Lecture 13

Eslam Khalaf

October 16, 2023

1 Electromagnetic gauge potential in quantum mechanics

We have already been discussing several examples that behave like the electromagnetic gauge field. Now we are going to discuss how the actual electromagnetic field behaves in the quantum theory. Recall that in Maxwell theory, the electromagnetic scalar and vector potentials are defined via

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}, \qquad \boldsymbol{E} = -\boldsymbol{\nabla}\phi - \frac{\partial}{\partial t}\boldsymbol{A} \tag{1}$$

The first equation follows from the Maxwell equation $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$ which forbids the existence of magnetic monopole. We will get back to this later and see that there is actually a loop hole that allows for the existence of monopoles despite Eq. 1 (provided that the vector potential $\boldsymbol{A}$ is not globally defined). The second equation ensures the Maxwell equation $\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}$. Importantly, the scalar and vector potentials yield a redundant description for the physics since the gauge transformation

$$\boldsymbol{A} \mapsto \boldsymbol{A} + \boldsymbol{\nabla}\Lambda, \qquad \varphi \mapsto \varphi - \frac{\partial}{\partial t}\Lambda \tag{2}$$

leaves the $\boldsymbol{E}$ and $\boldsymbol{B}$ fields invariant. For this reason, the gauge potentials are usually thought in the classical context as a convenient tool to introduce electromagnetism in the Hamiltonian and Lagrangian formalisms but not as an essential ingredient of the physics. The latter only depends on the physical fields $\boldsymbol{E}$ and $\boldsymbol{B}$. As we will see, the situation is different in the quantum theory where there is a sense in which gauge fields are fundamental objects.

In the Hamiltonian formulation of classical physics, the gauge fields $\boldsymbol{A}$ and V enter the Hamiltonian for a particle with charge q through

$$H = \frac{1}{2m}(\boldsymbol{p} - q\boldsymbol{A})^2 + q\varphi \tag{3}$$

This ensures that the Hamilton's equations of motion yields correct law for the force exerted on a charged particle (the Lorentz force)

$$\dot{\boldsymbol{x}} = \frac{1}{m}(\boldsymbol{p} - q\boldsymbol{A}), \qquad \dot{\boldsymbol{p}} = -\boldsymbol{\nabla}H = q(\boldsymbol{E} + \dot{\boldsymbol{x}} \times \boldsymbol{B}) \tag{4}$$

In the Lagrangian formulation, this corresponds to the Lagrangian

$$L = \boldsymbol{p} \cdot \dot{\boldsymbol{x}} - H = \frac{1}{2}m\dot{\boldsymbol{x}}^2 + q\boldsymbol{A} \cdot \dot{\boldsymbol{x}} - q\varphi \tag{5}$$

In the quantum theory, $\boldsymbol{p}$ and $\boldsymbol{A}$ are promoted to operators which, in general, do not commute (since $\boldsymbol{A}$ generally depends on position), but the Hamiltonian retains the same form

$$\mathcal{H} = \frac{1}{2m}(\hat{\boldsymbol{p}} - q\boldsymbol{A}(\hat{\boldsymbol{x}}))^2 + q\varphi(\hat{\boldsymbol{x}}) = \frac{1}{2m}\hat{\boldsymbol{\Pi}}^2 + q\varphi(\hat{\boldsymbol{x}}) \tag{6}$$

Here, we have introduced the so-called mechanical momentum operator $\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{p}} - q\hat{\boldsymbol{A}}$. We should now be careful on the definition of the different objects, their commutation relations and their behavior under gauge transformations. First, we note that the canonical momentum components $\hat{p}_i$ are defined as the generators of translation. Thus, they satisfy the commutation relations

$$[\hat{p}_i, \hat{p}_j] = 0, \qquad [\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij} \tag{7}$$

In contrast, the mechanical momentum $\hat{\boldsymbol{\Pi}}$ satisfy the commutation relations

$$[\hat{\Pi}_i, \hat{\Pi}_j] = iq\hbar\epsilon_{ijk}B_k, \qquad [\hat{x}_i, \hat{\Pi}_j] = i\hbar\delta_{ij} \tag{8}$$

