LECTURE II-2, PART II: SPONTANEOUS BREAKING OF GAUGE SYMMETRY

Edward Witten

Notes by Pavel Etingof and David Kazhdan

In this lecture we will consider gauge symmetry breaking.

2.1. Gauge symmetry. Recall what gauge symmetry is. We have a spacetime $X = \mathbb{R}_{time} \times X_0$. We have a compact gauge group G. We have a field theory where a field configuration is a connection is some principal bundle over $\mathbb{R}^d$ and possibly some matter fields.

Recall the Hamiltonian approach to gauge theory. Let $\tilde{M}_0$ be the space of solutions to the classical equations of motion. On $\tilde{M}_0$ we have an action of the group $\hat{G}$ of gauge transformations. Let $M_0 \subset \tilde{M}_0$ be the space of all solutions where the G-bundle is trivialized in the time direction, and the connection is trivial in that direction. Such solutions as usual are completely determined by the pair $A(t_0), \frac{dA}{dt}(t_0)$, where $t = t_0$ is a space cycle, and initial data for the matter fields. It is clear that any element of $\tilde{M}_0$ can be brought to M_0 by a gauge transformation, so M_0 still contains all solutions up to gauge transformations.

Suppose that $X_0 = \mathbb{R}^{d-1}$. In this case we may consider only trivial bundles, and connections which vanish at spatial infinity. In other words, if $\mathcal{A}$ is the space of connections A on the trivial G-bundle over $\mathbb{R}^{d-1}$ which vanish at ∞ then M_0 for pure gauge theory is $T^*\mathcal{A}$. If matter fields are present, then M_0 is a product of $T^*\mathcal{A}$ with some other space.

Define $\tilde{G}$ to be the group of elements of $Maps(\mathbb{R}^{d-1}, G)$ which have a limit at infinity, and $\tilde{G}_0$ to be the subgroup of $\tilde{G}$ consisting of functions which tend to 1 at ∞. We have $\tilde{G}/\tilde{G}_0 = G$. This quotient group is called the group of constant gauge transformations at ∞ and called G_∞.

The group $\tilde{G}$ acts symplectically on M_0. The physical phase space in gauge theory is the symplectic quotient $M = M_0//\tilde{G}_0$. Note that we only divide by $\tilde{G}_0$ and not by the whole group $\tilde{G}$, so that we inherit an action of the quotient G_∞ on M. It is this symmetry group whose breaking we will discuss.

2.2. Breaking of gauge symmetry and charges at infinity.

Definition. Suppose we have a (classical) gauge theory, and let $s \in M$ be its vacuum state. Let $H \subset G = G_\infty$ be the stabilizer of s. In this case we will say that at the vacuum state s the gauge symmetry is broken from G to H.

Thus, by symmetry breaking we mean essentially the same thing as for global symmetry: there is a symmetry of the Poisson algebra of functions on M which does not fix a particular vacuum state.

Important remark. The above expression "the same thing " should be taken with great care. There are some fundamental differences between the two situations, which will become clear below. They come from the fact that in the situation we are considering here, (unlike Lecture II1) the physical observables, being gauge invariant by definition, automatically commute with G and therefore do not, in general, separate points on M; i.e. not every function on M is "observable". In other words, the action of G on the "theory" (in the sense of Lecture II1) is trivial from the beginning.

Let us now compute the action of G (classically). First of all, we have a moment map $\mu : M_0 \to \tilde{\mathfrak{g}}_0^*$, where $\tilde{\mathfrak{g}}_0$ is the Lie algebra of $\tilde{G}_0$ – the algebra of functions from $\mathbb{R}^{n-1}$ to the Lie algebra $\mathfrak{g}$ of G which vanish at infinity. Thus, for any $\varepsilon \in \tilde{\mathfrak{g}}_0^*$ we have a Hamiltonian $Q(\varepsilon) \in C^\infty(M_0)$ defined by $Q(\varepsilon)(X) = \mu(X)(\varepsilon)$.

In fact, it is easy to compute $Q(\varepsilon)$ using Noether formalism. Namely,

$$Q(\varepsilon) = \int_{\mathbb{R}^{d-1}} Tr(\frac{\partial A}{dt}\nabla_A \varepsilon) d^{d-1}x + \text{ matter terms } , \tag{2.1}$$

On M, $Q(\varepsilon) = 0$ if ε vanishes at infinity. Thus, on M we have $\nabla_A^* \frac{dA}{dt} =$ matter terms . In particular, in pure gauge theory $\nabla_A^* \frac{dA}{dt} = 0$.

Taking this into account, we see that on M

$$Q(\varepsilon) = \int_{\mathbb{R}^{d-1}} Tr(\nabla_A(\varepsilon \frac{dA}{dt})) d^{d-1}x. \tag{2.2}$$

Using Stokes' formula, we can rewrite (2.2) as

$$Q(\varepsilon) = \lim_{r\to\infty} \int_{S^{n-2}(r)} *_{d-1} Tr(\varepsilon \frac{dA}{dt}) = \lim_{r\to\infty} \int_{S^{n-2}(r)} *_d Tr(\varepsilon F), \tag{2.3}$$

where F is the curvature of the spacetime connection corresponding to the given point of M_0, and $S^k(r)$ is the k-sphere of radius r. This formula defines the hamiltonians for the action of $G = G_\infty$ on M.

