Three Derivations of the Dirac Equation

Jeremy Atkins

2020-04-29

Abstract

We provide three derivations of the Dirac equation. The first builds off of facts we know about Weyl spinors, going from the Weyl Lagrangian to the Dirac Lagrangian, and seeing the Dirac equation pop out. The second uses group theory more closely, looking at the electron spinor at rest and boosting it to an arbitrary frame, from which the Dirac equation again falls out. Finally, we look at Dirac's original derivation, using only the Klein-Gordon equation and his intuition.

1 Introduction

The Dirac equation is one of the most brilliant equations in all of theoretical physics. It describes all relativistic spin-$\frac{1}{2}$ massive particles that are symmetric with respect to parity, and was the first fundamental equation to successfully account for relativity in quantum mechanics.

Dirac is said to have pulled the equation nearly fully-formed *ex nihilo*, by basically guessing the correct answer. While this method is useful as a glimpse at what the great physicists of the early 20th century were able to do, it doesn't give us much insight as to why this equation is actually right.

Here we work through three different derivations of Dirac's eponymous equation, the first two entirely using group theory to truly go from the basics to where we want to be. We save Dirac's coup for last.

2 Preliminaries

We will speed through some of the notation and basic concepts needed to understand the Dirac equation, with minimal detail. For more information, see the other papers, which go into them in more depth.

In general, we will be working with the Lorentz algebra, $SO(3,1)$, which has the generators J_x, J_y, and J_z for the rotations, and K_x, K_y, and K_z for the boosts. They have the commutation relations

$$\begin{aligned}[J_i, J_j] &= i\epsilon_{ijk} J_k \\ [J_i, K_j] &= i\epsilon_{ijk} K_k \\ [K_i, K_j] &= -i\epsilon_{ijk} J_k\end{aligned}$$

In fact, the Lorentz algebra falls apart into two pieces, two non-interacting $SU(2)$ subalgebras, which have the generators

$$J_{\pm,i} = \frac{1}{2}(J_i \pm iK_i)$$

We call the $+$ the right-handed spinor and the $-$ the left-handed spinor. This comes from the fact that $SO(4)$ is locally isomorphic to $SU(2)\otimes SU(2)$, and if we examine the Lie algebra for $SO(4)$, we find we obtain the above relations for $SO(3,1)$ by simply sending $K_i \to iK_i$. Thus, they should have similar structure (and we see that they do). As expected, the commutation relations for these pieces are

$$\begin{aligned}[J_{+,i}, J_{-,j}] &= 0 \\ [J_{\pm,i}, J_{\pm,j}] &= i\epsilon_{ijk} J_{\pm,k}\end{aligned}$$

The representations of $SU(2)$ are labeled by half-integers j, and have dimension $2j+1$, so since $SO(3,1)$ is just two of these multiplied, we can label its representations with (j^+, j^-), a pair of half-integers. This representation has dimension $(2j^+ + 1)(2j^- + 1)$.

The representations we need to focus on most are the 2-dimensional irreducible spinor representations, $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$, which are so-called because they both reduce to the 2D spinor of the rotation algebra upon restriction to $SO(3)$. We'll denote these objects with u and v, respectively. These are the *Weyl spinors*. Since in one sense, these are just the same spinors we're familiar with from $SO(3)$, they transform as we would expect under rotations: $u \to e^{i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} u$ and $v \to e^{i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} v$. However, it turns out that they transform oppositely under boosts, since v is really the complex conjugate of u: $u \to e^{\vec{\varphi}\cdot\frac{\vec{\sigma}}{2}} u$ and $v \to e^{-\vec{\varphi}\cdot\frac{\vec{\sigma}}{2}} v$ (note the lack of i's; that's because we're really working with iK, so they cancel).

Finally, we need the Lagrangian for these spinors. We can obtain this by using the four vector $\omega^\mu = (u^\dagger u, u^\dagger \vec{\sigma} u)$ to get the Lagrangian density

$$\mathcal{L} = iu^\dagger \sigma^\mu \partial_\mu u \tag{1}$$

with a similar one for v.

