CHAPTER 1

Introduction: conventional gauge theories

In this chapter we will give a crash course to gauge theories, put into a historical context.

1. Dirac and the dawn of quantum electrodynamics

In 1928, Dirac asked the question whether there exists a differential operator D such that its square is equal to the Laplace (d'Alembert) operator:

$$D^2 = \sum_{\mu} \eta^{\mu\nu} \partial_\mu \partial_\nu$$

where $\eta = \text{diag}(+1, -1, -1, -1)$ The motivation for this was to find a relativistic version of the Schrödinger equation:

eq:schr (1.1)
$$\sum_{j=1}^{3} \frac{1}{2m} \left(-i\hbar \frac{\partial}{\partial x_j} \right)^2 \psi(x,t) = i\hbar \frac{\partial}{\partial t} \psi(x,t)$$

Indeed, the fact that the left-hand-side involves a second-order differential operator and the right-hand-side a first-order, breaks the special relastivistic symmetry between space and time.

Two solutions can be found to this question, based on replacing the non-relativistic relation $E = p^2/2m$ by Einstein's relations:

einstein (1.2)
$$E = \sqrt{p^2c^2 + m^2c^4}$$

keeping in mind the quantum mechanical identification $p_j = -i\hbar\partial/\partial x_j$.

1.1. The Klein–Gordon equation

The first solution is to square the right-hand-side involving the time-derivative, leading to the **Klein–Gordon equation**:

$$\left(\sum_{j=1}^{3} \left(-i\hbar \frac{\partial}{\partial x_j} \right)^2 + m^2c^4 \right) \psi(x,t) = \left(i\hbar \frac{\partial}{\partial t} \right)^2 \psi(x,t)$$

or, equivalently, $(\Box + m^2c^2/\hbar^2)\psi(x,t) = 0$ with $\Box$ the d'Alembert operator:

$$\Box = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \sum_{j=1}^{3} \left(\frac{\partial}{\partial x_j} \right)^2 .$$

More compactly, $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ after writing $x_0 = ct$. The Klein–Gordon equation describes the relativistic motion of a free scalar particle with mass m.

1.2. The Dirac equation

Another solution to the above problem is to try to take the square root at the left-hand-side of (1.1). Dirac postulated a first-order differential operator H by setting

$$H = \alpha_0 mc^2 + c\sum_{i=1}^{3} \alpha_i \left(-i\hbar \frac{\partial}{\partial x_j}\right)$$

and then demanding that $H^2 = \sum_i c^2(-i\hbar\partial/\partial x_i)^2 + m^2c^4$, according to Einstein's relation (1.2) for the energy. One computes:

$$H^2 = \alpha_0^2 m^2 c^4 + \sum_i (\alpha_0\alpha_i + \alpha_i\alpha_0)mc^3(-i\hbar\partial_i) + \sum_i \alpha_i^2 c^2(-i\hbar\partial_i)^2 + \sum_{i>j}(\alpha_i\alpha_j + \alpha_j\alpha_i)c^2(-i\hbar\partial_i)(-i\hbar\partial_j)$$

which implies that the α_μ satisfy

$$\alpha_0^2 = 1 \qquad \alpha_0\alpha_i + \alpha_i\alpha_0 = 0$$
$$\alpha_i^2 = 1 \qquad \alpha_i\alpha_j + \alpha_j\alpha_i = 0 \quad (i \neq j)$$

Clearly, these relations cannot be satisfied by ordinary numbers, and the smallest representation of this (Clifford) algebra such that H is a symmetric operator is four dimensional.

