1. El electromagnetismo como teoría de gauge

En esta clase vamos a asumir que sabemos qué es el electromagnetismo, y vamos a mostrar que la teoría de un campo de Dirac acoplado al campo electromagnético posee invariancia de gauge. Después vamos a ver que si imponemos la invariancia de gauge al campo de Dirac, entonces podemos deducir la existencia de algo así como el electromagnetismo.

1.1. Electromagnetismo no relativista

El electromagnetismo es el fenómeno por el cual determinados cuerpos, llamados “cargados”, sufren una fuerza que depende de su posición y velocidad, de acuerdo con la fórmula de Lorentz

$$m\ddot{\boldsymbol{x}} = e\left[\boldsymbol{E} + \frac{1}{c}\boldsymbol{v}\times\boldsymbol{B}\right] \tag{1}$$

donde e es la “carga” del cuerpo en cuestión, $\boldsymbol{E}$ es el “campo eléctrico”, $\boldsymbol{B}$ es el “campo magnético”, $\boldsymbol{v} = \dot{\boldsymbol{x}}$ es la velocidad, y c es la velocidad de la luz en vacío. Por supuesto, que no tengamos que especificar la velocidad de la luz medida por quién es la inspiración inicial de la teoría de la Relatividad.

Uno de los temas más espinosos en el electromagnetismo es encontrar un sistema de unidades consistentes. Nosotros hemos elegido escribir la fuerza de Lorentz de tal manera que los campos eléctricos y magnéticos tengan las mismas unidades. Respecto de la carga, definimos que [1]

$$\frac{e^2}{\hbar c} = \alpha = \frac{1}{137} \tag{2}$$

es un número puro, la “constante de estructura fina”. Como $\hbar$ tiene unidades de acción, $[\hbar] = \mathrm{M\,L^2\,T^{-1}}$ eso determina las unidades de e. Observando que eE (lo mismo que eB) tiene unidades de fuerza, determinamos las unidades de los campos.

La dinámica de los campos está regida por las ecuaciones de Maxwell, las que a su vez se pueden dividir en dos grupos, las ecuaciones homogéneas

$$\begin{aligned} \nabla\cdot\boldsymbol{B} &= 0 \\ \nabla\times\boldsymbol{E} + \frac{1}{c}\frac{\partial\boldsymbol{B}}{\partial t} &= 0 \end{aligned} \tag{3}$$

y las inhomogéneas

$$\begin{aligned} \nabla\cdot\boldsymbol{E} &= 4\pi\rho \\ \nabla\times\boldsymbol{B} - \frac{1}{c}\frac{\partial\boldsymbol{E}}{\partial t} &= \frac{4\pi}{c}\boldsymbol{j} \end{aligned} \tag{4}$$

Donde ρ es la densidad de carga y $\boldsymbol{j}$ la corriente. Estas ecuaciones implican la conservación de la carga

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\boldsymbol{j} = 0 \tag{5}$$

Las ecuaciones homogéneas implican que los campos pueden ser escritos en términos de los potenciales

$$\begin{aligned} \boldsymbol{B} &= \nabla\times\boldsymbol{A} \\ \boldsymbol{E} &= -\frac{1}{c}\frac{\partial\boldsymbol{A}}{\partial t} - \nabla\phi \end{aligned} \tag{6}$$

La descripción en términos de potenciales es más compacta que la de los campos, pero tiene el problema de que no es unívoca. Distintas configuraciones de potenciales, aparentemente completamente diferentes, pueden estar describiendo los mismos campos, y por lo tanto la misma física. Esto ocurre cuando las dos configuraciones, digamos $(\boldsymbol{A}, \phi)$ y $(\boldsymbol{A}', \phi')$, pueden obtenerse una de la otra mediante una *transformación de gauge*

$$\begin{aligned} \boldsymbol{A}' &= \boldsymbol{A} + \nabla f \\ \phi' &= \phi - \frac{1}{c}\frac{\partial f}{\partial t} \end{aligned} \tag{7}$$

para una función f arbitraria.

Las ecuaciones inhomogéneas son por lo tanto las verdaderas ecuaciones de movimiento de la teoría. estas ecuaciones pueden deducirse de un principio variacional, adoptando como acción

$$S = \frac{1}{8\pi c} \int d^4x \ \left\{ \boldsymbol{E}^2 - \boldsymbol{B}^2 \right\} \tag{8}$$

1.2. Acoplamiento mínimo

La receta del "acoplamiento mínimo" es la forma más simple de acoplar una partícula al campo electromagnético. Si para la partícula neutra tenemos

$$H = \frac{1}{2m} \boldsymbol{p}^2 \tag{9}$$

para la partícula cargada escribimos

$$H = \frac{1}{2m} \left(\boldsymbol{p} - \frac{e}{c} \boldsymbol{A} \right)^2 + e\phi \tag{10}$$

las ecuaciones de Hamilton

$$\begin{aligned} \dot{x}^j &= \frac{\partial H}{\partial p_j} = \frac{1}{m} \left(p^j - \frac{e}{c} A^j \right) \\ \dot{p}_j &= -\frac{\partial H}{\partial x^j} = \frac{e}{c} \dot{x}_k A^k_{,j} - e\phi_{,j} \end{aligned} \tag{11}$$