3 Weyl Spinors Are Insufficient

Remember exactly what we're looking for here: we're trying to find a relativistic representation of spin-$\frac{1}{2}$ particles. It seems like we have what we want in the Weyl spinors: they're basically just an extension of the non-relativistic spinors to include time. Why don't they do what we want?

To see why, we need to look at the generators in differential form. These are

$$\begin{aligned} J_i &= -i\,\epsilon_{ijk}\,x_j\,\partial_k \\ iK_i &= t\,\partial_i + x_i\,\partial_t. \end{aligned}$$

Now, consider what happens when we apply a parity transformation to the generators, making $x_i \to -x_i$ and $\partial_i \to -\partial_i$. Then $J_i \to J_i$, but $K_i \to -K_i$. This is equivalent to exchanging $u \leftrightarrow v$. So just a Weyl spinor on its own won't work, if we want invariance under parity.

4 The Dirac Spinor

In fact, the solution is to simply stack u and v in a 4-dimensional reducible representation: $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$. This is the *Dirac spinor*.

To get the Dirac spinor's Lagrangian density, we can just add those obtained for u and v in (1), getting

$$\mathcal{L} = iu^\dagger \sigma^\mu \partial_\mu u + iv^\dagger \bar{\sigma}^\mu \partial_\mu v$$

where $\bar{\sigma}^\mu = (I, -\vec{\sigma})$. However, now that we have access to both u and v, we can construct new invariant terms. Since they transform oppositely under boosts,

$$u^\dagger v \to u^\dagger e^{-i\vec{\xi}^* \cdot \frac{\vec{\sigma}}{2}} e^{i\vec{\xi}^* \cdot \frac{\vec{\sigma}}{2}} v = u^\dagger v$$

where $\vec{\xi} = \vec{\theta} - i\vec{\varphi}$. It's easy to check that $v^\dagger u$ works similarly. Thus, we can add these terms to the Lagrangian density. To keep the Lagrangian Hermitian, we should add them in the combination $m(u^\dagger v + v^\dagger u)$, where m is a suggestively-named real parameter. Then our final Lagrangian density for the Dirac spinor is

$$\mathcal{L} = iu^\dagger \sigma^\mu \partial_\mu u + iv^\dagger \bar{\sigma}^\mu \partial_\mu v - m(u^\dagger v + v^\dagger u) \tag{2}$$

If we vary $\mathcal{L}$ with respect to $u^\dagger$, treating $u^\dagger$ and u as independent variables, and set the result equal to 0, we get $i\sigma^\mu \partial_\mu u = mv$, and doing similarly for $v^\dagger$ gets us $i\bar{\sigma}^\mu \partial_\mu v = mu$. Transforming to momentum space, these equations become

$$\sigma^\mu p_\mu u = (E - \vec{\sigma} \cdot \vec{p})u = mv \tag{3}$$

$$\bar{\sigma}^\mu p_\mu v = (E + \vec{\sigma} \cdot \vec{p})v = mu \tag{4}$$

Acting on (3) with $(E + \vec{\sigma} \cdot \vec{p})$ on the left and combining with (4), we get that $(E^2 - \vec{p}^2)u = m^2 u$, ie u vanishes unless $E^2 = \vec{p}^2 + m^2$, which is of course Einstein's famous relation! We see that adding that final term to the Lagrangian density has endowed our spinor with mass.

So, our spinor has mass, respects parity, and has two helicity states with $h = \pm 1$ (see previous lectures). This is all consistent with the electron!

5 The Dirac Equation Pops Up

Let's combine (3) and (4) into into a single matrix equation, and combine the two components $u(p)$ and $v(p)$ into a single Dirac spinor:

$$\begin{pmatrix} \bar{\sigma}^\mu p_\mu v(p) \\ \sigma^\mu p_\mu u(p) \end{pmatrix} = \begin{pmatrix} 0 & \bar{\sigma}^\mu p_\mu \\ \sigma^\mu p_\mu & 0 \end{pmatrix} \begin{pmatrix} u(p) \\ v(p) \end{pmatrix} \tag{5}$$