The **Dirac equation** is given by

$$H\psi = i\hbar\frac{\partial}{\partial t}\psi$$

Introducing the so-called **Dirac gamma matrices** $\gamma^\mu = (\alpha^0, \alpha^0\alpha^i)$ this is equivalent to

eq:dirac

$$\left(i\gamma^\mu\partial_\mu - \frac{mc}{\hbar}\right)\psi(t,x) = 0. \tag{1.3}$$

It describes the relativistic motion of a free electron, or, more generally, of a free fermion. The Dirac matrices satisfy

$$\{\gamma_\mu, \gamma_\nu\} = -2\eta^{\mu\nu}1_4$$

Of course, on a (pseudo)-Riemannian spin manifold one has the analogous Dirac equation written concisely as $(D-m)\psi = 0$ in terms of the **Dirac operator** $D = i\gamma\circ\nabla^S$ with ∇^S a spin connection, and a section ψ of a spinor bundle S. In this case, the square of the Dirac operator is not precisely the d'Alembertian, but we have

THEOREM 1 (Weitzenböck). *If $\Delta = g_{\mu\nu}\nabla^S_\mu\nabla^S_\nu$ is the Laplace–Beltrami operator, then*

$$D^2 = \Delta - \frac{1}{4}R$$

1.3. Principal of extremal action and electrodynamics

In physics, it is convenient to work with **action functionals** and obtain equation of motions – such as the Dirac equation – as extremas of this functional. Let us illustrate this in the equation of interest, that is, the Dirac equation. A fermionic action functional is given as the inner product on $\Gamma(M,S)$:

$$S_f[\psi] = (\psi, (D-m)\psi); \qquad (\psi \in \Gamma(M,S)).$$

In terms of the hermitian structure, we have $S_f[\psi] = \int_M \langle\psi, (D-m)\psi\rangle_x d\mu(x)$. Now, ψ extremizes the action S_f means that the directional derivative

$$S_f'[\psi][\chi] = \lim_{t\to 0}(S_f[\psi + t\chi] - S_f[\psi])/t,$$

vanishes for all $\chi \in \Gamma(M,S)$. One computes that this happens if and only if $(D-m)\psi = 0$. We conclude that the fermionic action S_f describes the physical system of a relativistic particle moving in spacetime M. More generally, the vanishing of the directional derivative of an action functional gives the equation of motion for the corresponding physical system.

REMARK 2. *Note that the Klein–Gordon equation can be obtained as the equation of motion of the action functional*

$$S_{kg}[\phi] = \frac{1}{2}\int_M \eta^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) - \frac{mc^2}{\hbar}\phi^2.$$

An interesting observation is that the above action S_f is invariant under the following **global** $U(1)$ symmetry:

$$\psi \to e^{i\theta}\psi; \qquad (\theta \in [0, 2\pi]).$$

Indeed, with $\langle\cdot,\cdot\rangle_x$ being anti-linear and linear in the first and second entry, respectively, we find that $S_f[e^{i\theta}\psi] = S_f[\psi]$.

Next, suppose that this symmetry transformation on ψ is position dependent, $\theta = \theta(x)$. Clearly, S_f is not invariant under this local $U(1)$-symetry unless we make the following **minimal replacement**:

$$\nabla^S \rightsquigarrow \nabla^S + ieA$$

Here we introduce a new field $A \in \Omega^1(M)$ that transforms under a $U(1)$-transformation as

$$A \mapsto A - e^{-1}d\theta$$

The fermionic action now depends on two fields A and ψ:

$$S_f[A,\psi] = \int_M \langle\psi, (i\gamma^\mu\nabla^S_\mu - e\gamma^\mu A_\mu - m)\psi\rangle_x d\mu(x)$$

Invariance of this functional under the $U(1)$-action follows:

$$S_f[A + ie^{-1}d\theta, e^{i\theta}\psi] = S_f[A,\psi] - \int_M \langle\psi, \gamma^\mu(\partial_\mu\theta)\psi\rangle_x d\mu(x) + \int_M \langle\psi, \gamma^\mu(\partial_\mu\theta)\psi\rangle_x d\mu(x)$$

The second term on the right-hand-side comes from the Leibniz rule for ∇^S on $e^{i\theta}\psi$, and the last term comes from the transformation of the A-field.

If we extremize the new action S_f' with respect to ψ we obtain the equation of motion

$$(i\gamma^\mu\partial_\mu - e\gamma^\mu A_\mu - m)\,\psi = 0$$

This describes the relativistic motion of an electron in the presence of an electromagnetic field A_μ. As such, A_μ satisfies Maxwell's equations; let us derive them here from the principle of extremal action.