Para calcular la aceleración tenemos que tener en cuenta que los potenciales están evaluados en la posición de la partícula, que depende el tiempo. Por lo tanto

$$\frac{dA^j}{dt} = \frac{\partial A^j}{\partial t} + \dot{x}^k A^j_{,k} \tag{12}$$

y obtenemos

$$\begin{aligned} m\ddot{x}^j &= \frac{e}{c} \dot{x}_k A^k_{,j} - e\phi_{,j} - \frac{e}{c} \left(\frac{\partial A^j}{\partial t} + \dot{x}^k A^j_{,k} \right) \\ &= e \left[E^j + \frac{1}{c} (\boldsymbol{v} \times \boldsymbol{B})^j \right] \end{aligned} \tag{13}$$

que reproduce la fuerza de Lorentz.

El acoplamiento mínimo se aplica igualmente a una partícula cuántica. Simplemente reemplazamos en la ecuación de Schrödinger

$$\partial_j = \frac{i}{\hbar} p_j \rightarrow \frac{i}{\hbar} \left(p_j - \frac{e}{c} A_j \right) = \partial_j - \frac{ie}{\hbar c} A_j \tag{14}$$

y agregamos el término del potencial escalar al Hamiltoniano. Esta construcción tiene una consecuencia observable inmediata: Si una partícula cargada sigue un camino Γ dado, su función de onda adquiere una fase adicional, respecto a la función de onda de la partícula neutra,

$$\psi \rightarrow \psi e^{(ie/\hbar c) \int_\Gamma dx^j A_j} \tag{15}$$

Esta fase no se anula si la partícula sigue un contorno cerrado, al contrario, tenemos

$$\psi \rightarrow \psi e^{(ie/\hbar c) \Phi} \tag{16}$$

donde Φ es el flujo del campo magnético concatenado por el contorno. Este corrimiento de fase es observable, y constituye el *efecto Aharonov-Bohm.*

1.3. Electromagnetismo relativista

En la teoría de la Relatividad, subsumimos el espacio y el tiempo en una variedad de cuatro dimensiones, el espacio de Minkowski, con coordenadas $x^\mu = \left(ct, x^k\right)$. Adoptamos la convención $(- + ++)$ para la métrica de Minkowski.

Para formular el electromagnetismo relativista, observamos que los potenciales pueden unificarse en un único potencial vector

$$A_\mu = \left(-\phi, A^j\right) \tag{17}$$

A partir de este vector, definimos el tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{18}$$

Explícitamente

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix} \tag{19}$$

Las ecuaciones de Maxwell homogéas (3) se convierten en

$$F_{\mu\nu,\rho} + F_{\nu\rho,\mu} + F_{\rho\mu,\nu} = 0 \tag{20}$$

Que es la condición necesaria y suficiente para que el tensor $F_{\mu\nu}$ admita la representación (18). Las ecuaciones de Maxwell inhomogéneas (4) se convierten en

$$F^{\mu\nu}_{\ ,\mu} = -\frac{4\pi}{c} j^\nu \tag{21}$$

Donde la corriente

$$j^\mu = \left(c\rho, j^k\right) \tag{22}$$

Como $F^{\mu\nu}$ es antisimétrico mientras que las derivadas parciales conmutan, debe ser

$$j^\mu_{,\mu} = 0 \tag{23}$$

que reproduce la conservación de la carga (5). También es obvio que los campos son invariantes frente a la transformación de gauge

$$A_\mu \to A_\mu + \partial_\mu f \tag{24}$$

Las ecuaciones de Maxwell inhomogéneas pueden deducirse de un principio variacional con acción

$$S = \frac{-1}{16\pi c} \int d^4x \; F^{\mu\nu} F_{\mu\nu} + S_M \tag{25}$$

donde S_M es la acción de los campos cargados. Resulta

$$j^\nu = c^2 \frac{\delta S_M}{\delta A_\nu} \tag{26}$$

1.4. El campo de Dirac cargado

Un campo de Dirac neutro está descripto por un espinor de cuatro dimensiones ψ, con el funcional de acción

$$S_M = \int d^4x \; \bar{\psi} \left(i\hbar\gamma^\mu \partial_\mu - mc\right) \psi \tag{27}$$

donde las γ^μ son las matrices de Dirac que obedecen el álgebra de Clifford

$$\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu} \mathbf{1} \tag{28}$$

y

$$\bar{\psi} = \psi^\dagger \gamma^0 \tag{29}$$

Aunque no vamos a necesitar una representación explícita de las matrices de Dirac, puede ser útil tener una a mano. Para eso introducimos las matrices de Pauli

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{30}$$

y escribimos las matrices de Dirac como bloques de 2×2

$$\gamma_0 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \quad \gamma_k = \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix} \tag{31}$$

Nótese que γ^0 es hermítica, y las otras son anti-hermíticas.