If we define $\psi = \begin{pmatrix} u \\ v \end{pmatrix}$ and

$$\gamma^\mu = \begin{pmatrix} 0 & \bar{\sigma}^\mu \\ \sigma^\mu & 0 \end{pmatrix}$$

we can write the left side of (11) as $\gamma^\mu p_\mu \psi(p)$. We can now package (3) and (4) together to get

$$(\gamma^\mu p_\mu - m)\psi(p) = \begin{pmatrix} \bar{\sigma}^\mu p_\mu v(p) - m\, u(p) \\ \sigma^\mu p_\mu u(p) - m\, v(p) \end{pmatrix} = 0$$

Transforming back to position space, we get

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0 \tag{6}$$

This is the famed *Dirac equation*, which models spin-$\frac{1}{2}$ relativistic massive particles. If we like, we can introduce Feynman's slash notation, where for a four vector a_μ, $\not{a} = \gamma^\mu a_\mu$. This lets us rewrite (6) as

$$(i\not{\partial} - m)\psi = 0 \tag{7}$$

6 Another Way, Without Detour

We have derived the Dirac equation by taking a detour through the Weyl equation. What if we didn't want to do that, and just want to start from $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ right away, as mandated by parity invariance?

Because of parity, we need to have a four-component spinor, but spin-$\frac{1}{2}$ particles only have two degrees of freedom. Recall that the state of a spin-$\frac{1}{2}$ particle can be written

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|\uparrow\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|\downarrow\rangle \tag{8}$$

The only degrees of freedom here are θ and φ. We somehow need to remove the extra two that we have, in a way that still respects Lorentz invariance.

We should be able to start from the rest frame, and then boost up to any other frame we want. So begin there, and take the electron's momentum $p_r = (m, \vec{0})$, where m is the electron mass. We essentially need to project out the two extra degrees of freedom, using a 4×4 projection operator P, such that $P\psi(p_r) = 0$

Because of parity, we should treat u and v on the same footing, since our answer should be invariant under their exchange. So set $u = v$ in the rest frame.

But see that we have now determined our operator (up to a constant)! Take

$$P = \frac{1}{2}\begin{pmatrix} I & -I \\ -I & I \end{pmatrix}$$

Then

$$P\psi(p_r) = \frac{1}{2}\begin{pmatrix} I & -I \\ -I & I \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{2}\begin{pmatrix} u - v \\ v - u \end{pmatrix} = 0.$$

Next, write $P = \frac{1}{2}(I - \gamma^0)$, where γ^0 is a suggestively named 4×4 matrix (we have of course never heard of "gamma matrices"). Then

$$\gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$$

Then finally, our projection equation $P\psi(p_r) = 0$ can be written

$$(\gamma^0 - I)\psi(p_r) = 0 \tag{9}$$

This doesn't look like much, does it? But in fact, this is the Dirac equation in a sneaky disguise!

(9) is satisfied by any spinor $\psi(p_r)$ in its rest frame. We can now use Lorentz transformations to boost our wavefunction to any frame we choose. Let's boost in the z direction, knowing that our choice of coordinate axes is arbitrary. At the boost angle φ, the electron acquires energy $E = m\cosh(\varphi)$ and 3-momentum $p = m\sinh(\varphi)$, where m is the electron mass. Note this identity: $me^{\varphi\sigma_3} = m(\cosh(\varphi) + \sigma_3 \sinh(\varphi) = E + \sigma_3 p)$.

To obtain the wavefunction for a moving particle, then, we simply apply a boost $e^{\varphi\frac{\sigma_3}{2}} = e^{i\varphi K_z}$ to $\psi(p_r)$. So act with the boost on (9), and sneakily multiply by m at the same time:

$$e^{i\varphi K_z} m(\gamma^0 - I)\psi(p_r) = m(e^{i\varphi K_z}\gamma^0 e^{-i\varphi K_z} - I)\psi(p) = 0 \tag{10}$$

We can write the first term in matrix form as

$$\begin{aligned} me^{i\varphi K_z}\gamma^0 e^{-i\varphi K_z} &= \begin{pmatrix} e^{\frac{1}{2}\varphi\sigma_3} & 0 \\ 0 & e^{-\frac{1}{2}\varphi\sigma_3} \end{pmatrix} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \begin{pmatrix} e^{-\frac{1}{2}\varphi\sigma_3} & 0 \\ 0 & e^{\frac{1}{2}\varphi\sigma_3} \end{pmatrix} \\ &= m \begin{pmatrix} 0 & e^{\varphi\sigma_3} \\ e^{-\varphi\sigma_3} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & E + \sigma_3 p \\ E - \sigma_3 P & 0 \end{pmatrix} \end{aligned}$$