The **curvature** of A is defined to be $F = dA \in \Omega^2(M)$. In local coordinates, we write:

$$F = \tfrac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu.$$

where $F_{\mu\nu} = \partial_\mu A_v - \partial_\nu A_\mu$. Maxwell's theory is described by the following action

$$S_{em}[A] := \int_M F \wedge (*F) \equiv \frac{1}{4}\int_M F_{\mu\nu}F^{\mu\nu}.$$

The so-called **Hodge star operator** $* : \Omega^r(M) \to \Omega^{n-r}(M)$ is given locally by

$$*(dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_r}) = \frac{\sqrt{|g|}}{(n-r)!}\epsilon^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{n-r}}dx^{\nu_1} \wedge \cdots \wedge dx^{\nu_{n-r}}$$

with $\epsilon_{\mu_1\cdots\mu_n} = \pm 1$ depending on whether $\mu_1\cdots\mu_n$ is an even or odd permutation of $12\cdots n$. The equation of motion for $S[A]$ is

$$d(*F) = 0$$

which together with the **Bianchi identity** $dF = 0$, which is always satisfied, forms **Maxwell's equation** for electromagnetism. More explicitly, we identify the components of $F_{\mu\nu}$ with the electric and magnetic field as

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

so that $d(*F) = 0 = dF$ become

$$\nabla \cdot \mathbf{B} = 0 \qquad\qquad \nabla \cdot \mathbf{E} = 0$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \qquad\qquad \nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t}$$

2. Non-abelian gauge theories

The above $U(1)$-symmetry principle led Yang and Mills to consider theories with a non-abelian symmetry group. At first, there seemed to be no direct physical application of such theories. However, at the beginning of the 1960s, Glashow, later joined by Weinberg and Salam, used a $U(1)\times SU(2)$-symmetry as underlying their electroweak theory. The **Standard Model of elementary particles** was finally completed by adding a $SU(3)$ quark color symmetry. This model has been tested up to previously unencountered precision, the only missing piece being the Higgs particle.

REMARK 3. *As said, the weak interactions correspond to a $SU(2)$ gauge group. Matter is supposed to be in a representation of this group. For example, the neutron and proton are supposed to be organized in the definining representation:*

$$g \cdot \begin{pmatrix} p \\ n \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \begin{pmatrix} p \\ n \end{pmatrix}$$

Historically, this occurence of $SU(2)$-symmetry was first motivated by the similarity in mass for the neutron and proton (940 and 938 $\mathrm{MeV}c^{-2}$, respectively). The slight difference in mass was later explained by interpreting the proton as a combination of two up and one down quark, and the neutron as one up and two down quarks. Then, the up and down quark are combined in a defining representation of $SU(2)$, which in addition both constitute a representation of $SU(3)$: the three colors of the quarks.

More generally, one considers matter as representations of a Lie group G, typically a matrix group such as $SU(N)$. One might consider N-vectors in the defining representation of $SU(N)$:

$$\Psi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix}$$

with each $\psi_i \in \Gamma(S)$ a spinor. The action functional that describes the dynamics of these N free massless particles is given by

$$S_f[\psi_1, \ldots, \psi_N] = \int_M \langle \Psi, i\gamma^\mu \nabla^S_\mu \Psi \rangle_x d\mu(x) = \sum_{j=1}^N \int_M \langle \psi_j, i\gamma^\mu \nabla^S_\mu \psi_j \rangle_x d\mu(x)$$

Indeed, the extremal points of this action are precisely sections ψ_i that satisfy the Dirac equation. This action has a global $SU(N)$-invariance, since $U \in SU(N)$ acting as

$$\Psi \mapsto U\Psi.$$

leaves S_f invariant. Again, by promoting this to a **local** $SU(N)$-symmetry – *i.e.* $U = U(x)$, requires replacing

$$\nabla^S \rightsquigarrow \nabla^S + A$$

The **gauge field** A is an element in $\Omega^1(M) \otimes su(N)$ that transforms under a $SU(N)$-transformation as

auge-transf (2.1)
$$U : A \mapsto UAU^* + UdU^* \tag{2.1}$$

The **curvature** of A is defined to be

$$F_A = dA + A \wedge A$$

and is an element in $\Omega^2(M) \otimes su(N)$. It transform as $F \mapsto UFU^*$ under a $SU(N)$-transformation.