The fact that the mechanical momentum $\hat{\boldsymbol{\Pi}}$ has the same commutation relations with the position operator $\hat{\boldsymbol{x}}$ means that it can be understood as the generator of a modified modified version of the translation operator

$$\tilde{T}_{\epsilon\boldsymbol{u}} \approx 1 - \frac{i\epsilon}{\hbar}\boldsymbol{u}\cdot\hat{\boldsymbol{\Pi}} + O(\epsilon^2) \tag{9}$$

which acts the same way as regular translations on $\hat{\boldsymbol{x}}$:

$$\tilde{T}_{\boldsymbol{u}}^{-1}\hat{\boldsymbol{x}}\tilde{T}_{\boldsymbol{u}} = \hat{\boldsymbol{x}} + \boldsymbol{u} \tag{10}$$

The main difference is that this modified translation operators do not commute for different vectors. To see this, consider

$$\tilde{T}_{\epsilon\boldsymbol{u}}\tilde{T}_{\epsilon\boldsymbol{v}}\tilde{T}_{\epsilon\boldsymbol{u}}^{-1}\tilde{T}_{\epsilon\boldsymbol{v}}^{-1} \approx e^{-\frac{i}{\hbar}\epsilon\boldsymbol{u}\cdot\hat{\boldsymbol{\Pi}}}e^{-\frac{i}{\hbar}\epsilon\boldsymbol{v}\cdot\hat{\boldsymbol{\Pi}}}e^{\frac{i}{\hbar}\epsilon\boldsymbol{u}\cdot\hat{\boldsymbol{\Pi}}}e^{\frac{i}{\hbar}\epsilon\boldsymbol{v}\cdot\hat{\boldsymbol{\Pi}}} = e^{-\frac{\epsilon^2}{\hbar^2}u_iv_j[\hat{\Pi}_i,\hat{\Pi}_j]} = e^{-i\frac{q\epsilon^2}{\hbar}\epsilon_{ijk}u_iv_jB_k} = e^{-i\frac{q}{\hbar}(\epsilon\boldsymbol{u}\times\epsilon\boldsymbol{v})\cdot\boldsymbol{B}} \tag{11}$$

The phase here is just the flux of the field $\boldsymbol{B}$ through the infinitesmal parallelogram spanned by the vectors $\epsilon\boldsymbol{u}$ and $\epsilon\boldsymbol{v}$. It turns out that this relation also holds for general vectors $\boldsymbol{u}$ and $\boldsymbol{v}$; $\tilde{T}_{\boldsymbol{u}}\tilde{T}_{\boldsymbol{v}}\tilde{T}_{\boldsymbol{u}}^{-1}\tilde{T}_{\boldsymbol{v}}^{-1} = e^{-i\frac{q}{\hbar}\Phi_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{B})}$ where $\Phi_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{B})$ is the flux of the fiedl $\boldsymbol{B}$ (which in general depends on the spatial position) through the parallelogram spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$. The operators $\tilde{T}$ are called magnetic translations.

Since the mechanical momentum $\hat{\boldsymbol{\Pi}}$ and the canonical momentum $\hat{\boldsymbol{p}}$ are related by $\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{p}} - q\hat{\boldsymbol{A}}$, at least one of them is gauge-dependent and is thus not physically measurable since its value depends on the unphysical gauge choice. To determine which one is physical, consider the Heisenberg EOM for the position operator

$$\frac{d\hat{x}_i}{dt} = \frac{1}{i\hbar}[\hat{x}_i, \mathcal{H}] = \frac{\hat{\Pi}_i}{m} \tag{12}$$

Since the position operator is a physical observable, it should not depends on the gauge. This means that $\hat{\boldsymbol{\Pi}}$ should also be gauge invariant. As a result, we see that the canonical momentum $\hat{\boldsymbol{p}}$, defined as the generator of regular (commuting) translations and satisfying the Heisenberg commutation relations is **not** gauge invariant and is thus not a direct physical observable.