Note that the factors of $\frac{1}{2}$ have disappeared from the exponentials. Thus, if we re-fold the σ_3 's into $\vec{\sigma}$'s, (10) becomes

$$\begin{pmatrix} -m & E + \vec{\sigma}\cdot\vec{p} \\ E - \vec{\sigma}\cdot\vec{p} & -m \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0 \tag{11}$$

If we define

$$\gamma^i = \begin{pmatrix} 0 & -\sigma_i \\ \sigma_i & 0 \end{pmatrix}$$

we see that the matrix in (11) can be written as $\gamma^0 p^0 - \gamma^i p^i - m$ (where the m is implied to be mI of course). So we can write (11) as $\gamma^\mu p^\mu - m)\psi(p) = 0$. Transforming to position space, this is just the Dirac equation $i\gamma^\mu \partial_\mu - m)\psi(x) = 0$.

Recognize that the two equations in (11) are simply the coupled equations $(E - \vec{\sigma} \cdot \vec{p})u = 0$ and $(E + \vec{\sigma} \cdot \vec{p})v = 0$ from our first derivation. So our derivations have basically converged on the same sort of process, just from different directions.

This derivation provides a deep way of looking at the Dirac equation: it seems it is simply a projection of a spinor boosted into an arbitrary frame. It provides a great example of the power of group theory in modern physics. All we need to know is how the electron field transforms under the rotation group, which tells us how it transforms under the Lorentz group. With this help, we don't need to be as brilliant as Dirac to derive his eminent equation. . .

7 Dirac's Brilliant Guess

. . . But it might be nice to know how he did it anyway.

He began with the Klein-Gordon equation:

$$(\partial^2 + m^2)\psi = 0 \tag{12}$$

which "merely" states that $p^2 = m^2$ in position space, and doesn't really work in practice for various reasons that we won't get into here. Dirac thought to possibly try an equation that was first-order in space and time derivatives, rather than second-order like the failed Klein-Gordon equation. As the story goes, he was sitting staring into a fireplace at Cambridge when he wrote down $c^\mu \partial_\mu - b)\psi = 0$.

If c^μ were four numbers, then c would define a 4-vector, and thus pick out a privileged direction, which would break Lorentz invariance, so that couldn't be right; they must be some other kind of object.

He then multiplied this equation on the left by $(c^\mu \partial_\mu + b)$, obtaining $\frac{1}{2}\{c^\mu, c^\nu\}\partial_\mu \partial_\nu - b^2)\psi = 0$. He then noticed that, if $\{c^\mu, c^\nu\} = -2\eta^{\mu\nu}$ and $b = m$, he would obtain the Klein-Gordon equation.

Clearly, to satisfy this anti-commutation relation, the c^μ would have to be matrices. By trial and error, he worked out that the smallest matrices that would work are 4×4, and thus ψ must have four components. He wrote

$c^\mu = i\gamma^\mu$, with γ^μ being 4×4 matrices satisfying the (Clifford) algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$, and thus obtained his equation: $(i\gamma^\mu\partial^\mu - m)\psi = 0$.

Thus, we have our third derivation.

8 Conclusion

We have shown three separate derivations of the Dirac equation, one of the most important discoveries of 20th century physics. From this equation, we finally had a description of spin-$\frac{1}{2}$ particles at relativistic speeds, which allowed us to better understand much of atomic and condensed matter physics.

Perhaps the most exciting of its consequences was its prediction of antimatter. If we slightly modify the Dirac equation to to include the electromagnetic potential, we get

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

Where e is the charge of the spin-$\frac{1}{2}$ particle under consideration. It turns out that we can *also* find a solution of an alternate version of this equation that as the sign of e flipped. This solution is related to ψ essentially by complex conjugation. Dirac was able to predict this merely from the form of his result. Truly a powerful relation.