What is a gauge theory?

In an earlier article, "A short introduction to quantum mechanics VIII: global gauge transformation," we have noted that the observed values in quantum mechanics are invariant under an overall, constant phase rotation called a "global gauge transformation." To explore this idea further, let's see whether the observed values change if a position-dependent phase is added to the wave function – that is, whether the observed values are invariant under the following transformation:

$$\psi \to e^{iq\theta(x)}\psi \tag{1}$$

where $\theta(x)$ is a real number that depends on position x, and q is a constant that turns out to be electric charge. Such a transformation is called a "gauge transformation."

Now, let's assume that we have a kinetic term in Schrödingers equation; i.e. a term that involves spacetime derivative(s), such as $\frac{\partial\psi}{\partial x^\mu}$ – which, in short-hand notation, is written $\partial_\mu\psi$ ($x^0 = t, x^1 = x, x^2 = y, x^3 = z$). To make the observed values invariant, the kinetic term $\partial_\mu\psi$ must transform in the same way as ψ (If you don't see why this is true, just accept it and see where it leads.) In other words, we must have the following:

$$\partial_\mu\psi \to e^{iq\theta(x)}\partial_\mu\psi \tag{2}$$

However, (1) doesn't imply (2) because

$$\partial_\mu(e^{iq\theta(x)}\psi) = iq\partial_\mu\theta(x)e^{iq\theta(x)}\psi + e^{iq\theta(x)}\partial_\mu\psi \neq e^{iq\theta(x)}\partial_\mu\psi \tag{3}$$

To remedy this situation, we introduce the covariant derivative D_μ, defined as follows:

$$D_\mu\psi = (\partial_\mu - iqA_\mu)\psi \quad ^1 \tag{4}$$

Here we have introduced A_μ, which is called a "connection." If A_μ varies as $A_\mu \to A_\mu + \partial_\mu\theta(x)$, while ψ varies as (1), then the extra term $iq\partial_\mu\theta(x)$ in (3) can be cancelled. Then, we get

$$D_\mu\psi \to e^{iq\theta(x)}D_\mu\psi \tag{5}$$

So, in Schrödinger's equation for the momentum operator, we have to introduce a covariant derivative instead of an ordinary derivative; a momentum operator with ordinary derivatives spoils gauge symmetry in Schröinger's equation. In other words, with ordinary derivatives, (5) is not satisfied, and the observed values are not invariant under (1). Given this, let's write Schrödinger's equation with covariant derivatives. Then we get:

[1] The usual convention is $D_\mu\psi = (\partial_\mu + iqA_\mu)\psi$; we use a different definition in the interest of convenience.

$$i\hbar D_t\psi = \frac{(-i\hbar D_x)(-i\hbar D_x) + (-i\hbar D_y)(-i\hbar D_y) + (-i\hbar D_z)(-i\hbar D_z)}{2m}\psi \tag{6}$$

This equation has gauge symmetry, since under a gauge transformation, both sides pick up the phase factor $e^{iq\theta(x)}$ which we then cancel.

The above equation can be rewritten as follows:

$$\begin{aligned}(i\partial_t + qA_t)\psi &= \frac{(p-qA)^2}{2m}\psi \\ i\partial_t\psi = H\psi &= \left[\frac{(p-qA)^2}{2m} + q(-A_t)\right]\psi\end{aligned} \tag{7}$$

where p is the momentum defined by ordinary derivatives. We have set $\hbar = 1$ for simplicity; this is equivalent to re-defining A with an $\hbar$ factor. Notice that the kinetic energy is given not by $\frac{p^2}{2m}$ as usual, but by $\frac{(p-qA)^2}{2m}$. However, if you remember your advanced classical mechanics or quantum mechanics, you will notice that this is the exact Hamiltonian for a particle in an electromagnetic field, where (A_x, A_y, A_z) is the vector potential $\vec{A}$, and $-A_t$ the electric potential ϕ. Therefore, we have an interpretation of the connection A_μ as the electromagnetic potential. Also, one can easily check that the electric field and the magnetic field are invariant under the gauge transformation. $A_\mu \to A_\mu + \partial_\mu\theta$ implies:

$$\begin{aligned}\phi &\to \phi - \partial_t\theta \\ \vec{A} &\to \vec{A} + \nabla\theta\end{aligned} \tag{8}$$

which leaves the following intact:

$$\begin{aligned}\vec{E} &= -\nabla\phi - \frac{\partial\vec{A}}{\partial t} \\ \vec{B} &= \nabla\times\vec{A}\end{aligned} \tag{9}$$

In this article, we introduced the idea of a covariant derivative by requiring that the theory has a gauge invariance or a gauge symmetry (i.e. the theory is invariant under a gauge transformation). Here, the gauge group is $U(1)$, as $e^{iq\theta(x)}$ is an element of $U(1)$. $U(1)$ is the group of 1×1 unitary matrices; i.e. the set of complex numbers with magnitude 1.) As this introduction of gauge invariance led to the correct Schrödinger equation in the presence of an electromagnetic field, we can say that the electromagnetic theory is a $U(1)$ gauge theory. Because $e^{iq\alpha(x)} \times e^{iq\beta(x)} = e^{iq\beta(x)} \times e^{iq\alpha(x)}$, the $U(1)$ group is Abelian (i.e. $AB = BA$ for all group members A and B). and we call the electromagnetic theory an Abelian gauge theory. Remarkably, all known physical theories are gauge theories; the electroweak theory has $SU(2)\times U(1)$ as a gauge group, and quantum chromodynamics has $SU(3)$ as a gauge group. ($SU(N)$ is the group of $N\times N$ unitary matrices with determinant 1. Such a matrix is called a "special unitary matrix.") Since the gauge groups of these theories are non-Abelian, ($AB \neq BA$ for group members A and B) we call such theories non-Abelian gauge theories, or Yang-Mills theories – named after the inventors Yang and

Mills who came up with the first non-Abelian gauge theory mathematically in the 1950s. The standard model unifies electroweak theory and quantum chromodynamics, so its gauge group is $SU(3) \times SU(2) \times U(1)$. Nobody knows why all known physical theories have gauge symmetries. Maybe God knows.