Yang and Mills then introduce an action functional for such a field, now carrying their name. For $A \in \Omega^1(M) \otimes su(N)$ the **Yang–Mills action functional** is given by

$$S_{ym}[A] = -\int_M \operatorname{Tr} F_A \wedge *F_A$$

One checks that it is invariant under the action of $U(x) \in G$, as in (2.1)

3. Yang–Mills gauge theory: mathematical setup

As the previous example should indicate, the proper mathematical setting for gauge theories is vector bundles and connections thereon.

DEFINITION 4. *Let $E \to M$ be a vector bundle. A* **connection** ∇^E *on E is a map*

$$\nabla^E : \Gamma E \to \Omega^1(M) \otimes_{C^\infty(M)} \Gamma E$$

such that the Leibniz rule is satisfied, i.e.

$$\nabla^E(f\eta) = f\nabla^E(\eta) + df \otimes_{C^\infty(M)} \eta; \qquad (f \in C^\infty(M), \eta \in \Gamma E).$$

The **curvature** F *of* ∇^E *is given by*

$$F = (\nabla^E)^2 : \Gamma E \to \Omega^2(M) \otimes_{C^\infty(M)} \Gamma E$$

In other words, $F \in \Omega^2(M) \otimes_{C^\infty(M)} \Gamma\mathrm{End}(E)$.

Locally, we can always write $\nabla^E = d + A$ with $A \in \Omega^1(M) \otimes_{C^\infty(M)} \Gamma\mathrm{End}(E)$. Similarly, $F \in \Omega^2(M) \otimes_{C^\infty(M)} \Gamma\mathrm{End}(E)$.

Suppose that there is a (smooth) action of a Lie group G on the fibers of $E \to M$. For instance, if E is an associated vector bundle to a G-principal bundle $P \to M$ it naturally comes with such an action. Indeed, one considers the associated fiber bundle

$$\mathrm{Ad}P = P \times_G G$$

so that the **gauge group** $\mathcal{G} := \Gamma\mathrm{Ad}P$ acts fiberwise on ΓE. The Lie algebra of $\mathcal{G}$ is the vector space of section $\Gamma\,\mathrm{ad}\,P$ where

$$\mathrm{ad}\,P = P \times_G \mathfrak{g}$$

In this case, it is also natural to assume that ∇^E comes from a connection on the principal bundle G, so that it is given by a $\mathfrak{g}$-valued one-form ω on P, satisfying:

$$\omega(X^*) = X; \qquad R_g^*\omega = \mathrm{Ad}_{g^{-1}}\omega \qquad (X \in \mathfrak{g}).$$

Of course, we can also write locally $\nabla^E = d + A$ with $A \in \Omega^1(M) \otimes \mathfrak{g}$. There is an action of the gauge group $\mathcal{G}$ on A:

$$A \mapsto uAu^* + udu^*; \qquad (u \in \mathcal{G})$$

identifying $\mathcal{G}$ locally with maps from M to G.

Next, we consider the **tensor product** of the bundle E with the spinor bundle $S \to M$. Essentially, one takes the fiberwise tensor product, and on $S \otimes E$ one can define the **tensor product connection**:

$$\nabla^{S\otimes E} = \nabla^S \otimes 1 + 1 \otimes \nabla^E$$

In local coordinates, one writes

$$\nabla^{S\otimes E} = d + \omega + A.$$

Definition 5. *The* **Dirac operator with coefficients in** E *is given (locally) as*

$$D_E = i\gamma^\mu \nabla_\mu^{S\otimes E} = i\gamma^\mu \nabla_\mu^S + i\gamma^\mu A_\mu$$

The corresponding fermionic action is

$$S_f[A,\psi] = (\psi, (D_E - m)\psi)$$

with ψ a section of $S \otimes E$.

The dynamics of the field A is described by the Yang–Mills action functional, now in its full form.