Para acoplar el campo de Dirac al electromagnetismo aplicamos la receta de acoplamiento mínimo

$$S_M = \int d^4x\, \bar{\psi} \left[i\hbar\gamma^\mu \left(\partial_\mu - \frac{ie}{\hbar c} A_\mu \right) - mc \right] \psi \tag{32}$$

La corriente asociada al campo de Dirac es

$$j^\mu = ec\, \bar{\psi}\gamma^\mu\psi \tag{33}$$

cuya conservación se obtiene de las ecuaciones de movimiento para ψ.

1.4.1. El campo de Klein-Gordon cargado

Si aplicamos la misma estrategia al campo de Klein-Gordon obtenemos

$$S_M = \int d^4x\ (-1) \left\{ \eta^{\mu\nu} \left(\partial_\mu\varphi^* + \frac{ie}{\hbar c} A_\mu \varphi^* \right) \left(\partial_\nu\varphi - \frac{ie}{\hbar c} A_\nu \varphi \right) + \frac{m^2c^2}{\hbar^2} \varphi^*\varphi \right\} \tag{34}$$

De donde la corriente es

$$j^\mu = \frac{-iec}{\hbar} \left[\phi^* \partial^\mu \phi - \phi \partial^\mu \phi^* \right] - \frac{2e^2}{\hbar^2} A^\mu \left| \phi \right|^2 \tag{35}$$

Para una configuración en que el campo escalar es constante, queda una corriente no nula, proporcional al tetrapotencial. Eso es raro, porque uno está acostumbrado a materiales conductores en que la corriente es proporcional al campo eléctrico, no al potencial.

Sin embargo, este tipo de corriente había sido propuesto por London como explicación fenomenológica del efecto Meissner en superconductores convencionales. Hoy diríamos que φ, que representa la amplitud del condensado de pares de Cooper, es el parámetro de orden de la transición a superconductividad, y que pasar de $\varphi = 0$ a $\varphi \neq 0$ representa un *rompimiento espontáneo de la simetría*. Vamos a volver sobre este tema.

1.5. El electromagnetismo como teoría de gauge

Si miramos la acción para el campo de Dirac neutro (27), está claro que es invariante frente a un cambio de la fase del espinor

$$\psi \to e^{-ie\theta}\psi; \quad \bar{\psi} \to e^{ie\theta}\bar{\psi} \tag{36}$$

Si pretendemos “gaugear” la simetría, es decir, que θ sea función de punto, entonces la invariancia se pierde, porque las derivadas de ψ no se transforman de manera homogénea

$$\partial_\mu e^{-ie\theta}\psi = e^{-ie\theta} \left[\partial_\mu\psi - ie \left(\partial_\mu\theta \right) \psi \right] \tag{37}$$

Sin embargo, nos queda el recurso de incorporar un “campo compensador” A_μ que se transforme de tal manera que absorba el término indeseado. Es decir, reemplazamos la derivada parcial por una *derivada covariante*

$$D_\mu\psi = \left(\partial_\mu - \frac{ie}{\hbar c} A_\mu \right) \psi \tag{38}$$

de tal manera que la derivada covariante se transforme como

$$D'_\mu e^{-ie\theta}\psi = e^{-ie\theta} D_\mu \psi \tag{39}$$

Lo que necesitamos es que

$$\frac{ie}{\hbar c} A'_\mu = \frac{ie}{\hbar c} A_\mu - ie\partial_\mu \theta \tag{40}$$

o sea que

$$A'_\mu = A_\mu - \hbar c \partial_\mu \theta \tag{41}$$

que es precisamente la regla para una transformación de gauge del campo electromagnético. Como vamos a ver en la próxima clase, la invariancia de gauge de la acción (32) implica que la corriente (33) se conserva. De paso, como la corriente no depende de las derivadas del espinor, se ve que es invariante de gauge.

A diferencia de las derivadas ordinarias, las derivadas covariantes no conmutan, pero el conmutador no es un operador diferencial

$$[D_\mu, D_\nu]\, \psi = -\frac{ie}{\hbar c} F_{\mu\nu} \psi \tag{42}$$

donde $F_{\mu\nu}$ es el tensor de campos (18). De las leyes de transformación de ψ y D_μ (o directamente de la de A_μ) se deduce que el tensor de campos es invariante de gauge. Como $F^{\mu\nu}_{\;\;,\mu\nu} = 0$ por simetría, es consistente acoplar este tensor a la corriente conservada.

A esta altura sería perverso no reconocer que imponer la invariancia de gauge del campo de Dirac nos obligó a inventar el electromagnetismo. En la segunda mitad del siglo pasado, este principio se usó sistemáticamente para desarrollar la teoría de las interacciones fuertes y débiles entre partículas subatómicas. El desarrollo de esta historia va a ser nuestro tema en la primera mitad del curso.

Referencias

[1] H. Bethe, G. Beck y W. Riezler, On the Quantum Theory of the Temperature of Absolute Zero, Die Naturwissenschaften 19, 39 (1931) (reimpreso en H. Bethe, *Selected Works of Hans A. Bethe*, World Scientific (1997), p. 185).