1.1 Solenoid flux and Aharonov-Bohm effect

The behavior of the Hamiltonian under gauge transformations enforces a corresponding change in the wavefunctions. In particular, consider a solution to the Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi,t\rangle = \mathcal{H}|\psi,t\rangle \tag{13}$$

Under the gauge transformation (2), the Hamiltonian transforms as

$$\mathcal{H} \mapsto \tilde{\mathcal{H}} = \frac{1}{2m}(-i\hbar\boldsymbol{\nabla} - q\boldsymbol{A} - q\boldsymbol{\nabla}\Lambda)^2 + q(\varphi + \frac{\partial}{\partial t}\Lambda) \tag{14}$$

It is straightforward to verify that the wavefunction

$$|\tilde{\psi},t\rangle = e^{\frac{iq}{\hbar}\Lambda}|\psi,t\rangle \tag{15}$$

satisfies the Schrödinger equation with $\mathcal{H}$ replaced by $\tilde{\mathcal{H}}$.

The way the vector potential enters the Hamiltonian in quantum mechanics has far-reaching consequences. Consider the following simple problem. Take an infinitely long solenoid with a total magnetic flux Φ. Although the field outside the solenoid is zero, the vector potential cannot be made zero since the flux piercing any closed loop that contains the solenoid is Φ. This means

$$\Phi = \int d\boldsymbol{S} \cdot \boldsymbol{B} = \int d\boldsymbol{l} \cdot \boldsymbol{A} \tag{16}$$

This, we can choose $\boldsymbol{A}$ in cylindrical polar coordinates to be

$$A_\phi = \frac{\Phi}{2\pi r} \tag{17}$$

outside the solenoid which guarantees (16) is satisfied. Now consider a particle confined to a ring of radius R that encloses the solenoid. It is straightforward to write the Schrodinger equation

$$\frac{1}{2m}(-i\frac{\hbar}{r}\frac{\partial}{\partial\phi} - q\frac{\Phi}{2\pi r})^2\psi_n(\phi) = E_n\psi_n(\phi) \tag{18}$$

whose solutions are

$$\psi_n(\phi) = e^{in\phi}, \qquad E_n = \frac{1}{2mr^2}(\hbar n - \frac{q\Phi}{2\pi})^2 \tag{19}$$

Remarkably, we find that the energy spectrum knows about the magnetic flux Φ although the particle only moves in a region where the magnetic field vanishes. This illustrates the crucial importance of the electromagnetic gauge potential in quantum mechanics. Although the gauge invariant and thus physical quantity is the $\boldsymbol{B}$ fields (or more generally the flux of the $\boldsymbol{B}$ field through different surfaces) a quantum mechanical particle depends **non-locally** on the physical fields/fluxes. If we insist that the state and energy of a quantum mechanical particle only depends on local fields it experiences, we have to use the gauge dependent vector potential $\boldsymbol{A}$ instead. This indicates the fundamental role of the vector potential $\boldsymbol{A}$ in the quantum theory.

A dramatic manifestation of this phenomenon is realized in the so-called Aharonov-Bohm effect. This effect is most clearly illustrated in the path integral language. Assuming the scalar potential φ vanishes, we see that the gauge potential enters the action as

$$S[\boldsymbol{A}] = S[\boldsymbol{A}=0] + q\int_0^T dt\dot{\boldsymbol{x}} \cdot \boldsymbol{A} = S[\boldsymbol{A}=0] + q\int_{\boldsymbol{x}_i}^{\boldsymbol{x}_f} d\boldsymbol{x} \cdot \boldsymbol{A} \tag{20}$$

We note that the dependence on the field is contained in the last term which is geometric, i.e. only depends on the path. As a result, the probability amplitude for a particle to go through a loop that encloses a solenoid flux Φ is modified due to the flux as $\langle x, T|x, 0\rangle \mapsto \langle x, T|x, 0\rangle e^{i\gamma}$ where

$$e^{i\gamma} = e^{\frac{iq}{\hbar}\oint_C d\boldsymbol{x}\cdot\boldsymbol{A}} = e^{\frac{iq}{\hbar}\Phi} \tag{21}$$