This formula shows that $Q(\varepsilon)$ vanishes for all gauge transformations (not necessarily vanishing at infinity) on a particular state if $F = o(r^{2-n})$, $r \to \infty$ on that state. However, if this is not the case, then $Q(\varepsilon)$ may be nonzero for a constant ε.

Example. Consider a $U(1)$ gauge theory with a charged complex scalar. The fields are a connection A on a hermitian line bundle and a section ϕ of this bundle. The Lagrangian is

$$\mathcal{L} = \frac{1}{4e^2} \int F^2 + \int |D_A \phi|^2 d^4x + \int \frac{\lambda}{8}(|\phi|^2 + v^2)^2 d^4x. \tag{2.3}$$

This is the most general renormalizable Lagrangian in these fields in 4 dimensions. Here e, λ, v are parameters and e^2, λ are positive while v^2 can be positive or negative. For simplicity we assume first that $v^2 \neq 0$.

This theory is not believed to exist in the UV, but we will regard it as an effective theory for some more fundamental theory.

Classically (and quantum mechanically for $e^2, \lambda << 1$) we have two cases.

1. $v^2 > 0$; the potential has a single minimum.
2. $v^2 < 0$; the potential has a circle of minima.

Let us consider how in these two cases the theory behaves in the infrared.

Figure 1. The potential for $v^2 > 0$.

Case 1. $v^2 > 0$. In this case the minimum of energy is attained when $\phi = 0$. First consider the case when the gauge coupling vanishes: $e^2 = 0$. In this case our theory is a direct product of a pure (free) abelian gauge theory and the ϕ^4 theory. Therefore, it has a unique vacuum, and the particles which occur at the lower part of the spectrum are a massless vector, or gauge boson (coming from gauge theory) and two massive real scalars (coming for ϕ^4 theory).

If we turn on small e^2 the situation should remain the same. Indeed, certainly nothing can happen to the massive scalars (the part of the Hilbert space with the nonzero charge, where these scalars are, has a mass gap, and massiveness is an open condition); moreover, their masses must be equal since there is a $U(1)$ symmetry at infinity (the $Q(\varepsilon)$ for constant ε) which prohibits the masses to differ. The fact that $Q(\varepsilon) \neq 0$ is clear since this is so at $e^2 = 0$, when $Q(\varepsilon)$ represents the $U(1)$ global symmetry.

Furthermore, the massless vectors cannot become massive. Indeed, recall that a massless vector means an irreducible representation of $SO(3,1)$ with $p^2 = 0$ and spin 1, i.e. the space of sections of a 2-dimensional equivariant vector bundle over the light cone. This vector bundle cannot be deformed to an equivariant vector bundle over the hyperboloid, since the stabilizer group $SO(3)$ of a point on the hyperboloid does not have an irreducible 2-dimensional representation. Thus, the quantum theory for small coupling will have the same particles – two massive scalars (the real and imaginary part of ϕ) and a massless vector (the gauge boson).

Remark. The above argument on non-deformability of a massless vector fails in 3 and 2 dimensions. For example, in 3 dimensions, the massless vector is just the space of functions over the cone, which can be successfully deformed into the space of functions over a hyperboloid. This actually happens when in pure $U(1)$ gauge theory one introduces a Chern-Simons term $c\int A \wedge dA$. The theory remains free but becomes massive, yielding one massive scalar. In the theory we are considering (for 3 dimensions), this cannot happen dynamically since the Chern-Simons term is odd under change of orientation, but in other theories this could happen.

In fact, quantum mechanically the operator $Q(\varepsilon)$ (for a suitable normalization of ε) has integer eigenvalues, and thus defines (in quantum theory) a $\mathbb{Z}$-grading of the corresponding Hilbert space. In particular, since $Q(\varepsilon) \neq 0$, there are sectors of the Hilbert space which cannot be reached from the vacuum by applying local operators. This shows that we have a fundamental violation of Wightman axioms: the representation of the operator algebra in the physical Hilbert space is not irreducible. However, the theory still has one vacuum only: the minimal energy in the sectors with nonzero charge $Q(\varepsilon)$ is positive.

Figure 2. The potential for $v^2 < 0$.

Case 2. $v^2 < 0$. Let $v^2 = -b^2$. Then classically we have a minimum of energy on the circle $|\phi| = b$. This implies that any finite energy configuration has the property $\phi = be^{i\theta_0}$ at infinity, where θ_0 is a constant. Therefore, by a gauge transformation which has a finite limit at infinity, we can arrange that ϕ is real and positive: $\phi = b + w$ where w is a new real variable. Writing the Lagrangian in terms of the new variables, we will get something with the following quadratic part:

$$L_{quadratic} = \frac{1}{4e^2} \int F^2 + \int d^4x((dw)^2 + M^2w^2) + \int d^4x b^2 A^2. \tag{2.4}$$

It is seen from (2.4) than now all fields are massive. Of course, Lagrangian (2.4) is not gauge invariant for A, since we have already "spent" the gauge symmetry on making ϕ real.

