ality) of the string. The fact that the symmetry is at most $\mathbb{Z}_N$ was related to the fact that the adjoint gluons in YM can screen any line of vanishing N-ality.

Adding additional fields in the adjoint representation of the gauge group, e.g. scalars or fermions, which naturally appear in supersymmetric extensions of YM theory, does not interfere with the 1-form symmetry which remains $\mathbb{Z}_N$.

However, things are different in the presence of dynamical matter fields in the fundamental representation. For instance, QCD has fundamental fermionic quarks. We can also contemplate theories with fundamental scalar fields, e.g. the standard model Higgs field is in the fundamental of the $SU(2)_W$ weak gauge group. Let us focus for concreteness on QCD.

Just like electrons in QED can end the fundamental $q_e = 1$ Wilson lines, thus breaking the electric $U(1)_e^{(1)}$ one-form symmetry of free Maxwell theory completely, the fundamental quarks in QCD completely break the $\mathbb{Z}_N^{(1)}$ center symmetry of pure $SU(N)$ YM theory.

This is because such quarks can end a fundamental Wilson line (and products of quarks can end any Wilson line with non-zero N-ality). Thus the quarks can screen all of these lines.

This has a number of physical ramifications:

- The stable strings of pure YM, which are charged under $\mathbb{Z}_N^{(1)}$, become unstable in the presence of finite-mass quarks. This is because the strings, which are electric flux tubes, can break/snap open by pair-creating $Q\overline{Q}$ pairs. The rate for this process is rapid for light quarks ($m_Q \ll \Lambda$) but is exponentially suppressed for heavy quarks. In the $m_Q \to \infty$ decoupling limit, the quarks disappear and the 1-form symmetry of pure YM re-emerges.

- Because of the screening, the expecation values of large Wilson loops now have perimeter law. However, because there is no one-form symmetry anymore, the expectation value of large Wilson loops (at zero temperature) or Polyakov loops (at finite T) is no longer an order parameter for symmetry breaking, and hence a non-zero vev does not necessarily imply a phase transition.

  This is most dramatic in real-world QCD at finite temperature: using numerical lattice simulations, it has been shown that QCD with physical, non-zero quark masses does not have a sharp phase transition as one dials the temperature T from the confining "regime" at very low T to the very hot quark-gluon plasma regime at very large T. It is tempting to think that these regimes are different phases, but they are like the liquid and vapor phase of water: qualitatively very different, but in principle part of the same thermodynamic phase.

- A closely related phenomenon is that it QCD with fundamental matter it is not straightforward to distinguish the confining phase from a Higgs phase. Perhaps the simplest example is $SU(2)$ gauge theory with a single Higgs field h^a in the doublet representation of $SU(2)$. In the positive mass-square phase for h^a we can integrate out the Higgs field and obtain pure YM gauge theory, which is gapped and trivial. If the Higgs mass is very large it has an emergent $\mathbb{Z}_2$ 1-form symmetry that is unbroken in the confining phase, but it is not exact.

  In the large negative mass-square phase for h, the theory is in a Higgs phase and the $SU(2)$ gauge group is completely Higgsed. **[Go through this in detail, it is like in the Standard Model!]** Thus both large-mass phases are gapped and trivial, with no unbroken symmetries. It is reasonable to conjecture that there is no phase transition in between. This has not been shown analytically for the model I discussed, but it has been looked ad with the lattice. Also, exact analytic proofs are available for some lattice gauge theories with matter – most famously due to Fradkin and Shenker (late 70s). The resulting standard lore that emerges from this is that in the presence of fundamental matter, Higgsing and confinement cannot be distinguished and should be continuously connected, without a phase transition. This lore has been verified in many examples.

  However, there are some curious exceptions involving symmetry-protected topological phases (SPTs), which you can learn about in some recent talks I have given.