Crucially, this phase does not depend on the details of the loop as long as it winds around the solenoid only once (for a path that winds n times, we get the phase $e^{in\gamma}$). The phase γ can be identified with the Berry phase we introduced last lecture if we think of slowly deforming the Hamiltonian to change the location of the particle so that it traces a loop. For example, we can take a confining potential which confines the particle in the vicinity of a point x_0 that is taken to change slowly in time to trace a loop enclosing the solenoid. An observable consequence of the Aharonov-Bohm effect is realized if we consider a setup where an electron is forced to take one of two paths that enclose the solenoid flux Φ. This can be realized in a modified double slit experiment where the flux is introduced somewhere behind the two-slit screen or in a metallic ring that encloses some flux. At $\Phi = 0$ and at any given point, there will be some probability (or intensity) that depends on the relative phase of the amplitude from the two slits/paths. As we change Φ, we are effectively changing the relative phase leading to a periodic pattern of constuctive/destructive interference with period $2\pi\hbar/q$.

1.2 Monopole quantization

Let me now discuss another important implication of the way the electromagnetic field affects quantum mechanical probability amplitudes. It is related to some of the properties for the spin Berry phase discussed earlier that seemed mysterious. First, the spin Berry phase could be phrased as the flux of a particle moving in the field of a magnetic monopole. Second, the corresponding gauge field always seemed to have a singularity. These two properties turned out to be tied.

Let us first try to understand the first property. Let us assume that we somehow have a magnetic monopole. Can we deduce anything about its properties? a brilliant argument by Dirac which we have already presented in a different context shows that indeed there is an important restriction on any monopole we could have. The field of a magnetic monopole of strength g is $\boldsymbol{B} = g\frac{\hat{r}}{4\pi \boldsymbol{r}^2}$ which represents a field pointing radially outwards and decaying as $1/r^2$. A particle with charge q moving on a closed loop on a circle of radius R acquires the Aharonov-Bohm phase $e^{\frac{iq}{\hbar}\int_{\mathcal{S}} d\boldsymbol{S}\cdot\boldsymbol{B}}$ where $\mathcal{S}$ is any area enclosing the loop. Following our previous discussions, we see that we have two possible areas which should give us the same phase which means that the integral $\frac{iq}{\hbar}\int d\boldsymbol{S}\cdot\boldsymbol{B}$ over a sphere of radius R has to be quantized leading to

$$qg = 2n\pi\hbar \tag{22}$$

This is a remarkable result! It tells us that if a monopole exists, its charge has to be quantized. Furthermore, repeating this argument for other particles with charge q', we deduce that all charges in the universe has to be quantized in terms of an elementary charge e, $q = ne$, such that the minimal monopole charge is $g = \frac{2\pi\hbar}{e}$.

In the argument above, we have avoided discussing the gauge potential. In fact, when trying to think of the gauge potential of a monopole, we immediately encounter an issue: a magnetic field constructed from a vector potential via $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$ cannot have a monopole i.e. a source of the magnetic field, since the divergence of a curl is always zero. This something we know from electromagnetism where the magnetic field lines always form closed loops and a magnet always have two poles. So how would we ever get a magnetic monopole? the loophole in this argument is that it assumes the field $\boldsymbol{A}$ is non-singular. For singular $\boldsymbol{A}$ fields, it turns out to be possible to describe a magnetic monopole as we have seen earlier. Let me now present a simpler calculation to illustrate this effect. Consider the magnetic vector potential defined in the 2D plane by

$$\boldsymbol{A} = \frac{\Phi}{2\pi}\boldsymbol{\nabla}\arg(x+iy) = -i\frac{\Phi}{2\pi}\boldsymbol{\nabla}\ln\frac{x+iy}{\sqrt{x^2+y^2}} = \frac{\Phi}{2\pi}\frac{(-y,x)}{x^2+y^2} \tag{23}$$

This field is singular at $x = y = 0$ and non-singular everywhere else. Consider the field $\boldsymbol{B}$ defined as $B\hat{z} = \boldsymbol{\nabla}\times\boldsymbol{A}$. Naively substituting (23), we get zero field which is not surprising since the curl of a gradient is zero. On the other hand, the flux through a circle of radius r surrounding 0 is given by

$$\oint d\boldsymbol{l}\cdot\boldsymbol{A} = \frac{\Phi}{2\pi}\oint d\phi = \Phi \tag{24}$$

This shows that we should be careful when dealing with singular gauge fields.