Thus, infrared limit of the corresponding quantum theory is trivial for small values of the couplings. In particular, there are no massless gauge bosons: they have been "eaten" by the ϕ-field. This situation is called Higgs phenomenon, or spontaneous breaking of gauge symmetry.

Note that in spite of the presence of a circle of zero energy states, our theory has only one vacuum. In other words, all points of the circle are regarded as the same state, on the grounds that they are gauge equivalent to each other and therefore define equivalent realizations (i.e. give the same expectation values of gauge invariant local operators) This is a fundamental difference between gauge and global symmetry breaking. In global symmetry breaking, the points of the circle represent different vacua (realizations) of the theory, since there exist non-symmetric operators which have different expectation values at different point of the circle.

Note also that the operator $Q(\varepsilon)$ doesn't act in the Hilbert space of states, since classically $Q(\varepsilon)$ generates a group which rotates the circle and permutes the zero energy states. In particular, in this case local operators act irreducibly in the Hilbert space, and there are no sectors which cannot be reached from the vacuum. This is the difference between case 2 and case 1: in case 1, as you remember, $Q(\varepsilon)$ acts in the Hilbert space nontrivially and defines a splitting into sectors.

Remark. If one tries to compute $Q(\varepsilon)$ in Case 2 (when the symmetry is broken) using formula (2.3), the answer will be zero since the integrand dies rapidly at infinity.

The particles which are found in the infrared in the situation of Case 2 are, according to (2.4), a massive vector (A) and a massive scalar (ϕ). There is only one scalar since ϕ is now real. Thus, at the level of representation theory the Higgs phenomenon arising in Case 2 boils down to a deformation of representations of the Poincare group: a massless vector plus a massless scalar is deformed to a massive vector. Recall for comparison that a massless vector separately cannot be deformed into a massive representation.

Finally, consider the special case $v^2 = 0$. In this case classically we have no symmetry breaking as for $v^2 > 0$, and the particles are a massless vector and two massless scalar. However, it is not expected to be the quantum answer, since this configuration is not stable under perturbations.

Remark. If the Lagrangian we start with is not IR free (say, it is the Lagrangian of an asymptotically free gauge theory) then the classical analysis we discussed above does not apply in quantum theory. In this case the infrared behavior of the theory is difficult to determine. In particular, it could happen that in the infrared the gauge group of the ultraviolet theory will be replaced with some completely different group, which is not even a subgroup in the original group.

2.3. Symmetry breaking and gauging. In conclusion, let us discuss the connection between global symmetry and gauge symmetry breaking. Suppose we have a Lagrangian L of a field theory (say in 4 dimensions) which has a global $U(1)$ symmetry. A typical example is when the theory contains some scalar fields ϕ_j which are sections of hermitian vector bundles, and $U(1)$ acts by multiplication of these sections by $e^{in_j\theta}$ This $U(1)$ symmetry can be gauged, by introducing a $U(1)$ gauge field A and new Lagrangian

$$L_{gauged} = \frac{1}{4e^2}\int F_A^2 + L_A, \tag{2.5}$$

where L_A is L in which all derivatives of ϕ_j are replaced by covariant derivatives.

The statement is that if L is infrared free then for small gauge couplings the symmetry breaking behavior of the theories defined by L and L_{gauged} is usually the same. Namely, if there is breaking of global symmetry for L then there is breaking of gauge symmetry for L_{gauged} and vice versa.

Indeed, let us consider both cases.

Case 1. No global symmetry breaking. In this case classically the minimum of energy is at $\phi_j = 0$, and thus there is a $U(1)$-invariant vacuum. In quantum theory, $U(1)$ acts in the Hilbert space, and there are no massless particles (Goldstone bosons) corresponding to $U(1)$. In this case, for small gauge coupling e, the matter part L_A of the Lagrangian almost decouples from the gauge part; so classically we get a massless gauge boson.

To consider the quantum mechanical situation, we assume that there are no massless particles in the ungauged theory. In this case, the above classical answer is also quantum mechanical for small couplings, by the non-defomability of representations from massless to massive. However, if massless particles (say Goldstone bosons corresponding to other symmetries which are broken) are present, this answer may not be true.

Case 2. Global symmetry breaking. In this case at the minimum of energy some of the ϕ_j is not zero. There is no invariant vacuum, and there is a Goldstone boson corresponding to this symmetry breaking. In this case, pick a vacuum state and a component $\phi_j^{(n)}$ which is not zero at this vacuum. In the gauged theory, we can perform a gauge transformation which will make this component real and positive. This shows that if there are no other massless particles (in particular, no other broken global symmetries), all fields in the theory will become massive. This happens classically, due to Higgs mechanism as in Case 2 above, and also quantum mechanically for small couplings. Thus, we have breaking of gauge symmetry.