Definition 6. *The* **Yang–Mills action functional** *is defined for a connection ∇^E locally of the form $d + A$ with $A \in \Omega^1(M) \otimes \mathfrak{g}$:*

$$S_{ym}[A] = -\int_M \mathrm{Tr}\, F \wedge *F.$$

where $F := (\nabla^E)^2$ is the curvature of ∇^E and Tr *the Killing form on $\mathfrak{g}$ (and minus the identity on the abelian part of $\mathfrak{g}$).*

The Yang–Mills action functional transforms under the action of $\mathcal{G}$ as

$$F \mapsto uFu^*$$

Thus, $S_{ym}[A]$ is invariant under the $\Gamma\mathrm{Ad}P$-action.

The equations of motion of the above action reads

$$[\nabla^E, *F] = 0$$

This is called the **Yang–Mills equation**. Note its similarity with the Bianchi identity, which is simply $[\nabla^E, *F] = 0$ and is always satisfied. This is the starting point of **instantons**, *i.e.* connections with a selfdual curvature $F = *F$. For these connections, the Bianchi identity implies the Yang–Mills equation so that instantons are minima of the Yang–Mills action. It was realized later, through the work of Donaldson, that the moduli space of instantons plays a key role in the classification of smooth structures on four-dimensional manifolds.

3.1. Higgs mechanism

Although the above is intriguing from a mathematical viewpoint, nature is (as usual) slightly more complicated. In fact, the $U(1) \times SU(2)$-symmetry discussed above is not observed in nature, only a residual $U(1)$-symmery (namely, electrodynamics). Let us describe the mathematical structure underlying this symmetry breaking.

Suppose that $H \subset G$, and that Φ is a scalar vector, that is, a section of $P \times_G V$. Then, one considers the action functional:

$$S_h[\Phi, A] = \frac{1}{2}\int_M g^{\mu\nu} \nabla_\mu^E \Phi \cdot \nabla_\nu^E \Phi - V(\Phi).$$

Here $V(\Phi)$ is a potential: a polynomial in the components Φ_i. The minima of this potential are supposed to be only invariant under a subgroup H, rather than under the full group G. Physically, this means that when the field Φ 'rolls down' to such a minimum, the symmetry group G is **spontaneously broken** to H.

Let us illustrate this in an example of great physical interest, namely the Weinberg–Salam electroweak theory. In this case, $G = U(1) \times SU(2)$ which will be broken to $H = U(1)$ as follows. The field Φ has two components, Φ_1 and Φ_2 on which $(e^{i\theta}, U) \in U(1) \times SU(2)$ acts by matrix multiplication:

$$\begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} \mapsto e^{i\theta} U \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix}$$

The potential in the above action S_h is taken of the form

$$V(\Phi) = \mu^2 |\Phi|^2 + \lambda |\Phi|^4$$

where, conventionally, $\mu^2 < 0$. This potential has the form of a mexican hat, with minima at $|\Phi|^2 = -\mu^2/2\lambda =: v$. After choosing a basis of V, we can assume that a minimum is of the form

$$\Phi_0 = \begin{pmatrix} v \\ 0 \end{pmatrix}$$

Clearly, this is not invariant any more under the $U(1) \times SU(2)$ action, however Φ_0 is invariant under the subgroup

$$H = \left\{ (e^{i\theta}, \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} \right\} \simeq U(1)$$

This is the celebrated **Higgs spontaneous symmetry breaking mechanism**, with the Higgs field h appearing as quantum fluctuations $v \rightsquigarrow v + h$. Note that when we put $\Phi = \Phi_0$ in the action S_h, one generates in this way terms of the form $v^2 g^{\mu\nu} A_\mu A_\nu$, which are interpreted as mass terms for the gauge fields. More precisely, the physical gauge fields are the photon A, the Z-boson and the $W^\pm$-bosons. They are a linear combination (a rotation) of the gauge fields $(B, W) \in \Omega^1(M) \otimes u(1) \oplus su(2)$ that would arise from the previous discussion:

$$\begin{pmatrix} A \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix}; \qquad W = \begin{pmatrix} W^3 & W^+ \\ W^- & -W^3 \end{pmatrix}$$

where θ_w is the so-called Weinberg angle.