The singularity of the gauge field can be understood as follows. Imagine we only have particles whose charge is a multiple of a certain fundamental charge e, $q = ne$. A consequence of the Aharonov-Bohm effect is that a flux of $2\pi\hbar/e$ is undetectable. This means that if we consider an infinitely thin solenoid enclosing a flux of $2\pi\hbar/e$, it will be undetectable unless a particle intersect with it. Now, the end points of such solenoid will looks like a source and sink of magnetic flux, i.e. a magnetic monopole and anti-monopole. If we send the anti-monopole to infinity, we basically have a description for a magnetic monopole. The invisible solenoid connecting the monopole to infinity is called the Dirac string and it coincides with.

While we may be able to get away with this notion of a singular gauge field, it is mathematically problematic. Furthermore, this singularity is unphysical since the magnetic field itself is only singular at the origin not at a line extending from the origin to infinity. To resolve this isse and find a more mathematically sound way to define the gauge field of a monopole, recall another example of an unnecessary singularity:

the polar coordinates on the sphere. The polar coordinates are not well-defined at the north and the south pole of the sphere since at $\theta = 0$, π, ϕ is undefined. On the other hand, there is nothing special about the north or the south pole and we should be able to define coordinates there without the problem. The caveat is that we need to define two different coordinate patches, rather than a single global coordinate chart. The resolution to our monopole gauge potential is very similar. Let us define the vector potential

$$A_\phi^N = \frac{g}{4\pi r}\frac{1-\cos\theta}{\sin\theta} \tag{25}$$

This potential is well-defined away from the south pole and it gives us the monopole field since

$$\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}^N = \frac{1}{r\sin\theta}\frac{d}{d\theta}(A_\phi^N\sin\theta)\hat{r} - \frac{1}{r}\frac{\partial}{\partial r}(rA_\phi^N)\hat{\theta} = g\frac{\hat{r}}{4\pi r^2} \tag{26}$$

We now define another gauge potential

$$A_\phi^S = -\frac{g}{4\pi r}\frac{1+\cos\theta}{\sin\theta} \tag{27}$$

which is singular at the north pole but not at the south pole and also yields the same magnetic field. The trick now is to define two patches: one covering the northern hemisphere and extending a bit in the southern hemisphere (but avoiding the south pole) where we take A_ϕ to be A_ϕ^N and the other covering the southern hemisphere and extending a bit in the northern hemisphere (while avoiding the north pole) where we take A_ϕ to be A_ϕ^S. We are free to do as long as we can define a gauge transformation connecting the two gauge field on the region where they overlap. It is straightforward to see that the two vector potentials are related by

$$A_\phi^N = A_\phi^S + \frac{1}{r\sin\theta}\partial_\phi\omega, \qquad \omega = \frac{g\phi}{2\pi} \tag{28}$$

However, we see there is a problem. The function ω is not really single valued on the equator (where the two patches are overlapping) since its value at $\phi = 0$ and $\phi = 2\pi$ is different. It seems we have moved the singularity somewhere else! However, remember that a gauge transformation acts on the wavefunctions via $e^{\frac{iq\omega}{\hbar}}$ (Eq. 15), thus it suffices that this phase is single-valued on the sphere. This is satisfies if

$$qg = 2\pi\hbar n \tag{29}$$

leading to the monopole quantization condition. In summary, we can define two patches with two associated gauge fields which are non-singular on the patch they are defined on. The non-trivial element of the construction is to ensure that the two patches are related by a valid (i.e. single-valued) gauge transformation where they overlap. This yields a consistency condition that imply the quantization of monopole charge. Some of you may recognize this as the construction of a non-trivial $\mathrm{U}(1)$-bundle on the sphere.